A MATHEMATICAL MODEL FOR THE
CONSTRUCTION OF CLADOGRAMS

by

George William Moore

Institute of Statistics
Mimeograph Series No. 731
Raleigh - February 1971
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1. INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Purpose of This Dissertation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 How to Read This Dissertation</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Chapter Organization</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. MATHEMATICAL MODELS IN BIOLOGY</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Structural Versus Empirical Models</td>
<td>5</td>
</tr>
<tr>
<td>2.2 The Axiomatic Method in Biology</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Biological Versus Mathematical Definitions</td>
<td>10</td>
</tr>
<tr>
<td>2.4 Model Theory in Biology</td>
<td>11</td>
</tr>
<tr>
<td>2.5 Flow of Reasoning in This Dissertation</td>
<td>15</td>
</tr>
<tr>
<td>2.6 Informal Set Theory</td>
<td>16</td>
</tr>
<tr>
<td>2.7 Set Relationships</td>
<td>20</td>
</tr>
<tr>
<td>2.8 Special Sets</td>
<td>22</td>
</tr>
<tr>
<td>2.9 A Precaution</td>
<td>24</td>
</tr>
<tr>
<td>2.10 Set Operations</td>
<td>25</td>
</tr>
<tr>
<td>2.11 Mathematical Relations</td>
<td>29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. PRINCIPLES OF GENEALOGY</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 First Principles</td>
<td>31</td>
</tr>
<tr>
<td>3.2 Ancestor or Equal to</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Pedigrees</td>
<td>36</td>
</tr>
<tr>
<td>3.4 Common Pedigrees</td>
<td>38</td>
</tr>
<tr>
<td>3.5 The Monoparental Biocosm</td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. OPERATIONAL TAXONOMIC UNITS AND ISOLATION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Operational Taxonomic Units</td>
<td>58</td>
</tr>
<tr>
<td>4.2 The Superpartition</td>
<td>64</td>
</tr>
<tr>
<td>4.3 Isolation</td>
<td>69</td>
</tr>
<tr>
<td>4.4 Sampling from Isolated Sets</td>
<td>72</td>
</tr>
<tr>
<td>4.5 Cladistic Partitions</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. DENDROGRAMS AND CLADOGRAMS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 The Dendrogram</td>
<td>77</td>
</tr>
<tr>
<td>5.2 Cardinality of a Binary Dendrogram</td>
<td>84</td>
</tr>
<tr>
<td>5.3 The Dendrogram as a Monoparental Biocosm</td>
<td>88</td>
</tr>
<tr>
<td>5.4 The Cladogram</td>
<td>97</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS (continued)

6. SOURCES OF INPUT DATA ............................................. 103
   6.1 The Phenetic Matrix ........................................... 103
   6.2 Comparison of Character States ............................... 107
   6.3 Direct Techniques for Obtaining a Phenetic Matrix .......... 109
   6.4 The Immunodiffusion Technique ............................... 119
   6.5 Solution for the Phenetic Matrix ............................ 128
   6.6 Inversion of the Primary Solution Matrix .................. 133

7. DIVERGENT EVOLUTION IN MOLECULAR CHARACTERS .................. 138
   7.1 Trouble with Genetic Load ................................... 138
   7.2 Low Selection at the DNA Level ............................. 141
   7.3 Low Selection at the Protein Level .......................... 145
   7.4 Low Selection and Divergent Evolution ..................... 147
   7.5 The Parsimony Hypothesis ................................... 148
   7.6 Divergence Hypothesis ...................................... 150
   7.7 The Superphenetic Matrix ................................... 154
   7.8 Formulas for a Superphenetic Matrix ....................... 157
   7.9 Nesting Levels in a Dendrogram ............................. 163
   7.10 Weighted Average Superphenetic Matrix .................... 170
   7.11 Choice of Cladogram Juncutures ............................ 172
   7.12 Divergence-in-Mean ....................................... 193
   7.13 Divergence and Uniform Evolution .......................... 198

8. THE STRUCTURE OF PHENOGRAM ALGORITHMS ......................... 203
   8.1 Iteration Structures ...................................... 203
   8.2 Refinements of a Partition .................................. 208
   8.3 Agglomerative and Divisive Iteration Structures .......... 211
   8.4 Phenogram Algorithms Generate Dendrograms ................ 213
   8.5 Cladogram Functions ....................................... 222

9. EVALUATION OF PHENOGRAM ALGORITHMS ............................ 224
   9.1 Strategy .................................................. 224
   9.2 Proofs ................................................... 228
   9.3 Counterexamples .......................................... 231
   9.4 Conclusions and Recommendations ............................ 241

10. LIST OF REFERENCES ............................................... 245

11. APPENDICES .................................................... 251
   11.1 Primitive and Defined Terms ................................ 251
   11.2 Flow of Reasoning in This Dissertation .................... 258
   11.3 Flow Chart for Unweighted Average Agglomerative .......... 262
        Algorithm .............................................. 262
## LIST OF TABLES

<table>
<thead>
<tr>
<th>#</th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Performance of phenogram algorithms. This table shows the weakest evolutionary hypothesis under which each phenogram algorithm is guaranteed to yield the cladogram</td>
<td>227</td>
</tr>
<tr>
<td>11.1</td>
<td>Set theory terms introduced in chapter two</td>
<td>251</td>
</tr>
<tr>
<td>11.2</td>
<td>Primitive and defined terms introduced in chapter three</td>
<td>252</td>
</tr>
<tr>
<td>11.3</td>
<td>Defined terms introduced in chapter four</td>
<td>253</td>
</tr>
<tr>
<td>11.4</td>
<td>Defined terms introduced in chapter five</td>
<td>254</td>
</tr>
<tr>
<td>11.5</td>
<td>Primitive and defined terms introduced in chapter six</td>
<td>255</td>
</tr>
<tr>
<td>11.6</td>
<td>Defined terms introduced in chapter seven</td>
<td>256</td>
</tr>
<tr>
<td>11.7</td>
<td>Defined terms introduced in chapter eight</td>
<td>257</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

2.1. The house of Abraham. The sons and grandsons of Abraham and their mothers if known. Source: Genesis, chapters 16-25. ................. 17

2.2. Sets: Abraham and his descendants. Set U is the set of Abraham and all his descendants. Sets A, B, C, D, and E are subsets of U. Dots (...) indicate additional members which are not named explicitly. .......... 21

2.3. The union of two sets. Set A U B is the union of sets A and B of Figure 2.2. Set C U E is the union of sets C and E of Figure 2.2. The sets A U B and C U E are highlighted by diagonal hatching. ......... 26

2.4. Set A N B is the intersection of sets A and B of Figure 2.2. Set C N E is the intersection of sets C and E of Figure 2.2. The sets A N B and C N E are highlighted by diagonal hatching. Note that A N B is the null set. ............... 27

2.5. The difference between two sets. Set C - E is set C minus set E, where sets C and E are from Figure 2.2. Set E - C is set E minus set C. The sets C - E and E - C are highlighted by diagonal hatching. Note that these two sets are not the same. .......... 28

3.1. Pedigrees. The pedigree of {Isaac}, denoted Pt({Isaac}), is the set of Isaac and all his ancestors. The pedigree of {Ishmael, Isaac}, denoted Pt({Ishmael, Isaac}), is the set of Ishmael, Isaac, and any ancestor of either Ishmael or Isaac. Dots (...) indicate additional members which are not named explicitly. ............... 37

3.2. Common and most recent common pedigrees. The common pedigree of {Ishmael, Isaac}, denoted CP({Ishmael, Isaac}), is the set of all mutual ancestor-or-equals of Ishmael and Isaac. Dots (...) indicate additional members which are not named explicitly. The most recent common pedigree of {Ishmael, Isaac}, denoted RP({Ishmael, Isaac}), is the set of only those members of the common pedigree which have no descendants in the common pedigree. ............... 40
LIST OF FIGURES (continued)

3.3. Evolution of horses. Simplified diagram of the evolution of the now-extinct Hipparion and the contemporary Equus. Source: Bastian (1964, p. 231) ........................................... 43

4.1. Partitions of X. Set X is the set of all the grandsons of Abraham. (A) The set X. (B) A partition of X, \( \Pi^1_X \). (C) A different partition of X, \( \Pi^2_X \) ....................................... 60

5.1. A dendrogram. Let \{A,B,C,D,E,F,G\} = \( \Pi_X \) be a system of OTUs (a partition of X) whose union is X. (A) A dendrogram on \( \Pi_X \) diagrammed in the conventional format. The set X is the trunk of this dendrogram and the set \( \Delta \), where \( \Delta = \{A,B,C,D,E,F,G,H,I,J,K,L,X\} \) is the set of junctures for this dendrogram. (B) The same dendrogram in Hennig's (1966) format .................................. 78

6.1. Complement fixation. A simplified account of the sequence of steps in the complement fixation technique ............................................ 116

6.2. Immunoelectrophoresis. Diagram of a completed immunoelectrophoresis plate ......................................................... 118

6.3. The trefoil Ouchterlony plate. (A) Charging the trefoil Ouchterlony plate. (B) Early diffusion pattern. (C) Final diffusion pattern ............................................ 122

6.4. Experimental design for sample immunodiffusion data. The homologous OTU (OTU1) is Homo sapiens. The heterologous OTUs are \( \text{OTU}_1 = Pongo pygmaeus, \) \( \text{OTU}_2 = Hylobates lar, \) and \( \text{OTU}_3 = Macaca mulatta. \) (A) The comparison matrix. (B) The protocol matrix ............................................ 127

6.5. The protocol and primary solution matrix for sample data. (A) The protocol matrix, E. (B) The primary solution matrix, S ............................................ 130

6.6. Revision of the protocol matrix. (A) The original protocol matrix, E. The corresponding primary solution matrix does not invert. (B) The revised protocol matrix, \( E' \), in which the (3,5)th comparison has not been made. The corresponding primary solution matrix does invert ............................................ 135
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.</td>
<td>Hypothetical parsimonious cladogram. (A) Diagram of patristic distances for a hypothetical cladogram having OTUs V, W, Y, and Z. (B) A phenetic matrix which corresponds to the diagram in part A</td>
<td>149</td>
</tr>
<tr>
<td>7.2.</td>
<td>Alternative parsimonious dendrogram. (A) An alternative diagram of patristic distances for a dendrogram having OTUs V, W, Y, and Z. (B) A phenetic matrix which corresponds to the diagram in part A. Note that the phenetic matrix for this figure is the same as the phenetic matrix for Figure 7.1, even though the dendrogram is different</td>
<td>151</td>
</tr>
<tr>
<td>7.3.</td>
<td>Minimal superphenetic matrix. (A) A hypothetical phenetic matrix. (B) A minimal superphenetic matrix calculated from the phenetic matrix in part A. (C) Sample calculations from part B</td>
<td>159</td>
</tr>
<tr>
<td>7.4.</td>
<td>Maximal superphenetic matrix. (A) A hypothetical phenetic matrix. (B) A minimal superphenetic matrix calculated from the phenetic matrix in part A. (C) Sample calculations from part B</td>
<td>161</td>
</tr>
<tr>
<td>7.5.</td>
<td>Unweighted average superphenetic matrix. (A) A hypothetical phenetic matrix. (B) An unweighted average superphenetic matrix calculated from the phenetic matrix in part A. (C) Sample calculations for part B</td>
<td>162</td>
</tr>
<tr>
<td>7.6.</td>
<td>Nesting levels. (A) The dendrogram of Figure 5.1. (B) Nonzero nesting levels. (C) Zero nesting levels. (D) Undefined nesting levels, because the latter set is not a juncture</td>
<td>164</td>
</tr>
<tr>
<td>7.7.</td>
<td>Weighted average superphenetic matrix. (A) A hypothetical phenetic matrix. (B) Assumed dendrogram. (C) Sample calculations from the weighted average superphenetic matrix</td>
<td>171</td>
</tr>
<tr>
<td>7.8.</td>
<td>Non-uniform evolution in a hypothetical cladogram. (A) Diagram of patristic distances for a hypothetical cladogram with OTUs W, Y, and Z. (B) The maximum parsimony phenetic matrix for the hypothetical cladogram. (C) The phenogram computed from the phenetic matrix by any of the agglomerative methods in chapter nine</td>
<td>200</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES (continued)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.9</td>
<td>Alternative parsimonious dendrograms. All dendrograms satisfy maximum parsimony assumptions on the phenetic matrix of Figure 7.8. Dendrograms (A) and (B) are not the &quot;true&quot; cladogram of Figure 7.8</td>
<td>202</td>
</tr>
<tr>
<td>8.1</td>
<td>Refinements of a partition. (Compare Figure 4.1.) (A) ( \Pi^1_X ) is the partition of ( X ) illustrated in Figure 4.1. (B) ( \Pi^2_X ) is a refinement of ( \Pi^1_X ). (C) ( \Pi^3_X ) is not a refinement of ( \Pi^2_X )</td>
<td>210</td>
</tr>
<tr>
<td>9.1</td>
<td>Minimal agglomerative: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) MNA phenogram. (D) Critical superphenetic calculations. ( \langle U U V U W, Y \rangle ) is minimal for UAA but ( \langle U U V U W, Z \rangle ) is minimal for MNA</td>
<td>233</td>
</tr>
<tr>
<td>9.2</td>
<td>Maximal agglomerative: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) MXA phenogram. (D) Critical superphenetic calculations. ( \langle U U V U W, Y \rangle ) is minimal for UAA but ( \langle U U V U W, Z \rangle ) is minimal for MXA</td>
<td>235</td>
</tr>
<tr>
<td>9.3</td>
<td>Weighted average agglomerative: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) WAA phenogram. (D) Critical superphenetic calculations. ( \langle U U V U W, Y \rangle ) is minimal for UAA but ( \langle U U V U W, Z \rangle ) is minimal for WAA</td>
<td>236</td>
</tr>
<tr>
<td>9.4</td>
<td>Maximal divisive: counterexample to equidivergence. (A) Prototype cladogram. (B) Equidivergent phenetic matrix. (C) All possible &quot;first steps&quot; in the divisive iteration structure give the same superphenetic value. Therefore, the MND phenogram is ambiguous</td>
<td>238</td>
</tr>
<tr>
<td>9.5</td>
<td>Minimal divisive: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) MND phenogram. (D) Critical superphenetic calculations. ( \langle U U V U W U Z, Y \rangle ) is greater than ( \langle U U V U W U Y, Z \rangle ) at the first divisive step, although the latter corresponds to the correct division</td>
<td>239</td>
</tr>
</tbody>
</table>
LIST OF FIGURES (continued)

9.6. Unweighted average divisive: counterexample to strict divergence. (A) Prototype cladogram. (B) Phenetic matrix. (C) UAD phenogram. (D) Critical superphenetic calculations. \( <U V W U Y Z > \) is greater than \( <U V W Y U Z > \) at the first divisive step, although the latter corresponds to the correct division division ........................................ 240

11.1. Flow of reasoning in chapters three and four ............. 258
11.2. Flow of reasoning in chapter five .......................... 259
11.3. Flow of reasoning in chapter seven ........................ 260
11.4. Flow of reasoning in chapters eight and nine ............ 261
11.5. Flow chart for unweighted average agglomerative algorithm.
Problem: Given N OTUs numbered 1 through N and an N by N phenetic matrix, P, find the unweighted average agglomerative phenogram ...................... 262
1. INTRODUCTION

1.1 Purpose of this Dissertation

In 1963, Sokal and Sneath published their controversial text, Principles of Numerical Taxonomy (PNT). On pages 189-194, PNT discusses the construction of tree-like diagrams for representing taxonomic relationships among sets of organisms, or taxonomic units. A dendrogram is any tree-like diagram. A phenogram is any dendrogram which has been calculated from data about the taxonomic units according to a computer algorithm. PNT presents six different phenogram algorithms, or methods for computing a dendrogram from input data.

The authors then evaluate these methods on the basis of their intuition and experience. Their evaluation reflects a rather prevalent attitude among biologists who use computers that if one runs enough data, one is able to judge the method.

For a biologist who understands his material and has a good feeling for what the computer does, this attitude is by and large correct. Trouble starts when this biologist confronts a new data source. Since he doesn't really understand what makes the technique "work", he may waste a great deal of time getting a "feel" for his new material. Often this is a consequence not of carelessness on the biologist's part, but because biological principles themselves are logically "mushy" in a few key places. This dissertation employs a manner of inquiry which overcomes the uncertainties in the intuitive approach: it derives phenogram algorithms from rigorously stated biological first principles, using the method of ordinary mathematical proof.
A cladogram is a dendrogram whose progressively thicker branches correspond to progressively more ancient common ancestors. It is distinguished from a phenogram in that a phenogram is the result of a well-defined sequence of calculations, whereas a cladogram is a fact of nature. The cladogram problem is the problem of specifying a process of calculation by which input data about taxonomic units can be used to infer a branching sequence of progressively more ancient common ancestors. For a biologist, the cladogram problem consists of two major questions: what kind of data to obtain and which phenogram algorithm to feed these data into. The purpose of this dissertation is to explore these two questions in detail. A model for cladogram formation in nature is set up using the "biocosm" formalism of Williams (1970); a formalism for computer phenogram algorithms is set up using the ordinary tools of set theory. Then the evolutionary model is used to rank the performance of seven different phenogram algorithms. I conclude that Sokal and Michener's (1958) "unweighted pair-group method" is the best choice for information about molecular characters summarized in matrix form.

1.2 How to Read this Dissertation

This dissertation uses methods of ordinary mathematical reasoning to evaluate the ability of seven phenogram algorithms to infer the true cladogram subject to three evolutionary hypotheses. Any biologist who has read Sokal and Sneath's (1963) PNT and has a general interest in numerical taxonomy might find this dissertation of sufficient interest to skim, skipping all formalized definitions, theorems, and proofs.
The mathematical portion of this dissertation is much more tedious than the descriptive text, and is not worth the expenditure of effort for most biological readers. Even a reader who intends to study the mathematical results might want to skip the proofs and the lemmas (specialized, nonintuitive theorems introduced solely to facilitate later proofs). A high school course in plane geometry, some college-level algebra, and a good deal of patience are the only prerequisites. Background in elementary set theory would be helpful, but not essential. All the mathematical and set theory terminology is explained in this dissertation, and collected for reference in Appendix 11.1. Each mathematical statement in this dissertation is prefaced by an informal resume of its contents and purpose, as an aid to this reader.

Finally, proofs are provided for all but the most obvious theorems, to verify the validity of each theorem.

1.3 Chapter Organization

This dissertation progresses in three major phases. The first phase develops enough mathematical machinery for the next two phases. The second phase discusses which data are appropriate for the solution of the cladogram problem. The third phase develops the seven phenogram algorithms of this dissertation and evaluates their performance with respect to three different evolutionary hypotheses.

Chapters two through five constitute the first phase of this dissertation. Chapter two develops the general area of mathematics which will be employed in this dissertation: the theory of sets and relations. Chapter three develops well-known concepts of genealogy
in a notation appropriate for subsequent chapters of this dissertation.
Chapter four develops the concepts of Operational Taxonomic Units (OTUs) and isolation between OTUs. Chapter five develops the concepts of dendrogram and cladogram.

Chapters six and seven constitute the second phase of this dissertation. Chapter six discusses several sources of molecular data commonly used for the cladogram problem, and shows how these sources can be reduced to a phenetic matrix. Chapter seven suggests that molecular data are an appropriate source for phenetic matrixes, because of the low level of selection ("non-Darwinian evolution") on individual molecular characters. Chapter seven suggests three evolutionary hypotheses for the evolution of molecular characters, and demonstrates some simple consequences of these hypotheses for use in later chapters.

Chapters eight and nine constitute the third phase of this dissertation. Chapter eight develops a mathematical machinery for phenogram algorithms, and chapter nine uses this machinery to evaluate each of seven phenogram algorithms with respect to the three evolutionary hypotheses described in chapter seven.

Chapter nine is followed by the list of references and three appendices. Appendix 11.1 is a glossary of all primitives and defined terms employed in this dissertation. Appendix 11.2 is a chart which details the flow of reasoning employed in this dissertation. Appendix 11.3 is a computer flow chart for the unweighted pair-group (Sokal and Michener, 1958) phenogram algorithm.
2. MATHEMATICAL MODELS IN BIOLOGY

2.1 Structural versus Empirical Models

Biology is a young science in its use of mathematical models. Few areas of biology depend upon mathematics as more than superficial aid for summarizing observations. Genetics is one of the few major areas of biology some of whose first principles are described by a genuine mathematical model. Most mathematical models in biology do an incomplete job. They are either

(i) empirical models
or
(ii) structural models
but rarely both.

An empirical model is a mathematical model which has been constructed largely to fit a particular class of data or observations. Characteristically, this kind of model is used for fitting curves through biological data under assumptions such as "linearity", "normal distribution", etc., without much biological basis for knowing whether these conditions indeed hold. An empirical model often solves for population or environmental parameters which have little or no basic biological meaning. Often, an empirical model gives the experimenter a false sense of meaning, when he has in fact done little more than translated one pile of numbers into another, smaller pile.

A structural model falls at the other extreme. This kind of model starts from "obvious" first principles and works forward toward observations. Unfortunately, there are many perfectly plausible candidates for the status of "first principles" which are false, and because the formulation of a structural model is often so remote
from actual observations, it is very difficult to separate the plausible and true statements from the plausible and false statements. Consequently, a combination of quite plausible basic assumptions may lead to ridiculous conclusions. Another common shortcoming among structural models is that they often lead to conclusions which either cannot be observed experimentally or are of no interest to the biologist.

Biology is still very much an experimental science, and most good biologists reserve their credence for results closely supported by good data, the more the better. There is a widespread failure among biologists to recognize that a good mathematical model should not only fit observations, but also arise from truly biological first principles.

In this dissertation, I develop Williams' (1970) structural model and demonstrate how this structural model leads to a well-known empirical model which is already in use among practicing biologists. The axioms for Williams' (1970) model are quite simple and straightforward. Stated informally: no organism is a parent of itself and ancestry is unidirectional. From these beginnings plus the axioms of set theory, I construct a system for evaluating seven phenogram algorithms, some of which are in current use by numerical taxonomists.

The achievement of this dissertation is that it starts from purely structural principles and derives computing algorithms which are already known to work. Until recent years, the construction of phenograms from biological data has depended almost exclusively upon empirical models. A bewildering variety of phenogram algorithms confronts the
interested biologist (Sørenson, 1948; Sneath, 1957; Sokal and Michener, 1958; Rogers and Tanimoto, 1960; Lockhart and Hartman, 1963; Edwards and Cavalli-Sforza, 1964; Camin and Sokal, 1965; Fitch and Margoliash, 1967; Moore et al., 1969; Dayhoff, 1969; Kluge and Farris, 1969), but there are virtually no guidelines as to the conditions under which these algorithms can be trusted to give meaningful results beyond the intuitive judgment of the investigator. There is even a widespread feeling that just because "the computer does it", the result is somehow more objective than in conventional biological studies (Sokal and Sneath, 1963, p. 49; Rogers et al., 1967).

The first attempts to understand the conditions under which phenogram algorithms yield a cladogram were computer simulations (Camin and Sokal, 1965; Farris, 1970). Computer simulations employ "made-up" data consistent with known properties, and evaluate a phenogram algorithm on the basis of its performance with these made-up data. This is a step ahead of the experiential approach of Sokal and Sneath (1963, pp. 189-194), who base their judgment on data from the real, biological world with unknown properties. Computer simulation can be used to reject an algorithm if the algorithm fails to reconstruct the original cladogram. But computer simulation can never be used to accept an algorithm, because it is always possible that the algorithm would fail for an untried data set, even though it had been successful on all data sets tried theretofore. The only way to show that an algorithm always works is by mathematical proof.

In 1968, Estabrook presented a method for restricting the choice of dendrograms to a small subset which must include the true
cladogram (see also Hendrickson (1968)). Estabrook solved the cladogram problem subject to the "maximum parsimony" hypothesis (see chapter seven) and the following additional restrictions:

(i) the (discrete) character states for each OTU are known

(ii) the ancestral sequence of character states is known and is irreversible

and

(iii) the ultimate ancestral state happened only once.

Estabrook's work marks a major turning point in the literature of numerical taxonomy, because it employs ordinary methods of mathematical proof (comparable to the methods of this dissertation), rather than the "heuristic" (i.e., inspirational, but not necessarily reliable) approaches of other authors. There are several shortcomings in Estabrook's paper. First, the restrictions placed upon the true cladogram are not satisfied in most taxonomic investigations; ordinarily, the investigator does not know the ancestral sequence of character states, and is not even sure that this sequence is irreversible.

Second, Estabrook does not demonstrate the uniqueness of his result; it is perfectly possible that there are two or more distinct dendrograms satisfying Estabrook conditions for a particular data set -- the true cladogram and other, quite different dendrograms.

For reasons detailed in chapter seven, this dissertation does not employ a maximum parsimony hypothesis of evolution; rather, it develops three "divergence" hypotheses which all correspond to the intuitive idea that a more ancient ancestral separation for a pair of OTUs results in a greater dissimilarity value for that pair of OTUs.
This dissertation demonstrates that any dendrogram generated by the appropriate phenogram algorithm is a true cladogram, and that this cladogram is unique. The phenogram algorithm which emerges as the best (that is, the unweighted pair-group method of Sokal and Michener (1958)) is guaranteed to arrive at the correct solution after a relatively small, finite number of steps.

2.2 The Axiomatic Method in Biology

The basic mathematical tool which is employed in this dissertation is the method of axiomatics. Axiomatics is different from conventional mathematical derivations in the biological literature in that it is entirely self-contained. Except for the axioms of set theory (which could be included if necessary), the only assumptions in this dissertation are summarized as two, simple axioms. All other statements in this dissertation have been proved in terms of these two axioms using only the methods of ordinary mathematical proof. Similarly, the only undefined terms, or primitives, in this dissertation are "the set of all organisms" and the relation "is a parent of". All other terms in this dissertation have been defined in terms of these two primitives using only the grammar of logic. The strict, deductive structure of an axiomatic theory allows one to move easily from a theorem to the prior statements which were used to prove it (and similarly from a defined term to the prior terms by which it was defined). Since every theorem is proved in terms of prior theorems and the unproved axioms, and similarly since every term is defined in terms of prior terms and the undefined primitives, an axiomatic theory is noncircular.
The methods of this dissertation are those of naive axiomatics, that is, the standards of logical statement and proof employed by most mathematicians most of the time. Naive axiomatics is contrasted to the more cumbersome formalized axiomatics of Woodger's (1937) *The Axiomatic Method in Biology.*

2.3 Biological versus Mathematical Definitions

It is important for the biological reader to distinguish between his intuitive concept of definition and the mathematical concept of definition. In biology, a new term is defined in terms of previously defined terms and concepts in general use. A biological definition is rarely required to correspond exactly to what the author really means. The usual practice is for the definition to "come close" to what the author really means. Then the author proceeds to enumerate the "exceptions" to the definition. In mathematics, a definition means exactly what it says and says exactly what it means. The term which is being defined and the phrase which is used to define it are completely interchangeable in any statement subsequent to the definition. This is often not possible for biological definitions. Because of the presence of exceptions in biological definitions, arbitrary substitution of the term-being-defined and the defining-terms may readily lead to a logical absurdity. Woodger (1952, pp. 219-252), for example, has taken two "respectable" biological definitions of species, translated them into logical terms, and proved they lead to an absurdity. The reason he can do this is that the biological definitions he uses were never really meant to be exactly true, merely approximately true.
Definitions in this dissertation are mathematical definitions. The biological reader should be continually on the alert for this distinction. Each defined term in this dissertation means nothing more nor less than the sequence of terms of which the definition consists. I have tried to define terms which correspond to intuitive biological notions, but in cases where I have failed to capture the intuitive biology, the mathematical content of the definition takes precedence. In chapter four, for example, I define "is isolated from" to correspond roughly to the biologist's concept of reproductive isolation. Because of shortcomings in the biological notion, I have found it necessary to adopt a more stringent definition of my own. This mathematical definition may or may not correspond to a particular biologist's notion of isolation, but when I employ the term "is isolated from" in subsequent discussion, I make reference to my definition only. If a biologist "disagrees" with my concept of "is isolated from" (i.e., does not feel that it corresponds to his intuitive notion of isolation), then any theorem which I prove using "is isolated from" may be false if he substitutes his intuitive notion of isolation for my notion of "is isolated from"; my theorems are guaranteed to be true only for the concept as it is defined.

2.4 Model Theory in Biology

Model theory (Robinson, 1965) is a branch of mathematics concerned with the correspondence of mathematical models to the "real world". A model theory theory is an axiomatic theory, such as will
be developed in this dissertation; biologists ordinarily call this a "model". A model theory model is some system in the real world to which the model theory theory corresponds; biologists ordinarily call this simply the "real world". Whenever we construct a model theory theory corresponding to some biological process, we would like this model theory theory to have a model theory model in the real, biological world.

The problem in actually executing this plan is that there are differing views as to what the real, biological world actually is:

(i) Some statements about the biological world are agreed upon by all because they are tautologous. For example, Axiom I (chapter three) states that "no organism is a parent of itself". All biologists agree to this (so long as "organism" and "is a parent of" are understood in the conventional sense) because it is inherent in what we mean by the term "parent".

(ii) Some statements about the biological world are true because they can be proven by experiment. For example, the statement "some viruses lack DNA" can be proven true by finding a single virus which lacks DNA: say, the Rous sarcoma virus (an RNA virus).
(iii) Unfortunately, many biological statements are neither inherently tautologous nor verifiable by experiment. For example, the statement "all viruses have either DNA or RNA or both" is accepted by most biologists, but it cannot be proved until every single virus particle (including every foreseeable particle which might be called a "virus") is analyzed.

(iv) A final shortcoming of "biological reality" as biologists think of it is that many important techniques required for biology are not really part of the traditional subject matter of biology. For example, a statement of the form "if x is true and y is true, then x is true" is not really a part of biology, although any biologist would accept it. This statement belongs to logic. In order to achieve anything more than the most trivial conclusions, we need not only logic, but also some of its more sophisticated derivatives (set theory, real number theory, linear algebra).

This dissertation studiously avoids unprovable biological statements, such as those of type (iii), but since you can't get something for nothing, the dissertation doesn't give you any very useful conclusions which apply unconditionally. For example, this
dissertation turns out to be fairly useless to any biologist not willing to accept some form of divergent evolution. All of the theorems in chapter nine have the form (roughly speaking): "if evolution is reasonably divergent, then such-and-such algorithm gives a true cladogram". The biologist who is willing to commit himself to one of the hypotheses of divergent evolution developed in this dissertation can use the theorems proved in chapter nine; the biologist who doesn't accept any of these hypotheses of divergence will find a chapter full of theorems which (though true) don't apply to his version of biological reality.

The model theory model for the axiomatic theory of this dissertation is any biological world in which Axioms I and II and the axioms of set theory are true. Axioms I and II should be obvious to any biologist; the axioms of set theory should be acceptable to biologists. The only "doubtful" axioms are ones which apply to infinite sets (conspicuously, the Zorn Lemma, chapter three); since no statement in this dissertation requires infinite sets (although many permit it), and since the finite counterparts of the infinite-set-axioms are undisputed, the set theory part of the real world should be acceptable to biologists. The most serious hazard in the use of set theory is that of misunderstanding (review Section 2.3; see Section 2.9, "A Precaution").

Suppose a theorem turns out to be false in the real world. According to model theory, the only way a theorem can be false in the real world is if one or more of the axioms used to prove that theorem is false (assuming that the proofs are free of
logical errors). The deductive structure of an axiomatic theory permits one to locate a false axiom or collection of axioms simply by working back from a false theorem which was proved from that axiom or axioms. The beauty of an axiomatic theory is that even if some theorem turns out to be false, then it is possible to rescue value from the axiomatic theory by locating the false axiom and changing it.

2.5 Flow of Reasoning in this Dissertation

Theorems are proved in this dissertation for three purposes: (i) to indicate the correspondence of the axiomatic theory to known situations in the real world, (ii) to prove the formal validity of the several computer phenogram algorithms under appropriate hypotheses of divergent evolution (the ultimate purpose of this dissertation), or (iii) to act as stepping stones toward the proof of theorems with purposes (i) or (ii). A great many theorems fall into class (iii), and thus are difficult to justify or explain intuitively at the time they are being proved. I have endeavored to place such theorems where they fit best, but often this is a rather lame effort. The least intuitive theorems are called "lemmas", and are proved immediately before they are used with a remark to the reader that their content may be obscure. Appendix 11.2 provides a map of the entire flow of reasoning employed in this dissertation.
2.6 Informal Set Theory

What is set theory? It is just that—the theory of "sets" or "bunches" or "collections" of objects, no matter what the objects may be. A bunch of grapes is a set. A pile of leaves is a set. A collection of coins is a set. However, there is no need for things to be organized into physically contiguous bunches to be considered a set. We can talk about the set of all automobiles in the United States, or the set of all red automobiles, or the set of all red 1959 Chevrolets. It is not likely that any of these sets will ever be assembled in a single, neatly-stacked bunch, yet we can still talk about them as "sets" in abstract discussions. We can even talk about sets like the "set of all married bachelors" or the "set of all female widowers". These sets are a little bit different than the previous ones, because they don't contain any members at all! Such sets are called "the empty set" or "the null set", and are just as valid for us to talk about as sets which contain very many members.

A convenient way to represent a set is as a line enclosure about the members of the set. Thus, the "set of the first two sons of Abraham" (see Figure 2.1) might be represented as follows:

Isaac

Ishmael
Source: Genesis, chapters 16-50.

Figure 5.1. The house of Abraham, the sons and grandsons of Abraham and their mothers if known.
These enclosure diagrams are called Euler-Venn diagrams. (Strictly speaking, we have not literally disinterred Isaac and Ishmael and drawn a line about them--instead, we have used their names. We shall employ this convention throughout the dissertation.) On the printed line, it is not very convenient to draw these enclosures, so we use "curly bracket" notation. The set of the first two sons of Abraham is the set, \{Isaac, Ishmael\}. Two conventions should be kept in mind when dealing with set notation: the order in which the members are listed doesn't matter, and repeated members don't count. Thus, the set of the first two sons of Abraham might equally well be represented as:

\{Ishmael, Isaac\}

or: \{Ishmael, Isaac, Isaac, Ishmael, Ishmael, Ishmael, Isaac\}

and it would still be set theoretically equivalent to our original notation.

For sets having lots of members, it isn't even very convenient to list all the members. For example, the "set of all living human beings" would be much too cumbersome to write out every time we needed it, even if we could get ahold of all those three billion names. Therefore, we use a notation called "definition by abstraction". The set of all living human beings is written:

\{x : x is a living human being\}
and is read "the set of all x such that x is a living human being". The letter x is called a "dummy variable", because it has no significance other than as an internal defining device. We could equally well use a building block:

\[
\{x : x \text{ is a living human being}\}
\]

or a Chinese character:

\[
\{\varphi: \varphi \text{ is a living human being}\}
\]

and we would still be talking about exactly the same set. Definition by abstraction is especially convenient for infinite sets, where the curly bracket notation would be actually impossible. For example:

\[
\{x : x \text{ is an even number}\}
\]

Finally, we can always give name to particular sets in order to avoid having continually to rewrite entire sets, either in curly bracket or definition by abstraction notation. For example, we set up the following shorthand for use in this chapter:

\[
U = \{x : x \text{ is Abraham or a descendant of Abraham}\}
\]

\[
A = \{\text{Abraham}\}
\]

\[
B = \{\text{Isaac, Jacob, Esau}\}
\]

\[
C = \{x : x \text{ is Isaac or a son of Isaac}\}
\]

\[
D = \{x : x \text{ is a descendant of Abraham}\}
\]

\[
E = \{\text{Abraham, Isaac, Jacob}\}
\]
2.7 Set Relationships

Figure 2.2 shows each of the sets U, A, B, C, D, and E in the Euler-Venn notation. These sets are not unrelated to one another: two of them are exactly the same, and almost all of them share one or more elements in common. Set U is called the "universe", because it contains all the members which we will be dealing with in our discussion of the sets A, B, C, D, and E.

The simplest operation in set theory is the membership operation. We say, for example, that "Isaac is a member of set B". The notation for this operation is:

Isaac ∈ B.

The opposite of membership is nonmembership. We say "Abraham is not a member of the set B", and employ the notation:

Abraham ∉ B.

Two sets are equal is they have exactly the same members. For example, set B equals set C, denoted:

B = C,

because every member of set B is a member of set C, and every member of set C is a member of set B. One set is a subset of another if every member of the first is also a member of the second. For example, set B is a subset of set D, denoted:

B ⊆ D,
Figure 2.2. Sets: Abraham and his descendants. Set $U$ is the set of Abraham and all his descendants. Sets $A$, $B$, $C$, $D$, and $E$ are subsets of $U$. Dots (...) indicate additional members which are not named explicitly.
because every member of set B is also a member of set D. It is also true that:

\[ B \subseteq C, \]

because every member of set B is also a member of set C. The relationship between sets B and D is a special subset relation, namely, is a proper subset of. We say that set B is a proper subset of set D, denoted:

\[ B \subset D, \]

because every member of set B is also a member of set D, and there is at least one member of set D which is not a member of set B.

Clearly, it is not true that \( B \subseteq C \); every member of set B is also a member of set C, but there is no member of set C which is not also a member of set B.

Throughout this dissertation, we shall use \( N(\cdot) \) to denote the cardinality, or number of members, of the set in parentheses. For example:

\[ N(A) = 1 \]
\[ N(B) = 3 \]

and
\[ N(\emptyset) = 0. \]

2.8 Special Sets

The set A is given a special name, because it has only one member: it is called a singleton. Any member of the universe
can be made into a singleton simply by enclosing it in curly brackets. Here are some of the singletons which can be created from our universe, U:

{Abraham}
{Isaac}
{Ishmael}
{Jacob}
{Esau}

A singleton is not the same thing as the single member it contains. For example, it is not true that Abraham is the same as {Abraham}. In fact, each of the following singletons are different from one another:

{Abraham}
{{Abraham}}
{ {{Abraham}} }
{ {{{Abraham}} } }

The first set is "the set of Abraham"; the second set is "the set of the set of Abraham"; etc.

One of the most important sets in set theory has no members at all. It is called the empty set, or the null set, and is denoted by either of the following notations:

{} or φ.
In set theory, all null sets are equal (because they all have exactly the same members, namely, no members at all). Some examples of the null set are:

\[
\{ x : x \text{ is a married bachelor} \} \\
\{ x : x \text{ is a female widower} \} \\
\{ x : x \text{ is a man who walked on the moon in 1900 A.D.} \}
\]

2.9 A Precaution

One of the hazards of the English language is a very imprecise little word: the word "in". This word is probably responsible for the infinite frustrations of the beginning student of set theory, and may even be responsible for the inability of taxonomists to apply set theory to their discipline (for example: Gregg's (1950) paradox). The following are all true statements:

<table>
<thead>
<tr>
<th>Formal Notation</th>
<th>Correct Reading</th>
<th>Imprecise Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isaac ∈ B</td>
<td>Isaac is a member of set B</td>
<td>Isaac is &quot;in&quot; B</td>
</tr>
<tr>
<td>{Isaac} ⊆ B</td>
<td>The set of Isaac is a subset of set B</td>
<td>Isaac is &quot;in&quot; B</td>
</tr>
<tr>
<td>{Isaac} ⊄ B</td>
<td>The set of Isaac is not a member of set B</td>
<td>Isaac is not &quot;in&quot; B</td>
</tr>
</tbody>
</table>

That is, Isaac is a member of set B, but \{Isaac\} is not. The set B has exactly three members: Isaac, Jacob, and Esau. \{Isaac\} is not the same thing as Isaac.
2.10 Set Operations

Set theory has a variety of operations which can be used to create new sets of desired properties. The union of two sets, denoted by the symbol $\cup$, is the set of members present in either of the original sets. For example (Figure 2.3):

read, "set A union set B" $A \cup B = \{\text{Abraham, Isaac, Jacob, Esau}\}$

read, "set C union set E" $C \cup E = \{\text{Abraham, Isaac, Isaac, Jacob, Jacob, Esau}\}$

$= \{\text{Abraham, Isaac, Jacob, Esau}\}$.

The intersection of two sets, denoted by the symbol $\cap$, is the set of members common to both of the original sets. For example (Figure 2.4):

read, "set C intersection set E" $C \cap E = \{\text{Isaac, Jacob}\}$

read, "set A intersection set B" $A \cap B = \{\}$ $= \emptyset$.

The differences between two sets, denoted by the symbol $-$, is the set of members present in the first but absent in the second. For example (Figure 2.5):

read, "set C minus set E" $C - E = \{\text{Esau}\}$

read, "set E minus set C" $E - C = \{\text{Abraham}\}$
Figure 2.3. The union of two sets. Set $A \cup B$ is the union of sets $A$ and $B$ of Figure 2.2. Set $C \cup E$ is the union of sets $C$ and $E$ of Figure 2.2. The sets $A \cup B$ and $C \cup E$ are highlighted by diagonal hatching.
Figure 2.4. Set $A \cap B$ is the intersection of sets $A$ and $B$ of Figure 2.2

Set $C \cap E$ is the intersection of sets $C$ and $E$ of Figure 2.2

The sets $A \cap B$ and $C \cap E$ are highlighted by diagonal hatching. Note that $A \cap B$ is the null set.
Figure 2.5. The difference between two sets. Set $C - E$ is set $C$ minus set $E$, where sets $C$ and $E$ are from Figure 2.2. Set $E - C$ is set $E$ minus set $C$. The sets $C - E$ and $E - C$ are highlighted by diagonal hatching. Note that these two sets are **not** the same.
Notice that "the first set minus the second set" is not the same as "the second set minus the first set". Unlike the previous two set operations, order is important for set difference. Set substitution is not the same as arithmetic subtraction.

2.11 Mathematical Relations

Another device we shall make extensive use of in this dissertation is the mathematical relation. Intuitively, a mathematical relation is a "relationship" shared by two objects. Let \( R \) be a relation and \( x \) and \( y \) be objects. Then the statement "\( x R y \)" means "object \( x \) has the relationship \( R \) with object \( y \)". We can illustrate this concept with an example from the next chapter of the dissertation. The symbol "\( \triangleright \)" is a mathematical relation signifying "is a parent of". If \( x \) and \( y \) are organisms, then the statement "\( x \triangleright y \)" means, "\( x \) is a parent of \( y \)". Observe that "\( \triangleright \)" is an arrow-like symbol, with the pointed end closest to the younger organism.

Now let's get some practice with the \( \triangleright \)-relation. Figure 2.1 shows the House of Abraham. From this diagram, we can write:

Sarah \( \triangleright \) Isaac

That is, \( x \triangleright y \) where \( x = \text{Sarah} \) and \( y = \text{Isaac} \). In this particular example, there is exactly one \( y \) such that Sarah \( \triangleright y \). That is, Sarah is a parent of exactly one organism. However, it is not
universally true that each organism \( x \) has exactly one \( y \) such that \( x \triangleright y \). That is, not every organism is a parent to exactly one organism. For example, let \( x = \) Abraham. Then:

- Abraham \( \triangleright \) Ishmael
- Abraham \( \triangleright \) Isaac
- Abraham \( \triangleright \) Zimran

etc.

However, the \( \triangleright \)-relation has at least one universal property: no organism is a parent of itself. This is the first axiom of Williams' (1970) biocosm and of this dissertation. That is:

For each organism \( x \), it is not true that \( x \triangleright x \).

The symbol "\( \neg \)" means "it is not true that". Thus,

For each organism \( x \), \( \neg (x \triangleright x) \).

For example:

- \( \neg (\text{Abraham} \triangleright \text{Abraham}) \)
- \( \neg (\text{Oedipus} \triangleright \text{Oedipus}) \).

The \( \triangleright \)-relation is not the only relation which will be used in this dissertation. We shall employ several additional relations, but they will always be explained when they are first introduced. All terminology is collected in Appendix II.1.
3. **PRINCIPLES OF GENEALOGY**

3.1 First Principles

The first principles of genealogy were known to the ancients and are common knowledge to biologists and educated laymen. This chapter develops these first principles in the formalized terminology of this dissertation. This chapter may appear to belabor the obvious with cumbersome mathematical notation, but this step is a necessary prelude to later chapters, in which the axiomatic counterparts to these "obvious" intuitive notions are used without further reference to their intuitive content. That is, we shall depend upon intuition to generate the axioms, and then depend only on the logical relations between these axioms in subsequent chapters.

In this dissertation, I shall employ Williams' (1970) biocosm, which is a simple, logical model that allows us to handle genealogies in an abstract framework. Her biocosm consists of two primitives, B and >, and two axioms (stated below), and is denoted by the shorthand, \( (B, >) \). The primitive set B is usually interpreted as the set of all organisms (which may be living now, or have lived, or shall live), and the primitive mathematical relation > is usually interpreted as the relationship is a parent of:

- **Primitive** B  The set of all organisms
- **Primitive** >  Is a parent of
For example, the statement "x is a member of B" means "x is a member of the set of all organisms", or simply, "x is an organism". Similarly, the statement "x > y" means "x is a parent of y".

The primitives B and > are not operational concepts. In other words, it is not necessary that the biologist have observed (or even be able to observe) the set of all organisms, nor that he know all possible parent relations. Rather, B and > must "make sense" throughout the entire scope of the biologist's investigation, in the sense that statements of the form "x is an organism" or "organism x is a parent of organism y" is either true or false, regardless of the biologist's immediate ability to verify this condition by experimental test. For example, if a biologist is not certain whether or not Axioms I and II are satisfied for viruses, then he should not apply the results of this dissertation to viruses.

Axioms I and II are the first two ("pre-Darwinian") axioms of Williams (1970) biocosm. Axiom I states that no organism is a parent of itself. That is, for any x a member of B, it is not true that x is a parent of x:

Axiom I Let x ∈ B. Then ~(x > x).

(Note: Any organism will be symbolized by a lower-case Roman letter, usually near the end of the alphabet.)
Using our primitive concept "is a parent of", we introduce our first derived concept, "is an ancestor of", which will be used in constructing Axiom II. Organism \( x \) is an ancestor of organism \( y \), denoted \( x \triangleright y \), if and only if (i) organism \( x \) is a parent of organism \( y \), or (ii) there is a sequence of organisms \( (z_1, \ldots, z_k) \) such that \( x \) is a parent of \( z_1 \), \( z_1 \) is a parent of \( z_2, \ldots \), and \( z_k \) is a parent of \( y \):

**Definition 3.1** Let \( x, y \in B \). Then \( x \) is an ancestor of \( y \), denoted \( x \triangleright y \), if and only if

(i) \( x \triangleright y \)

or

(ii) there exists a \( k \geq 1 \) and an ordered sequence of organisms \( (z_1, \ldots, z_k) \) such that \( x \triangleright z_1 \triangleright \ldots \triangleright z_k \triangleright y \).

Accordingly, we say that, by Definition 3.1 (i), Abraham is an ancestor of Isaac, denoted Abraham \( \triangleright \) Isaac, because Abraham \( \triangleright \) Isaac. However, it is also true that, by Definition 3.1 (ii), Abraham is an ancestor of Judah, denoted Abraham \( \triangleright \) Judah, because there exists an ordered sequence of organisms, namely (Isaac, Jacob), such that Abraham \( \triangleright \) Isaac \( \triangleright \) Jacob \( \triangleright \) Judah. Observe that is-an-ancestor-of requires only that there exist some sequence of organisms—it does not require that there is one and only one. There may be several possible sequences. For example, the statement Jocasta \( \triangleright \) Ismene can be equally well explained by the sequence Jocasta \( \triangleright \) Ismene or Jocasta \( \triangleright \) Oedipus \( \triangleright \) Ismene.
Axiom II states that if organism x is an ancestor of organism y, then organism y is not an ancestor of organism x:

**Axiom II** Let \( x, y \in B \). If \( x \succ y \), then \( \sim (y \succ x) \).

The next two theorems follow immediately. They are proved by Williams (1970). No organism is an ancestor of itself:

**Theorem 3.1** Let \( x \in B \). Then \( \sim (x \succ x) \).

If organism x is an ancestor of organism y, and organism y is an ancestor of organism z, then organism x is an ancestor of organism z:

**Theorem 3.2** Let \( x, y, z \in B \). If \( x \succ y \) and \( y \succ z \), then \( x \succ z \).

3.2 Ancestor or Equal To

There are numerous occasions in this dissertation in which we shall want to express the relationship that "organism x is either an ancestor of organism y or y itself". We already have a notation for "x is an ancestor of y", namely, \( x \succ y \). We employ the usual convention of set theory for expressing that "x is the same as y", namely, \( x = y \). (By convention, \( x = y \) is read "x equals y", but it means that "x is exactly the same organism as y".) The natural way to express a combination of these relationships (i.e. x is an ancestor of y or x is the same as y) is the symbol \( x \succ y \). We say that x is ancestor or equal to y, denoted \( x \succ y \), if and only if (i) x is an ancestor of y or (ii) x is the same as y:
Definition 3.2  Let \(x, y \in B\). Then \(x\) is ancestor or equal to \(y\), denoted \(x \triangleright y\), if and only if

1. \(x \triangleright y\)
2. \(x = y\)

Not surprisingly, each organism is ancestor or equal to itself (because it is the same as itself):

Theorem 3.3  Let \(x \in B\). Then \(x \triangleright x\).

Proof. Since \(x = x\), by Definition 3.2 (ii), \(x \triangleright x\).

If organism \(x\) is ancestor or equal to organism \(y\), and organism \(y\) is ancestor or equal to organism \(x\), then \(x\) is the same as \(y\):

Theorem 3.4  Let \(x, y \in B\). If \(x \triangleright y\) and \(y \triangleright x\), then \(x = y\).

Proof. By Definition 3.2, \(x \triangleright y\) implies that either \(x \triangleright y\) or \(x = y\). Assume the theorem is false. That is, \(x \triangleright y\). By Definition 3.2, either \(y \triangleright x\) or \(y = x\). Suppose \(y \triangleright x\). Contradiction of Axiom II. Suppose \(y = x\). Contradiction of Theorem 3.1.

If organism \(x\) is ancestor or equal to organism \(y\) and organism \(y\) is ancestor or equal to organism \(z\), then \(x\) is ancestor or equal to \(z\):
Theorem 3.3  Let $x, y, z \in B$. If $x \triangleright y$ and $y \triangleright z$, then $x \triangleright z$.

Proof. By Definition 3.2, either $x \triangleright y$ or $x = y$. Suppose $x = y$. then $x = y \triangleright z$ and $x \triangleright z$. Alternatively, suppose $x \triangleright y$. By Definition 3.2, either $y \triangleright z$ or $y = z$. Suppose $y = z$. Then $x \triangleright y = z$ and $x \triangleright z$. By Definition 3.2, $x \triangleright z$. Alternatively, suppose $y \triangleright z$. Then by Theorem 3.2, $x \triangleright z$. By Definition 3.2, $x \triangleright z$.

3.3 Pedigrees

The pedigree of a set of organisms is the set containing all the original organisms in the set and all their ancestors (Woodger, 1952, p. 247). Let $X$ be the original set of organisms and $Pd(X)$ be the pedigree of $X$. (Note: Any set of organisms will be symbolized by an upper-case Roman letter, usually near the end of the alphabet.) Then $Pd(X)$ is the set of all organisms $w$ such that $w$ is ancestor or equal to some member of $X$. We use definition by abstraction (Section 2.6) to define $Pd(X)$:

Definition 3.3  Let $\emptyset \neq X \subseteq B$. Then the pedigree of $X$, denoted $Pd(X)$, is the set

$$Pd(X) = \{w : w \in B \text{ and there exists an } x \in X \text{ such that } w \triangleright x\}$$

Referring once again to the House of Abraham (Figure 2.1), we observe that the pedigree (Figure 3.1) of the singleton $\{Isaac\}$ contains Isaac himself, Abraham, Sarah, and additional individuals...
Figure 3.1. Pedigrees. The pedigree of \{Isaac\}, denoted \text{Pd}(\{Isaac\}), is the set of Isaac and all his ancestors. The pedigree of \{Ishmael, Isaac\}, denoted \text{Pd}(\{Ishmael, Isaac\}), is the set of Ishmael, Isaac, and any ancestor of either Ishmael or Isaac. Dots (\ldots) indicate additional members which are not named explicitly.
not included in Figure 2.1 and not mentioned by name in Figure 3.1. The pedigree (Figure 3.1) of the set \{Ishmael, Isaac\} contains Ishmael himself, Isaac himself, Abraham, Sarah, Hagar, and additional individuals not included in Figure 2.1.

The pedigree concept satisfied the following two properties.

Any set $X$ is a subset of its own pedigree:

**Theorem 3.6** Let $\emptyset \neq X \subseteq B$. Then $X \subseteq \mathsf{Pd}(X)$.

**Proof.** Let $x \in X$. Since by Theorem 3.3, $x \not\succ x$, it follows that $x \in \mathsf{Pd}(X)$.

If $Y$ is a subset of $X$, then the pedigree of $Y$ is a subset of the pedigree of $X$:

**Theorem 3.7** Let $\emptyset \neq Y \subseteq X \subseteq B$. Then $\mathsf{Pd}(Y) \subseteq \mathsf{Pd}(X)$.

**Proof.** Let $z \in \mathsf{Pd}(Y)$. Then there is a $y \in Y$ such that $z \not\succ y$. By hypothesis, $Y \subseteq X$, hence $y \in X$. Since $z \not\succ y$, by Definition 3.3, $z \in \mathsf{Pd}(X)$. Hence $\mathsf{Pd}(Y) \subseteq \mathsf{Pd}(X)$.

### 3.4 Common Pedigrees

The **common pedigree** of a set of organisms $X$ is the set of organisms which are ancestor or equal to every organism in the original set. If organism $w$ is ancestor or equal to some, but not all, members of $X$, then $w$ is **not** in the common pedigree of $X$. We use the notation $\mathsf{CP}(X)$ to indicate the common pedigree of $X$. 
Definition 3.4 Let $\phi \neq X \subseteq B$. Then the common pedigree of $X$, denoted $\text{CP}(X)$, is the set

$$\text{CP}(X) = \{ w : w \in B \text{ and for every } x \in X, w \supseteq x \}$$

Intuitively, the "common pedigree of $X" is the same as what a biologist means by the "set of common ancestors of $X"; the slight distinction in the precise meaning of "common pedigree" from the biological notion of "common ancestors" is made for convenience in the mathematical proofs, and does not affect the major applications of this dissertation.

From the House of Abraham (Figure 2.1), we observe that the common pedigree (Figure 3.2) of the set \{Isaac, Ishmael\} contains Abraham and additional individuals not included in Figure 2.1 (i.e., all of Abraham's ancestors), but does not contain Sarah or Hagar.

The common pedigree of a nonempty set of organisms is a subset of the pedigree of that set of organisms:

Theorem 3.8 Let $\phi \neq X \subseteq B$. Then $\text{CP}(X) = \text{Pd}(X)$.

Proof. Let $w \in \text{CP}(X)$. By Definition 3.4, for every $x \in X$, $w \supseteq x$. Since $X \neq \phi$, there exists such an $x$. By Definition 3.3, $w \in \text{Pd}(X)$. Hence $\text{CP}(X) \subseteq \text{Pd}(X)$.

The pedigree of the common pedigree of a set of organisms $X$ is simply the common pedigree of $X$: 
Figure 3.2. Common and most recent common pedigrees. The common pedigree of \{Ishmael, Isaac\}, denoted \( \text{CP}\{\text{Ishmael, Isaac}\} \), is the set of all mutual ancestor-or-equals of Ishmael and Isaac. Dots (...) indicate additional members which are not named explicitly. The most recent common pedigree of \{Ishmael, Isaac\}, denoted \( \text{RP}\{\text{Ishmael, Isaac}\} \), is the set of only those members of the common pedigree which have no descendants in the common pedigree.
Theorem 3.9  Let \( \emptyset \neq X \subseteq B \). Then \( CP(X) = Pa(CP(X)) \).

Proof. By Theorem 3.8, \( CP(X) \subseteq Pa(CP(X)) \). Assume the theorem is false. That is, \( CP(X) \subsetneq Pa(CP(X)) \). Let \( z \in Pa(CP(X)) - CP(X) \). By Definition 3.3, there exists a \( y \in CP(X) \) such that \( z \triangleright y \). By Definition 3.4, for each \( x \in X, y \triangleright x \). By Theorem 3.5, for each \( x \in X, z \triangleright y \triangleright x \), or \( z \triangleright x \); hence \( z \in CP(X) \). Contradiction.

In Theorem 3.7, we observed that if organism set \( Y \) is a subset of organism set \( X \), then the pedigree of \( Y \) is a subset of the pedigree of \( Y \). That is, as a set of organisms gets "smaller", so does its pedigree. This is because an organism \( w \) need only be ancestor or equal to any member of \( X \) to qualify for the pedigree of \( X \). In Theorem 3.10, we observed that the opposite is true for the common pedigree of \( X \). An organism \( w \) must be ancestor or equal to every member of \( X \) to qualify for the common pedigree of \( X \). Accordingly, as a set of organisms gets "smaller", its common pedigree may get "larger". That is, if \( Y \) is a subset of \( X \), then the common pedigree of \( X \) is a subset of the common pedigree of \( Y \):

Theorem 3.10  Let \( \emptyset \neq Y \subseteq X \subseteq B \). Then \( CP(X) \subseteq CP(Y) \).

Proof. Consider any \( w \in CP(X) \). By Definition 3.4, for each \( x \in X, w \triangleright x \). In particular, for each \( y \in Y \subseteq X, w \triangleright y \). By Definition 3.4, \( w \in CP(Y) \). Hence \( CP(X) \subseteq CP(Y) \).
The common pedigree concept is too broad for many phyletic purposes. Consider, for example, the set $X$, where $X$ is the set of organisms comprising all members of the modern-day *Equus* genus and all members of the extinct *Hipparion* genus (Figure 3.3, adapted from Bastian (1964, p. 231)). The common pedigree of $X$ is the set of organisms which are ancestor or equal to every *Equus* and every *Hipparion*. Some members of *Eohippus*, as well as some members of *Merychippus* will be members of $\text{CP}(X)$. (Of course, not all members of either *Eohippus* or *Merychippus* will be members of $\text{CP}(X)$, because some members of *Eohippus* and *Merychippus* have not left descendants comprising all of $X$.) Often an evolutionary taxonomist makes reference to the immediate branching point (in this example, *Merychippus*) of two diverging lines of organisms without including prior members (*Eohippus*). This intuitive concept corresponds to the most recent common pedigree concept of this dissertation. The most recent common pedigree of a set of organisms, $X$, denoted $\text{RP}(X)$, is the set of organisms which are members of the common pedigree of the original set, but which have not had descendants which are also members of the common pedigree of $X$:

**Definition 3.5** Let $\emptyset \neq X \subseteq B$. Then the most recent common pedigree of $X$, denoted $\text{RP}(X)$, is the set

$$\text{RP}(X) = \{w : w \in \text{CP}(X) \text{ and there exists no } z \in \text{CP}(X) \text{ such that } w \geq z\}.$$
Figure 3.3. Evolution of horses. Simplified diagram of the evolution of the now-extinct *Hipparion* and the contemporary *Equus*. Source: Bastian (1964, p. 231).
The remaining theorems in this section elucidate the relationships between CP(X) and RP(X). For any nonempty X, RP(X) is a subset of CP(X):

**Theorem 3.11** Let $\phi \neq X \subseteq B$. Then $RP(X) \subseteq CP(X)$.

**Proof.** Consider any $w \in RP(X)$. By Definition 3.5, $w \in CP(X)$.
Hence $RP(X) \subseteq CP(X)$.

The proofs of Theorems 3.14 and 3.15 require two preliminary Lemmas (3.12 and 3.13) which do not have an intuitive content for biologists, as well as the use of the Zorn Lemma (Kelley, 1955), which is questioned by some mathematicians. This need not worry the biologist, because the Zorn Lemma is questioned only as it applies to infinite sets (in this case, infinite sets of organisms), and of course no set of organisms is infinite in the biological real world. Theorems 3.14 and 3.15 are proved in such a way that the number of organisms in set X is unspecified—it can be either finite or infinite. The only reason that I have bothered to prove Theorems 3.14 and 3.15 in the general case is because many models of population genetics presuppose an infinite population, and I seek to build a mathematical model which can readily be meshed with a variety of existing models. Thus, Theorems 3.14 and 3.15 apply both to biological reality and to existing models of population genetics which presuppose an infinite population.
Lemma 3.12  Let $\phi \not= X \subseteq B$, $y \in CP(X)$, $x \in X$, and $y \triangleright x$. Then there exists a $k \geq 0$ and a set $T = \{y = z_0, \ldots, z_{k+1} = x\}$ such that for $0 \leq i \leq k$, $z_i > z_{i+1}$ and $T \subseteq B$. Furthermore, $(T \cap CP(X), \triangleright)$ is a chain sensu Kelley (1955, p. 15), and $T \cap CP(X)$ has an upper bound sensu Kelley (1955, p. 13), $w \in T \cap CP(X)$.


Part II.  (Adapted from Williams (1970).) Consider $v_1, v_2, v_3 \in T \cap CP(X) \subseteq B$. If $v_1 \triangleright v_2$ and $v_2 \triangleright v_3$, then by Theorem 3.2, $v_1 \triangleright v_3$. Thus, $(T \cap CP(X), \triangleright)$ is a partial ordering sensu Kelley (1955, p. 13).

Part III.  (a) Choose any $z_i, z_j \in T \cap CP(X)$ such that $z_i \triangleright z_j$ and $z_j \triangleright z_i$. By Theorem 3.4, $z_i = z_j$.

(b) Choose any $z_i, z_j \in T \cap CP(X)$. Suppose $\sim (z_i \triangleright z_j)$. By Definition 3.2 (i), $z_i \not= z_j$ and by construction, $i \not= j$. By Definition 3.2 (ii), $\sim (z_i \triangleright z_j)$, and by Definition 3.1, there is no ordered sequence of organisms $(w_1, \ldots, w_\ell)$ such that $z_i \triangleright w_1 \triangleright \ldots \triangleright w_\ell > z_j$. Hence $i \not< j$. By the Law of Trichotomy for integers, $i > j$ and by construction, $z_j \triangleright z_{j+1} \triangleright \ldots \triangleright z_i$. By Definition 3.1 (ii), $z_j \triangleright z_i$. By Definition 3.2 (ii), $z_j \triangleright z_i$. Thus $\triangleright$ is a linear ordering for the set $T \cap CP(X)$ sensu Kelley (1955, p. 14).
Part IV. Since $\mathfrak{p}$ is a linear ordering for set $T \cap CP(X)$, $(T \cap CP(X), \mathfrak{p})$ is a chain sensu Kelley (1955, p. 15).

Part V. Choose the largest $h$, $0 \leq h \leq k+1$, such that $z_h \in T \cap CP(X)$. There must be such an $h$, because by hypothesis, $z_0 = y \in T \cap CP(X)$. Claim: $z_h = w$ is the upper bound of $T \cap CP(X)$.

Consider any $z_j \in T \cap CP(X)$. If $j > h$, then $z_j \notin T \cap CP(X)$, because $h$ is the largest integer such that $w = z_h \in T \cap CP(X)$. If $j = h$, then $z_j = w$. If $j < h$, then there is an ordered sequence

$$(z_j, z_{j+1}, \ldots, z_h = w)$$

such that for each $i$, $j \leq i \leq h-1$, $z_i > z_{i+1}$.

By Definition 3.1 (ii), $z_j \triangleright w$. By Definition 3.2 (ii), $z_j \triangleright w$.

Thus $w$ is an upper bound of $T \cap CP(X)$, sensu Kelley (1955, p. 13).

Lemma 3.13 Let $\phi \neq X = B$ and $CP(X) \neq \phi$. Fix $y \in CP(X)$ and consider any $x \in X$. Then $y \triangleright x$, as guaranteed in Definition 3.4. Let $k \geq -1$ and $(y = z_0, \ldots, z_{k+1} = x)$ be an ordered sequence of organisms such that if $k \geq 0$ and $0 \leq i \leq k$, $z_i > z_{i+1}$, as guaranteed in Definition 3.1. Let $T = \{z_0, \ldots, z_{k+1}\}$ and $R = T \cap CP(X)$. Let $\mathfrak{R}$ be the set of all such $R$ for every $x \in X$ and all possible ordered sequences.

Then the upper bound sensu Lemma 3.12 for each maximal element sensu Kelley (1955, p. 32) $R_{\max} \in \mathfrak{R}$ is a member of RP($X$).

Proof. Let $R_{\max} \in \mathfrak{R}$ be a maximal element of $\mathfrak{R}$ sensu Kelley (1955, p. 32). Then there is no $R \in \mathfrak{R}$ such that $R_{\max} \subset R$. Consider any
$R \in R$. By construction, $CP(X)$ and $T$ have the sense of Lemma 3.12.

By Lemma 3.12, $(R, \succ)$ is a chain and there is an upper bound, $w \in R$. Similarly, for the special case of $R_{max} \in R$, there is an upper bound

$w_{max} \in R_{max}$. Clearly, $w_{max} \in R_{max} \subseteq CP(X)$. By Definition 3.5, $w_{max} \in RP(X)$ unless there is a $v \in CP(X)$ such that $w_{max} \succ v$. Suppose there is such a $v \in CP(X)$. Locate an $x_{max} \in X$ which was used to construct $R_{max}$. Since $y \in R_{max}$ and $w_{max}$ is an upper bound of $R_{max}$, $y \succ w_{max}$. Let $k \geq -1$ and $(y = z_0, \ldots, z_{k+1} = w_{max})$ be the ordered sequence of organisms constructed for $R_{max}$. Let $\ell \geq k+2$ and

$(w_{max} = z_{k+1}, \ldots, z_{\ell} = v)$ be an ordered sequence of organisms such that if $k+1 \leq i \leq \ell-1, z_i \succ z_{i+1}$, guaranteed by Definition 3.1.

Since $v \in CP(X)$, by Definition 3.4, $v \succ x_{max}$. Let $m \geq \ell$ and

$(v = z_{\ell}, \ldots, z_m = x)$ be an ordered sequence of organisms such that if $\ell \leq i \leq m-1, z_i \succ z_{i+1}$, as guaranteed by Definition 3.1. Let

$T = \{z_0, \ldots, z_m\}$ and $R = T \cap CP(X)$. By construction, $R_{max} \subseteq R$ and $v \in R$. Suppose $v \in R_{max}$. Since $w_{max}$ is an upper bound of $R_{max}$,

$v \succ w_{max}$. By Definition 3.2, either $v = w_{max}$ or $v \succ w_{max}$. We already supposed that $w_{max} \succ v$. But $v = w_{max} \succ v$ contradicts Theorem 3.1 and $w_{max} \succ v$ contradicts Axiom II. Hence $v \notin R_{max}$ and $R_{max} \subseteq R$.

But then $R_{max}$ is not a maximal element. Contradiction.
Theorem 3.14 states that the common pedigree of a set of organisms is empty if and only if the most recent common pedigree is empty:

**Theorem 3.14** Let $\emptyset \neq X \subseteq B$. Then $CP(X) = \emptyset$ if and only if $RP(X) \neq \emptyset$.

**Proof.** Only If. Suppose $CP(X) = \emptyset$. By Theorem 3.10, $RP(X) \subseteq CP(X) = \emptyset$. From elementary set theory, the only subset of $\emptyset$ is $\emptyset$. Hence $RP(X) = \emptyset$.

If. Let $CP(X) \neq \emptyset$. Construct the set $\mathcal{R}$ sensu Lemma 3.13. By Lemma 3.12, each chain $R \in \mathcal{R}$ has an upper bound. By the Zorn Lemma sensu Kelley (1955, p. 33), there is a maximal element, $R_{\max} \in \mathcal{R}$. Let $w_{\max} \in R_{\max}$ be the upper bound guaranteed by Lemma 3.12. By Lemma 3.13, $w_{\max} \in RP(X)$. Hence $RP(X) \neq \emptyset$.

Theorem 3.15 states that the pedigree of the most recent common pedigree of $X$ is exactly the common pedigree of $X$:

**Theorem 3.15** Let $\emptyset \neq X \subseteq B$ and $RP(X) \neq \emptyset$. Then $Pd(RP(X)) = CP(X)$.

**Proof.** By hypothesis, $RP(X) \neq \emptyset$. By Theorem 3.11, $RP(X) \subseteq CP(X)$. By Theorem 3.7, $Pd(RP(X)) \subseteq Pd(CP(X))$. By Theorem 3.9, $Pd(CP(X)) = CP(X)$. Hence $Pd(RP(X)) = CP(X)$. Assume the theorem is false. That is, $Pd(RP(X)) \subset CP(X)$. Choose any $y \in CP(X)$ such
that \( y \not\in Pd(RP(X)) \). Let \( R \) be the set as constructed in Lemma 3.13, and find \( w_{\text{max}} \in R_{\text{max}} \) as in the proof of Theorem 3.14.

By Lemma 3.13, \( w_{\text{max}} \in RP(X) \). Since \( w_{\text{max}} \) is the upper bound of \( R_{\text{max}} \) and \( y \in R_{\text{max}}, y \triangleright w_{\text{max}} \). Since \( y \not\in Pd(RP(X)) \), there is no member \( w \in RP(X) \) such that \( y \triangleright w \). In particular, \( \sim (y \triangleright w_{\text{max}}) \). Contradiction.

Theorem 3.16 states that the common pedigrees of two sets of organisms, \( Y \) and \( Z \), are equal if and only if the most recent common pedigrees are equal:

**Theorem 3.16** Let \( \phi \neq Y \subseteq B \) and \( \phi \neq Z \subseteq B \). Then \( CP(Y) = CP(Z) \) if and only if \( RP(Y) = RP(Z) \).

**Proof.** Case I. Suppose \( CP(Y) = CP(Z) = \phi \). Then by Theorem 3.14, \( RP(Y) = RP(Z) = \phi \).

Case II. Suppose \( CP(Y) = CP(Z) \neq \phi \). If \( CP(Y) = CP(Z) \), then by Definition 3.14, \( RP(Y) = RP(Z) \). If \( RP(Y) = RP(Z) \), then by Theorem 3.15, \( CP(Y) = Pd(RP(Y)) = Pd(RP(Z)) = CP(Z) \).

### 3.5 The Monoparental Biocosm

A monoparental biocosm is a biocosm each of whose members has at most one parent. Our special interest in the properties of a monoparental biocosm arises from the fact that a cladogram itself has the overall appearance of a monoparental genealogy. Biologists use this analogy informally when speaking of one species as the ancestor of another species; I shall use it formally by showing
that the properties of monoparental organisms can be used to
describe the overall structure of a cladogram, that is, ancestry
relations among sets of organisms, even though all the individual
organisms within these sets may be biparental.

For the time being, we shall think in terms of monoparental organisms, even though we shall later apply concepts so developed to sets of organisms. A collection of asexually reproducing organisms is a monoparental biocosm. A collection of sexually interbreeding (i.e., biparental) organisms for which all events of female parentage are ignored (for example: inheritance of surnames or y chromosomes) is also a monoparental biocosm. We define a monoparental biocosm as any set of organisms such that each organism possesses at most one parent:

**Definition 3.6** Let \( (B, \succ) \) be a biocosm. Then \( (B, \succ) \) is monoparental if and only if for each \( x \in B \) there is at most one \( z \in B \) such that \( z \succ x \).

A monoparental biocosm has several special properties which we do not ordinarily associate with an arbitrary biocosm. Theorem 3.17 states that for two distinct organisms which are both ancestor or equal to a third, one of the two original organisms must be ancestor or equal to the other:
Theorem 3.17  Let \((B, \triangleright)\) be a monoparental biocosm and \(x, y, z \in B\).

If \(y \triangleright x\) and \(z \triangleright x\), then:

(i) \(z \triangleright y\)

(ii) \(z = y\)

or

(iii) \(y \triangleright z\).

Proof. Part I. Preliminaries. Since \(y \triangleright x\) and \(z \triangleright x\), by Definition 3.1, there are \(h \geq 0\) and \(k \geq 0\) and ordered sets of organisms

\((y = y_h, \ldots, y_0 = x)\) and \((z = z_k, \ldots, z_0 = x)\) such that

\(y_h \triangleright \ldots \triangleright y_0\) and \(z_k \triangleright \ldots \triangleright z_0\).

Part II. Suppose \(h \geq k\). Claim: for each \(i, 0 \leq i \leq k\),

\(y_i = z_i\). Proof by mathematical induction.

True for \(i = 0\): \(y_0 = x = z_0\).

True for \(i\) implies true for \(i+1\): By Definition 3.6

\(y_i = z_i\) has an identical parent, \(y_{i+1} = z_{i+1}\).

By Definition 3.1, \(y = y_h \triangleright y_k = z_k = z\).

Part III. Suppose \(k \geq h\). By a similar argument, \(z \triangleright y\).

Each collection of organisms in a monoparental biocosm has at most one member in its most recent common pedigree:

Theorem 3.18  Let \((B, \triangleright)\) be a monoparental biocosm, let \(\emptyset \neq X \subseteq B\), and \(RP(X) \neq \emptyset\). Then \(RP(X)\) is a singleton set.

Proof. Assume \(RP(X)\) is not a singleton set. Then there exist \(w \in RP(X)\) and \(z \in RP(X)\) such that \(w \neq z\). Consider any \(y \in X\).
By Definition 3.4, $w \triangleright y$ and $z \triangleright y$. By Theorem 3.17, $w \triangleright z$, $w = z$, or $z \triangleright w$. By assumption, $w \neq z$. If $w \triangleright z$, then $RP(X)$ is not a most recent common pedigree by Definition 3.5.

Similar argument if $z \triangleright w$.

Suppose we have a monoparental biocosm, such as the set of male members of the House of Abraham. For the pair of organisms "Abraham and Esau", we say that Abraham is the patriarch of Abraham and Esau because Abraham is the member of the set \{Abraham, Esau\} which is ancestor or equal to every member of that set. However, neither Jacob nor Esau is the patriarch of the pair of organisms "Jacob and Esau", because neither is ancestor or equal to both members of the set \{Jacob, Esau\}. In this case, we say that Isaac is the patriarch of Jacob and Esau, because Isaac is the most recent organism which is ancestor or equal to both Jacob and Esau. In general, the patriarch of a pair of organisms is a member of the most recent common pedigree for that pair of organisms. In a monoparental biocosm, by Theorem 3.18 this member must be unique. Hence, there is at most one patriarch for any pair of organisms in a monoparental biocosm, and we can talk about the patriarch of a pair of organisms. If there is no patriarch for a pair of organisms, then we say that the patriarch does not exist.
Since the patriarch concept is used extensively in this dissertation, it is convenient to have a notation for patriarch: \( y \circ z \), read "y dot z", denotes the patriarch of \( y \) and \( z \). The notation \( y \circ z \) is meaningless if \( y \) and \( z \) have no patriarch (for example, Adam o Eve is meaningless). In general, if \( y \) and \( z \) are members of a monoparental biocosm with a nonempty most recent common pedigree, then the unique member of that most recent common pedigree is called the patriarch of \( y \) and \( z \), denoted \( y \circ z \):

**Definition 3.7** Let \((B, \rightarrow)\) be a monoparental biocosm, and \( y, z \in B \) and \( RP([y, z]) \neq \emptyset \). Then \( w \) is the patriarch of \( y \) and \( z \), denoted \( w = y \circ z \), if and only if \( \{w\} = RP([y, z]) \).

The patriarch of two organisms \( y \) and \( z \) is the same organism as the patriarch of \( z \) and \( y \). This property is called commutativity.

**Theorem 3.19** Let \((B, \rightarrow)\) be a monoparental biocosm, and \( y, z \in B \), and let \( y \circ z \) exist. Then \( y \circ z = z \circ y \).

**Proof.** By Definition 3.7, the existence of \( y \circ z \) guarantees that \( RP([y, z]) \neq \emptyset \). Let \( w \) be the unique member of \( RP([y, z]) \) guaranteed by Theorem 3.18. Since \( \{y, z\} \) is an unordered set, \( \{z, y\} = \{y, z\} \). Hence \( RP([z, y]) = RP([y, z]) \) and \( w \) is the unique member of \( RP([z, y]) \) guaranteed by Theorem 3.18. By Definition 3.6, \( y \circ z = w = z \circ y \).
Theorem 3.20 states that, in a monoparental biocosm, if some organism is ancestor or equal to any two organisms, then it is ancestor or equal to the patriarch of those two organisms:

**Theorem 3.20** Let \((B, \succ)\) be a monoparental biocosm and \(x, y, z \in B\). If \(x \succ y\) and \(x \succ z\), then \(x \succ y \circ z\).

**Proof.** Strategy. (I) Show that \(y \circ z\) exists. (II) Show that \(x \succ y \circ z\).

**Part I.** Claim: \(y \circ z\) exists. By hypothesis, \(x \succ y\) and \(x \succ z\). By Definition 3.14, \(x \in CP([y, z])\). By Theorem 3.14, \(RP([y, z]) \neq \emptyset\). By Theorem 3.18, there is a \(w\) such that \([w] = RP([y, z])\). By Definition 3.7, \(w = y \circ z\).

**Part II.** Claim: \(x \succ w\). By Definition 3.5, \(w \succ y\). By hypothesis, \(x \succ y\). By Theorem 3.17, either \(x \succ w\) or \(x = w\) or \(w \succ x\). Assume the theorem is false. That is, \(w \succ x\). From Part I, \(x \in CP([y, z])\). Since \(w \succ x\) and \(x \in CP([y, z])\), by Definition 3.5, \(w \notin RP([y, z])\). Contradiction.

Suppose we want the patriarch of three organisms, \(x, y,\) and \(z\) (assuming that all patriarchs we need exist). It isn't really valid to use the notation \(x \circ y \circ z\), because the \(\circ\) operation is only defined over pairs of organisms. However, we could obtain the patriarch of \(x\) and \(y\) and call it \(x^*_y = x \circ y\);
then we could obtain the patriarch of \( x_\star \) and \( z \). Alternatively, we could first obtain the patriarch of \( y \) and \( z \), call it \( z_\star = y \circ z \), and then obtain the patriarch of \( x \) and \( z_\star \). How do we know that these two approaches for obtaining the patriarch of three organisms will yield the same final result? Theorem 3.21 gives us this guarantee. This property is called associativity.

**Theorem 3.21** Let \( (B, \triangleright) \) be a monoparental biocosm, and \( x, y, z \in B \), and let \( x \circ (y \circ z) \) and \( (x \circ y) \circ z \) be defined. Then \( x \circ (y \circ z) = (x \circ y) \circ z \).

**Proof.** Strategy. (I) Show that \( x \circ (y \circ z) \triangleright (x \circ y) \circ z \) or \( (x \circ y) \circ z \triangleright x \circ (y \circ z) \). (II) Show that strict ancestry leads to a contradiction.

**Part I.** Let \( w_1 = y \circ z \) and \( w_2 = x \circ y \) and \( w_3 = x \circ w_1 \) and \( w_4 = w_2 \circ z \). By Definition 3.5 \( w_3 \triangleright x \), \( w_3 \triangleright w_1 \), \( w_1 \triangleright y \), and \( w_1 \triangleright z \). By transitivity (Theorem 3.2), \( w_3 \triangleright x \), \( w_3 \triangleright y \), and \( w_3 \triangleright z \). By a similar argument, \( w_4 \triangleright x \), \( w_4 \triangleright y \), and \( w_4 \triangleright z \). By Theorem 3.17, either \( w_3 \triangleright w_4 \) or \( w_3 = w_4 \) or \( w_4 \triangleright w_3 \).

**Part II.** Assume the theorem is false. That is, \( w_3 \neq w_4 \).

Without loss of generality, let \( w_4 \triangleright w_3 \). Since \( w_3 \triangleright x \) and \( w_3 \triangleright y \), by Theorem 3.20 \( w_3 \triangleright w_2 \). Since \( w_3 \triangleright w_2 \) and \( w_3 \triangleright z \), by Theorem 3.20 \( w_3 \triangleright w_4 \). By assumption, \( w_3 \neq w_4 \); hence \( w_3 \triangleright w_4 \). Contradiction of Axiom II.
Suppose \( y \) and \( z \) are two organisms such that neither is ancestor or equal to the other, and suppose the patriarch of \( y \) and \( z \) is defined. If \( y \) is ancestor or equal to \( y_1 \) and \( z \) is ancestor or equal to \( z_1 \), then the patriarch of \( y_1 \) and \( z_1 \) is defined and the same as the patriarch of \( y \) and \( z \):

**Theorem 3.22**  Let \((B, \triangleright)\) be a monoparental biocosm and \( x, y, z \in B \) and \( x = y \circ z \) be defined and \( \sim (y \triangleright z) \) and \( \sim (z \triangleright y) \). If \( y \triangleright y_1 \) and \( z \triangleright z_1 \), then \( x = y_1 \circ z_1 \).

**Proof.** **Strategy.** (I) Show that \( y_1 \circ z_1 \) is defined. (II) Let \( w = y_1 \circ z_1 \). Show that \( w \triangleright x \) or \( x \triangleright w \). (III) Show that \( w \triangleright x \) leads to a contradiction. (IV) Show that \( x \triangleright w \) leads to a contradiction.

**Part I.** Claim: \( y_1 \circ z_1 \) exists. By Definition 3.4, \( x \triangleright y \) and \( x \triangleright z \). By hypothesis, \( y \triangleright y_1 \) and \( z \triangleright z_1 \). By transitivity (Theorem 3.2), \( x \triangleright y_1 \) and \( x \triangleright z_1 \). By Theorem 3.20, \( y_1 \circ z_1 \) exists and \( x \triangleright y_1 \circ z_1 \).

**Part II.** Let \( w = y_1 \circ z_1 \). Claim: \( w \triangleright x \) or \( x \triangleright w \). By Definition 3.4, \( w \triangleright y_1 \). By Part I, \( x \triangleright y_1 \). By Theorem 3.17, either \( w \triangleright x \), \( w = x \), or \( x \triangleright w \). By Definition 3.2, \( w \triangleright x \) or \( x \triangleright w \).

**Part III.** Suppose \( w \triangleright x \). By Definition 3.4, \( x \in CP([y_1, z_1]) \), so that by Definition 3.5, \( w \notin RP([y_1, z_1]) \). Contradiction.
Part IV. Suppose $x \triangleright w$. Substrategy: show that $w \triangleright y$.

By hypothesis, $y \triangleright y_1$. By Definition 3.4, $w \triangleright y_1$. By Theorem 3.17, $w \triangleright y$ or $y = w$ or $y \triangleright w$. Assume the contrary of the substrategy. That is, $y \triangleright w$. Then by transitivity, $y \triangleright z_1$.

By hypothesis, $z \triangleright z_1$. By Theorem 3.17, $y \triangleright z$, $y = z$, or $z \triangleright y$.

All three of these possibilities contradict the hypothesis.

End of substrategy. By a similar substrategy, $w \triangleright z$. By Definition 3.4, $w \in CP([y, z])$, so that by Definition 3.5, $x \notin RF([y, z])$. Contradiction.
4. OPERATIONAL TAXONOMIC UNITS AND ISOLATION

4.1 Operational Taxonomic Units

In the techniques of numerical taxonomy, one always begins with a set of ultimate units which the classification is supposed to classify. If one were constructing a simple family tree, these units might be single organisms; in constructing an ancestral diagram for a single species, the investigator might neither be willing nor even capable of drawing the intricate ancestral histories of each organism: he would be more likely to use populations as the ultimate units of his classification. An investigator classifying a family or particularly diverge genus of organisms might use species as his ultimate units. And so forth.

In the literature of numerical taxonomy, these ultimate units have a specialized name: they are called Operational Taxonomic Units, or OTUs. Boundaries of OTUs often correspond to boundaries of reproductive isolation, but they need not. The only requirements for a system of OTUs are as follows (Rogers et al., 1967).

**Informal Definition**

In any system of OTUs:

1. there is at least one and at most a finite number of OTUs
(ii) every organism belongs to exactly one OTU
(that is, no organism belongs to more than
one OTU and no organism is unclassified)

(iii) every OTU contains at least one organism
(that is, there are no empty OTUs)

and

(iv) every OTU contains only organisms we are
studying (that is, nothing extraneous).

In mathematics, a system of OTUs is an example of a
partition. The set of all organisms in a particular study will
typically be given the symbol $X$, and a partition of $X$(that is, a
set of OTUs) will be given the symbol $\Pi_X$. Figure 4.1 illustrates
the concept of a partition of $X$ where $X$ comprises the set of grand-
sons of Abraham as illustrated in Figure 2.1. Several features of
this example are of particular interest. The set $X$, shown in
Figure 4.1 A, is a simple set of organisms -- a nonempty subset of
the set $B$. In this case, $X$ has exactly fourteen members -- twelve
sons of Ishmael and two sons of Isaac. The partition $\Pi^1_X$, shown in
Figure 4.1 B, is a set of sets of organisms. (Note: Any set of
sets of organisms will be symbolized by an upper case Greek letter in
this dissertation.) The partition $\Pi^2_X$ is not a subset of $B$; it has
two members, each of which is a subset of $B$. Figure 4.1 C shows an
alternative partition, namely $\Pi^2_X$, and highlights the fact that there
is not necessarily a unique set of OTUs for an initial set $X$. In
Figure 4.1. Partitions of $X$. Set $X$ is the set of all the grandson grandsons of Abraham. (A) The set $X$. (B) A partition of $X$, $\Pi_X^1$. (C) A different partition of $X$, $\Pi_X^2$. 

**Table (A)**

<table>
<thead>
<tr>
<th>Nebaioth</th>
<th>Mishma</th>
<th>Tema</th>
<th>Esau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kedar</td>
<td>Dumah</td>
<td>Jetur</td>
<td></td>
</tr>
<tr>
<td>Adbeel</td>
<td>Massa</td>
<td>Naphish</td>
<td></td>
</tr>
<tr>
<td>Mibsam</td>
<td>Hadad</td>
<td>Kedemah</td>
<td>Jacob</td>
</tr>
</tbody>
</table>

**Table (B)**

<table>
<thead>
<tr>
<th>Nebaioth</th>
<th>Mishma</th>
<th>Tema</th>
<th>Esau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kedar</td>
<td>Dumah</td>
<td>Jetur</td>
<td></td>
</tr>
<tr>
<td>Adbeel</td>
<td>Massa</td>
<td>Naphish</td>
<td></td>
</tr>
<tr>
<td>Mibsam</td>
<td>Hadad</td>
<td>Kedemah</td>
<td>Jacob</td>
</tr>
</tbody>
</table>
this particular example, the partition in Figure 4.1 B is more
"biological" than the partition in Figure 4.1 C, because in the
former case, members within an OTU share a more recent common
ancestor (i.e., their respective parents, Ishmael and Isaac)
than members between OTUs (i.e., their grandparent, Abraham).

Now for a relatively minor point: \( \Pi_X \) is finite in
this dissertation, but X need not be. That is, there will be
at most a finite number of OTUs, but some OTUs may contain an
infinite number of members. As I mentioned in chapter three,
this merely means that I am leaving the cardinality (number of
members) of X unspecified: it doesn't matter if X is finite or
infinite, as long as it has at least one member. I don't believe
that any collections of biological organisms are truly infinite.
The main reason for this device is so that the models of population
genetics, some of which presuppose infinite, can be worked into my
model more easily by future investigators.

The formal definition of partition closely parallels the
informal definition of a set of OTUs:

Definition 4.1 \( \Pi_X \) is a partition (of X) if and only if there
exists a set X such that \( \emptyset \neq X \subseteq B \) and:

(i) \( 1 \leq N(\Pi_X) < \infty \)

(ii) for each \( x \in X \), there exists a unique \( P \in \Pi_X \)
such that \( x \in P \)
(iii) for each \( P \in \Pi_X \), \( P \neq \emptyset \)

and

(iv) for each \( P \in \Pi_X \), \( P \subseteq X \).

The following properties of partitions will be used repeatedly in the course of this dissertation: (i) every organism in an OTU is an organism in the overall study, (ii) no organism is in two different OTUs, and (iii) any organisms which are all members of one OTU are not all members of another OTU:

**Theorem 4.1** Let \( \Pi_X \) be a partition, Then

(i) For each \( P \in \Pi_X \), \( P \subseteq X \)

(ii) for distinct \( P_1, P_2 \in \Pi_X \), \( P_1 \cap P_2 = \emptyset \)

and

(iii) for distinct \( P_1, P_2 \in \Pi_X \), and set \( Y \) such that
\[
\emptyset \neq Y \subseteq P_1, \sim (Y \subseteq P_2)
\]

**Proof. Part I.** Choose any \( P \in \Pi_X \) and any \( x \in P \). By Definition 4.1 (iv), \( x \in X \). Hence \( P \subseteq X \).

**Part II.** Choose distinct \( P_1, P_2 \in \Pi_X \). Suppose \( P_1 \cap P_2 \neq \emptyset \).

Then there is an \( x \in P_1 \cap P_2 \). By Definition 4.1 (iv), \( x \in X \). By Definition 4.1 (ii), there is a unique member \( P \in \Pi_X \) such that \( x \in P \).

But \( x \in P_1 \) and \( x \in P_2 \). Contradiction.

**Part III.** Let distinct \( P_1, P_2 \in \Pi_X \) and \( \emptyset \neq Y \subseteq P_1 \). Since \( Y \neq \emptyset \), choose \( y \in Y \). Suppose \( Y \subseteq P_2 \). Then \( y \in P_1 \) and \( y \in P_2 \)

and hence \( y \in P_1 \cap P_2 \). By part II, \( P_1 \cap P_2 = \emptyset \). Contradiction.
A partition which will have special interest for us in this dissertation is the binary partition, or partition containing exactly two OTUs:

**Definition 4.2**  $\Pi_X$ is a binary partition if and only if $\Pi_X$ is a partition and $N(\Pi_X) = 2$.

Figure 3.1 furnishes an example of two different binary partitions of the same set $X$. Let $\Pi_X^1 = \{P_1, R_2\}$ and $\Pi_X^2 = \{R_1, R_2\}$. Then

\[ P_1 = \{Nebaioth, Kedar, Adbeel, Mbsam, Mishma, Dumah, Massa, Hadad, Tema, Jetur, Naphish, Kedemah\} \]
\[ P_2 = \{Essu, Jacob\} \]
\[ R_1 = \{Nebaioth, Kedar, Mishma, Dumah, Tema, Jetur, Esau\} \]
\[ R_2 = \{Adbeel, Mbsam, Massa, Hadad, Naphish, Kedemah, Jacob\} \]

The four possible sets $P_i \cap R_j$ for $i = 1, 2$ and $j = 1, 2$ are:

\[ P_1 \cap R_1 = \{Nebaioth, Kedar, Mishma, Dumah, Tema, Jetur\} \]
\[ P_1 \cap R_2 = \{Adbeel, Mbsam, Massa, Hadad, Naphish, Kedemah\} \]
\[ P_2 \cap R_1 = \{Essu\} \]
\[ P_2 \cap R_2 = \{Jacob\} \]

In this example, all four subsets have at least one member. In general, let $\{P_1, P_2\}$ and $\{R_1, R_2\}$ be two different binary partitions
of the same set \( X \). Theorem 4.2 states that at least three of
the sets \( P_i \cap R_j \) must be nonempty:

**Theorem 4.2** Let \( P_X = \{P_1, P_2\} \) and \( R_X = \{R_1, R_2\} \) be distinct
binary partitions of \( X \). Then at most one of the sets \( P_i \cap R_j \)
where \( i = 1, 2 \) and \( j = 1, 2 \) are empty.

**Proof.** Without loss of generality, let \( P_1 \cap R_1 \) be empty. Assume
the theorem is false. That is, one of the other three sets is
empty.

**Case I.** Suppose \( P_1 \cap R_2 = \emptyset \). By Definition 4.1 (iii),
\( P_1 \neq \emptyset \). But \( P_1 = P_1 \cap X = (P_1 \cap R_1) \cup (P_1 \cap R_2) = \emptyset \cup \emptyset = \emptyset \).
Contradiction.

**Case II.** Suppose \( P_2 \cap R_j \neq \emptyset \) for \( j = 1 \) or 2. By
Definition 4.1 (iii), \( R_j \neq \emptyset \). But \( R_j = X \cap R_j \)
\[ = (P_1 \cap R_j) \cup (P_2 \cap R_j) = \emptyset \cup \emptyset = \emptyset . \] Contradiction.

### 4.2 The Superpartition

Suppose we have a set of organisms, \( X \), consisting of
all living organisms which belong either to the genus Homo or
the genus Pan. We partition the set \( X \) at the species level to
obtain: \( \Pi_X = \{\text{Homo sapiens, Pan troglodytes, Pan paniscus}\} \).
Using the members of \( \Pi_X \) and the union operation, we can construct
higher taxonomic categories which are also sets of organisms.
For example, the set of all organisms which belong either to *Pan troglodytes* or *Pan paniscus* is also the set of organisms which belong to the genus *Pan*. Thus:

\[ Pan = \text{Pan troglodytes} \cup \text{Pan paniscus} \]

It is also possible to form hypothetical taxonomic categories by the same operations:

\[ \text{Homodytes} = \text{Homo sapiens} \cup \text{Pan troglodytes} \]
\[ \text{Homaniscus} = \text{Homo sapiens} \cup \text{Pan paniscus} \]

(The reason for forming hypothetical taxonomic categories is that, later in this dissertation, we shall compare many potential taxonomic categories to one another to determine which are the true ones.) We can form still a higher (hypothetical) taxonomic category by forming the union of all three sets:

\[ \text{Homopen} = \text{Homo sapiens} \cup \text{Pan troglodytes} \cup \text{Pan paniscus} \]

Of course, *Homopen* is just *X*. The entire collection of true and hypothetical taxonomic categories is called the superpartition of *X*, and is denoted throughout this dissertation as the symbol \( \Sigma \). For the above example:

\[ \Sigma = \{ \text{Homo sapiens, Pan troglodytes, Pan paniscus, Homodytes, Homaniscus, Pan, X} \} \]

In general:
Definition 4.3 Let $\Pi_X$ be a partition of $X$. Then the superpartition of $\Pi_X$, denoted $\Sigma$, is the set

$$\Sigma = \{ Y : \text{there is a } k \geq 1 \text{ and } P_1, \ldots, P_k \in \Pi_X \text{ such that } Y = P_1 \cup \ldots \cup P_k \}$$

The superpartition $\Sigma$ has exactly $2^{N(\Pi_X)} - 1$ members:

Theorem 4.3 Let $\Pi_X$ be a partition of $X$ and $\Sigma$ be the superpartition of $\Pi_X$. Then $N(\Sigma) = 2^{N(\Pi_X)} - 1$.

Proof. Let $\mathcal{P}$ be the power set of $\Pi_X$, that is, the set of all subsets of $\Pi_X$. Let $g : \mathcal{P} \to \Sigma \cup \{\emptyset\}$ be the function such that

$$g(\emptyset) = \emptyset$$

and for each $k \geq 1$ and $P_1, \ldots, P_k \in \Pi_X$,

$$g(\{P_1, \ldots, P_k\}) = P_1 \cup \ldots \cup P_k$$

By Definition 4.3, $S \in \Sigma$ if and only if there is a $k \geq 1$ and $P_1, \ldots, P_k \in \Pi_X$ such that $S = P_1 \cup \ldots \cup P_k$. That is, $g$ is one-to-one and onto. But from elementary set theory,

$$N(\mathcal{P}) = 2^{N(\Pi_X)}.$$ Hence

$$2^{N(\Pi_X)} - 1 = N(\mathcal{P}) - 1 = N(g(\mathcal{P})) - 1 = N(\Sigma \cup \{\emptyset\}) - 1 = N(\Sigma).$$
A nonempty set of organisms, Z, is a member of the superpartition if and only if every OTU is either a subset of Z or shares no organisms in common with Z. For example, let Z = Homo sapiens. Then Homo sapiens and Pan paniscus are subsets of Z, and Pan troglodytes shares no organisms in common with Z.

Theorem 4.4 Let \( \Pi_X \) be a partition and \( \Sigma \) be the superpartition of \( \Pi_X \) and \( \emptyset \neq Z \subseteq X \). Then \( Z \in \Sigma \) if and only if for each \( P \in \Pi_X \), either \( P \subseteq Z \) or \( P \cap Z = \emptyset \).

Proof. Only If. Consider any \( Z \) and \( P \) such that \( \emptyset \neq Z \subseteq X \) and \( Z \in \Sigma \) and \( P \in \Pi_X \). Assume the theorem is false. That is, \( P \cap Z \neq \emptyset \) and \( \sim (P \subseteq Z) \). Choose \( x \in P \cap Z \). By Definition 4.3, there is a \( k \geq 1 \) and members \( P_1, \ldots, P_k \in \Pi_X \) such that \( Z = P_1 \cup \ldots \cup P_k \). There exists an \( i \) such that \( 1 \leq i \leq k \) and \( x \in P_i \). Suppose \( P = P_i \). Then \( P \subseteq Z \), contradiction. Suppose \( P \neq P_i \). By construction, \( x \in P \cap P_i \) and \( P \cap P_i \neq \emptyset \). By Theorem 4.1 (ii), \( P = P_i \). Contradiction.

If. Consider any \( Z \) such that \( \emptyset \neq Z \subseteq X \) and for each \( P \in \Pi_X \), either \( Z \cap P = \emptyset \) or \( P \subseteq Z \). For each \( x \in Z \), by Definition 4.1 (ii), there exists a \( P \in \Pi_X \) such that \( x \in P \). Construct the set \( Y = P_1 \cup \ldots \cup P_k \) such that for each \( i, 1 \leq i \leq k, P_i \in \Pi_X \) and there is an \( x \in Z \) such that \( x \in P_i \). By Definition 4.3, \( Y \in \Sigma \). Since for any \( x \in Z \), \( x \in Y \), it follows
that $Z \subseteq Y$. Assume the theorem is false. That is, $Z \subset Y$.

Then there is a $y \in Y$ such that $y \notin Z$. Find a $P_i \subseteq Y$ such that $y \in P_i$. By construction, there must be a $z \in Z$ such that $z \in P_i$.

By hypothesis, for each $P_i \in \Pi_X$, either $Z \cap P_i = \emptyset$ or $P_i \subseteq Z$.

Since $z \in Z \cap P_i$, $Z \cap P_i \neq \emptyset$. Therefore $P_i \subseteq Z$. But $y \in P_i - Z$.

Contradiction.

Theorem 4.5 shows that the set of organisms common to two members of the superpartition is either empty or itself a member of the superpartition. For example, the set of organisms common to Homo assicus and Pan is the set Pan paniscus, that is, a member of $\Sigma$. But the set of organisms common to Homo sapiens and Pan is the empty set.

**Theorem 4.5** Let $\Pi_X$ be a partition of $X$. Let $\Sigma$ be the superpartition of $\Pi_X$. If $Y, Z \in \Sigma$, then either $Y \cap Z = \emptyset$ or $Y \cap Z \in \Sigma$.

**Proof.** Suppose $Y \cap Z \neq \emptyset$. Assume theorem is false. Then $Y \cap Z \notin \Sigma$. By Theorem 4.4, there must be some $P \in \Pi_X$ such that $P \cap (Y \cap Z) \neq \emptyset$ and $\sim (P \subseteq Y \cap Z)$. Let $x \in P \cap (Y \cap Z)$. By Definition 4.1 (ii), there is an $R \in \Pi_X$ such that $x \in R$. Since $x \in R \cap Y$, $R \cap Y \neq \emptyset$. Since $Y \in \Sigma$, by Theorem 4.4, $R \subseteq Y$.

By a similar argument, $R \subseteq Z$. That is, $R \subseteq Y \cap Z$. But $x \in P \cap R$, so that by Theorem 4.1 (ii), $P = R$. That is, $P \subseteq Y \cap Z$.

Contradiction.
4.3 Isolation

Consider any collection of organisms whose ancestors over the past few generations have engaged in unrestricted sexual interbreeding. If we examine the parent relationships among these organisms, we expect to find a complicated skein of interrelationships in which any organism shares a common ancestor with any organism not more than a few generations ago. Now suppose we place a barrier to interbreeding between a subset of these organisms and the rest of the collection. This barrier may be physical, behavioral, or physiological, and it is irrelevant whether or not organisms in either side of the barrier are unable to interbreed -- the essential point is that they do not interbreed. At the instant the barrier is imposed, we still observe no change in the skein of parental relationships; in the sense of this dissertation, the sets on either side of the barrier are not yet isolated. However, if we wait for a few generations, the organisms on one side of the barrier will share more recent common ancestries among themselves than with organisms from the other side of the barrier. Intuitively, we say that two sets of organisms are isolated from one another if every organism in one set (i.e., on one side of the barrier) shares a more recent common ancestor-or-equal with any organism in the same set than with any organism in the other set (i.e., on the other side of the barrier). We introduce the additional restriction that any organism from one set and any organism from the other set must share all the same common ancestor-or-equals as any other organism from the first set and
any other organism from the second set; this is to insure that all organisms in both sets descended from the same, initial interbreeding collection of organisms. Note that the above definition of isolation is "isolation in parentage" in the sense that it is defined solely on the basis of parent relationships. We summarize these properties in the following informal definition.

**Informal Definition.** If Y and Z are two nonempty sets of organisms, we say that Y is isolated from Z if and only if:

- (i) Each organism \( x \) which is an ancestor-or-equal of any member of Y and any member of Z must also be an ancestor-or-equal of any other member of Y and any other member of Z.

- (ii) There is an organism \( y_{\star} \) such that \( y_{\star} \) is an ancestor-or-equal of all members of Y and no members of Z.

and

- (iii) There is an organism \( z_{\star} \) such that \( z_{\star} \) is an ancestor-or-equal of all members of Z and no members of Y.

That is:

**Definition 4.4** Let \( \phi \neq Y \subseteq B \) and \( \phi \neq Z \subseteq B \). Then \( Y \) is isolated from \( Z \) if and only if there exist \( y_{\star}, z_{\star} \in B \) such that:

- (i) for each \( y_{1}, y_{2} \in Y \) and \( z_{1}, z_{2} \in Z \), \( CP(\{y_{1}, y_{2}\}) = CP(\{z_{2}, z_{2}\}) \)
(ii) for each \( y \in Y \) and \( z \in Z \), \( y_\star \triangleright y \) and \( \sim (y_\star \triangleright z) \)
and
(iii) for each \( z \in Z \) and \( y \in Y \), \( z_\star \triangleright y \) and \( \sim (z_\star \triangleright y) \).

Figure 4.1 illustrates the concept of isolated sets in the sense of this dissertation. The two members of \( \Pi K \) are isolated from one another because:

(i) Any member of the left set (sons of Ishmael) and any member of the right set (sons of Isaac) share the same common pedigree (namely, Abraham and all his ancestors) as any other member of the left set and any other member of the right set.

(ii) Every member of the left set has at least one common ancestor-or-equal (namely, Ishmael) which no member of the right set has.

(iii) Every member of the right set has at least one common ancestor-or-equal (namely, Isaac) which no member of the left set has.

The two members of \( \Pi K \) are not isolated from one another, however, because all three conditions of our definition are violated:

(i) The common pedigree of \{Nebaioth, Adbeel\} has a member not present in the common pedigree of \{Esau, Kedemah\} (namely, Ishmael), even though Nebaioth and Esau are both from the top set and Adbeel and Kedemah are both from the bottom set.
(ii) Every common ancestor-or-equal of the top set (namely, Abraham and his ancestors) is also a common ancestor-or-equal of the bottom set.

(iii) Every common ancestor-or-equal of the bottom set (namely, Abraham and his ancestors) is also a common ancestor-or-equal of the top set.

4.4 Sampling from Isolated Sets

Very often, a biologist will not have time to study all individuals from a set of organisms he is interested in. Alternatively, there may be a typical or borderline individuals for which the biologist is not certain as to which organism set they belong to; for a variety of reasons, he may wish to use only those individuals whose organism set identity is known. Whether for convenience or accuracy, the biologist will only take samples (i.e., subsets) of the organism sets under investigation. In the next theorem, we show that if a set $Y$ is isolated from a set $Z$, then any nonempty subset of $Y$ is isolated from any nonempty subset of $Z$:

Theorem 4.6 Let $\phi \neq Y \subseteq B$, $\phi \neq Z \subseteq B$, $\phi \neq Y_{1} \subseteq Y$, and $\phi \neq Z_{1} \subseteq Z$. If $Y$ is isolated from $Z$, then $Y_{1}$ is isolated from $Z_{1}$.
Proof. Strategy. Let \( y_* \in B \) and \( z_* \in B \) be as in Definition 4.4. Show that parts (i) - (iii) of Definition 4.4 are satisfied for sets \( Y_1 \) and \( Z_1 \).

Part I. Let \( y_1, y_2 \in Y_1 \) and \( z_1, z_2 \in Z_1 \). Since \( Y_1 \subseteq Y \) and \( Z_1 \subseteq Z \), it follows that \( y_1, y_2 \in Y \) and \( z_1, z_2 \in Z \). By Definition 4.4, \( CP(\{y_1, z_1\}) = CP(\{y_2, z_2\}) \).

Part II. Let \( y \in Y_1 \subseteq Y \) and \( z \in Z_1 \subseteq Z \). By Definition 4.4, \( y_* \uparrow y \) and \( \sim (y_* \uparrow z) \).

Part III. Let \( z \in Z_1 \subseteq Z \) and \( y_* \in Y_1 \subseteq Y \). By Definition 4.4, \( z_* \uparrow z \) and \( \sim (z_* \uparrow y) \).

In Theorem 3.10, we observed that if a set of organisms gets "smaller", then the common pedigree may get "bigger". That is, if we form a new set which is a subset of the original set, then the common pedigree of the original set is a subset of the common pedigree of the new set. Theorem 4.7 shows that if the original set consists of the union of two sets, \( Y \) and \( Z \), which are isolated from one another, and the new set contains members of both \( Y \) and \( Z \), then the common pedigree (and the most recent common pedigree) are proved:

**Theorem 4.7** Let \( \emptyset \neq Y \subseteq B \), \( \emptyset \neq Z \subseteq B \), and \( Y \) be isolated from \( Z \). Let \( \emptyset \neq Y_1 \subseteq Y \) and \( \emptyset \neq Z_1 \subseteq Z \). Then

(1) \( CP(Y \cup Z) = CP(Y_1 \cup Z_1) \)

and

(2) \( RP(Y \cup Z) = RP(Y_1 \cup Z_1) \).
Proof. Part I. Since \( Y \cup Z \supseteq Y_1 \cup Z_1 \), by Theorem 3.10,
\( \text{CP}(Y \cup Z) \subseteq \text{CP}(Y_1 \cup Z_1) \). Assume the theorem is false. That is, 
\( \text{CP}(Y \cup Z) \subset \text{CP}(Y_1 \cup Z_1) \). Then there is an \( x \in \text{CP}(Y_1 \cup Z_1) \) such that 
\( x \notin \text{CP}(Y \cup Z) \). Choose \( y_1 \in Y_1 \subseteq Y \) and \( z_1 \in Z_1 \subseteq Z \). Since 
\( x \in \text{CP}(Y_1 \cup Z_1) \), by Definition 3.4, \( x \supseteq y_1 \) and \( x \supseteq z_1 \). By Definition 3.4, \( x \in \text{CP}([y_1,z_1]) \). Since \( x \notin \text{CP}(Y \cup Z) \), by Definition 3.4, there 
is some organism \( w \in Y \cup Z \) such that \( \neg(x \supseteq w) \). Without loss of 
generality let \( w \in Y \). Choose \( z \in Z \). By Definition 3.4, 
\( x \notin \text{CP}([w,z]) \). By Definition 4.4, \( \text{CP}([y_1,z_1]) = \text{CP}([w,z]) \). 
Contradiction.

Part II. By Theorem 3.16, \( \text{CP}(Y \cup Z) = \text{CP}(Y_1 \cup Z_1) \) implies 
that \( \text{RP}(Y \cup Z) = \text{RP}(Y_1 \cup Z_1) \).

Theorem 4.8 shows that if the original set consists of 
members of both isolated sets \( Y \) and \( Z \), whereas the new set contains 
only members of \( Y \), then the common pedigree of the new set has addi-
tional members not present in the common pedigree of the original set,
but not vice versa:

Theorem 4.8 Let \( \emptyset \neq Y \subseteq B \) and \( \emptyset \neq Z \subseteq B \) and \( Y \) be isolated from \( Z \).
Let \( \emptyset \neq Y_1 \subseteq Y \) and \( \emptyset \neq Y_2 \subseteq Y \) and \( \emptyset \neq Z_1 \subseteq Z \). Then 
\( \text{CP}(Y_1 \cup Z_1) \subseteq \text{CP}(Y_2) \).

Proof. By Theorem 3.10, \( \text{CP}(Y \cup Z) \subseteq \text{CP}(Y) \). By Definition 4.4 (ii),
there exists a \( y^* \in B \) such that for each \( y \in Y \) and \( z \in Z \), \( y^* \supseteq y \) and 
\( \neg(y^* \supset z) \). By Definition 3.4, \( y^* \in \text{CP}(Y) \) and \( y^* \notin \text{CP}(Y \cup Z) \) Hence:

\[ \text{CP}(Y \cup Z) \subseteq \text{CP}(Y) \]  \( \text{(*)} \)
By Theorem 4.7:

\[ \text{CP}(Y_1 \cup Z_1) = \text{CP}(Y \cup Z) \]  

\[ (***) \]

By Theorem 3.10:

\[ \text{CP}(Y) \subseteq \text{CP}(Y_2) \]

\[ (***) \]

Combining expressions (*), (**), and (***) under the transitivity of \( \subseteq \), we have:

\[ \text{CP}(Y_1 \cup Z_1) = \text{CP}(Y \cup Z) \subseteq \text{CP}(Y) \subseteq \text{CP}(Y_2) \]

or

\[ \text{CP}(Y_1 \cup Z_1) \subseteq \text{CP}(Y_2) \]

### 4.5 Cladistic Partitions

A cladistic partition is the special case of a partition in which every OTU is isolated from every other OTU. For example, a set of organisms partitioned according to the boundaries of biological species constitute a cladistic partition of that set. As we shall see in chapter five (Theorem 5.12), cladistic partitions are the only kind of partitions which can be used to construct cladograms.

**Definition 4.5** \( \Pi_X \) is a **cladistic partition** (of \( X \)) if and only if \( \Pi_X \) is a partition (of \( X \)) and for distinct members \( P_1, P_2 \in \Pi_X \), \( P_1 \) is isolated from \( P_2 \).

A **binary cladistic partition** is a cladistic partition which has exactly two members:
Definition 4.6 \( \Pi_X \) is a binary cladistic partition (of \( X \)) if and only if \( \Pi_X \) is a cladistic partition (of \( X \)) and \( N(\Pi_X) = 2 \).

Theorem 4.9 gives us an important uniqueness property of a binary cladistic partition which we shall use extensively in later chapters. For any set of organisms \( X \), there is at most one binary cladistic partition:

**Theorem 4.9** Let \( \emptyset \neq X \subseteq B \). Then there is at most one binary cladistic partition of \( X \).

**Proof.** Suppose \( \{ Y_1, Y_2 \} \) and \( \{ Z_1, Z_2 \} \) are distinct binary partitions of \( X \) which are both cladistic partitions of \( X \). By Theorem 4.2, without loss of generality, let \( Y_1 \cap Z_1 \), \( Y_1 \cap Z_2 \), and \( Y_2 \cap Z_2 \) be nonempty. Since \( \{ Y_1, Y_2 \} \) is a binary cladistic partition of \( X \), it follows by Definition 4.5 that \( Y_1 \) is isolated from \( Y_2 \). Since \( Y_1 \cap Z_1 \subseteq Y_1 \) and \( Y_1 \cap Z_2 \subseteq Y_1 \) and \( Y_2 \cap Z_2 \subseteq Y_2 \), it follows by Theorem 4.8 that:

\[
\mathsf{CP}(Y_1 \cap Z_1) \cup (Y_1 \cap Z_2) = \mathsf{CP}(Y_1 \cap Z_1) \cup (Y_2 \cap Z_2).
\]

Since \( \{ Z_1, Z_2 \} \) is a binary cladistic partition of \( X \), it follows by Definition 4.5 that \( Z_1 \) is isolated from \( Z_2 \). Since \( Y_1 \cap Z_1 \subseteq Z_1 \) and \( Y_1 \cap Z_2 \subseteq Z_2 \) and \( Y_2 \cap Z_2 \subseteq Z_2 \), it follows by Theorem 4.7 (i) that:

\[
\mathsf{CP}(Y_1 \cap Z_1) \cup (Y_2 \cap Z_2) = \mathsf{CP}(Y_1 \cap Z_1) \cup (Y_2 \cap Z_2).
\]

Expressions (*) and (**) are contradictory.
5. DENDROGRAMS AND CLADOGRAMS

5.1 The Dendrogram

A dendrogram is any tree-like diagram for representing relationships among OTUs. Figure 5.1 shows such a diagram for the OTUs A, B, C, D, E, F, and G. The outermost twigs of a dendrogram correspond to individual OTUs, and are usually placed at the top of the drawing. Progressively thicker branches correspond to progressively more inclusive taxonomic categories; these intermediate branches include H, I, J, K, and L in Figure 5.1 A. The thickest branch, or trunk (called the "root" by Farris (1970)) is called X in Figure 5.1 A. An alternative convention (Hennig, 1966) for representing a dendrogram is shown in Figure 5.1 B.

In ordinary taxonomy, the "thicker branches" in a dendrogram are simply more inclusive sets, and do not necessarily correspond to a more ancient common ancestry. Hennig's dendrogram convention (Figure 5.1 B) is probably superior for strictly taxonomic purposes, because it is not reminiscent of family trees as is the ordinary convention (Figure 5.1 A). Since this dissertation is eventually concerned with ancestral relationships, we shall employ the ordinary tree convention (Figure 5.1 A).
Figure 5.1. A dendrogram. Let \([A, B, C, D, E, F, G] = \Pi_X\) be a system of OTUs (a partition of \(X\)) whose union is \(X\). (A) A dendrogram on \(\Pi_X\) diagrammed in the conventional format. The set \(X\) is the trunk of this dendrogram and the set \(\Delta\), where \(\Delta = \{A, B, C, D, E, F, G, H, I, J, K, L, X\}\) is the set of junctures for this dendrogram. (B) The same dendrogram in Hennig's (1966) format.
In this dissertation, an OTU is a set of organisms. Any higher taxonomic category is the set of all organisms subsumed under that taxonomic category. That is, a higher taxonomic category is the union of the OTUs which it subsumes. For Figure 5.1:

\[ H = A \cup B \]
\[ I = C \cup D \]
\[ J = F \cup G \]
\[ K = A \cup B \cup C \cup D \]
\[ L = E \cup F \cup G \]
\[ X = A \cup B \cup C \cup D \cup E \cup F \cup G \]

Each OTU and each of the higher taxonomic categories is called a juncture. The set of all junctures for a dendrogram is usually called \( \Delta \) ("delta"), or \( K \) ("kappa") when a cladogram is being discussed. For Figure 5.1:

\[ \Delta = \{A, B, C, D, E, F, G, H, I, J, K, L, X\} \]

The set of all OTUs of a dendrogram is always a partition of the trunk, since the trunk is always the union of all OTUs in a dendrogram. For Figure 5.1, \{A, B, C, D, E, F, G\} is a partition of \( X \).

Finally, we require a terminology which expresses the relationship of any juncture \( i \) to an immediately less inclusive juncture and \( ii \) to any less inclusive juncture. These relationships are called, respectively, \( i \) "is an immediate predecessor of" and \( ii \) "is a predecessor of". We say, for example, that \( K \) is an
immediate predecessor of $H$, denoted $K \triangleright H$, because $K \supset H$ and there is nothing else "in between" $K$ and $H$. On the other hand, $K$ is a predecessor of $B$, denoted $K \triangleright B$, because there is a sequence of immediate predecessor relationships leading from $K$ to $B$. The symbols for expressing immediate predecessor ($\triangleright$) and predecessor ($\supset$) relationships are meant to be reminiscent of the $\triangleright$ and $\supset$ symbols of Section 3.1 and the proper subset symbol, $\supset$.

On our sample dendrogram of Figure 5.1, some immediate predecessor relationships include:

$$K \triangleright H$$
$$X \triangleright D$$
$$I \triangleright D$$

Some predecessor relationships include:

$$J \supset F \text{ because } J \supset F$$
$$K \supset A \text{ because } K \triangleright H \text{ and } H \supset A$$
$$X \supset D \text{ because } X \triangleright K \text{ and } K \triangleright I \text{ and } I \supset D$$

In our axiomatic model, we define immediate predecessor and predecessor relationships as follows:

**Definition 5.1** Let $S, T \in \Delta$. Then $S$ is an immediate predecessor of $T$, denoted $S \triangleright T$, if and only if
(i) \( S \supset T \)
and
(ii) there is no \( R \in \Delta \) such that \( S \supset R \supset T \).

**Definition 5.2** Let \( S, T \in \Delta \). Then \( S \) is a predecessor of \( T \), denoted \( S \supset T \), if and only if

(i) \( S \supset T \)
or
(ii) there exists a \( k \geq 1 \) and an ordered set 
\( (S=R_0, \ldots, R_{k+1}=T) \) such that for \( 0 \leq i \leq k \),
\( R_i \in \Delta \) and \( R_i \supset R_{i+1} \).

We are now ready to consider the dendrogram as a whole. A **dendrogram** is an ordered pair \((\Delta, \mathbf{D})\) whose first member is its set of junctures and whose second member is the immediate predecessor relationship.

(i) Each juncture in the dendrogram is either \( X \) itself, or some OTU, or else it has \( X \) as predecessor and it is in turn the predecessor of some OTU.

(ii) Every pair of junctures has the property that they are the same, share no organism in common, or else one is the predecessor of the other. (In Figure 5.1, for example, \( L \) and \( D \) share no organisms in common (i.e., \( L \cap D = \emptyset \)) and also have no predecessor relationship; on the other hand, \( L \) and \( G \) share organisms in common (i.e., \( L \cap G \neq \emptyset \)) and it is true that \( L \supset G \).)

(iii) Every juncture except \( X \) itself has exactly one immediate predecessor. These assumptions are analogous to those of a **directed tree** of graph theory (Cres, pp. 34-42, 53-67).
(iv) The special case of a dendrogram which we will develop in this dissertation is the binary dendrogram. In a binary dendrogram, every juncture except an OTU is the predecessor of exactly two junctures. Furthermore, this dissertation will only develop the special case of a binary cladogram. Since a binary cladogram is a binary dendrogram whose branch-points correspond to events of isolation, we are in effect limiting the scope of this dissertation to events of binary ("dichotomous") isolation. Binary isolation is a major assumption in Hennig's (1966) *Phylogenetic Systematics*, and it is strongly criticized by Darlington (1970) as being biologically unsound. The difference between Hennig's attitude and the attitude of this dissertation is that Hennig assumes binary isolation as a property of evolution, whereas this dissertation restricts the applicability of its results to events of binary isolation. In other words, the results of this dissertation are true but irrelevant to nonbinary events of isolation such as the Goodman et al (1970) macaque complex or Selander's (1967) house sparrow complex. Since nonbinary isolation is a relatively common occurrence in the real world of biology, the results of this dissertation should be applied with great caution.

In effect, the binary isolation rule restricts the applicability of this dissertation to long-term events in evolution. For example, the true branching sequence of ancestral separations for the set of all contemporary primate genera is probably a binary dendrogram—not because the process of evolution originally happened that way, but because no more than two separate genera have managed to survive to the present day (the other evolutionary offshoots are by now extinct). The
usual practice is to regard a binary dendrogram as the simplest case, which is to be preferred unless there is strong evidence in favor of a more general framework. Biological investigations in which a binary dendrogram is clearly wrong (such as the macaque and sparrow complexes) cannot use the results of this dissertation.

The above properties of a binary dendrogram may be summarized informally as follows:

**Informal Definition** Let $X$ be a set of organisms and $\Pi_X$ be a set of OTUs (that is, a partition). A binary dendrogram is characterized as a set of junctures, $\Delta$, and an immediate predecessor operation, $\preceq$, subject to the restrictions:

(i) Each juncture is either $X$ or an OTU or else $X$ is its predecessor and it is the predecessor of some OTU.

(ii) Two junctures are either the same, share no organisms in common, or else one is the predecessor of the other.

(iii) Each juncture which is not $X$ has exactly one immediate predecessor.

(iv) Each juncture which is not an OTU is the predecessor of exactly two junctures.

Our formal definition of binary dendrogram closely parallels the informal definition:

**Definition 5.3** $(\Delta, \preceq)$ is a binary dendrogram on $\Pi_X$ if and only if $\Pi_X$ is a partition and
(i) \( S \in \Delta \) if and only if \( S = X, S \in \Pi_X \), or there exists a \( P \in \Pi_X \) such that \( X \triangleright S \triangleright P \).

(ii) \( S, T \in \Delta \) if and only if \( S = T, S \cap T = \emptyset, S \triangleright T \), or \( T \triangleright S \).

(iii) \( S \in \Delta - \{X\} \) if and only if there is a unique \( T \in \Delta \) such that \( T \triangleright S \).

(iv) \( S \in \Delta - \Pi_X \) if and only if there is a unique binary partition \( \{T_1, T_2\} \) of \( S \) such that \( \{T_1, T_2\} \in \Delta, S \triangleright T_1 \), and \( S \triangleright T_2 \).

5.2 Cardinality of a Binary Dendrogram

It is important in several of the major proofs of this dissertation to know exactly how many junctures are present in a dendrogram. This information is also valuable to a computer programmer, who has to know how much storage to allot for holding the results of a phenogram calculation. Theorem 5.5 gives the precise cardinality of a dendrogram, but several theorems must be proved before Theorem 5.5 can be proved. The strategy of proof is as follows. First, we show that if one juncture is a predecessor of another, then the second is a proper subset of the first. Then, we show that any juncture of any dendrogram must be a member of the superpartition, \( \Sigma \). Since \( \Sigma \) is finite (Theorem 4.3), the set of junctures, \( \Delta \), is finite, and we are allowed to perform the operations of ordinary arithmetic on \( N(\Delta) \). Only then can we prove Theorem 5.5.
Theorem 5.1 states that if one juncture is a predecessor of another juncture, then the second is a proper subset of the first.

**Theorem 5.1** Let $S, T \in \Delta$. If $S \supset T$, the $S \supset T$.

**Proof.** Let $S, T \in \Delta$ and $S \supset T$. Suppose $S \supset T$. Then by Definition 5.1 (i), $S \supset T$. Suppose $\sim (S \supset T)$. Then by Definition 5.2 (ii), there is a $k \geq 1$ and an ordered set $(S=R_0, \ldots, R_{k+1}=T)$ such that $S=R_0 \supset \ldots \supset R_{k+1}=T$. By definition 5.1 (i), $S=R_0 \supset \ldots \supset R_{k+1}=T$. By transitivity of $\supset$, $S \supset T$.

Theorem 5.2 shows that for each juncture and each OTU, either the juncture and the OTU share no organisms in common, or else the juncture is predecessor or equal to the OTU:

**Theorem 5.2** Let $(\Delta, \supset)$ be a binary dendrogram on $\Pi_X$. Then for each $S \in \Delta$ and $P \in \Pi_X$, either $S \cap P = \emptyset$ or $S \supset P$.

**Proof.** Strategy. Suppose $S \cap P \neq \emptyset$. By Definition 5.3 (ii), either $S \supset P$ or $P \supset S$. Assume theorem is false. That is, $P \supset S$.

By Theorem 5.1, $P \supset S$. Show contradiction. Summary of cases: (I) $S = X$; (II) $S \in \Pi_X$; (III) $S \in \Delta - (\Pi_X \cup \{X\})$.

**Case I.** Suppose $S = X$. By Theorem 4.1 (i), $P \subseteq X = S$.

Contradiction.

**Case II.** Suppose $S \in \Pi_X$. Since $S \neq \emptyset$, $P \cap S = S \neq \emptyset$. By Theorem 4.1 (ii), $P = S$. Contradiction.

**Case III.** Suppose $S \in \Delta - (\Pi_X \cup \{X\})$. By Definition 5.3 (ii), there exists a $Q \in \Pi_X$ such that $S \supset Q$. By Theorem 5.1, $P \supset S \supset Q$.

Since $Q \neq \emptyset$, $P \cap Q = Q \neq \emptyset$. By Theorem 4.1 (ii), $P = Q$. Contradiction.
Theorem 5.3 shows that any set of junctures is a subset of the superpartition, Σ.

**Theorem 5.3** Let \( \Pi_X \) be a partition and Σ be the superpartition on \( \Pi_X \). If \( (\Delta, D) \) is a binary dendrogram on \( \Pi_X \), then \( \Delta \subseteq \Sigma \).

**Proof. Strategy.** Consider any \( S \in \Delta \). Show that \( S \in \Sigma \). Summary of cases: (I) \( S \in \Pi_X \); (II) \( S \in \Delta - \Pi_X \).

**Case I.** Suppose \( S \in \Pi_X \). Then by Definition 4.3, \( S \in \Sigma \).

**Case II.** Suppose \( S \in \Delta - \Pi_X \). Consider any \( P \in \Pi_X \). By Theorem 5.2, either \( S \supset P \) or \( S \cap P = \emptyset \). By Theorem 5.1, either \( S \supset P \) or \( S \cap P = \emptyset \). By Theorem 4.4, \( S \in \Sigma \).

Theorem 5.4 shows that the number of junctures is finite:

**Theorem 5.4** Let \( (\Delta, D) \) be a binary dendrogram on \( \Pi_X \). Then \( N(\Delta) \leq 2^{N(\Pi_X)} - 1 \).

**Proof.** Let \( \Sigma \) be the superpartition on \( \Pi_X \). Then by Theorem 5.3, \( \Delta \subseteq \Sigma \). Hence \( N(\Delta) \leq N(\Sigma) \). By Theorem 4.3, \( N(\Sigma) = 2^{N(\Pi_X)} - 1 \). Hence \( N(\Delta) \leq 2^{N(\Pi_X)} - 1 \).

Theorem 5.5 shows that for any dendrogram, the number of junctures is one less than twice the number of OTUs:
Theorem 5.5 Let \((\Delta, \Phi)\) be a binary dendrogram on \(\Pi_X\). Then
\[ N(\Delta) = 2N(\Pi_X) - 1. \]

Proof. Strategy. (I) Let \(g: \Delta-\{X\} \to \Delta\) such that \(g(S) = T\) if and only if \(T \supset S\). Show that the range of \(g\) is \(\Delta-\Pi_X\). (II) Show that \(g\) is two-to-one. (III) Use \(N(\Delta-\{X\}) = 2N(g(\Delta-\{X\})) = 2N(\Delta-\Pi_X)\) to prove the theorem.

Part I. Construct the function \(g: \Delta-\{X\} \to \Delta\) such that \(g(S) = T\) if and only if \(T \supset S\). By Definition 5.3 (iii), the function \(g\) is everywhere defined. Consider any \(S \in \Delta-\{X\}\). Claim: \(g(S) \in \Delta-\Pi_X\). Let \(T = g(S)\). By Definition 5.3 (iii), \(T \in \Delta\). Assume claim is false. That is, \(T \in \Pi_X\). By construction, \(T \supset S\). By Definition 5.1 (i), \(T \supset S\).

Since \(S \neq \emptyset\), by Theorem 5.2, there is a \(P \in \Pi_X\) such that \(S \supset P\). Since \(P \neq \emptyset\), \(T \cap P = P \neq \emptyset\). By Theorem 4.1 (ii), \(T = P\). Contradiction.

Part II. Consider any \(T \in \Delta-\Pi_X\). Claim: there are exactly two \(S_1, S_2 \in \Delta-\{X\}\) such that \(g(S_1) = g(S_2) = T\). By Definition 5.3 (iv), there is a binary partition \(\{S_1, S_2\}\) of \(T\) such that \(S_i \in \Delta\) and \(T \supset S_i\), that is, \(g(S_i) = T\) for \(i = 1, 2\). By Definition 5.1 (i) and Definition 5.3 (i) and Theorem 5.1, \(X \supset T \supset S_i\); hence \(S_i \neq X\) and \(S_i \in \Delta-\{X\}\) for \(i = 1, 2\). Assume claim is false. That is, there is a distinct \(S_* \in \Delta-\{X\}\) such that \(T \supset S_*\). By Definition 5.1 (i), \(T \supset S_* \neq \emptyset\).

Since \(S_1 \cup S_2 = T\), either \(S_1 \cap S_* \neq \emptyset\) or \(S_2 \cap S_* \neq \emptyset\). Without loss of generality, let \(S_1 \cap S_* \neq \emptyset\). By Definition 5.3 (i), either \(S_1 \supset S_*\) or \(S_* \supset S_1\). If \(S_1 \supset S_*\), then \(T \supset S_1 \supset S_*\) and \(\sim(T \supset S_*);\) contradiction. If \(S_* \supset S_1\), the \(T \supset S_* \supset S_1\) and \(\sim(T \supset S_1);\) contradiction.
Part III. Since \( g: \Delta - \{X\} \to \Delta - \Pi_X \) is two-to-one and onto:

\[
N(\Delta - \{X\}) = 2 \cdot N(g(\Delta - \{X\})) = 2 \cdot N(\Delta - \Pi_X).
\]

By Definition 5.3 (i), \( \{X\} \cup \Pi_X \subseteq \Delta \). Hence:

\[
N(\Delta) - 1 = N(\Delta) - N(\{X\}) = 2 \cdot N(\Delta - \Pi_X) = 2 \cdot N(\Delta) - 2 \cdot N(\Pi_X).
\]

By Theorem 5.4, all values are finite and thus subject to ordinary algebraic manipulations to give:

\[
N(\Delta) = 2 \cdot N(\Pi_X) - 1.
\]

5.3 The Dendrogram as a Monoparental Biocosm

Up to this point, I have talked about \((B, \succ)\) as if \(B\) were the "set of all organisms" and \(\succ\) were the relationship "is a parent of". It is convenient to employ this usage, because then the biological reader can understand the theorems of this dissertation by relating these theorems back to his real-world experience; but it is not entirely accurate. As we explained in Section 2.4, \((B, \succ)\) is simply an ordered pair satisfying Axioms I and II---it is a model theory theory. The "real world" is a model theory model. If all the axioms of our model theory theory are satisfied in the model theory model, then any theorem which we have proved in the model theory theory has a corresponding true statement in our model theory model.

The beauty of model theory is that there may be more than one model theory model (i.e., more than one aspect of the "real world") to which the same model theory theory corresponds. We can
illustrate this point very simply. Suppose the axiomatic theory for integers (whole numbers) is the model theory theory for a model theory model consisting of a bushel of apples. The axioms of integers tell us, for example, that since

\[ 1 < 2 \]

it follows that

\[ 1 + 5 < 2 + 5. \]

In our model theory model, since

one apple is fewer than two apples

it follows that

one plus five apples is fewer than two plus five apples.

Another model theory model for our model theory theory might be a truck full of bushels-of-apples. Extending our model theory theory conclusions to this alternative model theory model, we conclude: since

one bushel is fewer than two bushels

it follows that

one plus five bushels is fewer than two plus five bushels.

Up to this point, we have used the "set of all organisms" and the relationship "is a parent of" as our sole model theory model
terms for the biocosm terms B and P, respectively. In this section, we shall demonstrate that the "set of all junctures" and the relationship "is an immediate predecessor of" in a dendrogram are equally suitable model theory model terms for the biocosm terms B and P, respectively. In fact, (Δ, D) is a model theory model for a monoparental biocosm. In other words, Axioms I and II and Definition 3.6 are all satisfied for (Δ, D). Accordingly, every theorem which we proved for (B, P) is also a true theorem for (Δ, D). That is, we get two sets of theorems for the price of one--many theorems which have already been used for the organism model theory model of the biocosm can be used once again for the juncture model theory model.

In this dissertation, (Δ, D) is a biocosm. Every member S ∈ Δ is a juncture and accordingly also a subset S ⊆ B. The theorems from chapters two and three can be used to describe the properties of S as a juncture (the juncture model theory model) and as a set of organisms (the organisms model theory model). The cladogram, for example, possesses both structural (tree-like) properties, which we think of in terms of the juncture model theory model and biological properties (i.e., the fact that branch-points correspond to events of isolation) which we think of in terms of the organism model theory model. This extraordinary "piggy-back" capability should be of special interest to biologists, who often deal with essentially parallel phenomena which take place at two or more levels of organization. Williams (1970) has noted, for example, that the set B in her biocosm could equally well be interpreted as the "set of all species" or the "set of all organisms" or the "set of all genes at a particular locus".
The organism model theory model is a general biocosm in which monoparental reproduction is perfectly permissible (although not required), whereas the juncture model theory model is the special case of a monoparental biocosms. This is not to say that all natural phenomena among sets of organisms reflect the monoparental special case I have chosen. Introgressive hybridization, for example, is a simple example of "biparentalism" among sets of organisms. Restriction of \((\Delta, \mathcal{D})\) to a monoparental biocosm in this dissertation reflects the current "state of the art" in numerical taxonomy. Computer algorithms for more complex cases haven't been developed.

Since we will move back and forth between the organism and juncture model theory models throughout the remainder of this dissertation, it is important that they not be confused. As we stated in Section 2.1, the \((\mathcal{B}, \mathcal{D})\) notation will retain its organism model theory model throughout. So will the symbols \(\mathcal{D}\) and \(\mathcal{Z}\).

Similarly, \((\Delta, \mathcal{D})\) (and \((K, \mathcal{D})\), when we are talking about a cladogram as opposed to just any dendrogram) will retain its juncture model theory model throughout. The symbol \(\mathcal{D}\) is \(\mathcal{D}\) for the juncture model theory model (note parallel Definitions 3.1 and 5.2) and the symbol \(\mathcal{Z}\), to be defined below, is \(\mathcal{Z}\) for the juncture model theory model (note parallel Definitions 3.2 and 5.4). All other symbols will be used for either model theory model, where the context should make it clear
which model theory model is intended. This is done for convenience and clarity, but it does not imply that the two usages mean the same thing! For example, if \( x \) and \( y \) are organisms and \( (B, \triangleright) \) is monoparental, then the patriarch of \( x \) and \( y \) (if it exists), denoted \( x \circ y \), is an organism. That is, \( x \circ y \in B \). On the other hand, if \( S \) and \( T \) are juncures \( (i.e., \) members of \( \Delta \)\)), then the patriarch of \( S \) and \( T \), denoted \( S \circ T \), is a juncture. That is, \( S \circ T \in \Delta \) and \( S \circ T \subseteq B \). We already noted (Section 2.9) that "\( S \circ T \in B \)" and "\( S \circ T \subseteq B \)" do not mean the same thing, even though they may correspond to an identical (sloppy) usage in English: "the patriarch of \( S \) and \( T \) is 'in' \( B \)."

Theorem 5.6 shows that \( (\Delta, \triangleright) \) is a biocosm:

**Theorem 5.6** Let \( (\Delta, \triangleright) \) be a binary dendrogram on \( \Pi_x \). Then \( (\Delta, \triangleright) \) is a biocosm.

**Proof.** **Part I.** Claim: Axiom I is satisfied. Assume claim is false. That is, there exists an \( S \in \Delta \) such that \( S \triangleright S \). Then by Definition 5.1 (i), \( S \triangleright S \). Contradiction.

**Part II.** Claim: Axiom II is satisfied. Assume claim is false. That is, there exist \( S, T \in \Delta \) such that \( S \circ T \) and \( T \circ S \). Then by Theorem 5.1, \( S \triangleright T \) and \( T \triangleright S \). Contradiction.

In Theorem 5.7, we show that each juncture is a subset of \( X \):
Theorem 5.7. Let $(\Delta, \mathcal{D})$ be a binary dendrogram on $\Pi_X$. Then for each $S \in \Delta$, $S \subseteq X$.

Proof. Assume theorem is false. That is, there is an $S \in \Delta$ such that $\sim(S \subseteq X)$. Clearly, $S \not\subseteq X$. By Definition 4.1 (iv), $S \in \Pi_X$. By elimination, $S \in \Delta - (\Pi_X \cup \{X\})$. By Definition 5.3 (i), $X \not\supseteq S$. By Theorem 5.1, $X \supseteq S$. Contradiction.

A binary dendrogram is a monoparental biocosm:

Theorem 5.8. Let $(\Delta, \mathcal{D})$ be a binary dendrogram on $\Pi_X$. Then $(\Delta, \mathcal{D})$ is a monoparental biocosm.

Proof. Strategy. By Theorem 5.6, $(\Delta, \mathcal{D})$ is a biocosm. We must prove that for each $S \in \Delta$ there is at most one $T \in \Delta$ such that $T \supseteq S$. (I) Consider $S = X$. Then there is no such $T$. (II) Consider $S \in \Delta - \{X\}$. Then there is exactly one such $T$.

Part I. Consider $S = X$. Claim: there is no $T \in \Delta$ such that $T \supseteq S$. Assume claim is false. That is, there is a $T \in \Delta$ such that $T \supseteq S$. By Definition 5.1 (i), $T \supseteq S$. By Theorem 5.7, $S = X \supseteq T$. Contradiction.

Part II. Consider $S \in \Delta - \{X\}$. By Definition 5.3 (iii), there is a unique $T \in \Delta$ such that $T \supseteq S$.

Juncture $S$ is predecessor or equal to juncture $T$, denoted $S \supseteq T$, if and only if $S$ is a predecessor or $T$ or $S$ is the same juncture as $T$. Is-predecessor-or-equal-to is the term in the juncture model theory model which corresponds to the is-ancestor-or-equal-to term in the organism model theory model:
Definition 5.4. Let \((\Delta, \triangleright)\) be a binary dendrogram and \(S, T \in \Delta\).

Then \(S\) is **predecessor or equal to** \(T\), denoted \(S \triangleright T\), if and only if:

(i) \(S \triangleright T\)

or

(ii) \(S = T\).

Theorem 5.9 shows that if juncture \(S\) is a predecessor of juncture \(T\), then \(S\) is a proper subset of \(T\), and vice versa:

Theorem 5.9. Let \((\Delta, \triangleright)\) be a binary dendrogram on \(\Pi_X\), and \(S, T \in \Delta\).

Then \(S \triangleright T\) if and only if \(S \supset T\).

**Proof. Only If.** Proved in Theorem 5.1

If. **Strategy.** Consider any \(S, T \in \Delta\) such that \(S \supset T\).

Assume theorem is false. That is, \(\neg(S \supset T)\). (I) Show by mathematical induction that for any \(k \geq 0\), there exists a sequence \(S=R_0, \ldots, R_{k+1} \supset T\) such that for \(0 \leq i \leq k\), \(R_i \in \Delta\) and \(R_i \supset R_{i+1}\). Since for \(0 \leq i < j \leq k + 1\), \(R_i \notin R_j\), \(N(\Delta)\) is infinite. Contradiction.

Part I. Consider any \(S, T \in \Delta\) such that \(S \supset T\). Assume theorem is false. That is, \(\neg(S \supset T)\). Claim: for any \(k \geq 0\), there exists a sequence \(S=R_0, \ldots, R_{k+1} \supset T\) such that for \(0 \leq i \leq k\), \(R_i \in \Delta\) and \(R_i \supset R_{i+1}\).

True for \(k=0\): Let \(R_0 = S\) and \(R_1 = R_k = T\). By hypothesis, \(R_0 \in \Delta\) and \(R_0 = S \supset T = R_1\).

True for \(k\) implies true for \(k+1\). Since true for \(k\), let \(S=R_0', \ldots, R_{k+1}' \supset T\) be the sequence guaranteed by the inductive hypothesis. That is, for \(0 \leq i \leq k\), \(R_i' \in \Delta\) and \(R_i' \supset R_{i+1}'\). By assumption, \(\neg(S \supset T)\).
By Definition 5.2 (ii), there exists an $h$, $0 \leq h \leq k$, such that
$\sim (R'_h \bowtie R'_{h+1})$. By Definition 5.1 (ii), there exists a $R'_* \in \Delta$ such that $R'_h \bowtie R'_* \bowtie R'_{h+1}$. Construct the sequence $S = R'_0, \ldots, R'_{k+2} = T$ as follows. For $0 \leq i \leq h-1$, let $R'_i = R'_i$; by the inductive hypothesis,
$R'_i \bowtie R'_{i+1}$. Let $R'_h = R'_h$ and $R'_{h+1} = R'_*$ and $R'_{h+2} = R'_h$; by construction,
$R'_h \bowtie R'_{h+1} \bowtie R'_{h+2}$. For $h+2 \leq i \leq k+1$, let $R'_{i+1} = R'_i$; by the inductive hypothesis,$R'_i \bowtie R'_{i+1}$.

**Part II.** Choose any $k \geq 0$ and construct the sequence
$S = R'_0, \ldots, R'_{k+1} = T$ as in Part I. Consider any $i, j$ such that
$0 \leq i < j \leq k+1$. By transitivity of $\bowtie$, $R'_j \bowtie R'_i$. By elementary set theory, $R'_j \neq R'_i$. Hence $N(\Delta)$ has at least $k+1$ distinct members. But $k$ is arbitrary, hence $N(\Delta)$ is infinite. Contradiction of Theorem 5.5.

In Theorem 3.18, we demonstrated that, in a monoparental biocosm, if the patriarch for a pair of organisms exists at all, then it must be unique. Since $(\Delta, \bowtie)$ is a monoparental biocosm, by Theorem 5.8, the same must be true for the juncture model theory model: if the patriarch for a pair of organisms exists at all, then it must be unique. In Theorem 5.10 we show that the patriarch for any pair of junctures in a dendrogram always exists. This property is called closure by mathematicians.
Theorem 5.10 Let \( (\Delta, \mathcal{D}) \) be a binary dendrogram on \( \Pi_X \). Then for each \( S, T \in \Delta \):

\[
\begin{align*}
(1) & \quad \text{RP}([S, T]) \neq \emptyset \\
(2) & \quad S \circ T \in \Delta
\end{align*}
\]

Proof. Part I. By Theorem 5.7, \( X \supseteq S \) and \( X \supseteq T \). By Theorem 5.9, \( X \supseteq S \) and \( X \supseteq T \). By Definition 3.4, \( X \in \text{CP}([S, T]) \); hence \( \text{CP}([S, T]) \neq \emptyset \).

By Theorem 3.18, \( \text{RP}([S, T]) \neq \emptyset \).

Part II. By Theorem 5.8, \( (\Delta, \mathcal{D}) \) is a monoparental biocosm.

By Theorem 3.18, \( \text{RP}([S, T]) = \{R\} \) is a singleton set. By Definition 3.7, \( S \circ T = R \in \Delta \).

Theorem 5.11 shows that if \( \{S_1, S_2\} \) is a binary cladistic partition of \( S \), then \( S \) is the patriarch of \( S_1 \) and \( S_2 \):

Theorem 5.11 Let \( (\Delta, \mathcal{D}) \) be a dendrogram on \( \Pi_X \) and \( S \in \Delta - \Pi_X \) and \( \{S_1, S_2\} \) be the binary partition guaranteed by Definition 5.3 (iv).

Then \( S = S_1 \circ S_2 \).

Proof. By Definition 5.3 (iv), \( S \supseteq S_1 \) and \( S \supseteq S_2 \). By Definition 5.2 (i), \( S \supseteq S_1 \) and \( S \supseteq S_2 \). By Definition 3.4, \( S \in \text{CP}([S_1, S_2]) \).

Assume the theorem is false. Then by Definition 3.6, \( S \notin \text{RP}([S_1, S_2]) \).

Since \( S \in \text{CP}([S_1, S_2]) \), by Definition 3.5, there must be a \( T \in \Delta \) such that \( S \supset T \supseteq S_1 \) and \( S \supset T \supseteq S_2 \). By Theorem 5.1, \( S \supset T \supseteq S_1 \) and \( S \supset T \supseteq S_2 \).

Case I. \( S \supset T \supseteq S_1 \) and \( S \supset T \supseteq S_2 \). But \( S \supset S_1 \). Contradiction of Definition 5.1 (ii).
Case II. \( S \supseteq T \supseteq S_1 \) and \( S \supseteq T \supseteq S_2 \). But \( S \nsubseteq S_2 \). Contradiction of Definition 5.1 (ii).

Case III. \( S \supseteq T = S_1 \) and \( S \supseteq T = S_2 \). Then \( S_1 = T = S_2 \), and \( N([S_1, S_2]) = N([T]) = 1 \). Contradiction of Definition 4.2.

5.4 The Cladogram

A binary cladogram is a binary dendrogram whose non-OTU junctures reflect events of isolation. We say that a binary dendrogram is a binary cladogram if and only if to each non-OTU juncture there is a binary partition of that juncture consisting of two sets which are isolated from one another. Suppose that the binary dendrogram in Figure 5.1 were a binary cladogram. Then, for example, the non-OTU juncture \( H \) must consist of a union of two junctures (namely, \( A \) and \( B \)) which are isolated from one another. The existence of some binary partition for \( H \) is guaranteed by Definition 5.3, part (iv) (definition of a binary dendrogram). A binary cladogram requires that the member of this binary partition be isolated from one another.

Definition 5.5 \((K, \mathcal{D})\) is a binary cladogram on \( \Pi_X \) if and only if \((K, \mathcal{D})\) is a binary dendrogram on \( \Pi_X \) and for each \( S \in K - \Pi_X \), the binary partition \([S_1, S_2]\) of \( S \) is a binary cladistic partition.

An immediate result of this isolation property is that if \((K, \mathcal{D})\) is a binary cladogram, then (i) all junctures with no organisms in common are isolated from one another, and in particular, (ii) all OTUs are isolated from one another.
Theorem 5.12  Let \((K, \mathbf{p})\) be a binary cladogram on \(\Pi_X\). Then:

(i) for each \(S, T \in K\) such that \(S \cap T = \emptyset\), \(S\) is isolated from \(T\).

and

(ii) \(\Pi_X\) is a cladistic partition

Proof.  Part I.  Consider any \(S, T \in K\) such that \(S \cap T = \emptyset\). By Theorem 5.10, there is a \(W \in K\) such that \(W = S \circ T\). Strategy:
show that \(W \in K - \Pi_X\), and hence there is a binary cladistic partition, \(\{W_1, W_2\}\), of \(W\). Show that without loss of generality, \(W_1 \supseteq S\) and \(W_2 \supseteq T\), and hence \(S\) is isolated from \(T\).

Subclaim: \(W \supset S\). By Definition 3.4, \(W \supset S\) and \(W \supset T\).
Assume the subclaim is false. That is, \(W = S\). Then \(S \supset T\) and by
Theorem 5.1, \(S \supset T\). Since \(T \neq \emptyset\), \(S \cap T = T \neq \emptyset\). Contradiction of hypothesis.

Subclaim: \(W \in K - \Pi_X\). Assume subclaim is false. That is, \(W \in \Pi_X\). Choose any \(x \in S \subseteq X\). By Definition 5.1 (ii), there exists a \(P \in \Pi_X\) such that \(x \in P\). Hence \(P \cap S \neq \emptyset\). By Theorem 5.2, \(S \supset P\).
By the previous subclaim, \(W \supset S\). By Theorem 5.1, \(W \supset S\). By the transitivity of \(\supset\), \(W \supset P\). Since \(P \neq \emptyset\), \(W \cap P = P \neq \emptyset\) and by Theorem 4.1 (ii), \(W = P\). Contradiction.

Let \(\{W_1, W_2\}\) be the binary cladistic partition of \(W\) guaranteed by Definitions 5.3 (iv) and 5.5. Since \(S \subseteq W = W_1 \cup W_2\), either \(S \cap W_1 \neq \emptyset\) or \(S \cap W_2 \neq \emptyset\). Without loss of generality, let \(S \cap W_1 \neq \emptyset\).

Subclaim: \(W_1 \supset S\). Since \(S \cap W_1 \neq \emptyset\), by Definition 5.3 (ii), either \(W_1 \supset S\) or \(S \supset W_1\). Assume the subclaim is false. That is,

Subclaim: W₂ ⊆ T. Since T ⊆ W = W₁ ∪ W₂, either T ∩ W₁ ≠ ∅ or T ∩ W₂ ≠ ∅. Assume that T ∩ W₁ ≠ ∅. Then by an analogous argument as in the previous subclaim, W₁ ⊆ T. By Definition 3.4, W₁ ∈ CP({S,T}). Since W ⊆ W₁, by Definition 3.5, W ≠ RP({S,T}). Contradiction of Definition 3.7. By elimination, T ∩ W₂ ≠ ∅. By an analogous argument to that in the previous subclaim, W₂ ⊆ T.

Claim: S is isolated from T. From the above subclaims, W₁ ⊆ S and W₂ ⊆ T. By Theorem 5.1, W₁ ⊇ S ≠ ∅ and W₂ ⊇ T ≠ ∅. By Definition 5.5, W₁ is isolated from W₂. By Theorem 4.6, S is isolated from T.

Part II. Consider any P, Q ∈ Pₓ and P ≠ Q. By Definition 5.3 (i), Pₓ ⊆ K. By Theorem 4.1 (ii), P ∩ Q = ∅. By Part I, P is isolated from Q. By Definition 4.5, Pₓ is a cladistic partition.

The second part of Theorem 5.12 has clear-cut experimental implications: one can't expect to find a cladogram unless all the OTUs one starts with are isolated from one another. In very broad studies, this is usually done by letting initial OTUs be distinct species (Fitch and Margoliash, 1967), distinct genera (Moore and Goodman, 1968), or even higher taxonomic categories. In finer investigations, cladistic partitions are formed on the basis of distinct racial groups (Fitch and Margoliash, 1967; Edwards and Cavalli-Sforza, 1965), which may interbreed slightly, but where the
occasional exceptions to the noninterbreeding rule are considered negligible. All of these instances are more or less special cases of the cladistic partition concept developed in chapter four. All situations in which distinct OTUs are not isolated from one another in a nonnegligible way (such as the Goodman et al. (1970) macaques and Selander's (1967) house sparrows) are not really suitable for the methodology developed in this dissertation.

Theorem 5.13 shows that no juncture besides members of the unique binary cladistic partition can be an immediate predecessor of the original non-OTU juncture:

**Theorem 5.13** Let \((K, \mathcal{B})\) be a binary cladogram on \(\Pi_X\). Let \(S \in K - \Pi_X\) and \([S_1, S_2]\) be the unique binary partition of \(S\) guaranteed by Definition 5.3 (iv). Then for any \(T \in K\) such that \(S \triangleright T,\ T \in \{S_1, S_2\}\).

**Proof.** Assume theorem is false. That is, assume \(S \triangleright T\) and \(T \neq S_1\) and \(T \neq S_2\). By Definition 5.1 (i), \(T \subset S\). By Definition 4.1 (ii), \(S \subset S_1 \cup S_2\). By elementary set theory, either \(T \cap S_1 \neq \emptyset\) or \(T \cap S_2 \neq \emptyset\). Without loss of generality, let \(T \cap S_1 \neq \emptyset\). By Definition 5.3 (ii), either \(T \supset S_1\) or \(S_1 \supset T\). If \(T \supset S_1\), then \(S \triangleright T \supset S_1\) and by Definition 5.1 (ii), \(\sim(S \triangleright S_1)\); contradiction. If \(S_1 \supset T\), then \(S \supset S_1 \supset T\) and by Definition 5.1 (ii), \(\sim(S \triangleright T)\); contradiction.

In Theorem 5.14, we show that a cladogram is unique. This theorem is of major importance in later chapters, because it guarantees the uniqueness of results obtained from all the computing algorithms.
in this dissertation. That is, we will show that our computing algorithms always yield a cladogram as a result. Since there is one and only one cladogram, our computing algorithms all generate unique results. In this respect, the computing algorithms developed in this dissertation differ from the other major class of computing algorithms (i.e., parsimony methods), for which uniqueness is not guaranteed (see chapter six).

Theorem 5.14 Let \( (K, \mathcal{D}) \) be a cladogram on \( \Pi_X \). Then \( (K, \mathcal{D}) \) is unique.

Proof. Strategy. Assume theorem is false. That is, there is a dendrogram \( (\Delta, \mathcal{D}) \) on \( \Pi_X \) such that \( \Delta \not\subset K \) and \( (\Delta, \mathcal{D}) \) is a cladogram on \( \Pi_X \). (I) Show that there is a \( T \in \Delta \) such that \( T \not\in K \). (II) Construct the function \( g: \Delta - \{X\} \to \Delta - \Pi_X \) in the sense of Theorem 5.5. Let \( T_0 = T \) and for each \( k \geq 1 \), \( T_k = g(T_{k-1}) \). Show that there must be some \( k \geq 1 \) such that \( T_k \in K \). That is, \( T_k \in K \) and \( T_{k-1} \not\in K \). Show contradiction.

Part I. Claim: there is a \( T \in \Delta \) such that \( T \not\in K \). Assume claim is false. That is, for each \( T \in \Delta \), \( T \in K \). Then \( \Delta \subset K \). By assumption that theorem is false, \( \Delta \not\subset K \). Hence \( \Delta \subset K \). Since both \( \Delta \) and \( K \) are finite, \( N(\Delta) < N(K) \). By Theorem 5.5, \( N(\Delta) = N(K) \). Contradiction.

Part II. Construct the function \( g: \Delta - \{X\} \to \Delta - \Pi_X \) in the sense of Theorem 4.6. Let \( T_0 = T \) and for any \( k \geq 1 \), \( T_k = g(T_{k-1}) \). Claim: there must be a \( k \geq 1 \) such that \( T_k \in K \). Assume claim is
false. That is, for each \( k \geq 1 \), \( T_k \notin K \). Show by mathematical induction that for each \( k \geq 0 \), \( g(T_k) \) is defined.

True for \( k = 0 \): \( T_0 = T \in \Delta - K \). Since by Definition 5.3 (i), \( X \in K \), \( T_0 = T \in \Delta - \{X\} \). By construction, \( g(T_k) \) is defined.

True for \( k \) implies true for \( k+1 \): Since \( g(T_k) \) is defined, \( g(T_k) = T_{k+1} \) exists. By assumption, \( T_k \notin K \). Since by Definition 5.3 (i), \( X \in K \), \( T_{k+1} \in \Delta - X \). By construction, \( g(T_{k+1}) \) is defined.

Subclaim: for any \( k \geq 0 \) and any \( i, j \) such that
\[ 0 \leq i < j \leq k \], \( T_j \notin T_i \). By the above mathematical induction, both \( T_j \) and \( T_i \) exists. By construction of \( g \), \( T_j \sqsupseteq \ldots \sqsupseteq T_i \). By Definition 5.1 (i), \( T_j \supseteq \ldots \supseteq T_i \). By transitivity of \( \supseteq \), \( T_j \supseteq T_i \). By elementary set theory, \( T_j \notin T_i \).

Thus, \( N(\Delta - K) \) must contain at least \( k \) distinct members for any \( k \geq 0 \), namely, an infinite number of members. By Theorem 5.5, \( N(\Delta - K) \leq N(\Delta) = 2^{N(\Pi_X)} - 1 \) is finite. Contradiction.

Part III. Choose the smallest \( k \geq 1 \) such that \( T_k \in K \).
That is, \( T_k \in K \cap \Delta \) and \( T_{k+1} \in \Delta - K \). Since the range of \( g \) is \( \Delta - \Pi_X \), \( T_k \in \Delta - \Pi_X \) and \( T_k \in K - \Pi_X \). Let \( \{S_1, S_2\} \) be the unique binary partition of \( T_k \) such that \( \{S_1, S_2\} \subseteq \Delta \), guaranteed by Definition 5.3 (iv). By Theorem 5.13, \( T_{k-1} \in \{S_1, S_2\} \). Let \( \{R_1, R_2\} \) be the binary partition of \( T_k \) such that \( \{R_1, R_2\} \subseteq K \), guaranteed by Definition 5.3 (iv). Since \( T_{k-1} \notin K \), \( T_{k-1} \notin \{R_1, R_2\} \). Since \( \Delta \) and \( K \) are both cladograms, by Definition 5.5, \( \{S_1, S_2\} \) and \( \{R_1, R_2\} \) are both binary cladistic partitions of \( T_k \). By Theorem 4.9, \( \{S_1, S_2\} = \{R_1, R_2\} \). But then \( T_{k-1} \notin \{S_1, S_2\} \). Contradiction.
6. SOURCES OF INPUT DATA

6.1 The Phenetic Matrix

A phenetic matrix is any matrix each of whose members correspond to a pair of OTUs and attach a number to the "amount of similarity" or "amount of dissimilarity" for that OTU pair. In this chapter, we shall discuss a variety of experimental methods for obtaining a phenetic matrix. The common objective of all these experimental methods is to obtain a matrix such that an entry corresponding to a pair of OTUs with a relatively ancient ancestral separation is usually smaller (or usually greater) than an entry corresponding to a pair of OTUs with a relatively recent ancestral separation. As we argue in the next chapter, phenetic matrices obtained from molecular data (DNA comparisons and polypeptide comparisons) are expected to have the correspondence between an ancestral separation and value of phenetic entries mentioned above. In this chapter, we shall discuss how experimental molecular data can be transformed into a phenetic matrix. Most of these transformation procedures are relatively straightforward, and can be accomplished by the experimentalist with a hand calculator. The transformation procedure for immunodiffusion data is more involved and usually requires a computer; it is discussed in detail in this chapter, but may be skipped without loss of continuity.

Phenetic matrices fall into two major categories: similarity matrices and dissimilarity matrices. A similarity matrix has the following properties:
(i) The value of each OTU compared to itself is the maximum value in the matrix (usually 1)

(ii) The value of one OTU compared to another equals the value of the second OTU compared to the first

and

(iii) In general, a relatively high value for a comparison between two OTUs reflects a recent ancestral separation, whereas a relatively low value for a comparison between two OTUs reflects an ancient ancestral separation.

A dissimilarity matrix, on the other hand, has the following properties:

(i) The value of each OTU compared to itself is the minimum value in the matrix (usually 0)

(ii) The value of one OTU compared to another equals the value of the second OTU compared to the first

and

(iii) In general, a relatively low value for a comparison between two OTUs reflects a recent ancestral separation, whereas a relatively high value for a comparison between two OTUs reflects an ancient ancestral separation.

In this dissertation, we shall deal exclusively with phenetic matrices of the dissimilarity type, with minimum value zero. This is purely a convenience, because any dissimilarity matrix can be transformed into a dissimilarity matrix with minimum value zero simply by subtracting the minimum value in the original dissimilarity matrix from each individual value in the original dissimilarity matrix. Furthermore, any similarity matrix can readily be transformed into a dissimilarity matrix by the following
manipulation. Subtract each element in the similarity matrix from the diagonal (i.e., maximum) value for the similarity matrix. The resulting phenetic matrix is of the dissimilarity type with zero diagonal and nonnegative offdiagonal elements. Finally, any matrix which satisfies all the properties of a dissimilarity matrix except property (ii) (i.e., that the \((i,j)\)th value equals the \((j,i)\)th value) can be transformed as follows. Take an arithmetic average of the \((i,j)\)th and the \((j,i)\)th values in the original matrix. Then replace the original \((i,j)\)th and \((j,i)\)th values with this arithmetic average. If a matrix already satisfies (ii), this averaging technique has no effect at all on the matrix. (Strictly speaking, this technique is applicable only when the \((i,j)\)th and \((j,i)\)th values are different because of a relatively small experimental error, which may not always be the case, especially with immunological data. Other, more suitable techniques for making an input matrix satisfy property (ii) lie beyond the scope of this dissertation.)

Each entry in a phenetic matrix is a real number which corresponds to a pair of OTUs. Since the number of OTUs in any study is finite, the OTUs can be numbered from 1 to \(n\). The phenetic matrix is an \(n \times n\) matrix whose rows and columns correspond to the numbered OTUs. The phenetic matrix is usually represented by a capital letter, say \(M\), and an individual entry is represented by the symbol \(m_{ij}\) where \(i\) corresponds to the \(i\)th OTU, \(j\) to the \(j\)th OTU, and \(m_{ij}\) means "the real value in the phenetic matrix which corresponds to the \(i\)th OTU compared to the \(j\)th OTU". In this dissertation, we shall employ an unconventional notation for representing the entries
of a phenetic matrix. The inconvenience in learning this notation is partially vindicated in later proofs, where the conventional notation proves to be relatively cumbersome. It is important to note that the two notations are formally the same; the entire dissertation could have been written in either notation and still be formally correct.

As we said before, each entry of a phenetic matrix is a real number which corresponds to a pair of OTUs. In the notation of this dissertation, the individual OTUs (that is, members of $\Pi_X$) are not numbered, but named by upper case Roman letters (since they are sets of organisms), say $P$ or $Q$. The entire phenetic matrix is represented by the symbol $\langle, \rangle$; the individual entry which corresponds to "OTU $P$ compared to OTU $Q$" is represented by the symbol $\langle P, Q \rangle$.

Properties (i) and (ii) of a (dissimilarity type, zero diagonal) phenetic matrix may be summarized in the following, informal definition:

**Informal Definition** Let $\langle, \rangle$ be a matrix defined over all pairs of OTUs. Then for any OTUs $P$ and $Q$:

(i) the phenetic value of an OTU compared to itself, such as $\langle P, P \rangle$ or $\langle Q, Q \rangle$, is zero

and

(ii) the phenetic value of one-OTU-compared-to-a-second equals the value of the second-compared-to-the-first, that is $\langle P, Q \rangle = \langle Q, P \rangle$, and both are nonnegative, that is $\langle P, Q \rangle = \langle Q, P \rangle \geq 0$.

Property (iii) is much more subtle, and its discussion is deferred to chapter seven. The formal definition of a dissimilarity type phenetic matrix closely parallels the informal definition:
Definition 6.1 \( \langle \cdot, \cdot \rangle \) is a phenetic matrix if and only if \( \Pi_X \) is a partition and \( \langle \cdot, \cdot \rangle \) is a real-valued function with domain \( \Pi_X \times \Pi_X \) and for each \( P, Q \in \Pi_X \):

(i) \( \langle P, P \rangle = \langle Q, Q \rangle = 0 \)

and

(ii) \( \langle P, Q \rangle = \langle Q, P \rangle \geq 0 \).

6.2 Comparisons of Character States

Any pair of OTUs which is to be compared must share a spectrum of characters in common. For example, suppose we were going to compare the author of this dissertation with a gorilla, Maximo, at the Detroit Zoological Park. These two individuals share the following characters in common: height, mass, hair color, and the 104th amino acid residue (Dayhoff's (1969) alignment) of the beta hemoglobin molecule (henceforth called "104\%h\%b"). However, these individuals do not display the same character states. Let us form the OTUs \{Moore\} and \{Maximo\}. Then we may summarize the character states for these OTUs in little, rectangular boxes as follows:

<table>
<thead>
<tr>
<th></th>
<th>Height</th>
<th>Mass</th>
<th>Hair Color</th>
<th>104%h%b</th>
</tr>
</thead>
<tbody>
<tr>
<td>{Moore}</td>
<td>1.91 m</td>
<td>110 kg</td>
<td>red</td>
<td>ARG</td>
</tr>
<tr>
<td>{Maximo}</td>
<td>1.52 m</td>
<td>273 kg</td>
<td>black</td>
<td>LYS</td>
</tr>
</tbody>
</table>

Two things should be immediately apparent regarding this short investigation we have just completed: first, the OTUs we have selected aren't really of very much general interest, and second, at least three of the characters we have chosen don't tell us much
about the common ancestry of our OTUs. In answer to the first objection, we may select the two OTUs *Homo sapiens* and *Gorilla gorilla*, which are of fairly general interest. Of course,

Moore e *Homo sapiens*

and

Maximo e *Gorilla gorilla*.

But now we have destroyed three of our characters, because "height", "mass", and "hair color" are not characters common to our two OTUs. The closest we are likely to come is "range of heights", "range of masses", and "range of hair color" -- or perhaps some other, appropriate statistical description.

Even our new characters -- range of heights, range of masses, and range of hair color -- are inappropriate, because they do not give us any insight into the ancestral relationships of man, gorilla, and other OTUs we may wish to study. In order to be able to get property (iii) for our dissimilarity matrix, we want characters which diverge slowly (so that the entire range of variation for that character will not be exhausted over a narrow ancestral range of OTUs) and which, once diverged, do not return (or at least do not return very often) to a more primitive state. Thus, if we sample a large enough spectrum of such characters, the number of characters at which character states differ will increase as the time of ancestral separation of OTUs becomes more ancient. This increase need not be linearly related (or any other specific functional relationship) to time, but must be some steadily increasing relationship with respect
to time. In chapter seven, we shall suggest why $\lambda\alpha^\beta^\gamma^\delta$, along with hundreds more such characters, might be considered especially appropriate for ancestral studies.

6.3 Direct Techniques for Obtaining a Phenetic Matrix

The objective of molecular methods in numerical taxonomy is to intercept the genetic message with enough accuracy and over a broad enough portion of the genome such that the convergence among a few characters is overwhelmed by the divergence among all the rest. Ideally, we would like to know the DNA nucleotide sequence for the entire genome of any desired individual; for practical purposes, this is impossible. For one thing, techniques are not available for obtaining the exact nucleotide sequence of DNA. However, there are methods for comparing two different strands of DNA, which we shall discuss below; furthermore, there are indirect methods for sampling some (but not all) of the nucleotide sequences and sequence differences on the DNA molecule.

DNA hybridization looks directly at the amount of correspondence between two single strands of the DNA molecule from different organisms. Amino acid sequencing looks at the primary structure of polypeptide molecules. This technique promises to be the leader in the molecular systematics field in years to come, but like DNA hybridization, is still rather time-consuming, and has been applied only to a few representatives covering a broad range of organisms with rather limited interest to systematics. Immunological techniques,
which sample small patches upon a polypeptide's surface, yield data that probably reflect changes among from one to four amino acid residues each at the primary structure level. Immunological techniques are far and away the easiest to perform on a routine basis in biological laboratories, and have yielded the most extensive results to date in molecular systematics. The principal techniques being used presently to draw systematic conclusions are summarized below.

DNA hybridization (McCarthy and Bolton, 1963; Britten and Kohne, 1968) exploits the fact that the double-stranded DNA molecule separates into single strands above a certain temperature and that these strands reassociate at a characteristic temperature. Re-association of the double-stranded DNA molecule is called annealing. The annealing temperature of single-stranded DNA from a single OTU source is relatively high, but the presence of impurities from other OTUs lowers the temperature. That is, if a single strand of DNA is forces to anneal with another strand which doesn't "match" (i.e., have a complementary sequence of nucleotides), then the annealing temperature is lowered. The amount by which the annealing temperature is lowered can be related to the degree of mismatching. Thus, if single-stranded human (Homo sapiens) DNA, say, is suspended in an agar medium (so that it can't reassociate with itself) and rhesus (Macaca mulatta) DNA is passed over the agar suspension, the annealing temperature will be lowered in proportion to the percent of
mismatches. The preliminary phenetic value for this comparison is simply this percentage of mismatches, and is denoted $\langle\text{Homo sapiens, Macaca mulatta}\rangle$.

In general, the $\langle\text{Macaca mulatta, Homo sapiens}\rangle$ value obtained from DNA hybridization studies (that is, where rhesus DNA is suspended in agar and human DNA is passed overhead) will differ from the $\langle\text{Homo sapiens, Macaca mulatta}\rangle$ value within the limits of experimental error. Thus, in order to satisfy property (ii) of a phenetic matrix, the averaging procedure from Section 6.1 should be followed.

**Amino Acid Sequencing** (Mahler and Cordes, 1966, pp. 71-85) exploits the availability of reagents which can split off the terminal amino acid from either the N-terminal or C-terminal end of a polypeptide chain. A single, terminal amino acid is split off and assayed by chromatographic techniques; the remaining polypeptide fragment is again split at its newly exposed terminal amino acid, and the assay procedure is repeated. A skilled, experienced worker can usually work from five to eight fragments in from either the N-terminal or C-terminal end, until the yield of retrieved polypeptide fragment is so low as to preclude further determinations.

Long polypeptides (longer than about ten residues) are broken into smaller fragments by digestive enzymes such as trypsin, chymotrypsin, and pepsin. Purified trypsin cleaves the polypeptide specifically at the C-terminal ends of all the ARG and LYS residues; chymotrypsin and pepsin are less specific, but tend to cleave at the C-terminal ends of aromatic amino acids, such as HIS, TYR, and PHE.
(Margoliash and Smith, 1961; Kreil and Tuppy, 1961). Since different digestive enzymes cleave the total polypeptide at different positions, it is possible to work four or five terminal residues in both directions from all the ARG and LYS residues with a trypsin digest; four or five terminal residues in both directions from all the HIS, TYR, and PHE residues with chymotrypsin of pepsin digests; and so forth. Once enough different regions of the total polypeptide have been worked out using the different digests, the structure of the total polypeptide can be reassembled.

Each individual amino acid sequence is displayed as the sequence of three-letter amino acid symbols from left (C-terminal) to right (N-terminal) on the printed page. Each pair of sequences is laid out by the investigator so as to minimize the number of mismatches between the two sequences; if necessary, one of the sequences may be cut and slid sideways to minimize the number of mismatches. The correspondence pattern between the two sequences is called the alignment for that pair of sequences. Each alignment position (i.e., individual pair of residues) is considered a separate character; amino acid sequences are one of the few forms of molecular data in which individual characters can be identified. The distance between two character states (i.e., two amino acid residues at a single alignment position) is usually taken as the minimum mutation distance, that is, the minimum number of nucleotide changes necessary to translate a codon (i.e., a nucleotide triplet) for one of the residues into a codon for the other residue. For example, residue MET has a minimum mutation distance of one from residue ILE, because
AUG is the MET codon and AUC is an ILE codon, and these codons differ at a single nucleotide (the third). The value for each entry of the phenetic matrix is simply the sum of the minimum mutation distances over all characters, divided by the total number of characters. Since the minimum mutation distance for any pair of amino acid residues satisfies property (ii), the entries into the phenetic matrix automatically satisfy property (ii) without further adjustment.

**Immunological methods** rely upon the ability of the rabbit (or some other suitable experimental animal) to produce specific antibodies against small patches (one to four amino acid residues in length: see Cocks and Wilson, 1969; Arnheim et al, 1969) on the surface of the polypeptide molecule. Since these surface patches are most likely hydrophilic, and thus do not determine the major structural underpinnings of the molecule, they are thought to be subject to low levels of selection (see chapter seven for details). All immunological methods have the common feature that a rabbit is stimulated to prepare antibodies against one or more proteins from OTU$_i$, and then the rabbit antibodies are reacted, directly or indirectly, with corresponding proteins from OTU$_j$. The amount of anti-OTU$_i$ which fails to react with OTU$_j$ protein is considered as the level of mismatches, and therefore measures the amount of dissimilarity between OTU$_i$ and OTU$_j$.

Immunological methods differ with respect to how they estimate the level of mismatches (or matches: one is just a constant minus the other). **Radioimmune inhibition** (Wang et al., 1968) employs a radioactively labelled sample of protein from OTU$_i$. First, anti-OTU$_i$ is reacted to exhaustion with OTU$_j$ protein. The remaining anti-OTU$_i$
is reacted with a known amount of radioactive OTU₁ protein. The amount of OTU₁ protein which is so removed by the (OTUᵢ-exhausted)-anti-OTUᵢ is assessed by radioactivity measurements, and considered as the amount of mismatches.

Complement fixation (Sarich and Wilson, 1966; Cushing and Campbell, 1957) estimates the level of mismatches between OTUᵢ protein and OTUᵢ protein by focussing upon a concomitant substance, called complement, K, which is simultaneously used up in any immunological reaction. (Strictly speaking, anti-OTUᵢ alone will not precipitate OTUᵢ protein, but anti-OTUᵢ + K will. We have ignored K heretofore, because it was assumed present in excess.) A good example of such a reaction is the hemolysis (disintegration) of sheep erythrocytes, SE. This reaction is thought to proceed as follows:

$$\text{Anti-SE} + \text{SE} + \text{K} \rightarrow \text{hemolyzed SE}$$

where the quantity of K which is used up matches the quantity of SE hemolyzed. The usual sources for the components in this reaction are as follows:

- washed sheep erythrocytes: only SE
- fresh serum from rabbit immunized with SE: anti-SE + K
- serum from rabbit immunized with SE, preheated to 60° C: only anti-SE
- washed sheep erythrocytes, pretreated with immunized-preheated rabbit serum ("sensitized sheep erythrocytes"): Anti-SE + SE
fresh guinea pig serum
protein from OTU₁
preheated rabbit antibody against OTU₁ protein:
protein from OTUₖ:

only K
OTU₁ protein
anti-OTU₁
only OTUₖ protein

In the first reaction of the complement fixation test (Figure 6.1), anti-OTU₁ is reacted with OTUₖ protein in the presence of a known amount (excess) of guinea pig serum:

\[
\text{anti-OTU}_1 + \text{OTU}_k \text{ protein } + \text{excess } K \rightarrow \text{(anti-OTU}_1 + \text{OTU}_k \text{ protein)} \text{ precipitate } + \text{leftover } K
\]

The amount of K used up (i.e., original K minus leftover K) is proportional to the amount of anti-OTU₁ which was able to crossreact with OTUₖ protein, namely, the matches. Sensitized sheep erythrocytes are added in known amount (excess) to obtain the following reaction:

\[
\text{leftover } K + \text{excess (anti-SE } + \text{ SE)} \rightarrow \text{ partially hemolyzed SE } + \text{ leftover (anti-SE } + \text{ SE)}
\]

This partially hemolyzed SE is directly proportional to the amount of leftover K, which can in turn be subtracted from the original amount of K, to estimate the level of matches. The amount of hemolyzed SE can be measured by centrifuging away the (anti-SE + SE) and measuring the hemolyzed SE spectrophotometrically at 670 nm. Thus, the amount of material which absorbs at 670 nm is proportional to the amount of hemolyzed SE, which can be used to estimate the amount of leftover K, which can in turn be subtracted from the original amount of K to estimate the level of matches.
Figure 6.1. Complement fixation. A simplified account of the sequence of steps in the complement fixation technique.
**Immunoelectrophoresis** (Williams and Grabar, 1955) takes advantage of the differing electrical charge of distinct components of a protein to identify the separate components. OTU$_i$ protein and OTU$_j$ protein are placed in adjacent small wells at the top of a long rectangular agar field, separated by a long thin vertical well in between. The antigens are electrophoresed to separate distinct components; then anti-OTU$_i$ is placed in this middle well. Anti-OTU$_i$ diffuses outward from the middle well until it meets the structural components which have been electrophoresed from OTU$_i$ protein and OTU$_j$ protein. Anti-OTU$_i$ will precipitate all components from OTU$_i$ protein, but only those components from OTU$_j$ protein which match those of OTU$_i$ protein. The components precipitated on the OTU$_i$ side but not on the OTU$_j$ side of the antibody well represent the number of component mismatches.

In all the immunological methods discussed in this chapter, the mismatch count is entered as the (i,j)th entry of the preliminary phenetic matrix. As in the DNA hybridization technique, the (j,i)th entry in this preliminary phenetic matrix is expected to be within experimental error of the (i,j)th value, but hardly ever exactly the same. Thus, in order to satisfy property (ii) of a phenetic matrix, the averaging procedure of Section 6.1 should be followed.
Figure 6.2. Immunoelectrophoresis. Diagram of a completed immunoelectrophoresis plate.
6.4 The Immunodiffusion Technique

The immunodiffusion technique has been applied to the largest variety of species of all the molecular methods for creating phenetic matrices. This technique, unlike the others discussed above, does not yield a direct readout of mismatch level; it requires substantial preliminary processing, which is usually done by computer.

For the remainder of this chapter, I shall present the details of solving for a phenetic matrix from a suitable experimental design of immunodiffusion comparisons. Proper derivation of an immunodiffusion theory from intuitive set theory foundations is beyond the scope of this dissertation and peripheral to the central purpose of this dissertation. Accordingly, I shall assume two additional primitives and several of their important properties (discussed more fully in Moore and Goodman, 1968) as axioms. This portion of the axiomatic development does not have a direct bearing on the main line of argument in this dissertation: it merely shows how to construct a phenetic matrix from a highly specialized source of data. Since a phenetic matrix (as defined in Definition 6.1, and independent of the specialized axioms below) from any source is suitable, the specialized theory to be developed below does not affect the main flow of thought in this dissertation. Readers who are not especially interested in the immunodiffusion technique may skip directly to chapter seven without loss of continuity.

In order to keep the experimental implications of our immunodiffusion theory clearly at hand, we shall develop a simple example alongside the theory. Four OTUs are used, with data from
Moore et al. (1969):

\[
\begin{align*}
\text{OTU}_1 &= \text{Pongo pygmaeus} \text{ (Orangutan)} \\
\text{OTU}_2 &= \text{Hylobates lar} \text{ (Gibbon)} \\
\text{OTU}_3 &= \text{Macaca mulatta} \text{ (Rhesus)} \\
\text{OTU}_4 &= \text{Homo sapiens} \text{ (Man)} 
\end{align*}
\]

The basic experimental device for the immunodiffusion technique is the trefoil Ouchterlony plate, illustrated diagrammatically in Figure 6.3. This plate consists of three wells, bottom, top left, and top right, surrounding a block of agar. The most general, unit experiment of the immunodiffusion technique is called the heterologous comparison, and is carried out as follows. Anti-\(\text{OTU}_n\) is placed in the bottom well and \(\text{OTU}_j\) protein and \(\text{OTU}_k\) protein are placed in the top left and top right wells, respectively. If allowed to stand in an appropriately moist atmosphere for several days, the substances in the three wells will diffuse toward one another into the agar block, until they meet and form a characteristic band of precipitation. At first, bands of precipitation will form on the lower left and lower right corners, where the reactants have had to diffuse the shortest distance before they meet (Figure 6.3 A).

The lower left band represents those antibodies which have precipitated components of \(\text{OTU}_j\) protein, namely, the components common to \(\text{OTU}_n\) and \(\text{OTU}_j\). Similarly, the lower right band represents those antibodies which have precipitated components of \(\text{OTU}_k\) protein, namely the components common to \(\text{OTU}_n\) and \(\text{OTU}_k\). As these two corner bands approach one another and join at the center of the agar block, however, an
Figure 6.3. The trefoil Ouchterlony plate. (A) Charging the trefoil Ouchterlony plate. (B) Early diffusion pattern. (C) Final diffusion pattern.
Components common to OTU_n and OTU_j but not OTU_k

Components common to OTU_n and OTU_k but not OTU_j
additional pattern forms. The components common to \( OTU_n \), \( OTU_j \), and \( OTU_k \) form a continuous band spanning from the lower left corner to the lower right corner. This common band is called the "A-band" because of its shape. However, the components common to \( OTU_n \) and \( OTU_j \) but not \( OTU_k \) will not arch over to the other side, but continue to grow outward toward the upper right corner of the plate in the form of a precipitation outgrowth, or left spur. Similarly, the components common to \( OTU_n \) and \( OTU_k \) but not \( OTU_j \) form the right spur, directed toward the upper left corner. An Euler-Venn diagram is used in Figure 6.3 to clarify these relationships: vertical hatching corresponds to the left spur and horizontal hatching to the right spur. Both the left and right spurs are judged by the investigator in terms of size and intensity, and given a numerical value ranging from 0 to 5. The heterologous comparison value is the value for the left spur minus the value for the right spur. Biologically, it tells us "how much more \( OTU_j \) has diverged from \( OTU_n \) than \( OTU_k \) has". It is entered as the \((j,k)\)th value in a comparison matrix.

We shall now discuss a specialized theory which can be used to transform these comparison matrices into a phenetic matrix. All primitives, axioms, definitions, and theorems which are specifically related to this immunodiffusion theory are indicated with a star(*).

The set of OTUs, \( \Pi_X \), is a finite set consisting of \( N(\Pi_X) \), or simply \( n \), members: \( \Pi_X = \{P_1, \ldots, P_n\} \). The phenetic matrix, \( \langle \cdot, \cdot \rangle \), is solved for one row at a time. Let \( P_n \) be the "homologous OTU" and \( P_l \), \( 1 \leq l \leq n-1 \), be the "heterologous OTUs". The procedure to be outlined below shows how to solve for \( \langle P_n, P_l \rangle \) for each \( 1 \leq l \leq n-1 \) (by Definition 6.1,
\langle P_n', P_n \rangle = 0). This gives us the nth row of the phenetic matrix.

Since the numbering convention for \( \Pi_X = \{P_1', \ldots, P_n'\} \) is arbitrary, one can simply renumber (giving a different row the special property of the nth, or homologous row). Eventually, all rows are exhausted and the entire phenetic matrix is solved. For brevity in stating the starred statements in this chapter, we shall uniformly assume that \( \Pi_X \) is a partition having \( n \) members \( P_1', \ldots, P_n' \), and that \( P_n' \) is the particular OTU which we have singled out as the "homologous OTU" (although any other OTU could equally well have been chosen as "homologous"). This assumption is implicit in the hypothesis of every starred primitive, axiom, definition, and theorem remaining in this chapter.

The comparison matrix, \( H = (h_{jk}) \), where \( 1 \leq j, k \leq n \), is one of our starred primitive concepts:

**Primitive**\( H = (h_{jk}) \) \( n \) by \( n \) comparison matrix.

In a previous publication (Moore and Goodman, 1968), Prof. Goodman and I proved the following properties of \( H \) as theorems based upon intuitive set theory foundations. Here, we state these properties as starred Axioms III* and IV*. Axiom III* (i) states that the comparison value of and OTU to itself is zero; Axiom III* (ii) states that the comparison value of one OTU to another is the negative of the comparison value of the second OTU to the first (this exchange amounts to switching the left-right conventions):

**Axiom III** Let \( 1 \leq j, k \leq n \). Then
(i) \( h_{jj} = h_{kk} = 0 \)
and
(ii) \( h_{jk} = -h_{kj} \)

Axiom IV* (i) states that the nth row of the comparison matrix is the nth row of the phenetic matrix: intuitively, this says that the number of antibody components common to \( \text{OTU}_n \) but not present in \( \text{OTU}_j \) is proportional to the number of mismatches between \( \text{OTU}_n \) and \( \text{OTU}_j \). Axiom IV* (ii) states that if one adds the heterologous comparison value \( h_{jk} \) to the \((n,j)\)th entry in the phenetic matrix, then the result is the \((n, k)\)th phenetic matrix value. This additive property of heterologous comparison value is observed experimentally within the limit of accuracy expected for the Ouchterlony technique (Moore and Goodman, 1968):

**Axiom IV**

Let \( 1 \leq j, k \leq n-1 \). Then

(i) \( \langle P_n, P_j \rangle = h_{nj} \)
and
(ii) \( \langle P_n, P_j \rangle + h_{jk} = \langle P_n, P_k \rangle \)

Only rarely does an experimenter obtain an entire matrix of heterologous comparisons, and as we shall show, if a simple solvability condition is satisfied, this enormous effort is not really necessary. A protocol matrix, \( E = ((e_{jk})) \) is a record of which comparisons were done and which were not. It is our other starred primitive:

**Primitive**

\( E = ((e_{jk})) \)

\( n \) by \( n \) protocol matrix
We adopt the convention that $e_{jk} = 0$ if the comparison from $OTU_j$ to $OTU_k$ (producing element $h_{jk}$) was not done and $e_{jk} = 1$ if the comparison from $OTU_j$ to $OTU_k$ (producing element $h_{jk}$) was done. Since the comparison of an OTU to itself yields no new information (since $h_{jj} = 0$ by Axiom III* (I)), we assume that the diagonal elements of the protocol matrix (that is, $e_{jj}$ for $1 \leq j \leq n$) are all zero. Since the $h_{jk}$ and $h_{kj}$ comparisons are redundant (since one is the negative of the other, by Axiom III* (ii)), we assume that symmetric elements ($e_{jk}$ and $e_{kj}$) are equal, that is $h_{jk}$ was done if and only if $h_{kj}$ was done. These properties are summarized in Axiom V*:

**Axiom V***  
Let $1 \leq j, k \leq n$. Then

(i) $e_{jj} = e_{kk} = 0$

and

(ii) $e_{jk} = e_{kj} \geq 0$

A comparison matrix for our sample set of OTUs is shown in Figure 6.4 A. Values for comparisons are given to two decimal places, based on averages of repeated trials; comparisons which were not made are indicated with a question mark. Figure 5.4 B shows a corresponding protocol matrix: 1's for comparisons made and 0's for comparisons not made.

The axioms for both starred primitives -- the comparison matrix and the protocol matrix -- correspond closely to the properties of experimental data from the real, biological world. Since all of the subsequent concepts (including a formula for obtaining the phenetic matrix) in the immunodiffusion theory can be derived from these axioms
Figure 6.4. Experimental design for sample immunodiffusion data. The homologous OTU (OTU₄) is *Homo sapiens*. The heterologous OTUs are OTU₁ = *Pongo pygmaeus*, OTU₂ = *Hylobates lar*, and OTU₃ = *Macaca mulatta*.  
(A) The comparison matrix. (B) The protocol matrix.
by well-defined mathematical operations, we have a system which, within experimental error, is guaranteed to reflect the properties of the real, biological world.

6.5 Solution for the Phenetic Matrix

The first step in solving a matrix of heterologous comparisons for the phenetic matrix involves the transformation of the protocol matrix, E, into a primary solution matrix, S. The protocol matrix, E, is an n by n square matrix (that is, one row and one column corresponding to each of the OTUs in our study, heterologous or homologous). By contrast, the primary solution matrix is an (n-1) by (n-1) square matrix, having a row and a column corresponding to each heterologous OTU, but no row or column for the homologous OTU. Each diagonal element in the primary solution matrix is minus one times the sum of all elements in the corresponding row (or column) of the protocol matrix. Each off-diagonal element in the primary solution matrix is identical to its corresponding element in the protocol matrix:

Definition 6.2* Matrix $S = (s_{jk})$ is the primary solution matrix (corresponding to the protocol matrix, E) if and only if for

$1 \leq j \neq k \leq n-1$:

(i) $s_{jj} = -\sum_{i=1}^{n} e_{ji}$

and

(ii) $s_{jk} = e_{jk}$

We need to define two more concepts before we proceed. The (n-1)-vector called $D = (d_{j})$ is the vector defined over the (n-1) heterologous OTUs such that the jth component is the scalar product of the jth row of the protocol matrix and the jth row of the comparison matrix:
Definition 6.3* \( D = (d_j) \) is the vector such that for \( 1 \leq j \leq n-1 \):

\[
d_j = \sum_{i=1}^{n} e_{ji} h_{ji}
\]

We can conveniently represent the \( n \)th row of the phenetic matrix (which we hope to solve for) as another \((n-1)\)-vector over the heterologous OTUs, namely \( C = (c_j) \):

Definition 6.4* \( C = (c_j) \) is the vector such that for \( 1 \leq j \leq n-1 \),

\[
c_j = P_n, P_j .
\]

Since the protocol matrix, \( E \), and the matrix of comparisons, \( H \), are available to us from our sample experimental data, we are in a position to obtain values for the \((n-1)\)-vector, \( D \):

\[
d_1 = (1)(-2.22) + (0)( ? ) + (0)( ? ) + (1)(2.50) = 0.28
\]
\[
d_2 = (1)(-3.10) + (0)( ? ) + (0)( ? ) + (1)(2.00) = -1.10
\]
\[
d_3 = (1)(-3.50) + (1)(-2.50) + (1)(-2.00) + (0)( ? ) = -8.00
\]

\[
D = \begin{bmatrix}
0.28 \\
-1.10 \\
-8.00
\end{bmatrix}
\]

We also have our 3 by 3 primary solution matrix \( S \) (Figure 6.5) as input data. According to Theorem 6.1* (below), the following equation is true:

\[
S C = D
\]
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>-3</td>
</tr>
</tbody>
</table>

Figure 6.5 The protocol and primary solution matrix for sample data. (A) The protocol matrix, E. (B) The primary solution matrix, S.
On the condition that $S$ has an inverse (this is the solvability condition to be discussed below), this formula becomes:

$$C = S^{-1} D$$

Since $S^{-1}$ (if it exists) can be calculated from $S$, and $D$ is already available from our protocol and comparison matrix data, we can solve for $C$:

Calculate $S^{-1}$:

$$S^{-1} = egin{bmatrix} -0.625 & -0.125 & -0.250 \\ -0.125 & -0.625 & -0.250 \\ -0.250 & -0.250 & -0.500 \end{bmatrix}$$

Solve for $C$:

$$C = \begin{bmatrix} -0.625 & -0.125 & -0.250 & 0.28 \\ -0.125 & -0.625 & -0.250 & -1.10 \\ -0.250 & -0.250 & -0.500 & -8.00 \end{bmatrix} = \begin{bmatrix} 1.9625 \\ 2.6525 \\ 4.2050 \end{bmatrix}$$

That is:

$$\langle P_4, P_1 \rangle = \langle \text{Homo sapiens}, \text{Pongo pygmaeus} \rangle = 1.9625$$

$$\langle P_4, P_2 \rangle = \langle \text{Homo sapiens}, \text{Hylobates lar} \rangle = 2.6525$$

$$\langle P_4, P_3 \rangle = \langle \text{Homo sapiens}, \text{Macaca mulatta} \rangle = 4.2050$$

Theorem 6.1* justifies this kind of calculation in the general case:
Theorem 6.1* Let $S = (s_{jk})$, $D = (d_j)$, and $C = (c_j)$ be defined as in Definitions 6.2*, 6.3*, and 6.4*, respectively. Then $S \cdot C = D$.

Proof. By Definition 6.4* and Axiom IV**(ii), $c_k - c_j = h_{jk}$ for $1 \leq j, k < n$. Multiplying by $e_{jk}$:

$$e_{jk} c_k - e_{jk} c_j = e_{jk} h_{jk}$$

Summing over all $1 \leq k \leq n$:

$$\sum_{k=1}^{n} e_{jk} c_k - c_j \sum_{k=1}^{n} e_{jk} = \sum_{k=1}^{n} e_{jk} h_{jk} \quad (*)$$

By Definition 6.2* (i):

$$s_{jj} c_j = -c_j \sum_{k=1}^{n} e_{jk}$$

By Definition 6.2* (ii) for $1 \leq j \neq k \leq n-1$:

$$s_{jk} c_k = e_{jk} c_k$$

By Definition 6.3*:

$$d_j = \sum_{k=1}^{n} e_{jk} h_{jk}$$

By Axiom V* (i):

$$e_{jj} c_j = 0.$$
Substituting into equation (*):

\[ \sum_{k=1}^{n} s_{jk} c_k + e_{jj} c_j + s_{jj} c_j = \sum_{k=1}^{n} s_{jk} c_j = d_j. \]

That is, \( S \cdot C = D \).

### 6.6 Inversion of the Primary Solution Matrix

In the sample data set from the previous section, we assumed that our primary solution matrix, \( S \), was invertible, and went glibly along and solved for \( C \). If \( S \) hadn't been invertible, one of two things would have happened: (i) In inverting \( S \), we would have been forced to divide by zero, which would of course terminate our calculations. (ii) Roundoff error in the course of calculation would disguise the necessity for dividing by zero (rather: we would divide by some extremely small number), and the resulting value for \( C \) would be meaningless. In practice, there is a very simple way to check a protocol matrix for whether or not its corresponding primary solution matrix will invert (i.e., to check whether we have gathered enough data). We simply check all heterologous OTUs to make sure they are what we will call connected to the homologous OTU.
Consider the hypothetical protocol matrix \( E \) in Figure 6.6.

We say that "OTU_1 is directly connected to OTU_5" because element \( e_{15} = e_{51} \) in the protocol matrix is greater than zero, i.e., the comparison between OTU_1 and OTU_5 has been made. Similarly:

- OTU_2 is directly connected to OTU_1 because \( e_{21} = e_{12} > 0 \)
- OTU_3 is directly connected to OTU_4 because \( e_{34} = e_{43} > 0 \).

In general, we say that a pair of OTUs is directly connected if and only if the element in the protocol matrix corresponding to that pair is greater than zero (i.e., the comparison for that pair was made):

**Definition 6.5** Let \( 1 \leq j, k \leq n \). Then \( P_j \) is directly connected to \( P_k \) if and only if \( e_{jk} > 0 \).

We say that a pair of OTUs, say OTU_j and OTU_k, is connected if and only if

(i) OTU_j is directly connected to OTU_k

or

(ii) OTU_j is directly connected to OTU_j (which is in turn directly connected to OTU_j, ..., which is in turn directly connected to OTU_k).

For example:

OTU_1 is connected to OTU_3 because:

- OTU_1 is directly connected to OTU_2,
- OTU_2 is directly connected to OTU_4,
- and OTU_4 is directly connected to OTU_3.
### Figure 6.6

Revision of the protocol matrix. (A) The original protocol matrix, $E$. The corresponding primary solution matrix does not invert. (B) The revised protocol matrix, $E'$, in which the $(3, 5)$th comparison has not been made. The corresponding primary solution matrix does invert.
In general terms:

**Definition 6.6** Let \( 1 \leq j, k \leq n \). Then \( P_j \) is connected to \( P_k \) if and only if:

(i) \( P_j \) is directly connected to \( P_k \)

or

(ii) there exists an \( l \geq 1 \) and a sequence 

\( j = j_0, j_1, \ldots, j_{l+1} = k \) of integers such that \( P_{j_0} \) is directly connected to 

\( P_{j_1}, \ldots, \) and \( P_j \) is directly connected to 

\( P_{j_{l+1}} \).

What does this have to do with the inversion of the primary solution matrix? According to Theorem 6.2*, below, the primary solution matrix inverts if and only if every heterologous OTU is connected to the homologous OTU. In our hypothetical example (Figure 6.6), neither OTU_3 nor OTU_4 are connected to OTU_5 (the homologous OTU); accordingly, Theorem 6.2* predicts that matrix \( E \) will not invert. We can settle upon one of two alternatives: either remove OTU_3 and OTU_4 from the protocol matrix (since OTU_1 and OTU_2 are both connected to OTU_5, what's left of the protocol matrix will yield an invertible primary solution matrix), or else add enough additional comparisons in the experimental study to make the necessary connections in the protocol matrix. For this particular example, the inclusion of a single additional comparison, the OTU_3 to OTU_5 comparison, is
sufficient to connect all heterologous OTUs to the homologous OTU in the revised protocol matrix, \( E' \). Accordingly, the revised primary solution matrix, \( S' \), is now invertible. In general:

**Theorem 6.2** Let \( S \) be the primary solution matrix. Then \( S^{-1} \) exists if and only if for each \( j, 1 \leq j \leq n-1, P_j \) is connected to \( P_n \).

**Proof.** (I wish to thank Prof. H. R. van der Vaart, who suggested this proof, and Prof. Carl Meyer and Mr. Thomas Curran, who suggested alternate proofs for this theorem.) Consult Bott and Mayberry (1954). Our matrix \( S \) (except for change of sign) is the "matrix \( A \)" sensu Bott and Mayberry (1954) and our \( P_n \) is the "point \( 0 \)" and the "carrier of \( A \)" contains at least one "rooted tree" if and only if each \( P_j, 1 \leq j \leq n-1, \) is connected to \( P_n \). By Theorem 1 sensu Bott and Mayberry (1954), det \( (A) \) is the sum of a product of positive values, and hence must be positive (i.e., \( A^{-1} \) exists), if and only if the carrier of \( A \) contains at least one rooted tree. Hence \( S^{-1} \) exists if and only if for each \( j, 1 \leq j \leq n-1, P_j \) is connected to \( P_n \).
7. DIVERGENT EVOLUTION IN MOLECULAR CHARACTERS

7.1 Trouble with Genetic Load

In chapter six, we observed that a dissimilarity matrix has the property that, on the whole, a greater dissimilarity value between two OTUs corresponds to a more ancient ancestral separation. We call this intuitive concept "evolutionary divergence", or simply "divergence". At the beginning of this chapter, we argue that phenetic matrices derived from molecular data are likely to have this property. This argument is based on certain discrepancies in the genetic load model and a modified version of the "non-Darwinian evolution" concept of King and Jukes (1969). At the end of this chapter, we present three, slightly different definitions of "divergence" (equidivergence, strict divergence, and divergence-in-mean).

In 1966, Lewontin and Hubby published two papers (Lewontin and Hubby, 1966; Hubby and Lewontin, 1966) in which they noted the presence of substantial frequencies of polymorphisms at an unexpectedly high number of distinct gene loci in Drosophila pseudoobscura populations. According to the usual models of population genetics, a polymorphism cannot persist indefinitely in a population unless it is specifically selected for. But if a polymorphism is selected for by an "overdominance" model, in which the heterozygotes are superior and at least some of the homozygotes
die by "genetic death", then the population must overproduce in each generation in order to maintain itself at a constant size over a long period of time. If polymorphisms are present at many different gene loci, then the genetic load will be relatively high, because almost every organism has one or more homozygous loci (Haldane, 1957). Using this crude model (namely, that all homozygous loci have a "lethality level" which is independent and cumulative when figured for several distinct loci simultaneously), Lewontin and Hubby (1966) estimated an astronomical level of overproduction ($10^{46}$-fold) which would be necessary in each generation of a *D. pseudoobscura* population, assuming a commonly used selective coefficient. Since a *D. pseudoobscura* female is not physiologically capable of producing $10^{46}$ offspring, something must be wrong with the overdominance-genetic load model of Lewontin and Hubby (1966), since these natural populations do indeed maintain themselves. A number of writers have suggested that the genetic load model is fundamentally incorrect (Brues, 1964; Milkman, 1967; Sved et al., 1967; King, 1967), but nobody has come up with a good replacement, so nobody really knows how far off the genetic load model really is. (Newton's concept of space-time is fundamentally incorrect in the context of Einsteinian relativity, but it still gives very good answers for large bodies travelling at slow speeds. It is not really fair to indict a flawed theory until a better theory is available.)
Another possibility is that the selective coefficient assumed in Lewontin' and Hubby's (1966) calculation was far too high. In other words, the selective forces operating on the characters which Lewontin and Hubby (1966) investigated (electrophoretic mobilities of various isozymes—i.e., molecular characters) might be much lower than the range of selective coefficients typically found for morphological characters investigated by classical geneticists (Mukai, 1964; however, see Johnson et al., 1969). This phenomenon has been rather flamboyantly termed "non-Darwinian evolution" by King and Jukes (1969), because the evolution of molecular characters is thought to take place in the absence of selective (i.e., Darwinian) forces. The terminology is unfortunate, because it is almost certainly wrong if interpreted in a strict sense (i.e., no selection instead of less selection), and it has unnecessarily upset a good many classical evolutionists. On the basis of evidence to be presented below, we shall argue that there is a good deal less selection among individual molecular characters than has been found for morphological characters (although this lowered selection must be interpreted in a broad context of heavy selection on classical characters), and that this lowered selection on individual molecular characters renders them especially suited to the construction of cladograms.
7.2 Low Selection at the DNA Level

The first parts of a model for low levels of molecular selection were pieced together in the aftermath of the discovery of the genetic code (Crick, 1966). Out of sixty-four possible nucleotide triplets (codons) on the mRNA molecule, sixty-one specify a unique amino acid (the other three codons specify punctuation in protein synthesis). Since only twenty amino acids are actually specified during protein synthesis, this means there are exactly forty-one redundant codons, or an average of three different codons for each amino acid. Synonymous codons (i.e., distinct codons which specify the same amino acid) differ in only a single nucleotide, usually the third. In general, "similar" codons (i.e., codons which differ by a single nucleotide, usually the third) which are not actually synonymous (i.e., specify the same amino acid) nonetheless specify similar amino acids (amino acids differing only slightly in size, polarity, acidity, etc.) (Dayhoff, 1969, pp. 85-87; Woese, 1968). Most amazing of all, the genetic code appears to be uniform (with a few minor alterations) across the entire phylogenetic scale—from man through Escherichia coli (but see Commoner (1968)).

This remarkable fact has not escaped the notice of molecular evolutionists. The incredible complexity and specificity of the genetic code is such that it is difficult to imagine it having evolved more than once in its present form. Therefore, all living
creatures—from man through *E. coli*—must be descendants of a single common ancestor which possessed the genetic code more or less in its present form. This does not eliminate the possibility of other lineages in evolutionary history with a differing genetic code—but either these lineages have passed on no descendants to the present day, or else biochemists haven't investigated them yet. Our current genetic code has clearly eliminated its competitors.

Sonneborn (1965) has further argued that our genetic code is "one of the best of all genetic codes" (apologies to Voltaire!), because of its redundancy. There is physical evidence (i.e., the structure of the tRNA molecule) that the third nucleotide on the codon is most likely to be misread. Therefore, a genetic code for which mistakes on this especially labile nucleotide will specify either synonymous or similar amino acids during protein synthesis will minimize the effects of DNA mutations on protein structure. Sonneborn (1965) states that natural selection favored this kind of genetic code; Sonneborn's idea has been modified by Crick (1967) and Woese (1968), but not changed in its essence.

King and Jukes (1969) use this line of evidence to argue that synonymous codons are entirely insensitive to natural selection (since mutation does not affect the ultimate gene product) and similar codons are almost insensitive. This argument isn't really
complete. As Richmond (1970) points in his critique of non-Darwinian evolution, King and Jukes' claim that synonymous codons lead to equivalent amino acid specification requires that a complete spectrum of tRNAs all be present, all in equal quantity, and that they all require equal energy for complexing their respective amino acids. The first two conditions are known to be false (Caskey et al., 1968). Richmond fails to demonstrate that the inequities among tRNAs in living systems are enough to invalidate the concept of less selective evolution among molecular characters, although he has successfully demolished the concept of no selective evolution (i.e., non-Darwinian evolution).

A strong item of evidence in favor of less selection at the DNA level than would be expected is the fact that evolutionary rates estimated from DNA hybridization data are ten times as fast as comparable rates estimated from amino acid sequence data. The essence of this argument is that there is a loss of specificity between the DNA and amino acid levels of expression (due to the redundancy of the genetic code) and hence many nucleotides on the DNA are free to mutate (giving synonymous or similar codons) without resulting in selectively important alterations at the amino acid level. Thus, DNA comparisons will manifest a higher evolutionary rate than amino acid comparisons. In fact, we expect a stepwise increase of apparent evolutionary rates as we progress from the morphological down to the DNA level of expression. For example, a protein which is functional
or non-functional at the morphological level is called one character--
present or absent; at the amino acid level, perhaps it is 100
characters (i.e., a protein consisting of 100 amino acid residues),
in which perhaps only twenty-five or thirty residues make the
difference between functionality and non-functionality, whereas the
remaining residues are essentially free to evolve (within the
limits of "similar" residues) unaffected by selective pressure.
Thus the same genetic material has an apparently lower evolu-
tionary rate at the morphological than at the amino acid level.
At the DNA level, the apparent evolutionary rate expands all the
more.

Richmond (1970) points out that much of the DNA from which
the evolutionary rates are calculated is regulatory DNA rather than
structural DNA, and thus would be missed in the calculations made
from proteins. His point would be well taken if regulatory DNA
were known to evolve much faster than structural DNA. However,
it is dangerous for Richmond to speculate about the evolutionary
rates of regulatory vs. structural DNA, since the evolutionary
processes affecting the former are so poorly understood.

Several additional arguments in favor of non-Darwinian
evolution at the DNA level are less compelling. The Treffers mutator
gene, present in E. coli, is capable of changing the G + C com-
position in the DNA as much as 0.7% (thousands of individual
nucleotides) without visibly attenuating the culture (Cox and
Yanofsky, 1967). This used as evidence to argue that nucleotide
changes do not seriously affect the viability of the entire organism. As Richmond points out, we have no way of knowing that a selective process didn't take place in the bacterial cultures. (That is: bacteria with increased G + C compositions and bad viability did appear, but they were not noticed amongst the other bacteria which managed to survive.) The Treffers mutator gene studies show that G + C composition changes don't lower the viability of all bacteria (evidence: the survivors), but it doesn't eliminate the possibility that G + C composition changes can lower the viability among some bacteria (Richmond's unnoticed non-survivors). The relevance of the Treffers mutator gene studies to (sexual) vertebrate evolution is further questioned by Leigh (1970), who suggests that G + C composition changes, though neutral to asexual organisms, might be disadvantageous to sexual organisms.

7.3 Low Selection at the Protein Level

A second level of less-selective-molecular-characters evolution is thought to occur between the primary and tertiary structures of protein molecules. For example, most of the amino acid residues (about 75 our or 100) of the eukaryote cytochrome c molecule are interchangeable without a serious alteration of the molecule's tertiary structure (Margoliash and Smith, 1965) or its
fundamental significance to the organism (various eukaryote cytochromes c are perfectly interchangeable among intact mitochondria in vitro (Jacobs and Sanadi, 1960)). Similar stories are emerging for the hemoglobins (Perutz et al., 1965) and the immunoglobulins (although this problem is aggravated by the technical difficulties in purifying the immunoglobulins and a poorly understood ontogeny (Hood and Talmage, 1970; Burnet, 1970)). Richmond notes that in certain genetic diseases, a single amino acid change results in a substantial alteration of tertiary protein structure (for example: sickle cell anemia). Here, the exception proves the rule. These genetic diseases occur at those few "favored" amino acid residues that really count in the tertiary structure. The poor survival value of such mutants results in their quick elimination from the evolutionary panorama, and leaves behind only those mutants which are subject to little or no selection. Richmond (1970) also notes that no gorilla hemoglobin (different from human hemoglobin by two residues) has ever been found in a human, or vice versa, despite their supposed functional identity. Gives the biochemists time! As soon as every biochemical laboratory has an amino acid analyzer as standard equipment, we may be shocked at the genetic diversity within our populations!
7.4 Low Selection and Divergent Evolution

In an ensemble of characters subject to low selective pressures, we expect evolutionary changes in characters to accumulate steadily through time (Zuckerkandl and Pauling, 1965). Characters in this ensemble which reconverge after their respective populations have ceased to interbreed should be relatively rare; on the whole, the longer two sets of organisms have been separated ancestrally, the more dissimilar they will be with respect to this ensemble of characters. This concept of increasing dissimilarity as a result of longer ancestral separation is called the divergence hypothesis. The divergence hypothesis is a sufficient condition for all but one of the cladogram computing algorithms developed in this dissertation (proofs in chapter nine).

Any biologist wishing to make use of the theorems in chapter nine must employ an experimental system in which he can find divergent characters. We have already developed an informal argument that an ensemble of molecular characters has this property (since they are apparently subject to a low level of selection); the theorems of chapter nine are equally applicable to any other ensemble of characters which have this divergence property. It is important to emphasize that this dissertation does not attempt to ascertain any "true" or "representative" picture of the evolution of all biological characters through time and space; rather, it attempts to use an
appropriate choice of characters for making a correct inference of ancestral relationships. Once having gained a reliable insight into the cladistic evolution of a given collection of OTUs, the investigator may wish to view a broader collection of characters, which will give him a more accurate picture of total character evolution, including convergent evolution. Such studies are beyond the scope of this dissertation, but can certainly be aided by an accurate knowledge of ancestral evolution, which this dissertation seeks to give.

7.5 The Parsimony Hypothesis

The major evolutionary hypothesis employed by numerical taxonomists is the hypothesis of evolutionary parsimony (Camin and Sokal, 1965; Sarich and Wilson, 1967a; Sarich and Wilson, 1967b; Estabrook, 1968; Hendrickson, 1968; Farris, 1970). Suppose we have a cladogram such as illustrated in Figure 7.1. The initial OTUs for this cladogram are the sets V, W, Y, and Z. Not only have we included the phenetic matrix for the cladogram, but we have labelled each branch between adjacent junctures of the cladogram with patristic distances (Farris, J. S., personal communication). The patristic distance between two non-adjacent junctures of the cladogram is the sum of patristic distances over the shortest sequence of junctures connecting the two non-adjacent junctures. In a maximum parsimony (Farris, 1970) cladogram (the "additive"
Figure 7.1. Hypothetical parsimonious cladogram. (A) Diagram of patristic distances for a hypothetical cladogram having OTUs V, W, Y, and Z. (B) A phenetic matrix which corresponds to the diagram in part A.
cladogram of Cavalli-Sforza and Edwards, 1967), the phenetic value for any pair of OTUs is exactly equal to the patristic value for that pair of OTUs. If no dendrogram exists from which it is possible to calculate a matrix of patristic distances for each pair of OTUs which exactly matches the phenetic matrix, then the object of a maximum parsimony approach is to arrive at a dendrogram which comes as "close" as possible, by some measure of closeness. Unfortunately, the maximum parsimony criterion alone may not be able to guarantee unambiguous results for a given phenetic matrix. For example, Figure 7.2 depicts a maximum parsimony dendrogram derived from the same phenetic matrix as our original, hypothetical cladogram and equally as parsimonious as the original. Since parsimony is not a sufficient condition for finding the true binary cladogram (by Theorem 5.14, there is at most one binary cladogram for a given set of OTUs), we shall not discuss the parsimony methods in this dissertation.

7.6 Divergence Hypotheses

The basic idea behind a divergence hypothesis is summarized in part (iii) of the description of a dissimilarity phenetic matrix (Section 5.1): two OTUs which are farther apart ancestrally have a correspondingly larger value in the phenetic matrix. Intuitively, we say that one pair of OTUs is "farther apart ancestrally" than a second pair of OTUs if and only if the patriarch of the first pair
Figure 7.2. Alternative parsimonious dendrogram. (A) An alternative diagram of patristic distances for a dendrogram having OTUs V, W, Y, and Z. (B) A phenetic matrix which corresponds to the diagram in part A. Note that the phenetic matrix for this figure is the same as the phenetic matrix for Figure 7.1, even though the dendrogram is different.
is more ancient ancestrally than the patriarch of the second pair. By "more ancient" ancestrally, we mean that the patriarch of the first pair is a predecessor of the patriarch of the second pair. For example, let the dendrogram of Figure 5.1 be a cladogram. Then the pair of OTUs A and D are "farther apart" ancestrally than, say, the pair of OTUs C and D, because the patriarch of A and D (namely, K) is a predecessor of the patriarch of C and D (namely, I). On the other hand, there is no meaningful way to judge whether F and G are "farther apart" than, say, C and D, because the patriarch of F and G (namely, J) and the patriarch of C and D (namely, I) bear no predecessor relationship to one another.

Definition 7.1 Let \((X, \cal D)\) be a cladogram on \(\Pi_{X}\) and \(\langle, \rangle\) be a phenetic matrix on \(\Pi_{X}\). Then \(\langle, \rangle\) is equidivergent if and only if for any \(S, T, U, V \in \Pi_{X}\):

(i) \(S \circ T \geq U \circ V\) implies that \(\langle S, T \rangle > \langle U, V \rangle\)

and

(ii) \(S \circ T = U \circ V\) implies that \(\langle S, T \rangle = \langle U, V \rangle\).

The most specific divergence condition to be developed in this dissertation is the condition of equidivergence. We say that a phenetic matrix for a cladogram is equidivergent if and only if (i) if the patriarch of one pair of OTUs is a predecessor of the patriarch of a second pair of OTUs, then the phenetic value of the first pair is greater than the phenetic value of the second
pair, and (ii) if the patriarch of one pair of OTUs is the same as the patriarch of a second pair of OTUs, then the phenetic value of the first pair equals the phenetic value of the second pair. If the two patriarchs are not the same and bear no predecessor relationship to one another, then no equality or inequality between the corresponding two phenetic values is specified. In the jargon of numerical taxonomy, a phenetic matrix is equidivergent if and only if it is the same as the corresponding cophenetic matrix (Sokal and Sneath, 1963, p. 202).

A slightly weaker condition is the condition of strict divergence. We say that a phenetic matrix is strictly divergent if and only if it fulfills at least the first condition of equidivergence, namely, if the patriarch of one pair of OTUs is a predecessor of the patriarch of a second pair, then the phenetic value of the first pair is greater than the phenetic value of the second pair. If the two patriarchs are either the same or have no predecessor relationship to one another, then no inequality or equality between the corresponding two phenetic values is specified.

Definition 7.2 Let \((K, \preceq)\) be a cladogram on \(\Pi_X\) and \(\langle,\rangle\) be a phenetic matrix on \(\Pi_X\). Then \(\langle,\rangle\) is strictly divergent if and only if for any \(S,T,U,V \in \Pi_X\); \(S \circ T \preceq U \circ V\) implies that \(\langle S,T \rangle > \langle U,V \rangle\).
From the definitions, it is immediately obvious (proof omitted) that equidivergence implies strict divergence (i.e., equidivergence is a special case of strict divergence):

**Theorem 7.1** Let \((\mathbf{K}, \Phi)\) be a cladogram on \(\Pi_X\) and \(\langle,\rangle\) be a phenetic matrix on \(\Pi_X\). Then \(\langle,\rangle\) is strictly divergent if \(\langle,\rangle\) is equidivergent.

The third and weakest divergence hypothesis, called divergence-in-mean, is best deferred until later in the chapter, where it will be easier to understand. Roughly speaking, it says that if the patriarch of one pair of OTUs is a predecessor of the patriarch of a second pair of OTUs, then the phenetic value of the first pair is "usually" greater than the phenetic value of the second pair.

### 7.7 The Superphenetic Matrix

The basic idea underlying all the phenogram algorithms developed in this dissertation is a decision procedure by which individual junctures from the true cladogram are distinguished from possible junctures which are not members of the true cladogram. That is, a set of possible junctures is formulated in which there is guaranteed to be at least one juncture which is a member of the true cladogram. Typically, a value is computed for each possible juncture, such that the juncture corresponding to the minimum (or maximum) value must be a member of the true cladogram.
The procedure by which the set of possible junctures is formulated is called the iteration structure of a phenogram algorithm, and will be developed in chapter eight. The procedure by which a value is attached to each pair of junctures is called the computation structure of a phenogram algorithm. Since computation structures are closely related to the formulation of the divergence-in-mean evolutionary hypothesis, we shall develop computation structures in this chapter.

A computation structure is a procedure for assigning a nonnegative real number to any pair of possible junctures whose union might be a member of the true cladogram. As we mentioned in Section 5.2, a cladogram has very few actual junctures, but many candidates for junctures. In order to distinguish the true junctures from the false candidates, we require a function which assigns a numerical value to all possibilities. Theorem 5.3 shows that the set of junctures for any dendrogram on \( \Pi_X \) is a subset of the superpartition, \( \Sigma \), on \( \Pi_X \). Therefore, any possible juncture of any dendrogram on \( \Pi_X \) must be a member of \( \Sigma \) and any pair of junctures of any dendrogram on \( \Pi_X \) must correspond to a pair of members of \( \Sigma \). The natural extension of a phenetic matrix, defined over all pairs of OTUs, is a superphenetic matrix, defined over all pairs of members of \( \Sigma \). For example, a phenetic matrix for the partition discussed in Section 4.2 would have dissimilarity values for:

\[
\begin{align*}
\langle \text{Homo sapiens, Pan troglodytes} \rangle \\
\langle \text{Homo sapiens, Pan paniscus} \rangle \\
\langle \text{Pan troglodytes, Pan paniscus} \rangle
\end{align*}
\]
The corresponding superphenetic matrix would also have values for:

\[(\text{Homo sapiens, Pan})\]
\[(\text{Homodytes, Pan paniscus})\]
\[(\text{Humniscus, Pan troglodytes})\]

A computer decision procedure, for example, might calculate the above three values, discover that \[(\text{Homo sapiens, Pan})\] was larger than the other two, and on that basis form a classification such that \text{Homo sapiens} comprised one category and the two species of \text{Pan} comprised the other.

Except for its domain of definition, a superphenetic matrix on \(\Pi_X\) is defined the same as a phenetic matrix on \(\Pi_X\): it is symmetric with zero diagonal and nonnegative offdiagonal elements.

**Definition 7.3** Let \(\Pi_X\) be a partition, \(\Sigma\) be the superpartition on \(\Pi_X\), and \((,\) be a real-valued matrix with domain \(\Sigma \times \Sigma\). Then \((,\) is a superphenetic matrix if and only if for distinct \(U, V \in \Sigma\):

1. \(\langle U, U \rangle = \langle V, V \rangle = 0\)
2. \(\langle U, V \rangle = \langle V, U \rangle \geq 0\)

Figures 7.3, 7.4, and 7.5 show phenetic matrices and corresponding superphenetic matrices.
7.8 Formulas for a Superphenetic Matrix

A phenetic matrix is a matrix of data, but the members of a superphenetic matrix which are not already members of the phenetic matrix are calculated from the phenetic matrix. The current literature suggests a variety of methods for calculating a superphenetic matrix; we shall mention four in this dissertation: minimal, maximal, unweighted average, and weighted average.

Suppose that $U_1$ and $U_2$ are two members of $\Sigma$ and we wish to calculate the superphenetic value of $U_1$ and $U_2$, namely, $\langle U_1, U_2 \rangle$.

If $P_1$ and $P_2$ are OTUs and $P_1$ is a subset of $U_1$ and $P_2$ is a subset of $U_2$, we say that $\langle P_1, P_2 \rangle$ "is a phenetic value for" the pair $(U_1, U_2)$. For example:

\[
\langle \text{Homo sapiens}, \text{Pan troglodytes} \rangle \text{ "is a phenetic value for"} \\
(\text{Homo sapiens and Pan})
\]

and

\[
\langle \text{Homo sapiens}, \text{Pan paniscus} \rangle \text{ "is a phenetic value for"} \\
(\text{Homo sapiens and Pan}).
\]

Each entry in a superphenetic matrix is calculated from the collection of phenetic values "for" the pairs corresponding to that entry in the superphenetic matrix.

Each entry in a minimal superphenetic matrix (MNSM) is the minimal of all phenetic values "for" the pair corresponding to that superphenetic entry (Sneath, 1957). A minimum superphenetic
matrix is sometimes called the "single linkage" formula, because only a single phenetic value (namely, the minimum) "for" the pair corresponding to a superphenetic entry must be below a certain constant to "link" that superphenetic entry at a level below that constant. Sokal and Sneath (1963, p. 192) recommend a MNSM when one desires to form "diffuse, easily-linked" junctures.

**Definition 7.4** Let \( \Pi_X \) be a partition, \( \Sigma \) be the superpartition on \( \Pi_X \), and \( \langle , \rangle \) be a superphenetic matrix on \( \Pi_X \). Then \( \langle , \rangle \) is a **minimal superphenetic matrix** (MNSM) if and only if for distinct \( U_1, U_2 \in \Sigma \):

\[
\langle U_1, U_2 \rangle = \min \langle P_1, P_2 \rangle \\
P_1, P_2 \in \Pi_X \\
P_1 U_1, P_2 U_2
\]

An example of a MNSM is illustrated in Figure 7.3.

Each entry in a **maximal superphenetic matrix** (MXSM) is the maximum of all phenetic values "for" the pair corresponding to that superphenetic entry (Sorenson, 1948). A maximum superphenetic matrix is sometimes called the "complete linkage" formula, because every phenetic value (namely, the maximum and all smaller phenetic values) "for" the superphenetic entry must lie below a certain constant in order to "link" the pair corresponding to that superphenetic entry at a level below that constant. Sokal and Sneath (1963, p. 192) recommend a MXSM when one desires to form "small, compact" junctures.
(A)  

<table>
<thead>
<tr>
<th>OTUs</th>
<th>W</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Y</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Z</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

(B)  

<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>Y</th>
<th>Z</th>
<th>WUY</th>
<th>WUZ</th>
<th>YUZ</th>
<th>WUZUZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Z</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WUY</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WUZ</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>YUZ</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WUZUZ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(C)  

\[ WUY, Z = \min(5, 6) = 5; \quad WUZ, Y = \min(4, 6) = 4; \quad WUZ, Z = \min(5) \]

\[ WUZ, Z = \min(5, 0) = 0. \]

Figure 7.3. Minimal superphenetic matrix. (A) A hypothetical phenetic matrix. (B) A minimal superphenetic matrix calculated from the phenetic matrix in part A. (C) Sample calculations from part B.
Definition 7.5  Let $\Pi_X$ be a partition, $\Sigma$ be a superpartition on $\Pi_X$, and $\langle , \rangle$ be a superphenetic matrix on $\Pi_X$. Then $\langle , \rangle$ is a maximal superphenetic matrix (MXSM) if and only if for distinct $U_1, U_2 \in \Sigma$:

$$\langle U_1, U_2 \rangle = \max_{P_1, P_2 \in \Pi_X} \langle P_1, P_2 \rangle$$

$$P_1 \cup U_1, P_2 \cup U_2$$

An example of a MXSM is illustrated in Figure 7.4.

Still another approach is to assign to each superphenetic entry the average of the phenetic values "for" the pair corresponding to that superphenetic entry (Sokal and Michener, 1958). A superphenetic matrix calculated in this fashion is called an unweighted average superphenetic matrix (UASM), sometimes called the "average linkage" formula, and is recommended by Sokal and Sneath (1963, p. 192) for forming junctures of "average" diffuseness.

Definition 7.6  Let $\Pi_X$ be a partition, $\Sigma$ be the superpartition on $\Pi_X$, and $\langle , \rangle$ be a superphenetic matrix on $\Pi_X$. Then $\langle , \rangle$ is an unweighted average superphenetic matrix (UASM) if and only if for distinct $U_1, U_2 \in \Sigma$:

$$\langle U_1, U_2 \rangle = \frac{\sum_{P_1, P_2 \in \Pi_X} \langle P_1, P_2 \rangle}{\sum_{P_1, P_2 \in \Pi_X} P_1 \cup U_1, P_2 \cup U_2}$$

An example of an UASM is illustrated in Figure 7.5.
(A) OTUs  W  Y  Z
    W  0  4  5
    Y  4  0  6
    Z  5  6  0

(B) W  Y  Z  WUY  WUZ  YUZ  WUYUZ
    W  0  4  5  4  5  5  6
    Y  4  0  6  4  6  6  6
    Z  5  6  0  6  5  6  6
    WUY  4  4  6  0  6  6  6
    WUZ  5  6  5  6  0  6  6
    YUZ  5  6  6  6  6  0  6
    WUYUZ  6  6  6  6  6  6  0

(c) \( \langle WUY, Z \rangle = \max(5, 6) = 6; \langle WUZ, Y \rangle = \max(4, 6) = 6; \)
\( \langle WUZ, Z \rangle = \max(5, 0) = 5. \)

Figure 7.4. Maximal superphenetic matrix. (A) A hypothetical phenetic matrix. (B) A maximal superphenetic matrix calculated from the phenetic matrix in part A. (C) Sample calculations from part B.
<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>OTUs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Y</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Z</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>Y</th>
<th>Z</th>
<th>WUY</th>
<th>WUZ</th>
<th>YUZ</th>
<th>WUYUZ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>3</td>
</tr>
<tr>
<td>Y</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>$\frac{3}{2}$</td>
</tr>
<tr>
<td>Z</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>3</td>
<td>$\frac{3}{2}$</td>
</tr>
<tr>
<td>WUY</td>
<td>2</td>
<td>2</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{2}$</td>
</tr>
<tr>
<td>WUZ</td>
<td>$\frac{1}{2}$</td>
<td>5</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{4}$</td>
<td>0</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{2}$</td>
</tr>
<tr>
<td>WUY</td>
<td>$\frac{1}{2}$</td>
<td>3</td>
<td>3</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{3}{4}$</td>
<td>0</td>
<td>$\frac{3}{2}$</td>
</tr>
<tr>
<td>WUYUZ</td>
<td>3</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>0</td>
</tr>
</tbody>
</table>

(C) \( \langle WUY, Z \rangle = \frac{5+6}{2} = \frac{11}{2} \); \( \langle WUZ, Y \rangle = \frac{4+6}{2} = 5 \); \( \langle WUZ, Z \rangle = \frac{5+0}{2} = \frac{5}{2} \).

Figure 7.5. Unweighted average superphenetic matrix. (A) A hypothetical phenetic matrix. (B) An unweighted average superphenetic matrix calculated from the phenetic matrix in part A. (C) Sample calculations for part B.
The fourth type of superphenetic matrix to be discussed in this dissertation is the weighted average superphenetic matrix (WASM), introduced by Sokal and Michener (1958). In order to get a handle on how the weighting procedure works, we must discuss the concept of nesting levels in a dendrogram.

7.9 Nesting Levels in a Dendrogram

Nesting levels in a dendrogram are nonpositive powers of two which tell us how "deep" a particular juncture is in reference to a particular initial OTU. Consider, for example, the dendrogram originally illustrated in Figure 5.1. (reproduced in Figure 7.6). The OTU F, for example, which is a subset of X, is "three deep" from the juncture X. That is, the sequence of junctures leading from F to X (not including F itself) is an ordered sequence of three elements: (J,L,X). Accordingly, we say that the nesting level of F in X is $2^{-3} = 1/8$. Figure 7.6 gives additional examples of nesting levels for the dendrogram of Figure 5.1. In general, if an OTU P is a subset of juncture S, we define the nesting level of P in S, denoted nest (P,S), as two to the minus power the number of members in the sequence of predecessors from P to S, not including P. If P is not a subset of S, then nest (P,S) equals zero.
Figure 7.6. Nesting levels. (A) The dendrogram of Figure 5.1. (B) Nonzero nesting levels. (C) Zero nesting levels. (D) Undefined nesting levels, because the latter set is not a juncture.
Definition 7.7 Let $\Pi_X$ be a partition and $(\Delta, \mathcal{D})$ be a dendrogram on $\Pi_X$. Let $P \in \Pi_X$, $S \in \Delta$, and $Y$ be the set of all $U$ such that $P \subseteq U \subseteq S$. Then, the nesting level of $P$ in $S$, denoted $\text{nest}(P, S)$, is

(i) $2^{-N(Y)}$ if $P \subseteq S$

and

(ii) 0 if $\neg(P \subseteq S)$.

Lemma 7.2 shows (i) that the nesting level of any OTU in itself is exactly one and (ii) that the sum of nesting levels of all OTUs in a particular OTU is one.

Lemma 7.2 Let $\Pi_X$ be a partition and $(\Delta, \mathcal{D})$ be a dendrogram on $\Pi_X$. Let $R \in \Pi_X$. Then:

(i) $\text{nest}(R, R) = 1$

and

(ii) $\sum_{P \in \Pi_X} \text{nest}(P, R) = 1$.

Proof. Part I. Clearly, $R \in \Pi_X \subseteq \Delta$ and $R \subseteq R$. Suppose there is a $U \in \Delta$ such that $R \subseteq U \subseteq R$. Then $R \subseteq R$, contradiction. Let $Y$ be the set of all $U$ such that $R \subseteq U \subseteq R$. Then $N(Y) = 0$ and by Definition 7.7 (i), $\text{nest}(R, R) = 2^{-0} = 1$. 
Part II. Consider any $P \in \Pi_X$ such that $P \neq R$. By Theorem 4.1 (ii), $P \cap R = \emptyset$. Since $R \neq \emptyset$, $\neg (R \subseteq P)$ and by Definition 7.7 (ii), $\text{nest}(P, R) = 0$. By Part I:

$$\sum_{P \in \Pi_X} \text{nest}(P, R) = \text{nest}(R, R) + \sum_{P \in \Pi_X \setminus \{R\}} \text{nest}(P, R) = \text{nest}(R, R) + 0 = 1.$$ 

Lemma 7.3 shows that for any juncture, the sum of nesting levels of all OTUs in that juncture is one:

**Lemma 7.3** Let $(\Delta, D)$ be a binary dendrogram on $\Pi_X$ and $S \in \Delta$. Then

$$\sum_{P \in \Pi_X} \text{nest}(P, S) = 1.$$ 

**Proof. Strategy.** (I) If $S \in \Pi_X$, then Lemma 7.3 is true by Lemma 7.2. (II) If $S \in \Delta - \Pi_X$, then by Definition 5.3 (iv), there exist $T_1, T_2 \in \Delta$ such that $\{T_1, T_2\}$ is a binary partition of $S$ and

$$\text{nest}(P, S) = \frac{1}{2}(\text{nest}(P, T_1) + \text{nest}(P, T_2)).$$

(III) If $T_1, T_2 \in \Pi_X$, then $\text{nest}(P, T_1) = 1$ and $\text{nest}(P, T_2) = 1$ and $\text{nest}(P, S) = \frac{1}{2}(1+1) = 1$. If either or both of $T_1$ and $T_2$ are not members of $\Pi_X$, then part II may be repeated. By Theorem 5.5, $N(\Delta)$ is finite, and hence this backward progression cannot continue indefinitely.
Part II. Claim: $\text{nest}(P, S) = \frac{1}{2}(\text{nest}(P, T_1) + \text{nest}(P, T_2))$.

Summary of cases: (1) $P \cap S = \emptyset$; (2) $P \cap S \neq \emptyset$ and $P \cap T_1 \neq \emptyset$; (3) $P \cap S \neq \emptyset$ and $P \cap T_1 = \emptyset$.

Case 1. Suppose $P \cap S = \emptyset$. Then by Definition 7.7 (ii), $\text{nest}(P, S) = 0$. By Theorem 4.1 (i), $P \subseteq (T_1 \cup T_2) \subseteq P \cap S = \emptyset$ and by Definition 7.7 (ii), $\text{nest}(P, T_1) = \text{nest}(P, T_2) = 0$. Therefore:

$$\text{nest}(P, S) = 0 = \frac{1}{2}(0 + 0) = \frac{1}{2}(\text{nest}(P, T_1) + \text{nest}(P, T_2)).$$

Case 2. Suppose $P \cap S \neq \emptyset$ and $P \cap T_1 \neq \emptyset$.

subclaim: $\text{nest}(P, T_2) = 0$. By Theorem 5.2, $S \subseteq P$ and $T_1 \subseteq P$. By Theorem 5.1, $P \subseteq T_1$. By Theorem 4.1 (ii), $T_1 \cap T_2 = \emptyset$. Thus, $P \cap T_2 \subseteq T_1 \cap T_2 = \emptyset$. Since $P \neq \emptyset$, $\sim(P \subseteq T_2)$.

By Definition 7.7 (ii), $\text{nest}(P, T_2) = 0$.

subclaim: $\text{nest}(P, T_1) = \frac{1}{2}\text{nest}(P, S)$. Let $Y \subseteq \Delta$ be the set of all $U$ such that $P \subseteq U \subseteq S$ and $Y' \subseteq \Delta$ be the set of all $U'$ such that $P \subseteq U' \subseteq T_1$.

subsubclaim: $Y = Y' \cup \{S\}$.

subsubsubclaim: $Y = Y' \cup \{S\}$. Consider any $U \in Y$ such that $U \neq S$. By construction, $\emptyset \neq P \subseteq U$. Hence $\emptyset \neq P \cap T_1 \subseteq U \cap T_1$. By Definition 5.3 (ii), either $T_1 \subseteq U$ or $U \supset T_1$. Assume the subsubsubclaim is false. That is, $U \supset T_1$.

By Theorem 5.1, $U \supset T_1$. By construction, $U \subseteq S$. Hence $S \supset U \supset T_1$. By Definition 5.3 (iv), $S \supset T_1$. Contradiction of Definition 5.1 (ii).
subsubsubclaim: \( Y' \cup \{s\} \subseteq Y \). Consider any \( U' \in Y' \).

By construction, \( P \subseteq U' \subseteq T_1 \). By Definition 5.3 (iv), \( S \not\supset T_1 \).

By Definition 5.1 (i), \( S \not\supset T_1 \). Hence \( P \subseteq U' \subseteq S \) and \( U' \in Y \).

Since the subsubclaim is satisfied (that is, \( = Y' \cup \{s\} \)), by Definition 7.7 (i):

\[
\text{nest}(P,S) = 2^{-N(Y)} = 2^{-N(Y')} - 1 = \frac{1}{2}(2^{-N(Y')}) = \frac{1}{2}\text{nest}(P,T_1)
\]

Therefore:

\[
\text{nest}(P,S) = \frac{1}{2}\text{nest}(P,T_1)
\]

\[
= \frac{1}{2}(\text{nest}(P,T_1) + 0)
\]

\[
= \frac{1}{2}(\text{nest}(P,T_1) + \text{nest}(P,T_2)).
\]

**Case 3.** Suppose \( P \cap S \neq \emptyset \) and \( P \cap T_1 = \emptyset \). By the distributive properties of \( \cap \) and \( \cup \):

\[
\emptyset \neq P \cap S = P \cap (T_1 \cup T_2) = (P \cap T_1) \cup (P \cap T_2) = P \cup (P \cap T_2) = P \cap T_2
\]

By reasoning analogous to Case 2:

\[
\text{nest}(P,S) = \frac{1}{2}
\]

\[
= \frac{1}{2}(\text{nest}(P,T_2) + 0)
\]

\[
= \frac{1}{2}(\text{nest}(P,T_2) + \text{nest}(P,T_1)).
\]

Lemma 7.4 shows that for any pair of junctures, the double sum of nesting levels over all OTUs in that pair of junctures is one:
Lemma 7.4 Let $(\Delta, \mathcal{D})$ be a binary dendrogram on $\Pi_X$ and $S, T \in \Delta$.

Then

$$\Sigma_{P_1, P_2 \in \Pi_X} \text{nest}(P_1, S) \text{nest}(P_2, T) = 1.$$ 

$$P_1 \subseteq S, P_2 \subseteq T$$

Proof. Since the first factor of the summand depends on $P_1$ only and the second factor of the summand depends on $P_2$ only, we can write the double sum as the product of two single sums:

$$\Sigma_{P_1, P_2 \in \Pi_X} \text{nest}(P_1, S) \text{nest}(P_2, T) = \Sigma_{P_1 \in \Pi_X} \text{nest}(P_1, S) \Sigma_{P_2 \in \Pi_X} \text{nest}(P_2, T)$$

$$P_1 \subseteq S, P_2 \subseteq T$$

By Lemma 7.3:

$$\Sigma_{P_1 \in \Pi_X} \text{nest}(P_1, S) = 1 \quad \text{and} \quad \Sigma_{P_2 \in \Pi_X} \text{nest}(P_2, T) = 1.$$ 

$$P_1 \subseteq S, \quad P_2 \subseteq T$$

Hence:

$$\Sigma_{P_1, P_2 \in \Pi_X} \text{nest}(P_1, S) \text{nest}(P_2, S) = (1)(1) = 1.$$ 

$$P_1 \subseteq S, P_2 \subseteq T$$
7.10 Weighted Average Superphenetic Matrix

Continuing the argument that was interrupted at the end of Section 7.7, we define a weighted average superphenetic matrix (WASM). As before, we wish to evaluate a superphenetic entry on the basis of known phenetic values "for" the pair corresponding to that superphenetic entry. There is an additional restriction on the weighted average technique: we must already have specified a dendrogram in advance. Thus we assume we know the dendrogram and we compute the WASM which corresponds to it. The WASM is then the average of all such phenetic values weighted by their appropriate nesting levels:

Definition 7.8 Let \((\Delta, \Pi)\) be a binary dendrogram on \(\Pi_X\) and \(\langle,\rangle\) be a superphenetic matrix. Then \(\langle,\rangle\) is a \textit{weighted average superphenetic matrix (WASM)} if and only if for distinct \(U_1, U_2 \in \Delta:\)

\[
\langle U_1, U_2 \rangle = \sum_{P_1, P_2 \in \Pi_X} \text{nest}(P_1, U_1) \text{nest}(P_2, U_2) \langle P_1, P_2 \rangle.
\]

Since the entire matrix would be too large to display conveniently, selected calculations from a hypothetical WASM are shown in Figure 7.7.
(A) OTUs

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>W</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>U</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>15</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>V</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>17</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>W</td>
<td>9</td>
<td>15</td>
<td>16</td>
<td>0</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Y</td>
<td>12</td>
<td>10</td>
<td>16</td>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Z</td>
<td>14</td>
<td>13</td>
<td>11</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

(B) T U V W Y Z

(C) $\langle TUU,U \rangle = \left(\frac{1}{2}\right)(1)(4)+\left(\frac{1}{2}\right)(1)(5) = 5$

$\langle TUU,U,V \rangle = \left(\frac{1}{4}\right)(1)(9)$

$+\left(\frac{1}{4}\right)(1)(15)+\left(\frac{1}{2}\right)(1)(17) = 14$

$\langle TUU,U,V \rangle = \left(\frac{1}{4}\right)(\frac{1}{2})(12)$

$+\left(\frac{1}{4}\right)(\frac{1}{2})(10)+\left(\frac{1}{2}\right)(\frac{1}{2})(16)$

$+\left(\frac{1}{4}\right)(\frac{1}{2})(14)+\left(\frac{1}{4}\right)(\frac{1}{2})(13)$

$+\left(\frac{1}{2}\right)(\frac{1}{2})(11) = 12$

$\langle TUU,W,U \rangle$ undefined, because $TUWW$ is not a juncture.

Figure 7.7. Weighted average superphenetic matrix. (A) A hypothetical phenetic matrix. (B) Assumed dendrogram. (C) Sample calculations from the weighted average superphenetic matrix.
7.11 Choice of Cladogram Junctures

In Section 7.7, we observed that the basic objective of a phenogram algorithm is to provide a decision procedure able to distinguish junctures of the true cladogram from alternative junctures. For example, suppose that Figure 5.1 is a cladogram and we know that L is a member of the set \( \Delta \) (i.e., the set of true cladogram junctures). Since L is not an OTU, there is a binary partition of L such that L is the immediate predecessor of the members of that binary partition. For this example, the correct (cladistic) binary partition is \( \{E, J\} \) or \( \{E, F \cup G\} \). However, it is also possible to construct the pairs \( \{E \cup F, G\} \) and \( \{E \cup G, F\} \). How do we distinguish \( \{E, J\} \) from its alternatives? Clearly, \( \{E, F \cup G\} \), \( \{E \cup F, G\} \), and \( \{E \cup G, F\} \) all correspond to defined pairs in the superphenetic matrix. The next two theorems specify which of the possible determinations of \( \langle E, F \cup G \rangle \) (that is, which of the possible superphenetic values corresponding to the correct, cladistic binary partition) will be the maximum value under several possible evolutionary hypotheses.

Theorem 7.5 shows that, under the hypothesis of equidivergent evolution, the binary cladistic partition of any non-OTU juncture has a lower unweighted average superphenetic value than any alternative binary partition of that juncture whose superphenetic value is defined.
Theorem 7.5. Let $(K, \mathcal{P})$ be a binary cladogram on $\Pi_X$ and $(,)$ be an unweighted average superphenetic matrix on $\Pi_X$ such that the phenetic matrix is equidivergent, and let $T \in K - \Pi_X$. If $\{T_1, T_2\}$ is a binary cladistic partition of $T$ and $\{U_1, U_2\} \subseteq \Sigma$ is a distinct binary partition of $T$, then $\langle T_1, T_2 \rangle > \langle U_1, U_2 \rangle$.

Remark. The reader, looking at Figure 5.1 A, will see that Theorem 7.5 asserts that for instance

$$\langle A \cup B, C \cup D \rangle > \langle A \cup C, B \cup D \rangle$$

where the superphenetic values are determined according to the UASM method. Thus he is asked to prove that

$$\frac{\langle A, C \rangle + \langle B, C \rangle + \langle A, D \rangle + \langle B, D \rangle}{4} \quad \frac{\langle A, B \rangle + \langle A, D \rangle + \langle C, B \rangle + \langle C, D \rangle}{4}$$

or that

$$\langle A, C \rangle + \langle B, D \rangle > \langle A, B \rangle + \langle C, D \rangle$$

This follows immediately from the equidivergence property, since $A \circ C \equiv A \circ B$ and $B \circ D \equiv C \circ D$. The point here is that the pairs, $(A, C)$ and $(B, D)$, that enter the calculation of the superphenetic value of the cladistic partition have patriarchs that are predecessors of the patriarchs of the pairs, $(A, B)$ and $(C, D)$, that enter the calculation for the alternative partition. By looking at a few
more cladistic and alternative partitions of non-OTU junctures, the reader will quickly develop a feeling for what's going on.

A general and formal proof now follows.

Proof. Strategy. (I) Show there is a constant \( k \) such that for each \( P_1, P_2 \in \mathbb{P}_k \), \( P_1 \subseteq T_1 \) and \( P_2 \subseteq T_2 \), it follows that \( (P_1, P_2) = k \). (II) Show that for each \( R_1, R_2 \in \mathbb{P}_k \) such that \( R_1 \subseteq U_1 \) and \( R_2 \subseteq U_2 \), \( (R_1, R_2) \leq k \). (III) Show that there exists a \( Q_1, Q_2 \in \mathbb{P}_k \) such that \( Q_1 \subseteq U_1 \), \( Q_2 \subseteq U_2 \), and \( (Q_1, Q_2) < k \). (IV) Show that \( (T_1, T_2) \) and \( (U_1, U_2) \) are defined. Assume the theorem is false. That is, \( (T_1, T_2) \neq (U_1, U_2) \). Show contradiction.

Part I. Claim: there exists a constant \( k \) such that for each \( P_1, P_2 \in \mathbb{P}_k \), \( P_1 \subseteq T_1 \) and \( P_2 \subseteq T_2 \), it follows that \( (P_1, P_2) = k \). Consider \( P_1, P_2, P_3, P_4 \in \mathbb{P}_k \), \( P_1, P_3 \subseteq T_1 \), and \( P_2, P_4 \subseteq T_2 \) such that \( P_1 \neq P_3 \) and \( P_2 \neq P_4 \). By Theorem 5.9, \( T_1 \supseteq P_1 \) and \( T_2 \supseteq P_2 \). By Theorem 4.1 (ii), \( T_1 \cap T_2 = \emptyset \). Since \( T_1 \neq \emptyset \) and \( T_2 \neq \emptyset \), by elementary set theory, \( \neg(T_1 \supseteq T_2) \) and \( \neg(T_2 \supseteq T_1) \). By Theorem 5.9, \( \neg(T_1 \supseteq T_2) \) and \( \neg(T_2 \supseteq T_1) \). By Theorem 3.22, \( T_1 \circ T_2 = P_1 \circ P_2 \). By a similar argument, \( T_1 \circ T_2 = P_3 \circ P_4 \). Thus, \( P_1 \circ P_2 = P_3 \circ P_4 \) and by Definition 7.1 (ii), \( (P_1, P_2) = (P_3, P_4) \).

Part II. Consider any \( R_1, R_2 \in \mathbb{P}_k \) such that \( R_1 \subseteq U_1 \) and \( R_2 \subseteq U_2 \). Claim: \( (R_1, R_2) \leq k \).
Summary of Cases. Subclaim: either \( R_1 \subseteq T_1 \) or \( R_1 \subseteq T_2 \).

Suppose that \( \neg (R_1 \subseteq T_1) \). Then there is an \( x \in R_1 \setminus T_1 \). By Theorem 4.4 (ii), \( U_1 \subseteq T \). Hence \( R_1 \subseteq U_1 \subseteq T \) and \( x \in (R_1 \cap T) - (R_1 \cap T_1) \)
\[ = R_1 \cap (T - T_1) = R_1 \cap T_1. \] Since \( R_1 \cap T_2 \neq \emptyset \), by Theorem 4.4,
\( R_1 \subseteq R_2 \cap T_2 \subseteq T_2 \). Similarly, either \( R_2 \subseteq T_1 \) or \( R_2 \subseteq T_2 \).

Case II-a Suppose \( R_1 \subseteq T_1 \) and \( R_2 \subseteq T_2 \). By Part I,
\[ \langle R_1, R_2 \rangle = k. \]

Case II-b Suppose \( R_1 \subseteq T_2 \) and \( R_2 \subseteq T_1 \). By Part I,
\[ \langle R_1, R_2 \rangle = \langle R_2, R_1 \rangle = k. \]

Case II-c Suppose \( R_1 \subseteq T_1 \) and \( R_2 \subseteq T_1 \). By Theorem 5.9,
\( T_1 \circ R_1 \) and \( T_1 \circ R_2 \). By Theorem 3.20, \( T_1 \circ R_1 \circ R_2 \). By Definition 5.3 (i), \( T \supseteq T_1 \). By Definition 5.2 (i), \( T \supseteq T_1 \). By Theorem 5.11,
\( T_1 \circ T_2 = T \). Consider any \( P_1, P_2 \in \Pi \) such that \( P_1 \subseteq T_1 \) and \( P_2 \subseteq T_2 \).
By Part I, \( P_1 \circ P_2 = T_1 \circ T_2 \). Combining these expressions:

\[ P_1 \circ P_2 = T_1 \circ T_2 = T \supseteq T_1 \circ R_1 \circ R_2 \]

or

\[ P_1 \circ P_2 \supseteq R_1 \circ R_2. \]

By Definition 7.1 (i), \( k = \langle P_1, P_2 \rangle > \langle R_1, R_2 \rangle \).

Case II-d Suppose \( R_1 \subseteq T_2 \) and \( R_2 \subseteq T_2 \). By an argument similar to Case II-c:

\[ P_1 \circ P_2 = T_1 \circ T_2 = T \supseteq T_2 \circ R_1 \circ R_2 \]

or

\[ P_1 \circ P_2 \supseteq R_1 \circ R_2. \]

By Definition 7.1 (i), \( k = \langle P_1, P_2 \rangle > \langle R_1, R_2 \rangle \).
Part III. Claim: there exists a \( Q_1, Q_2 \in \Pi_X \) such that
\( Q_1 \subseteq U_1 \) and \( Q_2 \subseteq U_2 \) and \( \langle Q_1, Q_2 \rangle < k \). By Theorem 4.2, at most
one of the sets \( T_i \cap U_j \neq \emptyset \) for \( i = 1, 2 \) and \( j = 1, 2 \).

Case III-a Suppose \( T_1 \cap U_1 = \emptyset \) or \( T_1 \cap U_1 = \emptyset \) or there
are no empty sets. Since \( T_2 \cap U_1 \neq \emptyset \), by Theorem 3.5, \( T_2 \cap U_1 \in \Sigma \).
By Theorem 4.3, there is a \( Q_1 \in \Pi_X \) such that \( Q_1 \subseteq T_1 \cap U_1 \). By a
similar argument, there is a \( Q_2 \in \Pi_X \) such that \( Q_2 \subseteq T_2 \cap U_2 \). By
Part II Case II-d, \( k > \langle Q_1, Q_2 \rangle \).

Case III-b Suppose \( T_2 \cap U_1 = \emptyset \) or \( T_2 \cap U_2 = \emptyset \). Analogous
to Case III-a, except that \( Q_1 \subseteq T_1 \cap U_1 \) and \( Q_2 \subseteq T_2 \cap U_2 \). By Part II
Case II-c, \( k > \langle Q_1, Q_2 \rangle \).

Part IV. By hypothesis, \( \{U_1, U_2\} \subseteq \Sigma \). By Theorem 5.3,
\( \{T_1, T_2\} \subseteq K \subseteq \Sigma \). By Definition 7.3, \( \langle T_1, T_2 \rangle \) and \( \langle U_1, U_2 \rangle \) are both
defined. Assume the theorem is false. That is, \( \langle T_1, T_2 \rangle \leq \langle U_1, U_2 \rangle \).
By Definition 7.6:

\[
\begin{align*}
\Sigma \langle P_1, P_2 \rangle & \leq \Sigma \langle R_1, R_2 \rangle \\
\frac{P_1, P_2 \in \Pi_X}{k} & \leq \frac{R_1, R_2 \in \Pi_X}{k} \\
\frac{P_1, P_2 \in \Pi_X}{P_1 \subseteq T_1, P_2 \subseteq T_2} & \leq \frac{R_1 \subseteq U_1, R_2 \subseteq U_2}{R_1, R_2 \in \Pi_X} \\
\frac{P_1, P_2 \in \Pi_X}{P_1 \subseteq T_1, P_2 \subseteq T_2} & \leq \frac{R_1 \subseteq U_1, R_2 \subseteq U_2}{R_1, R_2 \in \Pi_X}
\end{align*}
\]
Since all terms are positive, crossmultiplication does not alter the sense of the inequality:

\[ \sum_{R_1, R_2 \in \Pi_X} k_{R_1, R_2} \leq \sum_{R_1, R_2 \in \Pi_X} \langle R_1, R_2 \rangle \]  

(*)

\[ \sum_{R_1 \in U_1, R_2 \in U_2} k_{R_1, R_2} \leq \sum_{R_1 \in U_1, R_2 \in U_2} \langle R_1, R_2 \rangle \]

Let \( U = \{(R_1, R_2) : R_1, R_2 \in \Pi_X \text{ and } R_1 \subseteq U_1 \text{ and } R_2 \subseteq U_2\} \)

and \( U^* = \{(R_1, R_2) : (R_1, R_2) \in U \text{ and } \langle R_1, R_2 \rangle < k\} \).

By construction, inequality (*) is the same as:

\[ \sum_{(R_1, R_2) \in U^*} k_{(R_1, R_2)} + \sum_{(R_1, R_2) \in U - U^*} k_{(R_1, R_2)} = \sum_{(R_1, R_2) \in U} k_{(R_1, R_2)} \]

\[ < \sum_{(R_1, R_2) \in U} \langle R_1, R_2 \rangle = \sum_{(R_1, R_2) \in U} \langle R_1, R_2 \rangle + \sum_{(R_1, R_2) \in U^*} \langle R_1, R_2 \rangle \]  

(**)

\[ (R_1, R_2) \in U \]

\[ (R_1, R_2) \in U^* \]

\[ (R_1, R_2) \in U - U^* \]

By Part II:

\[ \sum_{(R_1, R_2) \in U - U^*} k_{(R_1, R_2)} = \sum_{(R_1, R_2) \in U^*} k_{(R_1, R_2)} \]

\[ (R_1, R_2) \in U - U^* \]

\[ (R_1, R_2) \in U^* \]

\[ (R_1, R_2) \in U - U^* \]

and equation (***) can be used to cancel inequality (**). By Part III, \( U^* \) is nonempty. Hence:

\[ \sum_{(R_1, R_2) \in U^*} k_{(R_1, R_2)} \leq \sum_{(R_1, R_2) \in U^*} \langle R_1, R_2 \rangle \]

\[ (R_1, R_2) \in U^* \]

\[ (R_1, R_2) \in U^* \]
By construction, for each member \((R_1, R_2) \in U^*, \langle R_1, R_2 \rangle < k\).

Summing over all members of \(U^*\):

\[
\Sigma \langle R_1, R_2 \rangle < \Sigma k \quad \text{as} \quad (R_1, R_2) \in U^* \quad (R_1, R_2) \in U^*
\]

Contradiction.

Under the hypothesis of strictly divergent evolution, the binary cladistic partition of any non-CTU juncture has a greater minimum superphenetic value than any alternative binary partition of that juncture whose superphenetic value is defined:

**Theorem 7.6** Let \((K, \mathfrak{M})\) be a binary cladogram on \(K_X\) and \(\langle , \rangle\) be a minimal superphenetic matrix on \(K_X\) such that the phenetic matrix is strictly divergent, and let \(T \in K - K_X\). If \(\{T_1, T_2\}\) is a binary cladistic partition of \(T\) and \(\{U_1, U_2\} \subseteq \Sigma\) is a distinct binary partition of \(T\), then \(\langle T_1, T_2 \rangle > \langle U_1, U_2 \rangle\).

**Remark.** Again looking at Figure 5.1 A, the reader will see that Theorem 7.6 contends that for instance

\[
\operatorname{Min}(\langle A, C \rangle, \langle B, C \rangle, \langle A, D \rangle, \langle B, D \rangle) > \operatorname{Min}(\langle A, B \rangle, \langle A, D \rangle, \langle C, B \rangle, \langle C, D \rangle)
\]

This follows from the strict divergence of the phenetic matrix since all pairs in the first member have patriarch \(K\), but the pairs in the second member have patriarchs \(K\), \(H\), and \(I\); for a
strictly divergent phenetic matrix, the phenetic value of a pair of OTUs with patriarch H or I is strictly less than the phenetic value of any pair of OTUs with patriarch K.

**Proof. Strategy.** (I) Construct a particular pair of sets $Q_1, Q_2 \in \Pi_X$ such that $Q_1 \subseteq U_1$ and $Q_2 \subseteq U_2$. (II) Show that for each $P_1, P_2 \in \Pi_X$ such that $P_1 \subseteq T_1$ and $P_2 \subseteq T_2$, $\langle Q_1, Q_2 \rangle < \langle P_1, P_2 \rangle$. (III) Show that $\langle T_1, T_2 \rangle > \langle U_1, U_2 \rangle$.

**Part I.** By Theorem 4.2, at most one of the sets $T_1 \cap U_j \neq \emptyset$ for $i = 1, 2$ and $j = 1, 2$.

**Case I-a.** Suppose $T_1 \cap U_1 = \emptyset$ or $T_1 \cap U_2 = \emptyset$ or there are no empty sets. Since $T_2 \cap U_1 \neq \emptyset$, by Theorem 3.5, $T_2 \cap U_1 \in \Sigma$.

By Theorem 4.3, there is a $Q_1 \in \Pi_X$ such that $Q_1 \subseteq T_2 \cap U_1$. By a similar argument, there is a $Q_2 \in \Pi_X$ such that $Q_2 \subseteq T_2 \cap U_2$.

**Case I-b.** Suppose $T_2 \cap U_1 = \emptyset$ or $T_2 \cap U_2 = \emptyset$. Analogous to Case I-a, except that $Q_1 \subseteq T_1 \cap U_1$ and $Q_2 \subseteq T_1 \cap U_2$.

**Part II.** Consider any $P_1, P_2 \in \Pi_X$ such that $P_1 \subseteq T_1$ and $P_2 \subseteq T_2$.

**Case II-a.** Suppose that $Q_1$ and $Q_2$ were obtained by Part I Case I-a. Then $Q_1 \subseteq T_1$ and $Q_2 \subseteq T_2$. By Theorem 5.9, $T_2 \triangleright Q_1$ and $T_2 \triangleright Q_2$. By Theorem 3.20, $T_2 \triangleright Q_1 \circ Q_2$. By Definition 5.3 (i), $T \triangleright T_2$. By Definition 5.2 (i), $T \triangleright T_2$. By Theorem 5.9, $T_1 \triangleright P_1$ and $T_2 \triangleright P_2$. By Theorem 4.1 (ii), $T_1 \cap T_2 = \emptyset$. Since $T_1 \neq \emptyset$ and $T_2 \neq \emptyset$, by elementary set theory, $\sim(T_1 \triangleright T_2)$ and $\sim(T_2 \triangleright T_1)$. 


By Theorem 5.9, \( T_1 \triangle R T_2 \) and \( T_2 \triangle R T_1 \). By Theorem 3.22,
\( T_1 \circ T_2 = P_1 \circ P_2 \). By Theorem 5.11, \( T_1 \circ T_2 = T \). Combining
these expressions:

\[ P_1 \circ P_2 = T_1 \circ T_2 = T \triangle R T_2 \triangle R Q_1 \circ Q_2 \]
or
\[ P_1 \circ P_2 \triangle R Q_1 \circ Q_2 \]

By Definition 7.2, \( \langle P_1, P_2 \rangle > \langle Q_1, Q_2 \rangle \).

**Case II-b** Suppose that \( Q_1 \) and \( Q_2 \) were obtained by Part I
Case I-b. By an analogous argument:

\[ P_1 \circ P_2 = T_1 \circ T_2 = T \triangle R T_2 \triangle R Q_1 \circ Q_2 \]
or
\[ P_1 \circ P_2 \triangle R Q_1 \circ Q_2 \]

By Definition 7.2, \( \langle P_1, P_2 \rangle > \langle Q_1, Q_2 \rangle \).

**Part III**. By Definition 7.4:

\[ \langle U_1, U_2 \rangle = \min \langle R_1, R_2 \rangle \leq \langle Q_1, Q_2 \rangle < \min \langle P_1, P_2 \rangle = \langle T_1, T_2 \rangle. \]

\[ R_1 \triangle R_2 \triangle R X \]

\[ P_1 \triangle P_2 \triangle R X \]

\[ R \leq U_1, R \leq U_2 \]

\[ P_1 \leq T_1, P_2 \leq T_2 \]

In the previous two theorems, we detailed conditions under which it is possible to distinguish cladogram from non-
cladogram junctures when a junction of the cladogram is being
partitioned into its cladistic binary partition. An alternative
approach is to union a pair of junctures to form another junction
which is also a juncture of the cladogram. For example, suppose that Figure 5.1 is the cladogram and we know that \{H, I, E, J\} is a partition of \( X \) and a subset of the set of cladogram junctures. Theorem 7.8 states that there must always be at least one pair of junctures in this partition whose union is also a juncture. For our example, there are two such pairs: (H and I) and (E and J), because \( H \cup I = K \) is a juncture and \( E \cup J = L \) is a juncture. Lemma 7.7 is used in the proof of Theorem 7.8.

**Lemma 7.7** Let \((K, \mathcal{D})\) be a cladogram on \( \Pi_X \) and \( \Pi_X \subseteq K \) be a partition and \( Z \subseteq K \) and \( K' \) be the set such that \( W \subseteq K' \) if and only if \( W \subseteq K \) and \( W \subseteq Z \). Then there exist distinct \( P_1, P_0 \in \Pi_X \cap K' \) such that \( P_1 \cup P_0 \subseteq K' \).

**Proof. Strategy.** (I) Suppose there are distinct \( P_1, P_0 \in \Pi_X \cap K' \) such that \( S = P_1 \circ P_0 \) and \( S \supseteq P_1 \) and \( S \supseteq P_0 \). Show that \( P_1 \cup P_0 = S \subseteq K' \). (II) Show the conditions of Part I do not hold.

Consider any distinct \( P_1, P_0 \in \Pi_X \cap K' \) such that \( \sim(P_0 \circ P_1 \circ P_1) \).

Show by mathematical induction that for each \( k \geq 1 \) there exists a sequence (a) \( P_0, \ldots, P_k \in \Pi_X \cap K' \) such that for \( 1 \leq i \leq k \),

(b) \( P_{i-1} \circ P_i \subseteq Z \), (c) \( P_{i-2} \cap (P_{i-1} \circ P_i) = \emptyset \), (d) \( P_{i-1} \circ P_i \supseteq P_i \),

and (e) \( \sim(P_{i-1} \circ P_i \circ P_i) \). (III) Show that \( N(\Pi_X) \) must be infinite. Contradiction of Definition 4.1 (i)
Part I. Suppose there are distinct $P_1, P_0 \in \mathbb{X} \cap K'$ such that $S = P_1 \circ P_0$ and $S \supset P_1$ and $S \supset P_0$. Claim: $P_1 \cup P_0 = S \in K'$. The existence of $S \in K$ is guaranteed by Theorem 5.10. Strategy: show that $S \supset P_1 \cup P_0$ and $S \supseteq P_1 \cup P_0 \subseteq Z$.

Subclaim: $S \supset P_1 \cup P_0$. By Definition 5.1 (i), $S \supset P_1$ and $S \supset P_0$. By elementary set theory, $S \supset P_1 \cup P_0$.

Subclaim: $S \subseteq P_1 \cup P_0 \subseteq Z$. Since $P_1 \subseteq Z$ and $P_0 \subseteq Z$, clearly $P_1 \cup P_0 \subseteq Z$. Since $P_0 \neq \emptyset$, choose any $x \in P_0$. By Definition 4.1 (ii), there is a $P \in \mathbb{X} \subseteq K$ such that $x \in P$. By Theorems 5.2 and 5.1, $P_0 \supseteq P$. Since $P_0 \supseteq P$ and $S \supset P_0$, by transitivity of $\supset$, $S \supset P$. Since $S \cap P = P \neq \emptyset$ and $S \neq P$, by Theorem 4.1 (ii), $S \notin \mathbb{X}$. By elimination, $S \in K - \mathbb{X}$. By Theorem 5.13 (ii), $\{P_1, P_0\}$ is a binary partition of $S$. By Definition 4.1 (ii), $S \subseteq P_1 \cup P_0$.

Part II. Assume that the conditions of Part I do not hold. Consider any distinct $P_1, P_0 \in \mathbb{X} \cap K'$ such that $\sim(P_0 \circ P_1 \supset P_1)$. Claim: for each $k \geq 1$ there is a sequence $P_0, \ldots, P_k \in \mathbb{X} \cap K'$ such that for each $1 \leq i \leq k$, $(P_{i-2} \circ P_{i-1}) \cap P_i = \emptyset$ and $P_{i-1} \circ P_i \supset P_i$ and $\sim(P_{i-1} \circ P_i \supset P_i)$.

Proof by mathematical induction.

True for $k = 1$: By construction, (a) $P_0, P_1 \in \mathbb{X} \cap K'$ and (e) $\sim(P_0 \circ P_1 \supset P_1)$. Condition (c) is true by default.

(d) Subclaim: $P_0 \circ P_1 \supset P_1$. By Definition 3.4, $P_0 \circ P_1 \supset P_1$. Assume the subclaim is false. That is, $P_0 \circ P_1 = P_1$. Then by substitution, $P_1 \supset P_0$. By Theorem 5.1, $P_1 \supset P_0 \neq \emptyset$. By Theorem 4.1 (ii), $P_1 = P_0$. But $P_1$ and $P_0$ are distinct. Contradiction.
(b) subclaim: \( P_0 \circ P_1 \subseteq Z \). Suppose subclaim is false. That is, \( \sim(P_0 \circ P_1 \subseteq Z) \). Then by Definition 5.3 (ii), either
\[ P_0 \circ P_1 \cap Z = \emptyset \text{ or } P_0 \circ P_1 \supseteq Z. \]
Suppose \( P_0 \circ P_1 \cap Z = \emptyset \). By previous subclaim, \( P_0 \circ P_1 \supseteq P_1 \). By Theorem 5.1, \( P_0 \circ P_1 \supseteq P_1 \not\subseteq \emptyset \). Hence \( \emptyset \neq P_1 \cap Z \subseteq (P_0 \circ P_1) \cap Z \). Contradiction. Alternatively,
\( P_0 \circ P_1 \supseteq Z \). Since \( Z \supseteq P_0 \) and \( Z \supseteq P_1 \), by Theorem 5.9, \( Z \supseteq P_0 \) and \( Z \supseteq P_1 \). By Definition 3.4, \( P_0 \circ P_1 \not\in CP([P_0, P_1]) \). By Definition 3.5, \( P_0 \circ P_1 \not\in RP([P_0, P_1]) \). Contradiction of Definition 3.7.

True for \( k \) implies true for \( k+1 \): Let \( P_0, \ldots, P_k \in \Pi_X' \cap K' \) be the sequence guaranteed by the inductive hypothesis. By Theorem 5.10, there is an \( S \in K \) such that \( S = P_{k-1} \circ P_k \). By the inductive hypothesis, \( Z \supseteq S \supseteq P_k \) and \( \sim(S \supseteq P_k) \). By Definition 5.2 (ii), there is a \( j \geq 1 \) and a sequence \( S = S_0, \ldots, S_j, S_{j+1} = P_k \). By construction, \( S_j \supseteq P_k \). By Definition 5.1, \( Z \supseteq S_j \supseteq P_k \). Choose any \( x \in S_j - P_k \). By Definition 4.1 (ii), there is a \( P_{k+1} \in \Pi_X' \) such that \( x \notin P_{k+1} \).

subclaim: \( S_j \supseteq P_{k+1} \). By Definition 5.3 (ii), either
\[ P_{k+1} \cap S_j = \emptyset \text{ or } P_{k+1} \supseteq S_j \text{ or } S_j \supseteq P_{k+1}. \]
The first possibility implies that \( x \notin P_{k+1} \cap S_j \); contradiction. By Theorem 5.1, the second possibility implies that \( P_{k+1} \supseteq S_j \) or \( P_{k+1} \supseteq S_j \supseteq P_k \); but then
\[ P_{k+1} \cap P_k = P_k \neq \emptyset \text{ and by Theorem 4.2 (ii), } P_{k+1} = P_k \text{; contradiction.} \]
By elimination, \( S_j \supseteq P_{k+1} \).

(a) subclaim: \( P_{k+1} \in \Pi_X' \cap K' \). By construction,
\( P_{k+1} \in \Pi_X' \). By previous subclaim, \( S_j \supseteq P_{k+1} \). By Theorem 5.1,
\( Z \supseteq S_j \supseteq P_{k+1} \). Hence \( P_{k+1} \in K' \) and subclaim is satisfied.
(b) subclaim: \( Z \models S_j = P_{k+1} \circ P_k \). By first subclaim, \( S_j \models P_{k+1} \). By construction, \( S_j \models P_k \). By Definition 3.4, \( S_j \in CP(\{P_k, P_{k+1}\}) \). Assume subclaim is false. Then \( S_j \notin RP(\{P_k, P_{k+1}\}) \) and by Definition 3.5, there is a 
\[ W \in CP(\{P_k, P_{k+1}\}) \text{ such that } S_j \models W \models P_k. \] 
By Theorem 5.1, \( S_j \models W \models P_k \). By Definition 5.1 (ii), \( \neg(S_j \models P_k) \). Contradiction.

(c) subclaim: \( P_{k-1} \cap (P_k \circ P_{k+1}) = P_{k-1} \cap S_j = \emptyset \). Assume the subclaim is false. Then by Definition 5.3 (ii), 
\[ S_j \models P_{k-1} \text{ or } P_{k-1} \models S_j. \] The first possibility implies that 
\[ S_0 \models S_j \models P_{k-1} \text{ and } S_0 \models S_j \models P_k \] 
and by Definition 3.5, \( S = S_0 \notin RP(\{P_{k-1}, P_k\}) \); contradiction. The second possibility implies by Theorem 5.1 that \( P_{k-1} \models S_j \models P_k \) or \( P_{k-1} \cap P_k = P_k \neq \emptyset \); then by Theorem 4.1 (ii), \( P_{k-1} = P_k \) and \( P_{k-1} \circ P_k = P_k \circ P_k = P_k \) and by the inductive hypothesis, \( P_k = P_{k-1} \circ P_k \models P_k \); contradiction of Theorem 3.1.

(d) subclaim: \( P_k \circ P_{k+1} = S_j \models P_{k+1} \). Demonstrated in the first and third subclaims.

(e) subclaim: \( \neg(P_k \circ P_{k+1} \models P_{k+1}) \), that is, \( \neg(S_j \models P_{k+1}) \). By construction, \( S_j \models P_k \). Since Part I does not hold, it cannot be true that \( S_j \models P_{k+1} \).

Part III. Claim: for any \( k \geq 1 \) and \( i, j \) such that 
\[ 0 \leq i \leq j \leq k, \ P_i \neq P_j. \] From Part II, 
\[ P_i \circ P_{i+1} \models P_{i+1} \circ P_{i+2} \models \ldots \models P_{j-1} \circ P_j \models P_j. \] From Theorem 5.1,
\[ P_i \circ P_{i+1} \supseteq P_{i+1} \circ P_{i+2} \supseteq \ldots \supseteq P_{j-1} \circ P_j \supseteq P_j. \] By transitivity of \( \supseteq \), \( P_i \circ P_{i+1} \supseteq P_j \). By Part II, \( P_i \cap P_j \subseteq (P_i \circ P_{i+1}) \cap P_j = \emptyset \).

By Theorem 4.1 (ii), \( P_i \neq P_j \).

Hence \( N(K) \) must be infinite, since for any \( k \geq 1 \), \( N(K) \geq k \).

Contradiction of Theorem 5.5.

**Theorem 7.8** Let \((K, \mathcal{P})\) be a cladogram on \( \Pi_X \) and \( \Pi_X^\prime \subseteq K \) be a partition. Then there exist \( P_1, P_0 \in \Pi_X^\prime \) such that \( P_1 \cup P_0 \in K \).

**Proof.** By Definition 5.3 (i), \( X \in K \). Let \( X \) be the set \( Z \) sensu Lemma 7.6. By Theorem 4.8, for each \( S \in K, S \subseteq X \). Hence \( K = K' \) sensu Lemma 7.6, and \( \Pi_X^\prime \cap K' = \Pi_X^\prime \). By Lemma 7.7, there exist \( P_1, P_0 \in \Pi_X^\prime \) such that \( P_1 \cup P_0 \in K' \).

Theorem 7.12 states that, under the hypothesis of strictly divergent evolution, there must be two members of any partition of \( X \) which is a subset of the set of junctures such that the union of those two members is a juncture and the superphenetic value for those two members is lower than the superphenetic value of any other two members of that partition whose union is not a juncture. For our example, suppose that \((H \text{ and } I)\) is the favored pair of members of \( \{H, I, E, J\} \). Then the superphenetic value of, say, \((I \text{ and } E)\) must be greater than the superphenetic value of \((H \text{ and } I)\) (since \( H \cup I \) is a juncture but \( E \cup I \) is not). The reason is, of course, that \( E \circ I \supset H \circ I \). Note that the assertion is that there exists a pair of members of our partition. Indeed, under the hypothesis of strict divergence \((E, J)\) could have been less than \((H, I)\). Theorem 7.12 does not regard such decisions.
Lemmas 7.9, 7.10, and 7.11 are used in the proof of
Theorem 7.12. Lemmas 7.9 and 7.10 are obvious properties of
averages stated without proof.

**Lemma 7.9** Let \((v_1, \ldots, v_n)\) and \((w_1, \ldots, w_n)\) be two ordered
n-tuples of nonnegative real numbers such that

\[
\sum_{i=1}^{n} v_i = 1.
\]

Then \(\min_{1 \leq i \leq n} (v_i) \leq \sum_{i=1}^{n} w_i v_i \leq \max_{1 \leq i \leq n} (v_i)\).

**Lemma 7.10** Let \((v_1, \ldots, v_n)\) be an ordered n-tuple of nonnegative
real numbers. Then

\[
\min_{1 \leq i \leq n} (v_i) \leq \frac{1}{n} \sum_{i=1}^{n} v_i \leq \max_{1 \leq i \leq n} (v_i).
\]

**Lemma 7.11** Let \((K, \mathcal{D})\) be a cladogram on \(\Pi_X\) and \(\langle, \rangle\) be

(i) a MNSM on \(\Pi_X\)

(ii) a MXSM on \(\Pi_X\)

(iii) a UASM on \(\Pi_X\)

or

(iv) a WASM on \(\Pi_X\).
such that the phenetic matrix is strictly divergent. If

\[
\max \langle P_1, P_2 \rangle < \min \langle R_1, R_2 \rangle,
\]

\[
P_1, P_2 \in X, \quad R_1, R_2 \in X,
\]

\[
P_1 \sqsubset S, P_2 \sqsubset T, \quad R_1 \sqsubset U, R_2 \sqsubset V
\]

then \(\langle S, T \rangle < \langle U, V \rangle\).

**Proof.** Case I. Let \(\langle , \rangle\) be a MWSM on \(\Pi_X\). By Definition 7.4:

\[
\langle S, T \rangle = \min \langle P_1, P_2 \rangle \leq \max \langle P_1, P_2 \rangle < \min \langle R_1, R_2 \rangle = \langle U, V \rangle
\]

\[
P_1, P_2 \in X, \quad P_1, P_2 \in X, \quad R_1, R_2 \in X,
\]

\[
P_1 \sqsubset S, P_2 \sqsubset T, \quad P_1 \sqsubset S, P_2 \sqsubset T, \quad R_1 \sqsubset U, R_2 \sqsubset V
\]

Case II.

\[
\langle S, T \rangle = \max \langle P_1, P_2 \rangle < \min \langle R_1, R_2 \rangle \leq \max \langle R_1, R_2 \rangle = \langle U, V \rangle.
\]

\[
P_1, P_2 \in X, \quad R_1, R_2 \in X, \quad R_1, R_2 \in X,
\]

\[
P_1 \sqsubset S, P_2 \sqsubset T, \quad R_1 \sqsubset U, R_2 \sqsubset V, \quad R_1 \sqsubset U, R_2 \sqsubset V
\]

Case III. By Definition 7.6:

\[
\langle S, T \rangle = \frac{\sum P_1, P_2 \in X \langle P_1, P_2 \rangle}{\sum P_1, P_2 \in X 1}
\]

\[
= \frac{P_1 \sqsubset S, P_2 \sqsubset T}{\sum P_1, P_2 \in X 1}
\]

\[
P_1 \sqsubset S, P_2 \sqsubset T
\]
Since the set of all such \( P_{1,2} \) must be finite (since \( \Pi_X \) is finite), this set can be arranged as an ordered \( n \)-tuple of nonnegative real numbers:

\[
\left( \langle P_{1,2} \rangle_1, \ldots, \langle P_{1,2} \rangle_n \right)
\]

Clearly, \( n = \sum_{P_{1,2} \in \Pi_X} \). By Lemma 7.10:

\[
\begin{align*}
\min \langle P_{1,2} \rangle & \leq \langle S,T \rangle \leq \max \langle P_{1,2} \rangle. \quad (\ast)
\end{align*}
\]

By an analogous argument:

\[
\begin{align*}
\min \langle R_{1,2} \rangle & \leq \langle U,V \rangle \leq \max \langle R_{1,2} \rangle. \quad (\ast\ast)
\end{align*}
\]

Combining expressions (\( \ast \)) and (\( \ast\ast \)):

\[
\langle S,T \rangle \leq \max \langle P_{1,2} \rangle < \min \langle R_{1,2} \rangle \leq \langle U,V \rangle.
\]
Case IV. Let \( \langle , \rangle \) be a WSM on \( \Pi_X \). By Definition 7.8:

\[
\langle S, T \rangle = \sum_{P_1, P_2 \in \Pi_X} \text{nest}(P_1, S) \text{nest}(P_2, T) \langle P_1, P_2 \rangle
\]

Since the set of \( P_1, P_2 \) must be finite (since \( \Pi_X \) is finite), this set can be arranged as an ordered n-tuple of real numbers:

\[
\langle \langle P_1, P_2 \rangle_1, \ldots, \langle P_1, P_2 \rangle_n \rangle
\]

By Lemma 7.4:

\[
\sum_{P_2 \subseteq S, P_2 \subseteq T} \text{nest}(P_1, S) \text{nest}(P_2, T) = 1.
\]

By Lemma 7.9:

\[
\text{min} \langle P_1, P_2 \rangle \leq \langle S, T \rangle \leq \text{max} \langle P_1, P_2 \rangle \quad (\ast)
\]

\[
\text{min} \langle P_1, P_2 \rangle \leq \langle U, V \rangle \leq \text{max} \langle R_1, R_2 \rangle \quad (\ast\ast)
\]

By an analogous argument:

\[
\text{min} \langle R_1, R_2 \rangle \leq \langle U, V \rangle \leq \text{max} \langle R_1, R_2 \rangle
\]

\[
\text{min} \langle R_1, R_2 \rangle \leq \langle U, V \rangle \leq \text{max} \langle R_1, R_2 \rangle
\]
Combining expressions (*) and (**):

\[
\langle S, T \rangle \leq \max \langle P_1, P_2 \rangle < \min \langle R_1, R_2 \rangle \leq \langle U, V \rangle
\]

\[
P_1, P \in \Pi_X \\
R_1, R \in \Pi_X \\
P \subseteq S, P \subseteq T \\
R \subseteq U, R \subseteq V
\]

**Theorem 7.12** Let \((K, \mathcal{P})\) be a cladogram on \(\Pi_X\) and \(\langle, \rangle\) be

(i) a MNSM on \(\Pi_X\)

(ii) a MXSM on \(\Pi_X\)

(iii) a UASM on \(\Pi_X\)

or

(iv) a WASM on \(\Pi_X\)

such that the phenetic matrix is strictly divergent. If \(\Pi_X^*\) is a partition such that \(\Pi_X^* \subseteq K\), then there exist distinct \(S, T \in \Pi_X^*\) such that \(S \cup T \in K\) and for distinct \(U, V \in \Pi_X^*\) such that \(U \cup V \not\in K\),

\[
\langle S, T \rangle < \langle U, V \rangle.
\]

**Proof. Strategy.** (I) Choose any of the four possible superphenetic matrices. Theorem 7.8 guarantees that there exist distinct \(S, T \in \Pi_X^*\) such that \(S \cup T \in K\). Choose such \(S, T\) with the lowest value for \(\langle S, T \rangle\). Choose any distinct \(U, V \in \Pi_X^*\) such that
U ∪ V ≠ K. (II) Find distinct S', T' ∈ \( \Pi_X \) such that
\[ \langle S, T \rangle \leq \langle S', T' \rangle \]
and
\[ \max \{ Q_1, Q_2 \} < \min \{ R_1, R_2 \} \]
\[ Q_1, Q_2 \in \Pi_X \quad R_1, R_2 \in \Pi_X \]
\[ Q_1 \subseteq S', Q_2 \subseteq T' \quad R_1 \subseteq U, R_2 \subseteq V \]

Apply Lemma 7.11. That is, \( \langle S, T \rangle < \langle S', T' \rangle < \langle U, V \rangle \).

**Part I.** Choose any of the four possible superphenetic matrices. Theorem 7.8 guarantees that there are distinct S, T ∈ \( \Pi_X \) such that S ∪ T ∈ K. Choose such S, T with the smallest value \( \langle S, T \rangle \). Now choose any distinct U, V ∈ \( \Pi_X \) such that U ∪ V ≠ K.

Let U ∪ V be Z sensu Lemma 7.7 and choose distinct \( P_1, P_0 \in \Pi \cap K \) such that \( P_1 \cup P_0 \in K \subseteq K \). By construction (Part I of Lemma 7.7),
\[ P_1 \circ P_0 = P_1 \cup P_0 \subseteq U \circ V \]
Claim: \( U \circ V \supseteq P_1 \circ P_0 \). By Theorem 5.9, \( U \circ V \supseteq P_1 \circ P_0 \). Assume claim is false. That is,
\[ U \circ V = P_1 \circ P_0 \]
By Part I of Lemma 7.7, \{P_1, P_0\} is a binary partition of U ∪ V. By Theorem 5.13, U, V ∈ \{P_1, P_0\} and by hypothesis, U and V are distinct. Hence U ∪ V = P_1 ∪ P_0 ∈ K.
Contradiction.

**Part II.** Let \( S' = P_0 \) and \( T' = P_1 \). Consider any \( Q_1, Q_2, R_1, R_2 \in \Pi_X \) such that \( Q_1 \subseteq S', Q_2 \subseteq T', R_1 \subseteq U \), and \( R_2 \subseteq V \).

Claim: \( S' \circ T' = Q_1 \circ Q_2 \).

Subclaim: \( S' \neq T' \). Assume subclaim is false. That is, \( S' = T' \). By Theorem 5.1, \( S' \subseteq T' \). By elementary set theory, \( S' \cap T' = T' \neq \emptyset \). By Theorem 4.1 (ii), \( S' = T' \). By construction, S' and T' are distinct. Contradiction.
subclaim: \( \sim(T' \sqsubseteq S') \). Argument analogous to previous subclaim. By Theorem 5.9, \( S' \sqsubseteq Q_1 \) and \( T' \sqsubseteq Q_2 \). By Theorem 3.22, \( S' \circ T' = Q_1 \circ Q_2 \) and claim is satisfied.

Claim: \( U \circ V = R_1 \circ R_2 \). Argument analogous to previous claim.

From Part I:

\[
R_1 \circ R_2 = U \circ V \sqsubseteq S' \circ T' = Q_1 \circ Q_2.
\]

By Definition 7.2:

\[
\langle Q_1, Q_2 \rangle < \langle R_1, R_2 \rangle.
\]

Since this is true over any such \( R_1, R_2, Q_1, Q_2 \):

\[
\max \langle Q_1, Q_2 \rangle < \min \langle R_1, R_2 \rangle
\]

By Lemma 7.11:

\[
\langle S', T' \rangle < \langle U, V \rangle
\]

By construction:

\[
\langle S, T \rangle \leq \langle S', T' \rangle
\]

By transitivity of \( < \):

\[
\langle S, T \rangle < \langle U, V \rangle.
\]
7.12 Divergence-in-Mean

The weakest of our divergence hypotheses is called divergence-in-mean. As we shall see in chapter eight, divergence-in-mean is such a weak divergence hypothesis that only one of the seven phenogram methods to be discussed is guaranteed to yield a true cladogram every time. The basic idea behind all the divergence hypotheses is that the farther apart two OTUs are ancestrally, the larger is their phenetic matrix value. That is, if one pair of OTUs has a patriarch which is the predecessor of the patriarch of a second pair of OTUs, then the phenetic value of the first pair is greater than the phenetic value of the second pair. Under the hypothesis of strict divergence, this condition must always hold. Under the hypothesis of divergence-in-mean, this condition need only hold "on the average". The question is: an average over which phenetic values?

The cladogram \((\Delta; \mathbf{X})\) of Figure 5.1 is a cladogram on \(\Pi_X\), where \(X = A \cup B \cup C \cup D \cup E \cup F \cup G\) and \(\Pi_X = \{A, B, C, D, E, F, G\}\). The set of OTUs, namely \(\Pi_X\), is a partition of \(X\) which is also a subset of \(\Delta\). The set \(\Pi_X\) is not the only such partition, however. The set \(\{H, I, E, F, G\}\) is a partition of \(X\) which is also a subset of \(\Delta\); the set \(\{H, I, E, J\}\) is a partition of \(X\) which is also a subset of \(\Delta\); and so forth. Consider, say, the set \(\{H, I, E, J\}\) which is a partition of \(X\) and a subset of \(\Delta\). By Theorem 7.7, there must be at least one pair of junctures in \(\{H, I, E, J\}\) whose union is also a juncture. In this case, \((E\text{ and } J)\) and \((H\text{ and } I)\) are such pairs.
of OTUs, because $E \cup J = L$ is a juncture and $H \cup I = K$ is a
juncture. Let $(E \text{ and } J)$ and $(H \text{ and } I)$ be called "good" pairs
and all other possible pairs of OTUs (such as $(H \text{ and } E)$, $(I \text{ and } J)$,
etc.) be "bad" pairs. Under the hypothesis of divergence-in-mean,
there must be at least one good pair such that for any bad pair,
the average of all phenetic values "for" the bad pair must exceed
the average of all phenetic values "for" the good pair. (Under
strict divergence, every phenetic value "for" the bad pair must
exceed every phenetic value "for" the good pair.) Suppose that
$(H \text{ and } I)$ is the good pair guaranteed for the partition $\{H,I,E,J\}$.
The phenetic values "for" $(H \text{ and } I)$ include:

$$\langle A, C \rangle$$
$$\langle A, D \rangle$$
$$\langle B, C \rangle$$
$$\langle B, D \rangle$$

Consider any bad pair from the partition $\{H,I,E,J\}$, say, $(I \text{ and } J)$.
Then phenetic values "for" $(I \text{ and } J)$ include:

$$\langle C, F \rangle$$
$$\langle C, G \rangle$$
$$\langle D, F \rangle$$
$$\langle D, G \rangle$$
Divergence-in-mean guarantees that:

\[ \text{average}(\langle C, F \rangle, \langle C, G \rangle, \langle D, F \rangle, \langle D, G \rangle) > \text{average}(\langle A, C \rangle, \langle A, D \rangle, \langle B, C \rangle, \langle B, D \rangle). \]

Strict divergence, of course, would guarantee that every one of the following inequalities hold:

\[ \langle C, F \rangle > \langle A, C \rangle \]
\[ \langle C, F \rangle > \langle A, D \rangle \]
\[ \langle C, F \rangle > \langle B, C \rangle \]
\[ \langle C, F \rangle > \langle B, D \rangle \]
\[ \langle C, G \rangle > \langle A, C \rangle \]

etc. (sixteen in all).

In general, for any cladogram \((K, \Xi)\) on \(\Pi_X\), the phenetic matrix is divergent-in-mean if and only if for each partition of \(X\) which is a subset of \(K\):

(i) there is at least one ("good") pair of junctures in the partition whose union is a juncture, and

(ii) at least one "good" pair of junctures has the property that the average of all phenetic values "for" any "bad" pair of junctures exceeds the average of all phenetic values "for" that "good" pair.
**Definition 7.9** Let \((K, \mathcal{D})\) be a cladogram on \(\Pi_X\) and \(\langle, \rangle\) be a phenetic matrix on \(\Pi_X\). Then \(\langle, \rangle\) is divergent-in-mean if and only if for each partition \(\Pi_X^*\) such that \(\Pi_X^* \subseteq K\) there exist distinct \(S, T \in \Pi_X^*\) such that \(S \cup T \in K\) and

\[
\sum_{P_1, P_2 \in \Pi_X} \langle P_1, P_2 \rangle = \sum_{R_1, R_2 \in \Pi_X} \langle R_1, R_2 \rangle
\]

\[
\sum_{P_1, P_2 \leq S, P_2 \leq T} 1 = \sum_{R_1, R_2 \leq U, R_2 \leq V} 1
\]

for all distinct \(U, V \in K\) such that \(U \cup V \notin K\).

Strict divergence is a special case of divergence-in-mean:

**Theorem 7.13** Let \((K, \mathcal{D})\) be a cladogram on \(\Pi_X\) and \(\langle, \rangle\) be a strictly divergent phenetic matrix on \(\Pi_X\). Then the phenetic matrix is divergent-in-mean.

**Proof.** Consider any partition \(\Pi_X^* \subseteq K\). By Theorem 7.12 (iii), there exist distinct \(S, T \in \Pi_X^*\) such that \(S \cup T \in K\) and for any distinct \(U, V \in \Pi_X^*\) such that \(U \cup V \notin K\):
By Definition 7.9, the phenetic matrix \( \langle , \rangle \) is divergent-in-mean.

Theorem 7.14 repeats the conclusion of Theorem 7.11 (iii) for the more restricted case of a divergent-in-mean phenetic matrix:

**Theorem 7.14** Let \((K, P)\) be a cladogram on \(\Pi_X\) and \(\langle , \rangle\) be an unweighted average superphenetic matrix on \(\Pi_X\) such that the phenetic matrix is divergent-in-mean. Let \(\Pi_X\) be any partition such that \(\Pi_X \leq K\). Then there exist distinct \(S, T \in \Pi_X\) such that \(S \cup T \in K\) and for distinct \(U, V \in \Pi_X\) such that \(U \cup V \notin K\), \(\langle S, T \rangle \prec \langle U, V \rangle\).
Proof. By Definition 7.6:

$$
\begin{align*}
\langle S, T \rangle &= \frac{\sum_{p_1, p_2 \in X} \mathbb{1}_{p_1 \leq S, p_2 \leq T}}{\sum_{p_1, p_2 \in X} \mathbb{1}_{p_1 \leq S, p_2 \leq T}} \\
\langle U, V \rangle &= \frac{\sum_{r_1, r_2 \in X} \mathbb{1}_{r_1 \leq U, r_2 \leq V}}{\sum_{r_1, r_2 \in X} \mathbb{1}_{r_1 \leq U, r_2 \leq V}}
\end{align*}
$$

By Definition 7.9, $$\langle S, T \rangle < \langle U, V \rangle$$.

7.13 Divergence and Uniform Evolution

A phenetic matrix is divergent if it satisfies the intuitive condition that greater ancestral separation of OTUs results in a greater value in the phenetic matrix. There is a tendency among some biologists to regard such hypotheses as hypotheses of "uniform evolution", because it is possible to construct dendrograms which are parsimonious but extremely nonuniform which do not satisfy even the divergence-in-mean
hypothesis. Consider, for example, the hypothetical cladogram illustrated in Figure 7.8. Clearly:

\[ W \circ Z = Y \circ Z \supseteq W \circ Y \]

But it is **not** true that

\[ \langle Y, Z \rangle > \langle W, Y \rangle \]

That is, it is **not** true that

\[ 9 > 14 \]

Thus, the hypothesis of divergence-in-mean is not satisfied. As a matter of fact, a phenogram calculated by any of the nonambiguous methods of chapter eight (bottom of Figure 7.8) gives a dendrogram different from the initial cladogram (top of Figure 7.8). We cannot reassure ourselves even by hoping that this is an unlikely case that would never happen in the real, biological world. Suppose \( W = \) birds, \( Y = \) crocodiles, and \( Z = \) alligators; then our initial cladogram is indeed a reasonable account of phylogeny. Birds diverged from crocodiles after both had diverged from alligators, but by taxonomists' criteria for reptiles, crocodiles and alligators are more similar to one another than either is to birds. It is valid to question whether any existing model of numerical taxonomy could accommodate such a problem. As one might expect from the discussion
Figure 7.8. Non-uniform evolution in a hypothetical cladogram.  
(A) Diagram of patristic distances for a hypothetical cladogram with OTUs W, Y, and Z. 
(B) The maximum parsimony phenetic matrix for the hypothetical cladogram. 
(C) The phenogram computed from the phenetic matrix by any of the agglomerative methods in chapter nine.
in Section 7.3, several solutions are consistent with the maximum parsimony hypothesis: for this example (Figure 7.9), every single one! Note that the most uniform of the three possible dendrograms (Figure 7.9 A) was selected by the computing methods of this dissertation.

For certain studies, the fact that computing algorithms based on divergence hypotheses consistently find a "most uniform" (and unique!) dendrogram is a desirable property, whether or not this dendrogram is finally regarded as the true cladogram. For example, Barnabas et al., (1970) used the computing techniques of chapter nine to create a dendrogram, and nonetheless found that some branches of the dendrogram had diverged more rapidly than other branches (using Fitch and Margoliash's (1967) patristic distance formula). Since the computing techniques of chapter nine tend to obscure nonuniform evolution (by computing the "most uniform" phenogram), the fact that these investigators were able to observe nonuniformities (i.e., differing rates of evolution) in a so-computed phenogram is strong evidence that they are really there. In other words, a phenogram computed under the divergence hypotheses might serve as a uniform reference point (much as the null hypotheses in statistics). If a so-computed phenogram shows other evidence of non-uniformity, then the uniformity hypothesis should be seriously questioned.
Figure 7.9. Alternative parsimonious dendrograms. All dendrograms satisfy maximum parsimony assumptions on the phenetic matrix of Figure 7.8. Dendrograms (A) and (B) are not the "true" cladogram of Figure 7.8.
8. THE STRUCTURE OF PHENOGRAM ALGORITHMS

8.1 Iteration Structures

All phenogram algorithms involve a decision procedure for distinguishing the true cladogram from possible dendrograms which are not the true cladogram.

Conceptually, the simplest technique is to construct all possible dendrograms over a partition \( \Pi_X \) and assign a suitable value to each dendrogram. The dendrogram having a minimum (or maximum) value is the true cladogram. This is the principle underlying the maximum parsimony hypothesis (each possible dendrogram for a data set has a length, or sum of patristic distances, and the dendrogram of minimum length is the true cladogram), although it is hardly ever used in practice (Estabrook, 1968; Farris, 1970). Fitch and Margoliash (1967) suggest calculating an "average percent standard deviation" (ASPD) for each possible dendrogram and considering the dendrogram of minimum ASPD as the true cladogram, although they don’t actually carry out this ambitious plan. For a collection of \( n \) OTUs, where \( n \geq 4 \), the number of possible binary dendrograms generously exceeds \( 2^n \). In their now-classic cytochrome c "cladogram" consisting of twenty OTUs (\( 2^{20} = 1,048,576 \)), Margoliash and Fitch have calculated the ASPD for less than 100 possible dendrograms. They justify their approach by suggesting they have
calculated ASPDs for "likely" dendrograms—that is, "nearby" permutations of an initial, unweighted pair-group dendrogram (Sokal and Michener, 1958)—but they do not demonstrate that the solution is contained in the small subset of possible dendrograms which they explored.

Since typical problems in numerical taxonomy rarely involve fewer than twenty OTUs, and since evaluating numerical coefficients over millions of possible dendrograms is not feasible as a routine computing expense, the exhaustive approach has no immediate future in numerical taxonomy.

A second approach is to limit the set of all dendrograms to a small subset which must contain the true cladogram. This approach is developed by Estabrook (1968) in his solution of the Camin and Sokal (1965) cladogram problem. Unfortunately, the Camin and Sokal (1965) cladogram problem demands a limitation to "good" characters (only characters which are guaranteed not to revert to a prior state during evolution) and a level of omniscience on the part of the investigator (complete knowledge of the ancestral sequence-of-states for each character) which are rarely met in ordinary taxonomic investigations.

As we observed in Section 7.7, the approach employed by this dissertation is to distinguish individual junctures of the true cladogram from junctures of alternative dendrograms—rather than to compare entire dendrograms to one another. A
known member of the true cladogram is used to generate a small subset of possible junctures, among which there is guaranteed to be at least one juncture of the true cladogram. A value is computed for each member of this subset (i.e., the superphenetic values of chapter seven) such that any juncture of minimum (or maximum) value must be a juncture of the true cladogram. This process is repeated until all junctures of the true cladogram have been obtained.

The sequence of steps by which known junctures of the cladogram are used to generate all junctures of the cladogram is called the iteration structure of a phenogram algorithm. This dissertation develops two different iteration structures: an agglomerative and a divisive type. An agglomerative iteration structure starts from the outermost branches of a cladogram and works toward the trunk. An agglomerative structure uses the partition $\Pi_X$ as the initial set of known junctures, and stops ("knows it is finished") when it reaches $\{X\}$. A divisive iteration structure, on the other hand, starts from the trunk of a cladogram and works toward the outermost branches. A divisive structure uses $\{X\}$ as the initial known juncture and stops when it reaches $\Pi_X$.

We can illustrate the agglomerative structure by reference to the dendrogram in Figure 5.1. For this dendrogram, the set of initial OTUs is an initial partition, $\Pi_X = \Pi^1_X = \{A,B,C,D,E,F,G\}$, of $X$. 
Step 1. Choose the two members of \( \Pi^1_X \) which are "closest together", that is, have a minimum value in the superphenetic matrix. This turns out to be members C and D. Form the union of these two members, that is \( I = C \cup D \). Replace C,D with I in the partition \( \Pi^1_X \) to form a new partition, \( \Pi^2_X = \{A,B,I,E,F,G\} \).

Step 2. As before, choose the two members of \( \Pi^2_X \) which have a minimum value in the superphenetic matrix. This turns out to be members F and G. Form the union of these two members, that is \( J = F \cup G \). Replace F,G with J in partition \( \Pi^2_X \) to form a new partition, \( \Pi^3_X = \{A,B,I,E,J\} \).

Iteration Rule. Repeat this procedure until the partition \( \{X\} \) is formed. This will result in the following sequence of partitions:

\[
\begin{align*}
\Pi^1_X &= \{A,B,C,D,E,F,G\} \\
\Pi^2_X &= \{A,B,I,E,F,G\} \\
\Pi^3_X &= \{A,B,I,E,J\} \\
\Pi^4_X &= \{H,I,E,J\} \\
\Pi^5_X &= \{H,I,L\} \\
\Pi^6_X &= \{K,L\} \\
\Pi^7_X &= \{X\}
\end{align*}
\]
Clearly:

\[ \Delta = \Pi_X^1 \cup \Pi_X^2 \cup \Pi_X^3 \cup \Pi_X^4 \cup \Pi_X^5 \cup \Pi_X^6 \cup \Pi_X^7 = \{A,B,C,D,E,F,G,H,I,J,K,L,X\} \]

That is, the union of the sequence of partitions is exactly the set of junctures, \( \Delta \), of the dendrogram. (Theorem 8.3). Note that the number of steps in the agglomerative sequence (i.e., seven) exactly equals the number of initial OTUs (Theorem 8.2).

A divisive structure operates in the opposite direction from an agglomerative structure. It is equally well illustrated from Figure 5.1. This time, select \( \{X\} \) as the initial partition, \( \Pi_X^1 \). That is, \( \Pi_X^1 = \{X\} = \{A \cup B \cup C \cup D \cup E \cup F \cup G\} \).

Step 1. Choose any member of \( \Pi_X^1 \) which is not an OTU. This must be \( X \). Choose the binary partition of \( X \) whose members are "farthest apart", that is, have a maximum value in the superphenetic matrix. This turns out to be the binary partition \( \{A \cup B \cup C \cup D \cup E \cup F \cup G\} = \{K,L\} \). In place of \( X \), insert members \( K,L \) to create a new partition: \( \Pi_X^2 = \{K,L\} \).

Step 2. Choose any member of \( \Pi_X^2 \) which is not an OTU, say \( L \). As before, choose the binary partition of \( L \) whose members have a maximum value in the superphenetic matrix. This turns out to be the binary partition, \( \{E,F \cup G\} = \{E,J\} \). In place of \( L \) insert members \( E,J \) to obtain a new partition: \( \{K,E,G\} \).
Iteration Rule. Repeat this procedure until only initial OTUs remain. This results in the following sequence of partitions:

\[\pi_X^1 = \{X\}\]
\[\pi_X^2 = \{K, L\}\]
\[\pi_X^3 = \{K, E, J\}\]
\[\pi_X^4 = \{H, I, E, J\}\]
\[\pi_X^5 = \{A, B, I, E, J\}\]
\[\pi_X^6 = \{A, B, C, D, E, J\}\]
\[\pi_X^7 = \{A, B, C, D, E, F, G\}\]

These partitions were created in a different manner from the agglomerative procedure, but their union is still exactly the set of junctures, \(\Delta\), of the dendrogram:

\[\Delta = \pi_X^1 \cup \pi_X^2 \cup \pi_X^3 \cup \pi_X^4 \cup \pi_X^5 \cup \pi_X^6 \cup \pi_X^7 = \{A, B, C, D, E, F, G, H, I, J, K, L, X\}\]

Note that the divisive sequence is simply an agglomerative sequence in reverse order (Theorem 8.4).

8.2 Refinements of a Partition

From chapter four, we recall that a partition of a set \(X\) is a system of OTUs. We observed that it is possible to
have more than one system of OTUs for the same set of organisms, that is, two different partitions for the same set X. When we have two partitions and the second differs from the first only by the fact that one or more OTUs from the first partition has been more finely sectioned in the second, then the second partition is called a refinement of the first partition. A refinement doesn't break down or relocate any of the previous boundaries—it merely creates additional boundaries where there were none before. Figure 8.1 shows three partitions of the set X, the set of grandsons of Abraham. Partition \( \Pi_3^X \) is a refinement of partition \( \Pi_1^X \), because all the boundaries of \( \Pi_1^X \) have been preserved. Partition \( \Pi_4^X \) is not a refinement of partition \( \Pi_1^X \), because the boundaries of \( \Pi_1^X \) have been relocated.

Taxonomy makes extensive use of the refinement concept. For example, a set of organisms, X, organized into a system of OTUs along generic boundaries is a partition of X; the same set of organisms organized into a system of OTUs along specific boundaries is also a partition of X. Since specific boundaries never relocate or break down any generic boundaries, the specific partition is a refinement of the generic partition.

In this dissertation, we use a restricted concept of refinement, called a 1-refinement, in which only one juncture is sectioned only once. In Figure 8.1, \( \Pi_3^X \) is a 1-refinement of
Figure 8.1. Refinements of a partition. (Compare Figure 4.1.)

(A) $\Pi_1^X$ is the partition of $X$ illustrated in Figure 4.1 B. (B) $\Pi_3^X$ is a refinement of $\Pi_1^X$. (C) $\Pi_4^X$ is not a refinement of $\Pi_3^X$. 

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nebaioth</td>
<td>Mishma</td>
<td>Tema</td>
</tr>
<tr>
<td>Kedar</td>
<td>Dumah</td>
<td>Jetur</td>
</tr>
<tr>
<td>Adbeel</td>
<td>Massa</td>
<td>Naphish</td>
</tr>
<tr>
<td>Mibsam</td>
<td>Hadad</td>
<td>Kedemah</td>
</tr>
</tbody>
</table>
Π¹_X, because only one OTU from Π¹_X (namely, \{Esau, Jacob\}) was
sectioned only once (to give \{Esau\} and \{Jacob\} in the new partition,
Π³_X). In our example from taxonomy, the specific partition would be
a 1-refinement of the generic partition only if every genus had exactly
one species, except for a single genus which had two species.

**Definition 8.1** Let Π¹_X and Π²_X be partitions. Then Π²_X is a
1-refinement of Π¹_X if and only if there are \(P_a, P_b \in \Pi^2_X\) such that

(i) \(P_a \cup P_b \in \Pi^1_X\)

and

(ii) \(\Pi^1_X - \{P_a \cup P_b\} = \Pi^2_X - \{P_a \cup P_b\}\)

8.3 Agglomerative and Divisive Iteration Structures

The reader will recall from Section 8.1 that an agglomerative
phenogram algorithm is a sequence of partitions of \(X\). Each step
(except the last) in an agglomerative phenogram algorithm is a
1-refinement of the next step:

\[\Pi^1_X \text{ is a 1-refinement of } \Pi^2_X\]

\[\Pi^2_X \text{ is a 1-refinement of } \Pi^3_X\]

\[\Pi^3_X \text{ is a 1-refinement of } \Pi^4_X\]

\[\Pi^4_X \text{ is a 1-refinement of } \Pi^5_X\]

\[\Pi^5_X \text{ is a 1-refinement of } \Pi^6_X\]

\[\Pi^6_X \text{ is a 1-refinement of } \Pi^7_X\]
By contrast, a divisive phenogram algorithm is a sequence of partitions of $X$ such that each step (except the first) is a 1-refinement of the previous step. The starting step for an agglomerative phenogram algorithm is $\Pi^1_X = \Pi_X$ and the stopping rule is $\Pi^n_X = \{X\}$. A divisive phenogram, on the other hand, has $\Pi^1_X = \{X\}$ as its starting step and $\Pi^m_X = \Pi_X$ as its stopping rule. (We use a different letter, $m$, for the last partition of a divisive algorithm, because we haven't yet proved that the number of steps in an agglomerative phenogram algorithm equals the number in a divisive phenogram algorithm.)

The following formal definitions of agglomerative and divisive binary phenogram algorithms parallel our informal discussion.

**Definition 8.2** AAlg is an **agglomerative binary phenogram algorithm** on $\langle, \rangle$ if and only if AAlg = $(\Pi^1_X, \ldots, \Pi^n_X)$ is an ordered sequence of partitions and $\langle, \rangle$ is a superphenetic matrix on $\Pi_X$ and:

(i) $\Pi^1_X = \Pi_X$

(ii) for each $i$, $1 \leq i \leq n-1$, there are distinct $S, T \in \Pi^i_X$ such that $\langle S, T \rangle = \min_{U, V \in \Pi^i_X} \langle U, V \rangle$ and

$$\Pi^{i+1}_X = (\Pi^i_X - \{S, T\}) \cup \{S \cup T\}$$

and

(iii) $\Pi^n_X = \{X\}$. 

**Definition 8.3** DAlg is a divisive binary phenogram algorithm on \(\langle,\rangle\) if and only if DAlg = \((\Pi^1_X, \ldots, \Pi^m_X)\) is an ordered sequence of partitions and \(\langle,\rangle\) is a superphenetic matrix on \(\Pi_X\) and:

(i) \(\Pi^1_X = \{X\}\)

(ii) for each \(i, 1 \leq i \leq m-1\), there is a binary partition \(\{T_1, T_2\} \subseteq \Sigma\) of \(T \in \Pi^i_X\) such that

\[
\langle T_1, T_2 \rangle = \max_{\{U_1, U_2\} \subseteq \Sigma} \langle U_1, U_2 \rangle \text{ and } \Pi^{i+1}_X = (\Pi^i_X - \{T\}) \cup \{T_1, T_2\}
\]

is a binary partition of \(T\)

and

(iii) \(\Pi^m_X = \Pi_X\).

**8.4 Phenogram Algorithms Generate Dendrograms**

In this section, we shall demonstrate that the agglomerative and divisive phenogram algorithms both generate dendrograms. In the examples from Section 8.1, the union of our sequence of partitions is in both cases the set of junctures for a dendrogram. To show that this is always true, we shall employ the following strategy. First, we shall show that the union of partitions from an agglomerative phenogram algorithm is the set of junctures of a dendrogram by showing that it satisfies the four criteria of Definition 5.3. Incidentally, we shall show that the number of partitions in the sequence, \(n\), equals the number of members in the partition, \(N(\Pi_X)\).
We shall then demonstrate that a divisive phenogram algorithm is simply an agglomerative phenogram algorithm in reverse order. Since order is irrelevant in specifying the union of members in a sequence, it follows that the union of partitions from a divisive phenogram algorithm is the set of junctures of a dendrogram. Incidentally, since the order of a sequence is irrelevant in specifying the number of members, it follows that the number of partitions in a divisive phenogram algorithm, \( m \), equals the number of members in the initial partition, \( N(\Pi_X) \).

Theorem 8.1 shows that the number of members in a partition is one less than the number of members in a \( 1 \)-refinement of that partition.

**Theorem 8.1** Let \( \Pi_X^1 \) and \( \Pi_X^2 \) be partitions and let \( \Pi_X^2 \) be a \( 1 \)-refinement of \( \Pi_X^1 \). Then \( N(\Pi_X^2) = N(\Pi_X^1) + 1 \).

**Proof.** From Definition 8.1 (ii), the reader will easily derive that:

\[
N(\Pi_X^2) - 2 = N(\Pi_X^2 - \{P_a, P_b\}) = N(\Pi_X^1 - \{P_a \cup P_b\}) = N(\Pi_X^1) - 1,
\]

from which the theorem follows immediately.

Theorem 8.2 shows that the number of steps, \( n \), in an agglomerative phenogram algorithm is exactly equal to the number of members, \( N(\Pi_X) \), in the partition \( \Pi_X \):
Theorem 8.2  Let $AA_{lg} = (\Pi^1_X, \ldots, \Pi^n_X)$ be an agglomerative phenogram algorithm on $\langle \rangle$. Then $n = N(\Pi_X)$.

Proof. By Theorem 8.1, since $\Pi^i_X$ is a 1-refinement of $\Pi^{i+1}_X$ (see Definition 8.2 (ii)):

$$N(\Pi^{i+1}_X) + 1 = N(\Pi^i_X), \; i = 1, \ldots, n-1.$$ 

Summation over $i$ from 1 to $n-1$ yields:

$$N(\Pi^n_X) + \sum_{k=2}^{n-1} N(\Pi^k_X) + n - 1 = N(\Pi^1_X) + \sum_{k=2}^{n-1} N(\Pi^k_X).$$

By Definition 8.2 (i), $\Pi^1_X = \Pi_X$, and by Definition 8.2 (iii), $\Pi^n_X = \{x\}$, so that $N(\Pi^n_X) = 1$. Therefore, the equality just obtained, 

$$N(\Pi^n_X) + n - 1 = N(\Pi^1_X),$$

reduces to $n = N(\Pi^1_X)$.

Theorem 8.3 shows that the union of the partitions in an agglomerative phenogram algorithm is the set of junctures of a binary dendrogram:
Theorem 8.3 Let \( A_{\text{Alg}} = (\Pi^1_X, \ldots, \Pi^n_X) \) be an agglomerative phenogram algorithm on \((,\). If \( \Delta^* = \bigcup_{i=1}^{n} \Pi^i_X \), then

(i) \( N(\Delta^*) = 2N(\Pi_X) - 1 \)

and

(ii) \((\Delta^*, \mathcal{D})\) is a binary dendrogram.

Proof. Strategy. (I-IV) Show that for each \( S \in \Delta^* \), \( S \) satisfies the sufficiency portion of parts (i) - (iv) of Definition 5.3. That is, there is a \( \Delta \) such that \( \Delta^* \subseteq \Delta \) and \((\Delta, \mathcal{D})\) is a dendrogram on \( \Pi_X \). (V) Then show that \( N(\Delta^*) = N(\Delta) \), so that \( \Delta^* = \Delta \), which implies the necessity portion of Definition 5.3 (i) - (iv).

Part 0. Preliminaries. Suppose that \( S \in \Delta^* \). Define the predecessor sequence of \( S \) as follows. Since \( S \in \Delta^* \), there is at least one \( i, 1 \leq i \leq n \), such that \( S \in \Pi^i_X \). Choose the smallest such \( i \). Create the ordered sequence \((S = S_1, \ldots, S_n)\) as follows. For each \( h, i \leq h \leq n-1 \), by Definition 8.2 (ii), \( S_h \) is a \( l \)-refinement of \( S_{h+1} \). Let \( P_a, P_b \) be as in Definition 8.1. If \( S_h = P_a \) or \( S_h = P_b \), then let \( S_{h+1} = P_a \cup P_b \). Otherwise, let \( S_{h+1} = S_h \). End of Definition.

Properties: Consider any \( j, k \) \( 1 \leq j \leq k \leq n \). By construction, \( S_j \subseteq S_{j+1} \). By transitivity of \( \subseteq \), \( S_j \subseteq S_k \). By Definition 8.1, \( S_k \in \Pi^k_X \). By Definition 8.2 (iii), \( S_n \in \{X\} \), so that \( S_n = X \). Suppose \( S_{j+1} \neq S_j \). Claim: \( S_{j+1} \not\supseteq S_j \). Assume the contrary. That is, there is an \( R \in \Delta^* \) such that \( S_{j+1} \supseteq R \supseteq S_j \). Since \( R \in \Delta^* \), there must be a
$g, 1 \leq g \leq n$, such that $R \in \Pi_X^g$. If $g \leq j$, then $R = S_j$ or $R \cap S_j = \emptyset$, contradiction. If $g > j$, then $R = S_g \supseteq S_{j+1}$ or $\emptyset = R \cap S_g \supseteq R \cap S_{j+1}$, contradiction.

**Part I.** Consider any $S \in \Delta$. Claim: (a) $S \in \Pi_X$ or (b) $S = X$ or (c) there is a $P \in \Pi_X$ such that $X \supseteq S \supseteq P$.

**Case I-a** Suppose that $S \in \Pi_X^1$. Then by Definition 8.2 (i), $S \subseteq \Pi_X^n$.

**Case I-b** Suppose that $S \in \Pi_X^n$. Then by Definition 8.2 (iii), $S = X$.

**Case I-c** Otherwise, $S \in \Pi_X^i$ where $2 \leq i \leq n-1$. Choose the smallest such $i$. Construct the predecessor sequence $(S = S_1, \ldots, S_n)$. By Part 0, $X = S_n \supseteq S_1 = S$. Since $\emptyset \neq S \subseteq X$, choose any $x \in S$. By Definition 4.1 (ii), there is a unique $P \in \Pi_X$ such that $x \in P$, and hence $S \cap P \neq \emptyset$. Construct the predecessor sequence $(P = P_1, \ldots, P_n)$. Since $\emptyset \neq P \cap S \subseteq P_1 \cap S$, by Theorem 4.1 (ii) (since $P_i, S_i$ are both members of partition $\Pi_X$), $P_1 = S_1 = S$. By Part 0, $S = P_1 \supseteq P_1 = P$.

**Part II.** Consider distinct $S, T \in \Delta$. Claim: $S \supseteq T$ or $T \supseteq S$ or $S \cap T = \emptyset$. Choose the smallest $j, k$, $1 \leq j, k \leq n$, such that $S \in \Pi_X^j$ and $T \in \Pi_X^k$. Suppose $j \leq k$. Construct the predecessor sequence $(S = S_j, \ldots, S_n)$. By Part 0, if $S_k = T$, then $T \supseteq S$. If $S_k \neq T$, then by Theorem 4.1 (ii), $S \cap T \subseteq S_k \cap T = \emptyset$. Suppose $k \leq j$. By similar reasoning, either $S \supseteq T$ or $S \cap T = \emptyset$. 


Part III. Consider any \( S \in \Delta^* - \{X\} \). Claim: there is a unique \( T \in \Delta^* \) such that \( T \triangleright S \).

Part III-a Claim: there is at least one such \( T \). Since \( S \not\in \{X\}, S \not\in \Pi^n_X \) and there is a \( k, 1 \leq k \leq n-1 \), such that \( S \in \Pi^k_X \). Choose the smallest such \( k \), and construct the predecessor sequence 
\( (S = S_k, \ldots, S_n) \). By Part 0, \( X = S_n \triangleright S_k = S \). Since \( S_n \triangleright S_k \), there is a sequence \( S_n \triangleright \ldots \triangleright T \triangleright S_k \).

Part III-b Claim: This \( T \) is unique. Assume the contrary. That is, there is a distinct \( T_* \in \Delta^* \) such that \( T_* \triangleright S_k \). By Definition 5.1, \( T_* \triangleright S_k \) and \( T \triangleright S_k \). Hence \( T_* \cap T \triangleright S_k \neq \emptyset \). By Part II, either \( T_* \triangleright T \) or \( T \triangleright T_* \). If \( T_* \triangleright T \), then by Theorem 5.1, \( T_* \triangleright T \triangleright S_k \), and by Definition 5.1 (ii), \( T_* \triangleright S_k \); contradiction. If \( T \triangleright T_* \), then by Theorem 5.1, \( T \triangleright T \triangleright S_k \), and by Definition 5.1 (ii), \( T \triangleright S_k \); contradiction.

Part IV. Consider any \( S \in \Delta^* - \Pi^X \). Claim: there is a unique binary partition \( \{T_1, T_2\} \) of \( S \) such that \( \{T_1, T_2\} \subseteq \Delta^* \) and \( S \triangleright T_i \) for \( i = 1,2 \).

Part IV-a Claim: there is a binary partition \( \{T_1, T_2\} \) of \( S \) such that \( \{T_1, T_2\} \subseteq \Delta^* \). Since \( S \not\in \Pi^X = \Pi^1_X \), choose the smallest \( j, 2 \leq j \leq n \), such that \( S \in \Pi^j_X \). By Definition 8.2, \( \Pi^{j-1}_X \) (which must exist, since \( j \geq 2 \)) is a \( 1 \)-refinement of \( \Pi^j_X \). By Definition 8.1, there are nonempty \( T_1, T_2 \in \Pi^{j-1}_X \subseteq \Delta^* \) such that \( T_1 \cup T_2 = S \). By Definition 1.1, \( \{T_1, T_2\} \) is a binary partition of \( S \).
Part IV-b Claim: $S \triangleright T_i$ for $i = 1,2$. Consider any $i$, $i = 1,2$. Since $T_i \neq \emptyset$, choose any $x \in T_i$. Since $\Pi_X$ is a partition, by Definition 4.1, there is a $P \in \Pi_X$ such that $x \in P$. Construct the predecessor sequence of $P$, namely, $(P = P_1, \ldots, P_n)$. Since $S$ and $P_j$ are both members of the partition $\Pi_X^j$, and since $x \in P \cap T_i \subseteq P_j \cap T_i \subseteq P_j \cap S$, by Theorem 4.1 (ii), $P_j = S$. By a similar argument, $T_i = P_{j-1}$. Since $j$ is the smallest $j$ such that $S \cap \Pi_X^j$, $S \not\subseteq \Pi_X^{j-1}$ and $T_i \neq S$. By Part 0, $S = P_j \triangleright P_{j-1} = T_i$.

Part IV-c Claim: $\{T_1, T_2\}$ is unique. Assume the contrary. That is, there is a distinct binary partition $\{U_1, U_2\}$ of $S$ such that $\{U_1, U_2\} \subset \Delta^*$ and $S \triangleright U_h$ for $h = 1,2$. By Theorem 4.2 three of four sets $T_i \cap U_h$ are nonempty. Without loss of generality, let $T_1 \cap U_1$ be nonempty. By Part II, either $T_1 \supset U_1$ or $U_1 \supset T_1$.

If $T_1 \supset U_1$, then by Theorem 5.1, $S \supset T_1 \supset U_1$ and by Definition 5.1, $\neg(S \triangleright U_1)$; contradiction. Similarly, if $U_1 \supset T_1$, then $\neg(S \triangleright T_1)$; contradiction.

Part V Claim: $N(\Delta^*) = 2N(\Pi_X) - 1$. By Theorem 8.2, $n = N(\Pi_X)$. Subclaim: for $1 \leq k \leq n = N(\Pi_X)$, $N(\bigcup_{i=1}^{k} \Pi_X^i) = N(\Pi_X) + k - 1$.

Proof by mathematical induction.

True for $k=1$: By Definition 8.2 (i),

$$N(\bigcup_{i=1}^{k} \Pi_X^i) = N(\Pi_X) = N(\Pi_X) = N(\Pi_X) + k - 1.$$
True for $k$ implies true for $k+1$: By Definition 8.2 (ii), $\Pi$ is a $l$-refinement of $\Pi^{k+1}_X$. Let $P_a, P_b \in \Pi^k_X$ be as in Definition 8.1. Then $\Pi^k_X \cup \Pi^{k+1}_X = \Pi^k_X \cup \{P_a \cup P_b\}$ and $\Pi^k_X$ and $\{P_a \cup P_b\}$ are disjoint. Therefore:

$$N(\bigcup_{i=1}^{k+1} \Pi^i_X) = N(\bigcup_{i=1}^{k} \Pi^i_X) \cup (\Pi^{k+1}_X)$$

$$= N(\bigcup_{i=1}^{k} \Pi^i_X) \cup \{P_a \cup P_b\}$$

$$= N(\bigcup_{i=1}^{k} \Pi^i_X) + N(\{P_a \cup P_b\})$$

$$= N(\bigcup_{i=1}^{k} \Pi^i_X) + 1$$

By our inductive assumption,

$$N(\bigcup_{i=1}^{k+1} \Pi^i_X) = N(\bigcup_{i=1}^{k} \Pi^i_X) + 1 = (N(\Pi^k_X) + k - 1) + 1$$

$$= N(\Pi^k_X) + (k + 1) - 1.$$

Theorems 8.4 and 8.5 parallel the conclusions of Theorems 8.2 and 8.3, respectively, for divisive binary phenogram algorithms. Theorem 8.4 shows that the number of steps, $m$, in a divisive binary phenogram algorithms is exactly equal to the number of OTUs:
Theorem 8.4  Let $DAlg = (\Pi_X^1, \ldots, \Pi_X^m)$ be a divisive binary phenogram algorithm on $\langle \rangle$. Then $m = N(\Pi_X)$.

Proof. For $i=1, \ldots, n-1$, $\Pi_X^{i+1}$ is a $1$-refinement of $\Pi_X^i$, so that

$$N(\Pi_X^{i+1}) = N(\Pi_X^i) + 1,$$

from which, by summation,

$$N(\Pi_X^m) = N(\Pi_X^1) + m - 1.$$

Since $\Pi_X^1 = \{X\}$, $N(\Pi_X^1) = 1$, and $\Pi_X^m = \Pi_X$, the theorem follows.

Theorem 8.5 shows (i) that the number of members in the union of partitions in a divisive binary phenogram algorithm is one less than two times the number of OTUs and (ii) that this union of partitions is the set of junctures for a binary dendrogram. Since the proof of Theorem 8.5 is so similar to that of Theorem 8.3, only a sketch of proof is provided.

Theorem 8.5  Let $DAlg = (\Pi_X^1, \ldots, \Pi_X^m)$ be a divisive binary phenogram algorithm on $\langle \rangle$. If $\Delta^* = \bigcup_{i=1}^{n} \Pi_X^i$, then

(i) $N(\Delta^*) = 2N(\Pi_X) - 1$

and

(ii) $(\Delta^*, D)$ is a binary dendrogram on $\Pi_X$. 
Sketch of Proof. (I-IV) Show that for each $S \in \Delta^*$, $S$ satisfies the sufficiency portion of parts (i) - (iv) of Definition 5.3. That is, there is a $\Delta$ such that $\Delta^* \subseteq \Delta$ and $(\Delta, \mathcal{P})$ is a binary dendrogram on $\Pi_X$. (V) Then show that $N(\Delta^*) = N(\Delta)$, so that $\Delta^* = \Delta$.

8.5 Cladogram Functions

A phenogram algorithm, as discussed in previous sections, is merely a sequence of partitions which assembles all the junctures of a dendrogram. The objective of this dissertation, however, is not just to construct any dendrogram: we want the true cladogram, or tree structure of ancestral relationships.

The basic idea underlying all the phenogram algorithms developed in this dissertation is a decision procedure by which individual junctures from the true cladogram are distinguished from possible junctures which are not members of the true cladogram. An agglomerative phenogram algorithm is a sequence of "steps", such that each "step" is a partition of the initial set $X$, and each partition (except the last) is a 1-refinement of the next partition in the sequence. That is, $\Pi_{X}^{i+1}$ consists of all elements of $\Pi_{X}^{i}$ with the exception of two, $P_a$ and $P_b$ say, but including the element $P_a \cup P_b$. How do we decide which $P_a$ and $P_b$ to remove from $\Pi_{X}^{i}$? This decision is made by finding which superphenetic value $\langle P_a, P_b \rangle$ is minimum over all
pairs of \( P_a, P_b \in \Pi^j_X \). Any superphenetic matrix which guarantees that a true cladogram juncture will always be selected in an agglomerative binary phenogram algorithm is called an agglomerative cladogram function:

**Definition 8.4** Let \((K, d)\) be a cladogram on \( \Pi_X \) and \( \langle , \rangle \) be a superphenetic matrix on \( \Pi_X \). Then \( \langle , \rangle \) is an agglomerative cladogram function if and only if for each agglomerative phenogram algorithm, 

\[
A_{\text{Alg}} = (\Pi^1_X, \ldots, \Pi^n_X), \text{ on } \Pi_X, \bigcup_{i=1}^n \Pi^i_X = K.
\]

A divisive cladogram algorithm is a sequence of "steps" such that each step is a partition of the initial set \( X \), and each partition (except the first) is a \( l \)-refinement of the previous partition in the sequence. That is, \( \Pi^{j+1}_X \) consists of all elements of \( \Pi^j_X \) with the exception of one, \( P \) say, but including the two elements, \( P_a \) and \( P_b \), of a binary partition, \( \{P_a, P_b\} \) of \( P \). How do we decide how to partition \( P \)? This decision is made on the basis of a computation over all partitions \( \{P_a, P_b\} \) of \( P \) such that \( \{P_a, P_b\} \subseteq \Sigma \). Any superphenetic matrix which guarantees that true cladogram junctures will always be selected in a divisive binary phenogram algorithm is called a divisive cladogram function:

**Definition 8.5** Let \((K, d)\) be a cladogram on \( \Pi_X \) and \( \langle , \rangle \) be a superphenetic matrix on \( \Pi_X \). Then \( \langle , \rangle \) is a divisive cladogram function if and only if for each divisive phenogram algorithm, 

\[
D_{\text{Alg}} = (\Pi^1_X, \ldots, \Pi^m_X), \text{ on } \Pi_X, \bigcup_{i=1}^m \Pi^i_X = K.
\]
9. EVALUATION OF PHENOGRAM ALGORITHMS

9.1 Strategy

This is the final chapter of this dissertation. For each one of seven phenogram algorithms we shall set up sufficient conditions such that it generates a true cladogram. In chapter seven, we introduced four calculating methods for superphenetic matrices:

- minimal
- maximal
- unweighted average
- weighted average

In chapter eight, we introduced two iteration structures:

- agglomerative
- divisive

Ostensibly, it should be possible to use each possible superphenetic matrix with each iteration structure, giving eight possible algorithms:

1. minimal agglomerative (MNA)
2. maximal agglomerative (MXA)
3. unweighted average agglomerative (UAA)
4. weighted average agglomerative (WAA)
5. minimal divisive (MND)
6. maximal divisive (MXD)
7. unweighted average divisive (UAD)
8. weighted average divisive (WAD)
Actually, the weighted average divisive (WAD) algorithm is not feasible, because nesting levels for calculating a weighted average must be reckoned from the initial OTUs through a known sequence of (true cladogram) junctures which are subsets of the juncture whose nesting level is currently being evaluated (see Definition 7.7). This is possible in an agglomerative iteration structure but not in a divisive iteration structure. All other algorithms are possible. Algorithms MNA, MXA, UAA, WAA, and UAD are currently available in the numerical taxonomy literature, and usually go by the following names:

MNA: Single Linkage Method (Sneath, 1957)
MXA: Complete Linkage Method (Sorensen, 1948)
UAA: Unweighted Pair-Group Method (Sokal and Michener, 1958)
WAA: Weighted Pair-Group Method (Sokal and Michener, 1958)
UAD: Moore-Goodman Method (Moore et al., 1969)

Methods MND and MKD were discovered by creating the classification of the existing methods and noting the "missing" methods.

In chapter seven, we developed three different conditions of evolutionary divergence, listed here in ascending order of weakness (i.e., the former implies the latter):

- equidivergence
- strict divergence
- divergence-in-mean
In this chapter, we shall evaluate the performance of each of the seven methods with respect to the three evolutionary hypotheses. The results of this evaluation are summarized in Figure 9.1. The stated divergence hypotheses represent the weakest condition under which a true cladogram is guaranteed; counterexamples can be found for the next weaker condition. Thus, a minimal divisive phenogram algorithm, for example, gives valid cladograms under the hypothesis of strict divergence, but not under the next weaker hypothesis of divergence-in-mean.

The strategy of this chapter is as follows: in Section 9.2, we shall prove that the conditions listed in Table 9.1 are indeed sufficient; in Section 9.3, we shall demonstrate that the next weaker condition results in a counterexample. Since the unweighted average agglomerative (UAA) algorithm is valid under the weakest of the three divergence conditions (and hence under the other two, stronger ones), we shall use a UAA phenogram as a standard of comparison, against which the performance of other phenogram algorithms will be judged. Any disagreement between a UAA phenogram and a phenogram calculated by another method shows where the other method has failed.

Since the unweighted average agglomerative algorithm is the "best" algorithm under the criteria of this dissertation, it is summarized in the form of a simple computer flow chart in Appendix 11.3.
<table>
<thead>
<tr>
<th>Structure</th>
<th>Agglomerative</th>
<th>Divisive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>Strict Divergence</td>
<td>Strict Divergence</td>
</tr>
<tr>
<td>Maximal</td>
<td>Strict Divergence</td>
<td>(Ambiguous)</td>
</tr>
<tr>
<td>Unweighted Average</td>
<td>Divergence-in-Mean</td>
<td>Equidivergence</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>Strict Divergence</td>
<td>(Non-existent)</td>
</tr>
</tbody>
</table>

Table 9.1. Performance of phenogram algorithms. This table shows the weakest evolutionary hypothesis under which each phenogram algorithm is guaranteed to yield the cladogram.
9.2 Proofs

Theorem 9.1 shows that an unweighted average superphenetic matrix is an agglomerative computational function under the condition of divergence-in-mean:

Theorem 9.1. Let \((K, D)\) be a cladogram on \(\Pi_X\) and \(\langle,\rangle\) be an unweighted average superphenetic matrix on \(\Pi_X\) such that the phenetic matrix is divergent-in-mean. Then \(\langle,\rangle\) is an agglomerative cladogram function.

Proof. Strategy. (I) Show by mathematical induction that for

\[1 \leq i \leq n, \quad \Pi_X^i \subseteq K.\]  

(II) Let \(K^* = \bigcup_{i=1}^{n} \Pi_X^i\). By Part I, \(K^* \subseteq K\). Since \(N(K^*) = N(K)\), \(K^* = K\).

**Part I.** Claim: For \(1 \leq i \leq n\), \(\Pi_X^i \subseteq K\). Proof by mathematical induction.

True for \(i = 1\): By Definition 8.2 (i), \(\Pi_X^1 = \Pi_X \subseteq K\).

True for \(i\) implies true for \(i+1\): Since true for \(i\), \(\Pi_X^i \subseteq K\).

Consider the \(S, T \in \Pi_X^i\) such that \(\Pi_X^{i+1} = (\Pi_X^i - \{S, T\}) \cup \{S \cup T\}\). By Definition 8.2(ii), \(\langle S, T \rangle = \min \{\langle U, V \rangle\}_{U \in \Pi_X^i, V \notin \Pi_X^i, U \neq V}\). By Theorem 7.14, \(S \cup T \in K\).

Hence \(\Pi_X^{i+1} \subseteq K\).

**Part II.** Let \(K^* = \bigcup_{i=1}^{n} \Pi_X^i\). Claim: \(K^* = K\). By Part I, \(K^* \subseteq K\). By Theorem 8.3, part V, \(N(K^*) = 2N(\Pi_X) - 1\). By Theorem 5.5, \(N(K) = 2N(\Pi_X) - 1\). Since \(N(K^*) = N(K)\) and both are finite, \(K^* = K\).

By Definition 8.4, \(\langle,\rangle\) is an agglomerative cladogram function.
Theorem 9.2 shows that a minimal, maximal, or weighted average superphenetic matrix is an agglomerative cladogram function under the condition of strict divergence:

**Theorem 9.2** Let \((K, \mathcal{D})\) be a cladogram on \(\Pi_X\) and \((\cdot, \cdot)\) be

(i) a MNSM on \(\Pi_X\)

(ii) a MXSM on \(\Pi_X\)

or

(iii) a WASM on \(\Pi_X\)

such that the phenetic matrix is strictly divergent. Then \((\cdot, \cdot)\) is an agglomerative cladogram function.

**Proof.** Verbatim the same as the proof of Theorem 9.1, except that "Theorem 7.14" (at the end of Part I) should now be replaced by "Theorem 7.12".

Theorem 9.3 shows that an unweighted average superphenetic matrix is a divisive cladogram function under the condition of equidivergence:

**Theorem 9.3** Let \((K, \mathcal{D})\) be a cladogram on \(\Pi_X\) and \((\cdot, \cdot)\) be an unweighted average superphenetic matrix on \(\Pi_X\) such that the phenetic matrix is equidivergent. Then \((\cdot, \cdot)\) is a divisive cladogram function.

**Proof.** Strategy (I) Show by mathematical induction that for

\[ 1 \leq i \leq n, \quad \Pi_X^i \subseteq K. \] (II) Let \(K^* = \bigcup_{i=1}^{n} \Pi_X^i\). By Part I, \(K^* \subseteq K\). Since \(N(K^*) = N(K)\), \(K^* = K\).
Part I. Claim: For $1 \leq i \leq n$, $\Pi^i_X \subseteq K$. Proof by mathematical induction.

True for $i = 1$: By Definition 8.3 (i), $\Pi^1_X = \{x\} \subseteq K$.

True for $i$ implies true for $i+1$: Since true for $i$, $\Pi^i_X \subseteq K$.

Consider the $T \in \Pi^i_X$ and the binary partition $\{T_1, T_2\}$ of $T$ such that $\Pi^{i+1}_X = (\Pi^i_X \setminus \{T\}) \cup \{T_1, T_2\}$. By Definition 8.3 (ii),

$\langle T_1, T_2 \rangle = \max_{\{U_1, U_2\} \subseteq \Sigma} \langle U_1, U_2 \rangle$.

By Theorem 7.5, $\{T_1, T_2\} \subseteq K$ is a binary partition of $T$.

Hence $\Pi^{i+1}_X \subseteq K$.

Part II. Let $K^* = \bigcup_{i=1}^{n} \Pi^i_X$. By Part I, $K^* \subseteq K$. By Theorem 8.5 part V, $N(K^*) = 2N(\Pi^1_X) - 1$. By Theorem 5.5, $N(K) = 2N(\Pi^1_X) - 1$. Since $N(K^*) = N(K)$ and both are finite, $K^* = K$. By Definition 8.5, $\langle , \rangle$ is a divisive cladogram function.

Theorem 9.4 shows that a minimal superphenetic matrix is a divisive cladogram function under the condition of strict divergence:

**Theorem 9.4** Let $(K, \Phi)$ be a cladogram on $\Pi^1_X$ and $\langle , \rangle$ be a minimal average superphenetic matrix on $\Pi^1_X$ such that the phenetic matrix is strictly divergent. Then $\langle , \rangle$ is a divisive cladogram function.

**Proof.** Verbatim the same as the proof of Theorem 9.3, except that "Theorem 7.5" (at the end of Part I) should now be replaced by "Theorem 7.6".
9.3 Counterexamples

Now that we have shown which divergence conditions are sufficient for the various phenogram algorithms to yield true cladograms, we shall produce counterexamples to show that any weaker divergence conditions result in dendrograms which are not true cladograms. For example, to show that strict divergence, but not divergence-in-mean, is sufficient for the minimal agglomerative (MNA) algorithm, we shall produce a phenetic matrix which is divergent-in-mean but not strictly divergent, and show that the result of a minimal agglomerative calculation is different from the result of an unweighted average agglomerative (UAA) calculation on the same phenetic matrix. Since, by Theorem 9.1, divergence-in-mean is a sufficient condition for a true cladogram by the UAA method, we are assured that the UAA calculation is "correct" and the MNA calculation is "incorrect". I shall use this standard of comparison for all my counterexamples.

Agglomerative counterexamples can give us a good deal of insight into the differences between various phenogram algorithms, because our prototype algorithm, the UAA algorithm, is also an agglomerative algorithm. Thus we can follow the construction of the prototype and the pathological phenogram step-by-step from the initial OTUs, and focus our attention upon the first step in which
the prototype differs from the pathological phenogram. Consider, for example, the MNA algorithm, illustrated in Figure 9.1. The initial collection of OTUs is the partition:

\[ \{U,V,W,Y,Z\} \]

In both the UAA and MNA algorithms, the first step unions the sets U and V to form the new partition:

\[ \{U \cup V,W,Y,Z\} \]

In both the UAA and MNA algorithms, the second step unions the sets \( U \cup V \) and W to form the new partition:

\[ \{U \cup V \cup W,Y,Z\} \]

Trouble arises in the third step. Here we have a choice of three possible pairs of sets which can be unioned:

\[ U \cup V \cup W,Y \]

\[ U \cup V \cup W,Z \]

or

\[ Y,Z \]

The superphenetic entry \( \langle U \cup V \cup W,Z \rangle \) is minimal (the criterion for "closest" pair in an agglomerative algorithm) in the prototype algorithm, whereas the entry \( \langle U \cup V \cup W,Y \rangle \) is minimal in the MNA
Figure 9.1. Minimal agglomerative: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) MNA phenogram. (D) Critical super-phenetic calculations. \langle U \cup V \cup W, Y \rangle is minimal for UAA but \langle U \cup V \cup W, Z \rangle is minimal for MNA.
algorithm. The reason is obvious: UAA averages over the values for all appropriate OTU pairs in the phenetic matrix, whereas MNA takes only the value for the minimum pair. A single, low-valued "oddball" point can spoil the entire calculation in the MNA algorithm, whereas this oddball averages out in the UAA algorithm.

The MXA and WAA algorithms, illustrated in Figures 9.2 and 9.3 respectively, fail for essentially similar reasons: the difference at a crucial step between an unweighted average and, respectively, a maximum or weighted average.

Divisive counterexamples give us less insight into the differences between phenogram algorithms, because the standard of comparison (UAA) is an agglomerative algorithm. It is not possible to compare the execution of the prototype and pathological algorithms step-by-step, because an agglomerative algorithm starts from the OTUs and works toward the trunk, whereas a divisive algorithm starts from the trunk and works toward the OTUs. Divisive algorithms perform much worse than agglomerative algorithms. A weighted average divisive algorithm doesn't even exist, because nesting levels must be reckoned from the OTUs portion of the dendrogram to the current level of calculation. Since a divisive algorithm starts from the trunk and works toward the OTUs, nesting levels (which are needed at each step of the calculation) can never be reckoned until the end of the calculation. A maximal divisive algorithm is also very pathological: it
Figure 9.2. Maximal agglomerative: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) MXA phenogram. (D) Critical super-phenetic calculations. \( \langle U \cup V \cup W, Y \rangle \) is minimal for UAA but \( \langle U \cup V \cup W, Z \rangle \) is minimal for MXA.
Figure 9.3. Weighted average agglomerative: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) WAA phenogram. (D) Critical superphenetic calculations. \( \langle U U V U W, Y \rangle \) is minimal for UAA but \( \langle U U V U W, Z \rangle \) is minimal for WAA.
gives ambiguous results, even for equidivergent phenetic matrices. Consider, for example, the simple, equidivergent phenetic matrix illustrated in Figure 9.4. In the first divisive step, the only member of the initial partition \{V U W U Y U Z\} can be partitioned in any of seven ways (Figure 9.4 C), each of which has exactly the same superphenetic value. Clearly, there is no way to decide which step to take next, since all possible first steps are equally "correct" under the criterion of the MXD algorithm. Thus the MXD algorithm is ambiguous, and satisfies none of the three divergence hypotheses.

As we showed in Theorem 9.3, a minimal divisive (MND) algorithm is reliable under the hypothesis of strict divergence (and hence under equidivergence). In Figure 9.5, we see how a divergent-in-mean (but not strictly divergent) phenetic matrix leads to a different phenogram in MND as compared to our UAA prototype. In the first step of the MND phenogram algorithm, \langle U U V U W U Z, Y \rangle takes on a maximum value (the criterion for "farthest" pair in a divisive algorithm) over all possible superphenetic entries which results in a final phenogram different from the prototype.

As we showed in Theorem 9.4, an unweighted average divisive (UAD) algorithm is suitable under the hypothesis of equidivergence. In Figure 9.6, we see how a strictly divergent (but not equidivergent) phenetic matrix leads to a different phenogram
(A) \(\begin{array}{cccc}
V & W & Y & Z \\
\end{array}\)

(B)\[
\begin{array}{cccc}
V & W & Y & Z \\
0 & 1 & 11 & 11 \\
1 & 0 & 11 & 11 \\
11 & 11 & 0 & 9 \\
11 & 11 & 9 & 0 \\
\end{array}
\]

(C)\[
\begin{align*}
\langle V, W U Y U Z \rangle &= \max(1,11,11) = 11 \\
\langle W, V U Y U Z \rangle &= \max(1,11,11) = 11 \\
\langle Y, V U W U Z \rangle &= \max(11,11,9) = 11 \\
\langle Z, V U W U Y \rangle &= \max(11,11,9) = 11 \\
\langle V U W, Y U Z \rangle &= \max(11,11,11,11) = 11 \\
\langle V U Y, W U Z \rangle &= \max(1,11,11,9) = 11 \\
\langle V U Z, W U Y \rangle &= \max(1,11,11,9) = 11 \\
\end{align*}
\]

Figure 9.4. Maximal divisive: counterexample to equidivergence. (A) Prototype cladogram. (B) Equidivergent phenetic matrix. (C) All possible "first steps" in the divisive iteration structure give the same super-phenetic value. Therefore, the MND phenogram is ambiguous.
Figure 9.5. Minimal divisive: counterexample to divergence-in-mean. (A) Prototype cladogram. (B) Phenetic matrix. (C) MND phenogram. (D) Critical super-phenetic calculations. \( \langle U U V U W U Y, Z \rangle \) is greater than \( \langle U U V U W U Y, Z \rangle \) at the first divisive step, although the latter corresponds to the correct division.
Figure 9.6. Unweighted average divisive: counterexample to strict divergence. (A) Prototype cladogram. (B) Phenetic matrix. (C) UAD phenogram. (D) Critical superphenetic calculations. \( \langle U U V U W, Y U Z \rangle \) is greater than \( \langle U U V U W, Y U Z \rangle \) at the first divisive step, although the latter corresponds to the correct division.
in the UAD as compared to our UAA prototype. In the first step of the UAD phenogram algorithm, \( (U \cup V \cup W \cup Y, Z) \) takes on a maximum value over all possible superphenetic entries which result in a final phenogram different from the prototype.

9.4 Conclusions and Recommendations

The most important contribution of this dissertation is its development of a unified formal structure for pedigrees, isolation, and cladograms. The intuitive structure of these concepts is well-known to biologists (Dobzhansky, 1937; Mayr, 1942; Mayr, 1966; Hennig, 1966) and some formal development of hierarchies (Gregg, 1967; Woodger, 1937; Jardine and Sibson, 1967; Buck and Hull, 1966; Estabrook, 1966) and cladograms (Wirth et al., 1966; Estabrook, 1968; Hendrickson, 1968, Farris, 1970) has appeared recently in the literature. However, this dissertation is, to my knowledge, the first attempt to unify these diverse concepts in a single, formal framework.

The most incomplete aspects of this dissertation are its developments of immunological methods and divergent evolution. In my development of immunological methods, I employed concepts such as antigens, antisera, complement, etc., only as intuitive concepts to provide motivation for the properties subsumed under Axioms III*, IV*, and V*. Whereas these assumed properties are sufficient for the purposes of this dissertation, one would hope that they could
be derived from first principles of immunology, which are recognizable to immunologists as such. In future investigations, the immunological system might well be formulated in terms of set theory, as my own investigations with Professor Goodman (Moore and Goodman, 1968) and the independent researches of Kirsch (1967; 1969) suggest.

All the cladogram algorithms developed in this dissertation are proved under various hypotheses of divergence, as defined in chapter seven. Divergence hypotheses were stated with a short, intuitive buildup, but without any derivation from first principles. It would be interesting to see what manner of first principles generate a divergence hypothesis. Alternatively, it would be interesting to see what kinds of first principles give rise to a parsimony hypothesis, and how the first principles of divergence and parsimony might differ.

The parsimony hypothesis is defective in the context of this dissertation, because some phenetic matrices consistent with parsimony give ambiguous solutions. Parsimony is one of the leading concepts in numerical taxonomy today, and despite its shortcomings, it must eventually find its way into numerical taxonomy axiomatics. Since parsimony guarantees neither unique nor a class of "close" solutions, it may be that some compromise between the divergence and parsimony concepts will emerge as a paradigm of cladogram construction from phenetic matrices.
If the parsimony concept could be restricted so as to produce a unique solution, (see, for example, (Farris, 1970)), and if the requisite evolutionary hypothesis turns out to be different from the divergence hypotheses developed in this dissertation, it should nonetheless be possible to move back and forth between these two concepts by way of an appropriate transformation. One envisions a transformation which could transform a parsimony phenetic matrix into a corresponding divergence phenetic matrix, and vice versa. We already have the potential for doing this sort of thing on a limited scale: that is, transforming a phenetic matrix subject only to divergence-in-mean into a new phenetic matrix subject to equidivergence. At worst, one could apply the unweighted average agglomerative algorithm to the initial divergent-in-mean phenetic matrix, work out the correct cladogram, and design a new phenetic matrix which is equidivergent with respect to the now-known cladogram. The so-transformed matrix can now be used on any of six algorithms (instead of only one, as before).

Clearly, the proposed transformation defeats its purpose (once you have the true cladogram, why bother working out a transformed equidivergent phenetic matrix?), but the mere fact that such a transformation must exist may entice another biomathematician to seek a computationally efficient transformation.
Alternatively, some future biomathematician may wish to propose yet additional evolutionary hypotheses and discover which algorithms can be proved in the basic framework of this dissertation. I have purposely structured my axiom system so that the most solid biological concepts are deeply imbedded in the axiomatic foundations, whereas the more dubious statements are developed late in the dissertation, where they can be altered without major revisions in the entire dissertation. It should thus be possible for some later investigator to restructure the later chapters of this dissertation to meet current biological thought, while still being able to use intact the formalization developed in early chapters.
LIST OF REFERENCES


11. APPENDICES

11.1 Primitive and Defined Terms

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler-Venn diagram</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>curly bracket notation</td>
<td>{ }</td>
<td>18</td>
</tr>
<tr>
<td>definition by abstraction</td>
<td>{;}</td>
<td>18</td>
</tr>
<tr>
<td>is a member of</td>
<td>∈</td>
<td>20</td>
</tr>
<tr>
<td>is not a member of</td>
<td>⊈</td>
<td>20</td>
</tr>
<tr>
<td>equals</td>
<td>=</td>
<td>20</td>
</tr>
<tr>
<td>is a subset of</td>
<td>⊆</td>
<td>20</td>
</tr>
<tr>
<td>is a proper subset of</td>
<td>⊂</td>
<td>22</td>
</tr>
<tr>
<td>cardinality</td>
<td>N( )</td>
<td>22</td>
</tr>
<tr>
<td>null set; empty set</td>
<td>{ }, φ</td>
<td>23</td>
</tr>
<tr>
<td>union</td>
<td>U</td>
<td>25</td>
</tr>
<tr>
<td>intersection</td>
<td>∩</td>
<td>25</td>
</tr>
<tr>
<td>difference</td>
<td>-</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 11.1. Set theory terms introduced in chapter two.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Definition No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>set of all organisms</td>
<td>B</td>
<td>(primitive)</td>
<td>31</td>
</tr>
<tr>
<td>is a parent of</td>
<td></td>
<td>(primitive)</td>
<td>31</td>
</tr>
<tr>
<td>is an ancestor of</td>
<td></td>
<td>3.1</td>
<td>33</td>
</tr>
<tr>
<td>is ancestor or equal to</td>
<td></td>
<td>3.2</td>
<td>35</td>
</tr>
<tr>
<td>pedigree</td>
<td>Pd</td>
<td>3.3</td>
<td>36</td>
</tr>
<tr>
<td>common pedigree</td>
<td>CP</td>
<td>3.4</td>
<td>36</td>
</tr>
<tr>
<td>most recent common pedigree</td>
<td>RP</td>
<td>3.5</td>
<td>42</td>
</tr>
<tr>
<td>monoparental</td>
<td></td>
<td>3.6</td>
<td>50</td>
</tr>
<tr>
<td>is a patriarch of</td>
<td>o</td>
<td>3.7</td>
<td>53</td>
</tr>
</tbody>
</table>

Table 11.2. Primitive and defined terms introduced in chapter three.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Definition No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>partition (of X)</td>
<td>$\Pi_\chi$</td>
<td>4.1</td>
<td>61</td>
</tr>
<tr>
<td>binary partition</td>
<td></td>
<td>4.2</td>
<td>63</td>
</tr>
<tr>
<td>superpartition</td>
<td>$\Sigma$</td>
<td>4.3</td>
<td>66</td>
</tr>
<tr>
<td>is isolated from</td>
<td></td>
<td>4.4</td>
<td>70</td>
</tr>
<tr>
<td>cladistic partition</td>
<td></td>
<td>4.5</td>
<td>75</td>
</tr>
<tr>
<td>binary cladistic partition</td>
<td></td>
<td>4.6</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 11.3. Defined terms introduced in chapter four.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Definition No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>is an immediate predecessor of</td>
<td>( D )</td>
<td>5.1</td>
<td>80</td>
</tr>
<tr>
<td>is a predecessor of</td>
<td>( D )</td>
<td>5.2</td>
<td>81</td>
</tr>
<tr>
<td>dendrogram</td>
<td>((\Delta, D))</td>
<td>5.3</td>
<td>83</td>
</tr>
<tr>
<td>is predecessor or equal to</td>
<td>( D )</td>
<td>5.4</td>
<td>94</td>
</tr>
<tr>
<td>cladogram</td>
<td>((K, D))</td>
<td>5.5</td>
<td>97</td>
</tr>
</tbody>
</table>

Table 11.4. Defined terms introduced in chapter five.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Definition No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>phenetic matrix</td>
<td>⟨,⟩</td>
<td>6.1</td>
<td>107</td>
</tr>
<tr>
<td>comparison matrix</td>
<td>H</td>
<td>(primitive)</td>
<td>124</td>
</tr>
<tr>
<td>protocol matrix</td>
<td>E</td>
<td>(primitive)</td>
<td>125</td>
</tr>
<tr>
<td>primary solution matrix</td>
<td>S</td>
<td>6.2*</td>
<td>128</td>
</tr>
<tr>
<td>D-vector</td>
<td>D</td>
<td>6.3*</td>
<td>129</td>
</tr>
<tr>
<td>C-vector</td>
<td>C</td>
<td>6.4*</td>
<td>129</td>
</tr>
<tr>
<td>is directly connected to</td>
<td></td>
<td>6.5*</td>
<td>134</td>
</tr>
<tr>
<td>is connected to</td>
<td></td>
<td>6.6*</td>
<td>136</td>
</tr>
</tbody>
</table>

Table 11.5. Primitive and defined terms introduced in chapter six.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Definition No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>equidivergent</td>
<td></td>
<td>7.1</td>
<td>152</td>
</tr>
<tr>
<td>strictly divergent</td>
<td></td>
<td>7.2</td>
<td>153</td>
</tr>
<tr>
<td>superphenetic matrix</td>
<td>⟨,⟩</td>
<td>7.3</td>
<td>156</td>
</tr>
<tr>
<td>minimal superphenetic matrix</td>
<td></td>
<td>7.4</td>
<td>158</td>
</tr>
<tr>
<td>maximal superphenetic matrix</td>
<td></td>
<td>7.5</td>
<td>160</td>
</tr>
<tr>
<td>unweighted average superphenetic matrix</td>
<td></td>
<td>7.6</td>
<td>160</td>
</tr>
<tr>
<td>nesting level of -- in -- nest(,)</td>
<td></td>
<td>7.7</td>
<td>165</td>
</tr>
<tr>
<td>weighted average superphenetic matrix</td>
<td></td>
<td>7.8</td>
<td>170</td>
</tr>
<tr>
<td>divergent-in-mean</td>
<td></td>
<td>7.9</td>
<td>196</td>
</tr>
</tbody>
</table>

Table 11.6. Defined terms introduced in chapter seven.
<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol, if any</th>
<th>Definition No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>is a l-refinement of</td>
<td></td>
<td>8.1</td>
<td>211</td>
</tr>
<tr>
<td>agglomerative binary phenogram algorithm</td>
<td></td>
<td>8.2</td>
<td>212</td>
</tr>
<tr>
<td>divisive binary phenogram algorithm</td>
<td></td>
<td>8.3</td>
<td>213</td>
</tr>
<tr>
<td>agglomerative cladogram function</td>
<td></td>
<td>8.4</td>
<td>223</td>
</tr>
<tr>
<td>divisive cladogram function</td>
<td></td>
<td>8.5</td>
<td>223</td>
</tr>
</tbody>
</table>

Table 11.7. Defined terms introduced in chapter eight.
11.2 Flow of Reasoning in this Dissertation

Figure 11.1. Flow of reasoning in chapters three and four.
Figure 11.2. Flow of reasoning in chapter five.
Figure 11.3. Flow of reasoning in chapter seven.
Figure 11.4. Flow of reasoning in chapters eight and nine.
11.3 Flow Chart for Unweighted Average Agglomerative Algorithm

START

Initialize:
Let \( W(K) = 1 \) for \( 1 \leq K \leq N \)

Find the smallest element of \( P \),
i.e., \( P(I,J) \) where \( I \neq J \) and
\( W(I) > 0 \) and \( W(J) > 0 \)

Print \( I,J \)

Set \( P(I,K) \) and \( P(K,I) \) equal to
\( (P(I,K)*W(I) + P(J,K)*W(J))/(W(I) + W(J)) \)
for \( 1 \leq K \leq N \)

Set \( P(J,K) \) and \( O(K,J) \) equal to 0
for \( 1 \leq K \leq N \)

Set \( W(I) \) equal to \( W(I) + W(J) \)

Set \( W(J) \) equal to 0

If there are more than two values
for \( K \) at which \( W(K) > 0 \), then go to SEARCH

STOP

Figure 11.5. Flow chart for unweighted average agglomerative algorithm. Problem: Given \( N \) OTUs numbered 1 through \( N \) and an \( N \) by \( N \) phenetic matrix, \( P \), find the unweighted average agglomerative phenogram.