INTRODUCTION TO PHYLOGENETICS:
A STUDY OF MAXIMUM PARSIMONY AND MAXIMUM LIKELIHOOD METHODS

By

NAOMI R. IUHASZ

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2010
To my sister, Andreea Iuhasz
ACKNOWLEDGMENTS

I take this opportunity to thank my supervisor, Dr. Sergei Pilyugin, for his expertise, kindness, and most of all, for his patience. I thank him also for many insightful conversations during the development of the ideas in this thesis, for the questions that challenged my understanding of the material, and for helpful comments on the text. I would like to express sincere gratitude to my thesis co-advisor Dr. Edward Braun, for taking me under his wing long before I knew anything about bio-mathematics and for his constant guidance throughout the journey that lead to this thesis. I also thank Dr. Rebecca Kimball and the entire Braun-Kimball lab for allowing an outsider to enter their niche and to learn from their expertise. I thank Dr. Murali Rao for his valuable suggestions. I am greatly indebted to Prof. Billy Gunnells for challenging me to pursue my interests in biology. My thanks and gratitude goes to Prof. Claire Kurtgis-Hunter for recognizing my potential and for being the first to encourage me to pursue a graduate degree.

Above all, I thank my family who stood beside me and encouraged me constantly. My thanks to my grandparents for their love and sound advice, to my aunt for taking on the responsibility to see me through college, and especially to my father who is my role model. Special thanks and appreciation to Camilo for his constant support and unending encouragement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>6</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>7</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>8</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>9</td>
</tr>
<tr>
<td>2 PRELIMINARIES</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Definition of Terms</td>
<td>12</td>
</tr>
<tr>
<td>2.2 X-trees</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Tree Shapes</td>
<td>15</td>
</tr>
<tr>
<td>2.4 X-splits</td>
<td>18</td>
</tr>
<tr>
<td>2.5 Characters and Convexity</td>
<td>23</td>
</tr>
<tr>
<td>2.6 Character Compatibility</td>
<td>27</td>
</tr>
<tr>
<td>3 MAXIMUM PARSIMONY</td>
<td>34</td>
</tr>
<tr>
<td>3.1 Classical Parsimony</td>
<td>34</td>
</tr>
<tr>
<td>3.2 Optimization on a Fixed Tree</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Tree Rearrangement Operations</td>
<td>46</td>
</tr>
<tr>
<td>3.4 Relevance</td>
<td>49</td>
</tr>
<tr>
<td>4 MAXIMUM LIKELIHOOD</td>
<td>51</td>
</tr>
<tr>
<td>4.1 Basic Principles</td>
<td>52</td>
</tr>
<tr>
<td>4.2 Models of Sequence Evolution</td>
<td>54</td>
</tr>
<tr>
<td>4.3 Calculating Change Probabilities</td>
<td>58</td>
</tr>
<tr>
<td>4.4 Differences in Perspective between Parsimony and Likelihood</td>
<td>60</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>62</td>
</tr>
<tr>
<td>BIOGRAPHICAL SKETCH</td>
<td>65</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1 Characters $\chi_1$, $\chi_2$, and $\chi_3$</td>
<td>32</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>(a) An X-tree. (b) A binary phylogenetic X-tree.</td>
<td>13</td>
</tr>
<tr>
<td>2-2</td>
<td>The two tree shapes for B(6).</td>
<td>16</td>
</tr>
<tr>
<td>2-3</td>
<td>A rooted tree shape.</td>
<td>17</td>
</tr>
<tr>
<td>2-4</td>
<td>Edges $e_1$ and $e_2$ induce ${1, 2, 3, 4}</td>
<td>{5, 6, 7, 8, 9}$ and ${1, 2, 3, 4, 5, 6, 7}</td>
</tr>
<tr>
<td>2-5</td>
<td>One iteration of tree popping.</td>
<td>23</td>
</tr>
<tr>
<td>2-6</td>
<td>An X-tree and a mapping $\chi_t$.</td>
<td>24</td>
</tr>
<tr>
<td>2-7</td>
<td>Vertex disjoint subtrees.</td>
<td>26</td>
</tr>
<tr>
<td>2-8</td>
<td>Compatibility of characters. (a) A restricted chordal completion of $\text{int}(C)$. (b) A maximal clique tree representation of G. (c) An X-tree on which each character in C is convex.</td>
<td>33</td>
</tr>
<tr>
<td>3-1</td>
<td>(a) The X-tree. (b) A minimum extension of a character. (c) Edges forming a cut-set.</td>
<td>37</td>
</tr>
<tr>
<td>3-2</td>
<td>A maximum-sized set of edge-disjoint proper paths.</td>
<td>39</td>
</tr>
<tr>
<td>3-3</td>
<td>An Erdős -Székely path system.</td>
<td>40</td>
</tr>
<tr>
<td>3-4</td>
<td>Deficiencies of forward pass reconstruction.</td>
<td>43</td>
</tr>
<tr>
<td>3-5</td>
<td>(a) The X-tree. (b) Forward pass. (c) Backward pass.</td>
<td>45</td>
</tr>
<tr>
<td>3-6</td>
<td>A schematic representation on the generic TBR operation.</td>
<td>46</td>
</tr>
<tr>
<td>3-7</td>
<td>A schematic representation on the generic SPR operation.</td>
<td>47</td>
</tr>
<tr>
<td>3-8</td>
<td>A schematic representation on the generic NNI operation.</td>
<td>47</td>
</tr>
<tr>
<td>4-1</td>
<td>Overview of the calculation of the likelihood of a tree. (a) Hypothetical sequence alignment. (b) An unrooted tree for the four taxa in (a). (c) Tree after rooting at arbitrary interior vertex, in this case (6). (d) Likelihood of character $\chi$.</td>
<td>54</td>
</tr>
</tbody>
</table>
Abstract of Thesis Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Master of Science

INTRODUCTION TO PHYLOGENETICS: A STUDY OF MAXIMUM PARSIMONY AND MAXIMUM LIKELIHOOD METHODS

By

Naomi R. Iuhasz

May 2010

Chair: Sergei S. Pilyugin
Major: Mathematics

In this thesis we investigate the conceptual framework of phylogenetics together with two of the most popular methods of inferring phylogenetic relationships. Some notation and background concepts of graph theory are introduced. The general class of X-trees is presented together with its properties, shapes of trees, and X-splits. The functions of characters applied to trees make the connection between mathematics and biology. The notions of character convexity and compatibility are formalized. We also introduce and analyze the maximum parsimony and maximum likelihood methods of creating phylogenies, with comparison and contrast between the methods.
CHAPTER 1
INTRODUCTION

The subject of this thesis was motivated by the undergraduate research conducted by this author in the field of computational biology. The premise of the undergraduate research was to re-evaluate the general assumption that the rate of molecular evolution and the rate of morphological evolution are in effect dissociated from each other. If there is a correlation between the two rates, there would be a universal rate of evolution for organisms and the expected number of changes (either morphological or molecular) during a certain period of evolutionary history would be proportional to that universal rate. Even with a universal rate, some variation in the number of changes is expected. This variance would reflect a Poisson process in the ideal case. The research tested whether the negative binomial method is an improved alternative over the Poisson method to accommodate the variation in the number of changes. Also, it tested whether there is a need to assume more variance to fit morphological change to the universal rate than to fit molecular change. Both negative binomial and Poisson methods were applied to phylogenetic trees that were computed using one of the widely known software packages for inferring phylogenetic trees, PAUP*. This thesis is a continuation of the undergraduate research in that it investigates the mathematical concepts on which phylogenetic trees are based on and how they are inferred from the available data.

Phylogenetics is the biological discipline which studies the evolutionary relatedness among different organisms based on molecular and morphological information. This relatively new branch of biology has its roots in Charles Darwin’s theory of evolution. The quest to use present-day characteristics of a group of species to infer the historical relationships between them and their evolution from a common ancestor has been the subject of numerous studies. These relationships are consistently represented by an evolutionary (phylogenetic) tree, structure first proposed by Darwin himself. Initially, the relationships were drawn after studying the morphological characteristics of the species.
However, such criterion of comparison has its limitations due to simplistic assumptions about evolutionary processes and difficulty of comparison between very distantly related species or morphologically identical yet different species. The field of phylogenetics flourished with the discovery and study of molecular data which began in the late 1960s. Protein and genetic sequences provide an immensely richer pool of information which can be explored through a steadily growing number of methods and techniques. [32]

The field of phylogenetics developed with a strong interdisciplinary foundation as it incorporates elements of mathematics, statistics, and computer science with biology. Here we will look at the mathematical aspect of the field. The reconstruction and analysis of phylogenetic trees involves almost exclusively discrete mathematics, mainly graph theory and probability theory. [32]

Inferring a phylogenetic tree is an estimation procedure since the true tree is essentially unknowable. The estimation models employed calculate a "best estimate" of an evolutionary history based on the incomplete information contained in the data. Phylogenetic inference methods seek to accomplish this goal in one of two ways: by defining a specific algorithm that leads to the determination of a tree or by defining a criterion for comparing alternative phylogenies to one another and deciding which is better. While in a purely algorithmic method the algorithm defines the tree selection criterion, in the criterion-based methods the algorithms are merely tools used to evaluate and compare trees. Purely algorithmic methods tend to be computationally fast because they proceed directly towards the final solution without evaluating large numbers of competing trees. These methods include all forms of pair-group cluster analysis and some other distance methods such as neighbor joining, not discussed in this thesis. The second class of methods first defines an optimality criterion for evaluating a given tree and then uses specific algorithms for computing the value of the objective function and for finding the trees that have the best value according to the criterion. The price of this logical clarity is that the criterion-based methods tend
to be much slower than those from the first class. However, since criterion methods can assign a score to every tree examined, phylogenies can be ranked in order of preference. The two main criterion-based models, the maximum parsimony and the maximum likelihood models will be examined in this thesis.[34]
CHAPTER 2
PRELIMINARIES

2.1 Definition of Terms

Various diagrams used to illustrate evolutionary relationships among organisms resemble the structure of a tree; therefore, graphs are an elegant way to portray and study these relationships. A graph $G$ is an ordered pair $(V, E)$ consisting of a non-empty set $V$ of vertices and a multiset $E$ of edges each of which is an element of $\{\{x, y\} : x, y \in V\}$. A graph $H$ is a subgraph of a graph $G$ if $V(H)$ and $E(H)$ are subsets of $V(G)$ and $E(G)$ respectively. If $V'$ is a non-empty subset of $V(G)$, then the subgraph of $G$ that has vertex set $V'$ and the edge set consisting of those edges of $G$ that have both ends in $V'$ is the subgraph of $G$ induced by $V'$, and is denoted by $G[V']$.

Since graphs and trees in particular are used extensively in several disciplines, such as mathematics, biology, and computer science, there are several names attributed to each component. For this reason, we will present the most common definitions but use only one name for each component consistently throughout this thesis. Most graphs discussed in this thesis correspond to unrooted trees, also known as unrooted phylogeny. In these structures, the location of the common ancestor is not identified. Vertices are also called nodes or points. The terminal nodes, also called leaves or external nodes, correspond to the contemporary taxa or species being assessed. The branch points within the interior of the tree, corresponding to past, intermediary species are called internal nodes or interior vertices. Hereafter, we will employ the terms vertices, leaves, and interior vertices respectively. The branches connecting pairs of nodes are also called edges, links, or segments. Branches incident to a leaf are called exterior or peripheral branches and those connecting two interior vertices care interior edges or interior branches. In this thesis we will use edges, interior or exterior. The degree of a vertex $v$, denoted by $d(v)$, is the number of edges that are incident with $v$. Traditionally, a tree is defined as a connected graph with no cycles. The phylogenetic
research works extensively with binary trees, that are the trees in which every interior vertex has degree three. Unless mentioned otherwise, all graphs are connected and all trees are binary.

2.2 X-trees

From a biological standpoint, a ‘phylogenetic tree’ represents the standard graphical depiction of evolutionary relationships. However, we need to define a more general class of objects to thoroughly investigate the mathematical processes involved. For this purpose, we will introduce the concept of an ‘X-tree’ as described in [32].

**Definition 2.2.1.** An X-tree \( T \) is an ordered pair \( (T; \phi) \), where \( T \) is a tree with vertex set \( V \) and \( \phi : X \to V \) is a map with the property that, for each \( v \in V \) of degree at most two, \( v \in \phi(X) \). An X-tree is also called a semi-labeled tree (on \( X \)).

**Definition 2.2.2.** A phylogenetic (X-) tree \( T \) is an X-tree \( (T; \phi) \) with the property that \( \phi \) is a bijection from \( X \) into the set of leaves of \( T \). If in addition, every interior vertex of \( T \) has degree three, \( T \) is a binary phylogenetic (X-) tree.

![Figure 2-1](image)

Figure 2-1. (a) An X-tree. (b) A binary phylogenetic X-tree.

Figure 2-1 (a) is an example of an X-tree, and Figure 2-1 (b) is its binary phylogenetic X-tree equivalent. For a phylogenetic X-tree \( T = (T; \phi) \), \( X \) can be viewed as the set of leaves of the tree \( T \). Next, we present two important properties of binary phylogenetic trees.
Proposition 2.2.3. Let $T$ be a binary phylogenetic X-tree and let $n = |X|$. Then, for all $n \geq 2$, $T$ has $2n - 3$ edges and $n - 3$ interior edges, and $2n - 2$ vertices and $n - 2$ interior vertices.

Proof. Use induction on $n$. For $n = 2$, there are 2 vertices, no interior vertices, 1 edge, and no interior edges. Result holds.

Suppose result holds for some $n$. Then we have $2n - 3$ edges, $n - 3$ interior edges, $2n - 2$ vertices, and $n - 2$ interior vertices. Now check for $n + 1$: We add a leaf to the tree. This means we need to add an interior vertex also. So the new tree has $2n$ vertices and $n - 1$ interior vertices. One edge is destroyed and three edges are created to link the leaf and vertex to the tree. If $n$ is odd, the edge destroyed is not interior, but it is replaced by an interior edge. If $n$ is even, the edge destroyed is interior and it is replaced by two interior edges. Hence, we have $2n - 3 - 1 + 3 = 2(n + 1) - 3$ edges and $n - 3 + 1 = (n + 1) - 3$ interior edges. Hence, the claim holds for $n + 1$. ∎

Proposition 2.2.4. Let $B(n)$ denote the collection of all binary phylogenetic trees with label set $\{1, 2, \ldots, n\}$ and let $b(n) = |B(n)|$. If $n \in \{1, 2\}$, then $b(n) = 1$. For all $n \geq 3$,

$$b(n) = 1 \times 3 \times 5 \times \ldots \times (2n - 5) = \frac{(2n - 4)!}{(n - 2)!2^{n-2}}.$$ 

Proof. Use induction on $n$. For $n = 3$, we get $b(3) = 1$, so the result holds. Let

$$a(n) = \frac{(2n - 4)!}{(n - 2)!2^{n-2}}, \quad a(3) = \frac{2!}{1! \cdot 2} = 1 = b(3).$$

Suppose the result holds for $n - 1, n \geq 4$. Let $\psi : B(n) \to B(n - 1)$ such that $\psi(T)$ is the binary phylogenetic tree in $B(n - 1)$ that is obtained from $T \in B(n)$ by deleting the leaf labeled $n$ and its incident edge, and then suppressing the resulting degree-two vertex. By construction, $\psi$ is onto. We obtain a binary phylogenetic tree in $B(n)$ by connecting a new leaf to an edge of a tree in $B(n - 1)$. From Proposition 2.2.3, a binary phylogenetic tree in $B(n - 1)$ has $2(n - 5)$ edges, so the leaf can be added in $2(n - 5)$ ways. Hence, each binary phylogenetic tree in $B(n - 1)$ in the range of $\psi$ is the
image of $2n - 5$ trees in $B(n)$. We know $b(n - 1) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n - 7)$. Therefore, $b(n) = b(n - 1) \cdot (2n - 5) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n - 7)(2n - 5)$. Also,

$$a(n) = a(n - 1) \cdot \frac{(2n - 5)(2n - 4)}{(n - 2)^2} = a(n - 1) \cdot (2n - 5) = b(n - 1) \cdot (2n - 5) = b(n).$$

The number of all possible phylogenetic trees with a given label set is important since we often need to search for the optimal tree. However, an exhaustive search is practically impossible when $n$ is large. For this reason, new methods must be employed to find the "best" tree without searching through all possibilities. These methods utilize various optimality criteria to compare and rate alternative trees. We will discuss them in detail in the next chapters.

### 2.3 Tree Shapes

A tree shape is a phylogenetic tree in which we ignore the labels on the leaves. It is useful to be able to determine the number of possible phylogenetic trees given by $n$-leaved trees that share the same shape.

**Definition 2.3.1.** *Two phylogenetic trees $\mathcal{T}_1$ and $\mathcal{T}_2$ with equal label sets are shape equivalent if $T(\mathcal{T}_1)$ is isomorphic to $T(\mathcal{T}_2)$. [32]*

Unrooted binary phylogenetic trees with $n \in \{2, 3, 4, 5\}$ have exactly one shape. Starting with $n = 6$, however, we encounter multiple tree shapes. For example, Figure 2-2 shows the two shapes of the collection of binary phylogenetic trees with 6 leaves ($B(6)$). Determining the number of tree shapes for phylogenetic trees with $n$ leaves is not trivial and was not explored since it is not relevant to the topic of this thesis. For details, see [7] for unrooted binary phylogenetic trees and [16] for rooted binary trees.

Concerning tree shapes, we have examined how to count the number of phylogenetic trees on a given label set with a specific tree shape $\tau$. For this purpose, first we need to introduce some elements of group actions on sets.
Figure 2-2. The two tree shapes for B(6).

**Definition 2.3.2.** An action of a group $G$ on a set $M$ is a map $M \times G \to M$ such that, for all $m \in M$ and for all $g_1, g_2 \in G$,

(a) $(m, 1_G) = m$ and

(b) $(m, g_1 g_2) = ((m, g_1), g_2)$.

We define the relation $m_1 \sim m_2$ if there is an element $g \in G$ such that $(m_1, g) = m_2$. It is easily seen that $\sim$ is an equivalence relation on $M$. We denote the equivalence class of $m$ under this relation by $\overline{m}$.

**Lemma 2.3.3** (Burnside’s Lemma). Let $G$ be a finite group acting on a finite set $M$ and let $m \in M$. Then,

$$|\overline{m}| = \frac{|G|}{|G_m|},$$

where $G_m = \{g \in G : (m, g) = m\}$ is a subgroup of $G$. [32]

$m$ is also known as the orbit of $m$ and $G_m$ as the centralizer of $m$. In our context, $M$ is the collection of all phylogenetic trees with the label set $\{1, 2, \ldots, n\}$ and $G$ is the (symmetric) group $S_n$ of all $n!$ permutations of $\{1, 2, \ldots, n\}$. Let $g \in G$ and $T \in M$. The action of $g$ on $T$ maps $T$ to the phylogenetic tree obtained from $T$ by permuting the label set according to $g$. So, if $T$ has tree shape $\tau$, then the number of phylogenetic trees having tree shape $\tau$ is

$$|\overline{T}| = \frac{n!}{|G_\tau|}.$$
where $\mathcal{G}_T$ is the collection of permutations of $\{1, 2, \ldots, n\}$ that leaves $T$ unchanged.

[19] provides the following formulae for determining $\mathcal{G}_T$.

**Proposition 2.3.4.** Let $T$ be a rooted phylogenetic tree. For each interior vertex $\nu$ of $T$, let $D(\nu)$ denote the collection of maximal rooted phylogenetic subtrees that lie below $\nu$. Now, let us impose the rooted shape equivalence relation on $D(\nu)$ and let $n_1(\nu), n_2(\nu), \ldots$ denote the sizes of the resulting equivalence classes. Then,

$$|\mathcal{G}_T| = \prod_{\nu=\tilde{V}(T)} \prod n_i(\nu)!,$$

where $\tilde{V}(T)$ is the set of interior vertices of $T$.

![Figure 2-3. A rooted tree shape.](image)

To illustrate this proposition, consider a rooted phylogenetic tree $T$ having the shape shown in Figure 2-3. Applying Proposition 2.3.4, we get

$$|\mathcal{G}_T| = (1!)^2 \times (3!) \times (2!) \times (2!) = 48.$$  

(The interior vertices have been numbered to show the order in which they were considered.) The formula for $|\mathcal{G}_T|$ is simpler for a rooted binary phylogenetic tree. Let $s(\tau)$ denote the number of interior vertices $\nu$ of a rooted binary tree with shape $\tau$, if the two maximal rooted subtrees that lie below $\nu$ have the same shape.
Corollary 2.3.5. For a rooted binary phylogenetic tree $T$ of shape $\tau$,

$$|G_T| = 2^{s(\tau)}.$$

Thus, the number of rooted binary phylogenetic trees having shape $\tau$ is $n!2^{-s(\tau)}$.

The corresponding formula for $|G_T|$ is slightly more complicated because an unrooted tree can have an additional symmetry when two adjacent vertices are interchanged. An unrooted phylogenetic tree can have an edge for which the two rooted subtrees obtained by deleting this edge have the same shape. This edge is called a central edge, and a phylogenetic tree can have at most one such edge. Let $c(T) = 1$ if $T$ has a central edge and 0 if it does not. Then, for an unrooted phylogenetic tree $T$,

$$|G_T| = 2^{c(T)} \prod_{v=\hat{V}(T)} \prod n_i(v)!.$$

### 2.4 X-splits

The concepts in this section have played an important role in the mathematical development of phylogenetics. [32]

**Definition 2.4.1.** An X-split is a partition of $X$ into two non-empty sets. We denote the X-split whose blocks are $A$ and $B$ by $A\mid B$.

Since we label the two components $A$ and $B$ arbitrarily, the X-split $B\mid A$ is equivalent to $A\mid B$. Next, we define the entire collection of X-splits associated with every X-tree. Let $T = (T; \phi)$ be an X-tree and let $e$ be an edge of $T$. Then $T \setminus e$, the tree obtained from $T$ by deleting $e$, is composed of two components that we will name $V_1$ and $V_2$. Hence, $\phi^{-1}(V_1) | \phi^{-1}(V_2)$ is the X-split corresponding to $e$ in $T$. This X-split is unique to edge $e$. We denote by $\Sigma(T)$ the collection of X-splits that correspond to the edges of $T$, and we refer to it as the X-split of $T$ or induced by $T$ as in [32]. To better explain this concept, we employ the following example:
Consider the X-tree shown in Figure 2-4, where \( X = \{1, 2, \ldots, 9\} \). The X-splits corresponding to edges \( e_1 \) and \( e_2 \) are \( \{1, 2, 3, 4\}\{5, 6, 7, 8, 9\} \) and \( \{1, 2, 3, 4, 5, 6, 7\}\{8, 9\} \) respectively.

**Definition 2.4.2.** A pair of X_splits \( A_1 | B_1 \) and \( A_2 | B_2 \) are compatible if at least one of the sets \( A_1 \cap A_2, A_1 \cap B_2, B_1 \cap A_2, \) and \( B_1 \cap B_2 \) is the empty set.

This definition, given by [32], is justified by the Splits-Equivalence Theorem, first presented by [3]. It is the most important concept regarding X_splits. Its proof, also from [32], makes use of the following three Lemmas.

**Lemma 2.4.3.** Let \( T = (T; \phi) \) be an X-tree, and let \( \sigma_1 \) and \( \sigma_2 \) be distinct elements of \( \Sigma(T) \). Then \( X \) can be partitioned into three sets \( X_1, X_2, \) and \( X_3 \) such that \( \sigma_1 = X_1 | (X_2 \cap X_3) \) and \( \sigma_2 = (X_1 \cap X_2) | X_3 \). Furthermore, the intersection of the vertex sets of the minimal subtrees of \( T \) induced by \( T(\phi(X_1)) \) and \( T(\phi(X_2)) \) is empty.

**Proof.** Let \( e_1 = \{u_1, v_1\} \) and \( e_2 = \{u_2, v_2\} \) be the unique edges corresponding to \( \sigma_1 \) and \( \sigma_2 \) respectively. Obviously, there is a path \( P \) in \( T \) such that \( e_1 \) and \( e_2 \) are the first and last edges, respectively, that are traversed by \( P \). Without loss of generality, assume \( u_1 \) and \( u_2 \) are initial and terminal vertices of \( P \), respectively. Observe that \( u_1 \neq u_2 \), but \( v_1 \) and \( v_2 \) may not be distinct. Let \( V_1, V_2, \) and \( V_3 \) denote the vertex set of the components of \( T \setminus \{e_1, e_2\} \) containing \( u_1, v_1, \) and \( u_2 \) respectively. Choose \( X_i = \phi^{-1}(V_i) \) for each \( i = \{1, 2, 3\} \). Then \( \sigma_1 \) and \( \sigma_2 \) are distinct. \( \square \)

![Figure 2-4. Edges e₁ and e₂ induce \( \{1, 2, 3, 4\}\{5, 6, 7, 8, 9\} \) and \( \{1, 2, 3, 4, 5, 6, 7\}\{8, 9\} \) X-splits respectively.](image-url)
To illustrate Lemma 2.4.3, consider the X-tree shown in Figure 2-4. Let $\sigma_1$ and $\sigma_2$ be the X-splits corresponding to the edges $e_1$ and $e_2$, respectively (obviously distinct). Choosing $X_1 = \{1, 2, 3, 4\}$, $X_2 = \{5, 6, 7\}$, and $X_3 = \{8, 9\}$ provides a partition of $X$ into three sets. Moreover, $\sigma_1 = X_1 | (X_2 \cap X_3)$ and $\sigma_2 = (X_1 \cap X_2) | X_3$.

The next Lemma is a general property of trees. Let $T$ be a tree and let $f$ be a function from a finite set $Y$ into the vertex set $V$ of $T$. Color the elements of $Y$ either red or green. Next, assign a coloring to the elements of $V$ in $f(Y)$ in the following way. Let $v$ be an element of $f(Y)$. If all elements of $f^{-1}(v)$ are of the same color, assign that color to $v$ itself; otherwise, assign both red and green to $v$. [32] refers to this coloring as the coloring of $V$ induced by $f$. A subgraph of $T$ is monochromatic if all of its colored vertices are of one particular color.

**Lemma 2.4.4.** Let $T = (V, E)$ be a tree, and let $f$ be a mapping from a finite set $Y$ into $V$. Consider the coloring of $V$ induced by $f$. Suppose that, for each edge $e \in E$, exactly one of the components of $T \setminus e$ is monochromatic. Then, there exists a unique vertex $v \in V$ for which each component of $T \setminus v$ is monochromatic.

**Proof.** First, show there exists at least one such vertex. Let $e \in E$. Then, one component is monochromatic. Assign an orientation from the end of $e$ that is incident with the monochromatic component of $T \setminus e$ to the other end of $e$. Then, there exists $v \in V$ with out-degree zero; otherwise, we would have a directed path of infinite length. Deleting $v$ produces monochromatic components. Now, show there can be at most one such vertex $v$. Suppose for the sake of a contradiction that there is another vertex $v' \in V$ with the claimed property. Select an edge $e$ in the path connecting $v$ and $v'$. Then exactly one of the two components of $T \setminus e$ is not monochromatic. Without loss of generality, this component contains $v$. But this contradicts the assumption that each component of $T \setminus v'$ is monochromatic as the component containing $v$ is not monochromatic. $\square$
Lemma 2.4.5. Let $A|B$ be an $X$-split. Suppose that $T = (T; \phi)$ is an $X$-tree such that $A|B$ is not a split of $T$, but $A|B$ is compatible with each $X$-split of $T$. Then, there exists a unique vertex $v$ of $T$ such that for each component of $(V', E')$ of $T \setminus v$ either $\phi^{-1}(V') \subseteq A$ or $\phi^{-1}(V') \subseteq B$.

Proof. Color the elements of $A$ red and the elements of $B$ green, and consider the corresponding coloring of the vertices of $T$ induced by $\phi$. Then, for each edge $e$ of $T$, exactly one of the components of $T \setminus e$ is monochromatic under the coloring of the vertices of $T$ by $\phi$. Applying Lemma 2.4.4 with $f = \phi$ and $Y = X$, there exists a unique vertex $v$ of $T$ for which each component of $T \setminus v$ is monochromatic. Therefore, $A|B$ satisfies the condition described in the Lemma.

Theorem 2.4.6 (Splits-Equivalence Theorem). Let $\Sigma$ be a collection of $X$-splits. Then, there is an $X$-tree $T$ such that $\Sigma = \Sigma(T)$ if and only if the splits in $\Sigma$ are pairwise compatible. Moreover, if such a tree exists, then $T$ is unique up to isomorphism.

Proof. First, suppose $\Sigma = \Sigma(T)$. Let $\sigma_1$ and $\sigma_2$ be distinct elements of $\Sigma$. By Lemma 2.4.3, there is a partition of $X$ into three sets $X_1$, $X_2$ and $X_3$ such that $\sigma_1 = X_1|(X_2 \cup X_3)$ and $\sigma_2 = (X_1 \cup X_2)|X_3$. Since $X_1 \cap X_2 = \emptyset$, the $x$-splits $\sigma_1$ and $\sigma_2$ are compatible; therefore, the $X$-splits of $\Sigma$ are pairwise compatible.

Conversely, suppose that $\Sigma$ is a pairwise compatible collection of $X$-splits. We use induction on the cardinality of $\Sigma$ to prove that $\Sigma = \Sigma(T)$ for some $X$-tree $T$ and that the choice of $T$ is unique up to isomorphism. If $|\Sigma| = 0$, then the tree $T$ with a single vertex labeled $X$ is the unique tree for which $\Sigma = \Sigma(T)$.

Now suppose that $|\Sigma| = k + 1$, where $k \geq 0$, and that the existence and uniqueness properties hold for $|\Sigma| = k$. Let $A|B \in \Sigma$. Since $\Sigma - \{A|B\}$ is pairwise compatible, it follows by our induction assumption that there is, up to isomorphism, a unique $X$-tree $T' = (T', \phi')$ with $\Sigma - \{A|B\} = \Sigma(T)$. By Lemma 2.4.5, there is a unique vertex $v'$ of $T'$ such that, for each component $(V', E')$ of $T' \setminus v'$, either $\phi'^{-1}(V') \subseteq A$ or $\phi'^{-1}(V') \subseteq B$. 
Let $T$ be the tree obtained from $T'$ by replacing $v'$ with two new adjacent vertices $v_A$ and $v_B$, and attaching the subtrees that were incident with $v'$ to the new vertices in such a way that the subtree consisting of vertices in $\phi'(A)$ and $\phi'(B)$ are attached to $v_A$ and $v_B$ respectively. Let $\phi : X \to V(T)$ be the map defined as follows:

$$
\phi(x) = \begin{cases}
\phi'(x), & \text{if } \phi'(x) \neq v', \\
v_A, & \text{if } \phi'(x) = v' \text{ and } x \in A, \\
v_B, & \text{if } \phi'(x) = v' \text{ and } x \in B.
\end{cases}
$$

It is easily checked that $(T; \phi)$ is an X-tree, and that if we denote $T = (T; \phi)$, then we have $\Sigma = \Sigma(T)$. Moreover, since $T'$ is the unique X-tree for which $\Sigma - \{A|B\} = \Sigma(T')$, it is easily seen that $T$ is the only such X-tree satisfying $\Sigma = \Sigma(T)$ up to isomorphism. This completes the proof of the Splits-Equivalence Theorem.

One application of Lemma 2.4.5, first described by [27], is the ability to reconstruct an X-tree $T$ from $\Sigma(T)$ called tree popping. We order the elements $\sigma_1, \sigma_2, \ldots, \sigma_k$ arbitrarily, where $k = |\Sigma(T)|$, and we construct a sequence $T_0, T_1, \ldots, T_k$ of X-trees such that, for all $i \in \{1, 2, \ldots, k\}$, $\Sigma(T_i) = \{\sigma_1, \sigma_2, \ldots, \sigma_i\}$. Thus, $\Sigma(T_k) = \Sigma(T)$. In this construction, $T_0$ is the X-tree consisting of only one vertex labeled $X$, and, for all $i$, $T_i$ is the X-tree obtained from $T_{i-1}$ by introducing an edge corresponding to the X-split $\sigma_i$. This introduction is described in the induction step of the Splits-Equivalence Theorem.

To illustrate one such iteration in the tree popping method, let $X = \{1, 2, \ldots, 7\}$, and let $\sigma_1 = \{7\}|(X - \{7\}), \sigma_2 = \{1, 2\}|(X - \{1, 2\}), \sigma_3 = \{4\}|(X - \{4\})$, and $\sigma_4 = \{6, 7\}|(X - \{6, 7\})$. Applying the tree popping method in the chosen order we get the X-tree $(T, \phi)$ shown in Figure 2-5 (a) after three iterations. Now consider $\sigma_4$ and color the elements $\{6, 7\}$ red and the elements of $(X - \{6, 7\})$ green. Since the vertex labeled 3, 5, 6 is not monochromatic anymore, we separate that vertex into two monochromatic vertices as in Figure 2-5 (b).
2.5 Characters and Convexity

The concept of ‘characters’ is essential to any work in the domain of Phylogenetics. In biology, it refers to the attributes of the species being considered and are the data typically used to reconstruct phylogenetic trees. However, mathematically, characters are functions. In this section, we will formalize the notion and examine the mathematical properties of characters needed to construct phylogenetic trees. The following section was first presented in [32].

**Definition 2.5.1.** A character on $X$ is a function $\chi$ from a non-empty subset $X'$ of $X$ into a set $C$ of character states. $C$ is referred to as the state set of $\chi$. The character $\chi$ is said to be trivial if there is at most one element $\alpha \in C$ for which $|\chi^{-1}(\alpha)| \geq 2$; otherwise, $\chi$ is non-trivial. If $X' = X$, we say $\chi$ is a full character. If $|\chi(X')| = r$ we say $\chi$ is an $r$-character state. A character $\chi$ on $X$ is a binary character if $\chi$ is a two-state full character.

The biological interpretation of characters may vary. They can be morphological (e.g. fur versus feathers), behavioral, physiological, biochemical, embryological, or molecular. The undergraduate research that led to this thesis dealt with molecular and morphological characters. The next definition introduces the concept of convexity, which has a fundamental biological interpretation that we will discuss later in this section.

**Definition 2.5.2.** Let $\chi$ be a character on $X$ from $X'$ into a set $C$ of character states. We say that $\chi$ is convex on an $X$-tree $(T; \phi)$ with $T = (V, E)$ if there is a function $\overline{\chi} : V \rightarrow C$ satisfying the following properties:
(C1) $\chi \circ (\phi|X') = \chi$ and

(C2) for each $\alpha \in C$, the subgraph of $T$ induced by $\{v \in V : \chi(v) = \alpha\}$ is connected.

It follows immediately that a binary character $\chi$ on $X$ is convex on an $X$-tree $T$ precisely if the bipartition of $X$ induced by $\chi$ is an $X$-split of $T$.

The following is an example for compatibility.

![Figure 2-6. An X-tree and a mapping $\chi_t$.](image)

Let $X$ be the set $\{1, 2, \ldots, 7\}$. Let $C = \{\alpha, \beta, \gamma, \delta, \eta\}$ be the set of character states. Let $\chi : X \rightarrow C$ be the full character on $X$ defined by $\chi(2) = \chi(3) = \alpha, \chi(4) = \chi(5) = \beta, \chi(1) = \gamma,$ and $\chi(6) = \chi(7) = \delta$. So, $\chi$ is a four-state character. Now, consider the $X$-tree $T = (T; \phi)$ with $T = (V, E)$ shown in Figure 2-6 (a). For each $t \in \{\alpha, \beta, \gamma\}$, let $\chi_t$ be the map from $V$ into $C$ specified in Figure 2-6 (b). Clearly $\chi_t$ satisfies both (C1) and (C2) for all $t = \alpha, t = \beta, \text{ or } t = \gamma$. Hence, $\chi$ is convex on $T$ for any character state $t \in \{\alpha, \beta, \gamma\}$.

The next proposition provides two alternative descriptions of convexity.

**Proposition 2.5.3.** Let $T = (T; \phi)$ be an $X$-tree with $T = (V, E)$ and let $\chi : X' \rightarrow C$ be a character on $X$. The following statements are equivalent:

(i) $\chi$ is convex on $T$;

(ii) the members of $\{T(\alpha) : \alpha \in C\}$ are pairwise vertex disjoint; and

(iii) for all distinct $\alpha, \beta \in \chi(X')$, there exists an $X$-split $A|B$ of $T$ such that $\chi^{-1}(\alpha) \subseteq A$ and $\chi^{-1}(\beta) \subseteq B$. 

24
(\(T(\alpha)\) denotes the minimal subtree of \(T\) containing \(\phi(\chi^{-1}(\alpha))\) - the labels in \(X\) that were assigned character state \(\alpha\)).

**Proof.** It is sufficient to prove \((i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)\).

\((i) \Rightarrow (ii)\): Suppose \(\chi\) is convex on \(T\). Then there exists \(\overline{\chi} : V(T) \rightarrow C\) satisfying properties (C1) and (C2). Let \(\alpha_1, \alpha_2 \in C\). Then, by (C2), for any \(i \in \{1, 2\}\), \(T(\alpha_i)\) is a subtree of the subgraph of \(T\) induced by \(\{v \in V(T) : \overline{\chi} = \alpha_i\}\). So, \(T(\alpha_1)\) and \(T(\alpha_2)\) are vertex disjoint.

\((ii) \Rightarrow (iii)\): Suppose the elements of \(\{T(\alpha) : \alpha \in C\}\) are pairwise disjoint. Let \(\alpha\) and \(\beta\) be two distinct elements of \(\chi(X') \subseteq C\). Then, from property (C1), \(T(\alpha) \neq T(\beta)\).

Therefore, there exists a path from \(T(\alpha)\) to \(T(\beta)\) such that \(v_\alpha \in T(\alpha)\) and \(v_\beta \in T(\beta)\) are the beginning and the ending vertex of the path, and for any edge \(e\) of the path, \(e \notin E(T(\alpha))\) and \(e \notin E(T(\beta))\). Take an edge \(e\) in this path and consider its corresponding X-split. Let \(\chi^{-1}(\alpha) \subseteq A\) and \(\chi^{-1}(\beta) \subseteq B\) for any distinct \(\alpha\) and \(\beta\) in \(\chi(X')\). The (C1) property is clear; otherwise, an X-split wouldn’t be possible. Now, suppose for the sake of a contradiction that there exists an \(\alpha \in \chi(X')\) such that the subgraph of \(T\) induced by \(\{v \in V(T) : \overline{\chi}(v) = \alpha\}\) is not connected. Without loss of generality, suppose that \(T(\alpha)\) has two components, say \(M\) and \(N\). Then, the path between \(M\) and \(N\) contains a vertex \(v\) such that \(\chi(v) \neq \alpha\). Let \(\beta = \chi(v)\). Let \(e\) be the edge between \(M\) and \(v\). Take the X-split corresponding to \(e\).

Then, \(\chi^{-1}(M) \subseteq A\), but \(\chi^{-1}(N) \subseteq B\), which is a contradiction to the initial assumption. This completes the proof.

We can illustrate Proposition 2.5.3 using the X-tree in Figure 2-6. The vertex disjoint subtrees are shown in Figure 2-7.
Convexity is a fundamental concept to phylogenetics because of its biological meaning. Let a rooted phylogenetic X-tree $T = (T; \phi)$ be describing the evolution of the set $X$ of extant species from and ancestral species which we will introduce as the root $\rho$ of $T$. Now, suppose that each species (or vertex $v$) has an associated character state in $C$. We can regard the character state as ‘evolving’ from $\rho$ towards the elements of $X$ on $T$. For all $v \in V(T)$, let $c(v)$ denote the character state assigned to $v$. We can define the assumption that each time a species changes its character state, the new state it acquires appears for the first time in the tree in the following way:

**Definition 2.5.4.** We say that a character $c$ is homoplasy-free if neither of the following occur:

(i) Suppose $v_1, v_2, \ldots, v_k$ is a path in $T$ directed away from the root $\rho$. $c$ is said to exhibit reverse transition if for some $i \in \{2, 3, \ldots, k - 1\}$, $c(v_1) = c(v_k) \neq c(v_i)$. This corresponds to a new character state arising but then reverting back to an earlier state.

(ii) Suppose that $v_1, v_2, \ldots, v_k$ and $w_1, w_2, \ldots, w_l$ are paths in $T$ directed away from the root $\rho$ and that $v_1 = w_1$. $c$ is said to exhibit convergent transition if $c(v_k) = c(w_l) \neq c(v_1)$. This corresponds to the same state arising in different parts of the tree independent of each other.

Reverse and convergent transitions are known to occur in biology in many types of characters. We will now explain the the connection between these concepts and
convexity. Let $T = (T; \phi)$ be a rooted phylogenetic X-tree with $T = (V, E)$ and root $\rho$. Suppose each vertex $v$ of $T$ has a character state in $C$. Consider the associated phylogenetic X-tree $T^-\rho$. If we look only at the values of $c$ at the leaves of $T$, we obtained an induced full character $\chi$ on $X$ by setting $\chi(x) = c(\phi(x))$ for all $x \in X$.

This character is describing the character states of the present-day species. If $c$ is homoplasy-free, then $\chi$ is convex on $T^-\rho$ since $\chi: V \rightarrow C$ defined as $\chi(u) = c(u)$ for all $u \in V$, satisfies conditions (C1) and (C2). Conversely, if $\chi$ is convex on a phylogenetic X-tree $T_1$ with $T_1 = (V_1, E_1)$ and a corresponding function $\chi_1: V_1 \rightarrow C$ that satisfies conditions (C1) and (C2), then for all choices of a root $\rho$, we can extend $\chi_1$ to a map from $V_1 \cup \{\rho\}$ to $C$ that is homoplasy-free. It is important to note, however, that even if $c$ is not homoplasy-free on a rooted phylogenetic tree $T$ it is entirely possible that the associated character $\chi$ may be convex on $T^-\rho$. The concept of homoplasy is quantified in the next chapter.

### 2.6 Character Compatibility

**Definition 2.6.1.** A collection of characters on $X$ is said to be compatible if there exists an X-tree on which all the characters in the collection are convex.

In other words, a collection of characters is compatible if they could all have evolved on some tree without any reverse or convergent transitions. The condition of compatibility is the same for phylogenetic X-trees and binary phylogenetic X-trees. Determining whether a collection of characters is compatible and, if so, constructing the tree on which they are all complex is known as the character compatibility problem or, more recently in computer science circles, as the perfect phylogeny problem.

In the case of binary characters, the Splits-Equivalence Theorem (2.4.6) states that a collection of binary characters is compatible if and only if the characters are pairwise compatible. Hence, there is a unique minimal tree on which binary characters are convex. For non-binary characters, however, this observation does not apply. Semple and Steel offer a framework in [32] to ascertain compatibility using chordal graphs, which
we will now describe. We first need to introduce a series of definitions for the terms used.

**Definition 2.6.2.** Let \( S = \{ S_1, S_2, \ldots, S_k \} \) be a family of sets. The intersection graph of \( S \), denoted \( \text{int}(S) \), is the graph that has vertex set \( S \) and an edge between \( S_i \) and \( S_j \) precisely if \( S_i \cap S_j \neq \emptyset \), for \( i, j \in \{ 1, 2, \ldots, k \} \), and distinct.

**Definition 2.6.3.** A graph \( G \) is chordal, also called triangulated, if every induced subgraph of \( G \) that is a cycle has at most three edges. Equivalently, a graph is chordal if every cycle with at least four vertices has an edge (called a "chord") connecting two non-consecutive vertices in the cycle.

**Definition 2.6.4.** A chordalization (also called triangulation) of a graph \( G = (V, E) \) is a graph \( G' = (V, E') \) with the properties that \( G' \) is chordal and \( E \subseteq E' \).

**Definition 2.6.5.** For a character \( \chi : X' \rightarrow C \) on \( X \), let \( \pi(\chi) \) denote the partition of \( X' \) corresponding to \( \{ \chi^{-1}(\alpha) : \alpha \in C \} \). Let \( \mathcal{C} \) be a collection of characters on \( X \) and let \( T = (T; \phi) \) be an \( X \)-tree. Next, define two graphs, each of which has vertex set

\[
\bigcup_{\chi \in \mathcal{C}} \{(\chi, A) : A \in \pi(\chi)\}.
\]

(i) The partition intersection graph of \( \mathcal{C} \) is the graph that has the vertex set mentioned above and an edge joining two vertices precisely if the intersection of the second coordinates is non-empty. We denote this graph by \( \text{int}(\mathcal{C}) \).

(ii) The subtree intersection graph of \( T \) induced by \( \mathcal{C} \) is the graph that has the vertex set mentioned above and an edge \( \{(\chi, A), (\chi', B)\} \) if the intersection of the vertex sets of \( T(\phi(A)) \) and \( T(\phi(B)) \) is non-empty. This graph is denoted by \( \text{int}(\mathcal{C}, T) \).

**Definition 2.6.6.** A vertex of graph \( G \) is simplicial if its neighbors together with itself induce a clique (a graph in which each pair of distinct vertices is joined by one edge, also known as a complete graph).

**Definition 2.6.7.** We say that \( G \) has a perfect elimination ordering if the vertices of \( G \) can be ordered as \( v_1, v_2, \ldots, v_k \) so that for each \( i \in \{ 1, 2, \ldots, k \} \), \( v_i \) is a simplicial vertex of the subgraph of \( G \) induced by \( \{ v_i, \ldots, v_k \} \).
Definition 2.6.8. A graph \( G \) is a restricted chordal completion of \( \text{int}(C) \) if \( G \) is a chordalization of \( \text{int}(C) \) and, for all edges \( \{ (\chi, A), (\chi', B) \} \) of \( G, \chi \neq \chi' \).

The following theorem was stated by [32] with various parts of the equivalences due to [4], [13], [14], [30], and [36].

Theorem 2.6.9. Let \( G \) be a graph. Then the following statements are equivalent:

(i) \( G \) is chordal;

(ii) \( G \) is a subtree intersection graph;

(iii) \( G \) has a perfect elimination ordering;

(iv) there exists a tree \( T \) whose vertex set \( K \) is the set of maximal cliques of \( G \) and, for each vertex \( v \) in \( G \), the subgraph of \( T \) induced by the elements of \( K \) containing \( v \) is a subtree of \( T \).

The tree described in (iv) of Theorem 2.6.9 is referred to as a maximal clique tree representation of \( G \).

Theorem 2.6.10, indicated by [4] and [27], and formally proved by [33], is the main result of this section.

Theorem 2.6.10. Let \( \mathcal{C} \) be a collection of characters on \( X \). Then, \( \mathcal{C} \) is compatible if and only if there exists a chordal completion of \( \text{int}(\mathcal{C}) \).

Proof. Suppose \( \mathcal{C} \) is compatible. Then there exists an \( X \)-tree \( T \) on which every character in \( \mathcal{C} \) is convex. By Theorem 2.6.9, (i) \( \Leftrightarrow \) (ii), \( \text{int}(\mathcal{C}, T) \) is chordal. The edge set of \( \text{int}(\mathcal{C}) \) is a subset of the edge set of \( \text{int}(\mathcal{C}, T) \) and every character in \( \mathcal{C} \) is convex on \( T \).

Therefore, \( \text{int}(\mathcal{C}, T) \) is a restricted chordal completion of \( \text{int}(\mathcal{C}) \).

To prove the converse, suppose that \( G \) is a restricted chordal completion of \( \text{int}(\mathcal{C}) \). From Theorem 2.6.9, (i) \( \Leftrightarrow \) (iv), there exists a tree \( T' \) whose vertex set \( K \) is the set of the maximal cliques of \( G \), and for each vertex \( (\chi, A) \) the subgraph of \( T' \) induced by the elements of \( K \) containing \( (\chi, A) \) is a subtree of \( T' \). To complete the proof, we construct an \( X \)-tree via \( T' \) on which every character in \( \mathcal{C} \) is convex. Define \( \phi : X \rightarrow K \) such that, for any \( x \in X \), \( \phi(x) \) contains the vertices of the maximum-sized clique in \( G \) in which \( x \) is
an element of the second coordinate of every vertex. Observe that $int(C)$ is a subgraph of $G$, so a vertex of $G$ is in this clique precisely if this vertex contains $x$. Note that such a map may not be unique. Define $T$ to be the tree obtained from $T'$ by suppressing all vertices of degree two that are not identified by an element of $X$. It is easily checked that all degree-one vertices of $T'$ are identified by an element of $X$, and so $T = (T; \phi)$ is an $X$-tree.

Now, show that every character in $C$ is convex on $T$. Let $A_1, A_2$ be members of $\pi(x)$ for some $x \in C$. Then, the subtrees $T'_1$ and $T'_2$ of $T'$ induced by the elements of $K$ containing $(x, A_1)$ and $(x, A_2)$ respectively, do not intersect. Since the elements of $A_i$ can only be identified with vertices in $T'_i$, for each $i \in \{1, 2\}$, it follows that the intersection of the vertex sets of $T(A_1)$ and $T(A_2)$ is empty. Thus, every element of $C$ is convex on $T$, and therefore, $C$ is compatible by definition. 

**Corollary 2.6.11.** Two characters $\chi$ and $\chi'$ on $X$ are compatible if and only if $int(\{\chi, \chi'\})$ is acyclic.

**Proof.** Suppose $int(\{\chi, \chi'\})$ is acyclic. Then $int(\{\chi, \chi'\})$ is chordal. Hence, by Theorem 2.6.10 $\chi$ and $\chi'$ are compatible.

Conversely, suppose $int(\{\chi, \chi'\})$ contains a cycle. Let $G$ be a chordalization of $int(\{\chi, \chi'\})$. Then, $G$ must contain a three-cycle $(x-x'-x$ or $x-x'-x')$. This implies $G$ is not a restricted chordal completion of $int(\{\chi, \chi'\})$. Therefore, by Theorem 2.6.10, $\chi$ and $\chi'$ are not compatible. 

**Corollary 2.6.12.** Let $C$ be a collection of binary characters on $X$. Then, $C$ is compatible if and only if $int(C)$ is chordal.

**Proof.** If $int(C)$ is chordal, then $int(C)$ is a restricted chordal completion of itself. Therefore, by Theorem 2.6.10, $C$ is compatible.

Now, suppose $C$ is compatible. Then, by the Splits-Equivalence Theorem 2.4.6, there is a unique $X$-tree $T$ such that $\Sigma(T)$ is equal to the set of $X$-splits induced by the
elements in $C$. We next show $\text{int}(C) = \text{int}(C, T)$ by verifying the claim that the edge sets of $\text{int}(C)$ and $\text{int}(C, T)$ are equal.

Let $(\chi, A)$ and $(\chi', B)$ be distinct vertices of $\text{int}(C)$. If $\chi = \chi'$, then $A|B$ is an $X$-split induced by $T$ and the claim clearly holds. Next, assume that $\chi \neq \chi'$. If $A \cap B \neq \emptyset$ then the claim trivially holds. Assume $A \cap B$ is empty. Then, $A|(X - A)$ and $B|(X - B)$ are distinct $X$-splits induced by $T$. Therefore, by Lemma 2.4.3 there is a partitioning of $X$ into the three sets $X_1, X_2, X_3$ so that $X_1 \in \{A, X - A\}$, $X_3 \in \{B, X - B\}$ and the intersection of the vertex sets of $T(\phi(X_1))$ and $T(\phi(X_3))$ is empty. The only possible choices for $X_1$ and $X_3$ are $A$ and $B$, respectively. The claim now readily follows from this case.

This corollary, however, does not extend to collections of two-state characters on $X$.

We will now provide the framework of how to construct a maximal clique tree representation of a chordal graph following [15]. Suppose $G = (V, E)$ is a chordal graph. Let $v_1, v_2, \ldots, v_k$ be a perfect elimination ordering of the vertices of $G$, where $k = |V|$. Since every chordal graph has at least one simplicial vertex ([6]) and every vertex-induced subgraph of a chordal graph is chordal, obtaining such an ordering is elementary. Let $i \in \{1, 2, \ldots, k - 1\}$ and let $K_i$ denote the vertex set of the maximal clique of $G[\{v_i, v_{i+1}, \ldots, v_k\}]$ that contains $v_i$. Define $T_k$ as the tree consisting of the single vertex $v_k$. So, $T_k$ is a maximal clique representation of $G[\{v_k\}]$. In general, for all $i$, define $T_i$ to be the tree obtained from $T_{i+1}$ as follows:

(i) if $K_i - v_i$ is a vertex of $T_{i+1}$, then replace $K_i - v_i$ with $K_i$ to get $T_i$;

(ii) otherwise, join a new vertex $K_i$ to a vertex of $T_{i+1}$ containing $K_i - v_i$ to get $T_i$.

It easily checked that $T_i$ is a maximal clique tree representation of $G[\{v_i, v_{i+1}, \ldots, v_k\}]$ for any $i$. Thus, $T_i$ is a maximal (not necessarily unique) clique tree representation of $G$.

We now illustrate these concepts with an example. Suppose that $X = \{1, 2, 3, 4, 5, 6\}$ and let $C = \{\chi_1, \chi_2, \chi_3\}$ be a set of characters on $X$ with $\chi_1$, $\chi_2$, and $\chi_3$ as defined in Table 2-1. Then $\pi(\chi_1) = \{\{2\}, \{1, 4\}, \{3, 5, 6\}\}$, $\pi(\chi_2) = \{\{1, 2\}, \{3, 5\}, \{6\}\}$, and $\pi(\chi_3) = \{\{2, 3\}, \{4, 6\}\}$. Let $G$ denote the graph shown in Figure 2-8 (a). Since the first
coordinates of the end vertices of each of the dashed lines are distinct, \( G \) is a restricted chordal completion of \( \text{int}(\mathcal{C}) \), with \( \text{int}(\mathcal{C}) \) being the graph induced by the solid lines of this graph. Hence, by Theorem 2.6.10, \( \mathcal{C} \) is compatible.

Table 2-1. Characters \( \chi_1, \chi_2, \) and \( \chi_3. \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \chi_1(x) )</th>
<th>( \chi_2(x) )</th>
<th>( \chi_3(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \alpha )</td>
<td>( \beta )</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>( \alpha'' )</td>
<td>( \beta )</td>
<td>( \gamma )</td>
</tr>
<tr>
<td>3</td>
<td>( \alpha' )</td>
<td>—</td>
<td>( \gamma' )</td>
</tr>
<tr>
<td>4</td>
<td>( \alpha )</td>
<td>—</td>
<td>( \gamma' )</td>
</tr>
<tr>
<td>5</td>
<td>( \alpha' )</td>
<td>( \beta' )</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>( \alpha' )</td>
<td>( \beta'' )</td>
<td>( \gamma' )</td>
</tr>
</tbody>
</table>

We next construct an X-tree on which all of the characters in \( \mathcal{C} \) are convex. Let the following sequence be the perfect elimination ordering of \( G \) that we use:

\[(\chi_2, \{6\}), (\chi_1, \{2\}), (\chi_1, \{1, 4\}), (\chi_3, \{4, 6\}), (\chi_2, \{1, 2\}), (\chi_1, \{3, 5, 6\}), (\chi_3, \{2, 3\}), (\chi_2, \{3, 5\}).\]

Following the process described immediately prior to this example, we can construct the maximal clique tree representation \( T = (\mathcal{K}, E) \) of \( G \) shown in Figure 2-8 (b), where \( \mathcal{K} \) is the collection of maximal cliques of \( G \) and, for each vertex \( v \) in \( G \), the subgraph induced by the elements of \( \mathcal{K} \) containing \( v \) is a subtree of \( T \). In this case, this is the only maximal clique representation of \( G \). Lastly, to obtain the desired X-tree, we define a map \( \phi : X \rightarrow \mathcal{K} \) so that, for each element \( x \) in \( X \), \( \phi(x) \) contains the vertices of the maximum sized cliques in \( G \) in which \( x \) is an element of the second coordinate of every vertex. The X-tree in Figure 2-8 (c) is the tree on which all characters of \( \mathcal{C} \) are convex on it.

The problem of compatibility is relatively easily determined for binary characters, but it becomes a much more difficult problem to solve in other cases. [32] states that determining if an arbitrary collection \( \mathcal{C} \) of characters is compatible is NP-complete even if all characters in \( \mathcal{C} \) are two-state. Moreover, pairwise compatibility is not sufficient for compatibility of the entire collection of characters, also shown by [32]. It is, nonetheless, a useful method for a bounded number of distinct states or full characters on \( X \).
Figure 2-8. Compatibility of characters. (a) A restricted chordal completion of \( \text{int}(\mathcal{C}) \). (b) A maximal clique tree representation of \( G \). (c) An X-tree on which each character in \( \mathcal{C} \) is convex.

[1] discovered a number of polynomial-time algorithms, and so did [26].
CHAPTER 3
MAXIMUM PARSIMONY

The maximum parsimony method is one of the more popular techniques used for reconstructing phylogenetic trees from characters. The concept behind this method is to fit character data to a semi-labeled tree in a way that minimizes convergent and reverse transitions. This way of thinking is based on 'Ockham’s Razor' principle which says that a simple explanation is more likely and should be chosen over a more complex one. Here, the complexity is measured by the number of reverse and convergent transitions with a homoplasy-free tree being the ideal case. Moreover, such transitions are generally considered to be relatively rare and therefore a case with fewer transitions may be a more probable situation. In this chapter, we will explore the fundamental concepts of parsimony, looking mostly at classical parsimony. This method has direct connections to graph theory. We will deal mostly with full characters and phylogenetic X-trees in this chapter. Also, we will mostly deal with sequences of characters instead of sets of characters to allow for some data to have a character appear more than once.

3.1 Classical Parsimony

Definition 3.1.1. For a graph \( G = (V, E) \) and a function \( f \) on \( V \), the changing set of \( f \) is the subset \( Ch(f) = \{ \{u, v\} \in E : f(u) \neq f(v) \} \) of edges of \( G \). The changing number of \( f \), denoted \( ch(f) \), is the cardinality of \( Ch(f) \).

Definition 3.1.2. Let \( \chi : X' \to C \) be a character on \( X \) and let \( T = (T; \phi) \) be an X-tree. An extension of \( \chi \) to \( T \) is a function \( \overline{\chi} : V(T) \to C \) for which \( \overline{\chi} \circ (\phi|X') = \chi \). The parsimony score of \( \chi \) on \( T \) is the minimum value of \( ch(\overline{\chi}) \) over all extensions \( \overline{\chi} \) of \( \chi \) to \( T \). We denote this score by \( l(\chi, T) \). Furthermore, if \( \overline{\chi} \) is an extension of \( \chi \) to \( T \) and \( ch(\overline{\chi}) = l(\chi, T) \), then \( \overline{\chi} \) is called a minimum extension of \( \chi \) to \( T \).

Definition 3.1.3. Let \( C = \{\chi_1, \chi_2, \ldots, \chi_k\} \) be a sequence of characters on \( X \). The parsimony score of \( C \) on an X-tree \( T \), denoted by \( l(C, T) \), is the sum of the individual...
parsimony scores of the characters of $T$; thus,

$$l(C, T) = \sum_{i=1}^{k} l(\chi_i, T).$$

An $X$-tree $T'$ that minimizes $l(C, T)$ is called the maximum parsimony tree for $C$ and the corresponding value of $l(C, T)$ is $l(C)$.

**Proposition 3.1.4.** Let $\chi$ be an r-state character on $X$ and let $T$ be an $X$-tree. Then $l(\chi, T) \geq r - 1$. Moreover, $l(\chi, T) = r - 1$ if and only if $\chi$ is convex on $T$.

**Proof.** Let $T$ be the underlying tree of $T$ and let $\overline{\chi}$ be a minimum extension of $\chi$ to $T$.

Let $T_{\overline{\chi}}$ denote the tree obtained from $T$ by contracting every edge in $E(T) - Ch(\overline{\chi})$, and consider the mapping on $V(T_{\overline{\chi}})$ induced by $\overline{\chi}$. Since the cardinality of the image of this mapping is $r$, we have $|V(T_{\overline{\chi}})| \geq r$. Therefore, as $|E(T_{\overline{\chi}})| = ch(\overline{\chi})$ and $T_{\overline{\chi}}$ is a tree, we have $l(\chi, T) = ch(\overline{\chi}) \geq r - 1$. Furthermore, if $\chi$ is convex on $T$, then $|V(l(\chi, T))| = r$. Hence, equality holds if and only if $\chi$ is convex on $T$. $\square$

In Chapter 2 we discussed reverse and convergent transitions and homoplasy-free characters. We can now count the number of such transitions, defined as $h(\chi, T)$, using the following formulas given by [32].

Let $\chi : X' \rightarrow C$ be a character on $X$ and let $T$ be an $X$-tree. Let

$$h(\chi, T) = l(\chi, T) - r + 1.$$ $h(\chi, T)$ is sometimes referred to as the homoplasy of $\chi$ on $T$. From Proposition 3.1.4 we know that $h(\chi, T)$ is non-negative, and is equal to zero precisely when $\chi$ is convex on $T$ ($\chi$ is homoplasy-free). Now let $C$ be a sequence $(\chi_1, \chi_2, ..., \chi_k)$ of characters on $X$ and let $T$ be a maximum parsimony tree for $C$. The quantity

$$h(C) = \sum_{i=1}^{k} h(\chi_i, T).$$
is the called the total homoplasy of $C$ and measures the number of reverse and convergent transitions that need to be postulated if all the characters in $C$ evolved on a common X-tree. The following corollary is an immediate consequence of Proposition 3.1.4.

**Corollary 3.1.5.** Suppose that $C = (\chi_1, \chi_2, \ldots, \chi_k)$ is a sequence of characters on $X$. Then, $h(C) \geq 0$ with equality precisely if $C$ is compatible.

The parsimony score of a character on a semi-labeled tree can be viewed in terms of sets of edges separating vertices assigned different character states or in terms of a maximal system of paths under certain restrictions. We will consider both ways of viewing parsimony.

**Definition 3.1.6.** Let $\chi : X' \to C$ be a character on $X$ and let $T = (T; \phi)$ be an X-tree with $T = (V, E)$. A subset $E_1$ of $E$ is a cut-set for $\chi$ on $T$ if, for each pair $x, y \in X'$ with $\chi(x) \neq \chi(y)$, the vertices $\phi(x)$ and $\phi(y)$ lie in different components of the disjoint union of trees $T \setminus E_1$.

A minimum cut-set for $\chi$ on $T$ is a cut-set for $\chi$ of minimum size and the size of such a cut-set is denoted by $\text{cut}(\chi, T)$. The set of all minimum cut-sets for $\chi$ on $T$ is $\text{Cut}(\chi, T)$.

The following example illustrates these concepts. Let $T$ be the X-tree shown in Figure 3-1 (a) with $X = \{1, 2, 3, 4, 5, 6, 7\}$. Let $\chi : X \to \{\alpha, \beta, \gamma\}$ be the character defined by $\chi(2) = \chi(6) = \chi(7) = \alpha, \chi(1) = \chi(4) = \chi(5) = \beta$, and $\chi(3) = \gamma$. One can easily check that $l(\chi, T) = 3$. Figure 3-1 (b) indicates a minimum extension $\chi$ of $\chi$ to $T$ as well as the three corresponding edges in $Ch(\chi)$. Note that this minimum extension is not unique for this tree and character set $\chi$. In Figure 3-1 (c), the set $\{e_1, e_2, e_3, e_4\}$ of edges of $T$ is a cut-set for $\chi$, but it is not the changing set of any extension of $\chi$. This shows that an arbitrary cut-set $E_1$ for $\chi$ does not necessarily correspond to $Ch(\chi)$ for some extension $\chi$ of $\chi$. However, Lemma 3.1.7 shows this is not the case if $E_1$ is a minimum cut-set for $\chi$. 

36
Lemma 3.1.7. Let $\chi$ be a character on $X$, let $T = (T; \phi)$ be an X-tree with $T = (V, E)$, and let $E_1$ be a subset of $E$. If $E_1$ is a minimum cut-set for $\chi$, then $E_1 = ch(\chi)$ for a unique extension $\overline{\chi}$ of $\chi$.

Proof. Suppose $E_1$ is a minimum cut-set for $\chi$. If $V'$ is the set of vertices of a component of the disjoint union of trees $T \setminus E_1$, then as $E_1$ is a cut-set for $\chi$ we must have $|\chi(\phi^{-1}(V'))| \leq 1$. Furthermore, $|\chi(\phi^{-1}(V'))| \neq 0$; otherwise, by selecting an edge $e$ in $E_1$ that is incident with a vertex in $V'$, the set $E_1 - \{e\}$ is a cut-set for $\chi$, which contradicts the minimality of $E_1$. Hence, $|\chi(\phi^{-1}(V'))| = 1$. We next define an extension $\overline{\chi}_1$ of $\chi$ using the following criteria. Each vertex $v$ of $V$ lies in exactly one of the components of $T \setminus E_1$. If the vertex set of that component is $V'$, set $\overline{\chi}_1$ equal to the unique element in $\chi(\phi^{-1}(V'))$. Now, each edge $\{u, v\}$ in $E_1$ satisfies $\overline{\chi}_1(u) \neq \overline{\chi}_1(v)$ (otherwise $E_1$ would not be minimal), and so we have $Ch(\overline{\chi}_1) = E_1$. Furthermore, it is easily checked that every
extension $\overline{\chi}$ of $\chi$ that satisfies $Ch(\overline{\chi}) = E_1$ must agree with $\overline{\chi}_1$ at all vertices in $V$. So, $\overline{\chi}_1$ is unique.

The following proposition shows that the parsimony score of a character $\chi$ on $X$ on an X-tree $T$ is equal to the size of a minimum cut-set for $\chi$ on $T$ and that any two minimum extensions that induce the same changing set are equal.

**Lemma 3.1.8.** Let $\chi$ be a character on $X$ and let $T$ be an X-tree. Then, $\text{cut}(\chi, T) = l(\chi, T)$. Furthermore, the map $\psi$ from the set of minimum extensions of $\chi$ on $T$ into $\text{Cut}(\chi, T)$ defined by $\psi(\overline{\chi}) = Ch(\overline{\chi})$, for all such extensions, is a bijection.

**Proof.** Let $\overline{\chi}$ be a minimum extension of $\chi$ on $T$. Then, $ch(\overline{\chi}) = l(\chi, T)$, and since $Ch(\overline{\chi})$ is a cut-set for $\chi$ on $T$, $\text{cut}(\chi, T) \leq l(\chi, T)$. Now, suppose that $E_1 \in \text{Cut}(\chi, T)$. Then, by Lemma 3.1.7,

$$\text{cut}(\chi, T) = |E_1| \geq \min \{ \text{ch}(\overline{\chi}_1 : \overline{\chi}_1 \text{ is an extension of } \chi \} = l(\chi, T),$$

establishing $\text{cut}(\chi, T) \geq l(\chi, T)$, and thereby the first part of the proposition.

We now prove the second part. Since $Ch(\overline{\chi})$ is a cut-set for $\chi$ on $T$ and since $ch(\chi, T) = l(\chi, T)$, it follows from the first part of the proposition that $Ch(\overline{\chi}) \in \text{Cut}(\chi, T)$. Moreover, by Lemma 3.1.7, the map $\psi$ is a bijection.

Having described the parsimony score in terms of the sets of edges that separate vertices with different character states, we now examine the parsimony score in terms of a maximal system of paths. We first do this for two-state characters using the following classical graph theory result proven by [28].

**Lemma 3.1.9** (Menger’s Lemma). Let $G = (V, E)$ be a graph, and let $V_1$ and $V_2$ be disjoint subsets of $V$. Then, the maximum number of edge-disjoint paths in $G$ with the property that each path has one endpoint in $V_1$ and the other endpoint in $V_2$ is equal to the minimum number of edges whose removal form $G$ leaves the vertices in $V_1$ in different components from the vertices in $V_2$. 

38
Menger’s Lemma 3.1.9 can be applied to binary characters. First, however, we need to introduce the following concept. For an X-tree $T = (T; \phi)$, a path $P$ in $T$ is a proper path relative to a character $\chi$ on $X$ if, for some $x, y \in X$, $P$ connects $\phi(x)$ and $\phi(y)$, and $\chi(x) \neq \chi(y)$. Combining Lemma 3.1.9 with Proposition 3.1.8 we obtain the following corollary.

**Corollary 3.1.10.** Let $\chi$ be a two-state character on $X$ and let $T = (T; \phi)$ be an X-tree. Then, the maximum parsimony score $l(\chi, T)$ is equal to the maximum number of edge-disjoint proper paths of $T$ relative to $\chi$.

To illustrate Corollary 3.1.10, consider the phylogenetic X-tree $T$ shown in Figure 3-2 (a) and the binary character $\chi : X \rightarrow \{\alpha, \beta\}$, where $\chi(2) = \chi(3) = \chi(4) = \chi(7) = \alpha$ and $\chi(1) = \chi(5) = \chi(6) = \beta$. Each dashed path in Figure 3-2 (b) is proper relative to $\chi$. Furthermore, as these paths are edge disjoint and as $|\chi^{-1}(\beta)| = 3$, the set of these paths is a maximum-sized set of edge-disjoint proper paths of $T$ relative to $\chi$. Hence, by Corollary 3.1.10, the maximum parsimony score is $l(\chi, T) = 3$.

Figure 3-2. A maximum-sized set of edge-disjoint proper paths.

Erdős and Székely [8] have extended the concept in Corollary 3.1.10 to arbitrary characters by permitting paths to intersect, provided some conditions are met.

Suppose there is a phylogenetic X-tree $T = (T; \phi)$ and a character $\chi$ on $X$. A collection $\mathcal{D}$ of directed paths in $T$ is an Erdős -Székely path system for $\chi$ on $T$ if it satisfies the following two conditions:
(i) If $P \in D$, then $P$ connects two leaves $\phi(x)$ and $\phi(y)$ of $T$ for which $\chi(x) \neq \chi(y)$.

(ii) Let $P$ and $P'$ be paths in $D$ that share some edge. Then, $P$ and $P'$ traverse this edge in the same direction and, if $\phi(x)$ and $\phi(y)$ denote the terminal vertices of $P$ and $P'$, $\chi(x) \neq \chi(y)$.

Consider the phylogenetic X-tree $T$ shown in Figure 3-3 (a) and the full character $\chi: X \to \{\alpha, \beta, \gamma\}$, where $\chi(1) = \chi(3) = \alpha$, $\chi(5) = \chi(6) = \beta$, and $\chi(2) = \chi(4) = \gamma$. An Erdös -Székely path system for $\chi$ on $T$ is shown in Figure 3-3 (b). Note that the path system is not unique.

Figure 3-3. An Erdös -Székely path system.

**Theorem 3.1.11.** Let $\chi$ be a character on $X$ and let $T$ be a phylogenetic X-tree. Then the parsimony score $l(\chi, T)$ is equal to the maximum size of an Erdös-Székely path system for $\chi$ on $T$.

Theorem 3.1.11 is due to Erdös and Székely [8], who also provided a polynomial-time algorithm to construct an explicit Erdös-Székely path system for a given character on $X$ and phylogenetic X-tree. The result of Theorem 3.1.11 can be extended to X-trees as shown by [32]. Let $T = (T; \phi)$ be an X-tree and suppose that, for some $x \in X$, $\phi(x)$ is an internal vertex on $T$. Let $V_1, V_2, \ldots, V_k$ denote the vertex sets of the components of $T \setminus \phi(x)$ and, for each $i \in \{1, 2, \ldots, k\}$, let $T_i$ denote the subgraph of $T$ induced by $V_i \cup \{\phi(x)\}$. Let $X_i = \phi^{-1}(V(T_i))$, let $\phi_i = \phi|X_i$, and let $T_i = (T_i; \phi_i)$ for all $i$. It is now
easily seen that
\[ l(\chi, T) = \sum_{i=1}^{k} l(\chi_i, T_i). \]
where \( \chi_i = \chi_i|X_i \) for all \( i \). This process, repeated for all \( x \) for which \( \phi(x) \) is an interior vertex, provides a collection of phylogenetic trees whose individual parsimonies are summed to get the parsimony score of \( \chi \) on \( T \).

### 3.2 Optimization on a Fixed Tree

The next step in understanding the maximum parsimony method, as presented in [32], is sometimes called the 'small parsimony problem'. Given a fixed parsimony tree, we wish to compute the parsimony score of a sequence of characters.

Suppose we have a character \( \chi : X \to C \) and a phylogenetic X-tree \( T \) with \( v \) internal vertices. There are several ways of computing \( l(\chi, T) \). The first one is the brute-force approach of searching all \(|C|^v\) extensions of \( \chi \) to find the minimum extension of \( \chi \) on \( T \). However, this method is impractical for more than 20 internal vertices since the number of such extensions grows exponentially. A second, more accessible approach is computing the parsimony score using a general dynamic programming approach which requires \( O(|X| \times |C|^2) \) time. This method does not use any fixed underlying mathematical concepts and is not be examined in this thesis. The third method is a classic in the field, being described in its entirety in [11]. The Fitch-Hartigan algorithm requires \( O(|X| \times |C|) \) time. The algorithm's framework is as follows.

If \( T \) is unrooted, we introduce a root arbitrarily, since the final output of the method is independent of the root chosen and the location where it is inserted. The method has two main passes of the tree. The initial pass, described previously ([11]), is called a forward pass and allows us to compute the parsimony score. We assign non-empty subsets of \( C \) and corresponding integers to the vertices of \( T \) recursively from leaves to the root. The subsequent pass, the backward pass from the root back to the leaves, constructs a minimum extension of \( \chi \) to \( T \).
Algorithm 3.2.1. Let \( \chi : X \to C \) be a character and let \( T \) be a rooted phylogenetic \( X \)-tree with root \( \rho \). Define two maps \( \psi : V(T) \to (2^C - \{\emptyset\}) \) and \( l : V(T) \to \{0, 1, 2, \ldots\} \) recursively as follows. Let \( v \in V(T) \). If \( v \) is a leaf of \( T \) labelled by \( x \in X \), set \( \psi(v) = \{\chi(x)\} \) and \( l(v) = 0 \). For the first half of the algorithm, we analyze all paths from the leaves to the root \( \rho \), considering preceding vertices to be descendants of the following vertices in each path. For each \( i \in \{1, 2, \ldots, 2n - 1\} \), let \( v_{i1}, v_{i2}, \ldots, v_{ik} \) denote the immediate descendants of \( v_i \), \( n \) being the number of vertices of \( T \). For every vertex \( v_i \) of \( T \), let \( f(v_i) = \max_{\alpha \in C} |\{j : \alpha \in \psi(v_{ij})\}| \).

Now, for each \( i \in \{1, 2, \ldots, 2n - 1\} \), set \( \psi(v_i) \) to be the set of character states of \( C \) that appear in \( f(v_i) \) of the sets \( \psi(v_{i1}), \psi(v_{i2}), \ldots, \psi(v_{ik}) \). In other words,

\[
\psi(v_i) = \begin{cases} 
\bigcap_{j=1}^{k} \psi(v_{ij}), & \text{if } \bigcap_{j=1}^{k} \psi(v_{ij}) \neq \emptyset, \\
\bigcup_{j=1}^{k} \psi(v_{ij}), & \text{otherwise}.
\end{cases}
\]

Also, define the quantity

\[ l(v_i) = \sum_{j=1}^{k} l(v_{ij}) + k - f(v_i). \]

Next, associate with each vertex \( v \) of \( V(T) \) the ordered pair \((\chi(v), l(v))\).

This assignment is called the forward pass and the parsimony score is given by

Theorem 3.2.2. [17]

Theorem 3.2.2. Let \( \chi : X \to C \) be a character and let \( T \) be a rooted phylogenetic \( X \)-tree with root \( \rho \). Suppose that we have completed the forward pass on the vertices of \( T \).

Then

(i) \( l(\rho) = l(\chi, T) \), and

(ii) \( \psi(\rho) = \{\alpha \in C : \text{there is a minimum extension } \overline{\chi} \text{ of } \chi \text{ to } T \text{ with } \overline{\chi}(\rho) = \alpha\} \).
The character-state sets \( \psi(\rho) \) obtained for each vertex \( v \) of \( T \) through forward pass cannot be used for the minimal extension because of the deficiencies of the preliminary phase described in [11]. These deficiencies are shown in Figure 3-4.

![Figure 3-4](image_url)

Figure 3-4. Deficiencies of forward pass reconstruction.

Figure 3-4 (a) shows a preliminary phase reconstruction of a position from three leaves. The set \{\( \alpha\beta \)\} at the lower ancestral vertex represents the impossibility to decide whether the ancestral character state was \( \alpha \) or \( \beta \). The only certainty is that a replacement is required. The third vertex requires the ultimate ancestor to be an \( \alpha \). Therefore, by the assumptions of parsimony, the lower ancestral vertex has to be an \( \alpha \) as in Figure 3-4 (b). The elimination of \( \beta \) from the first ancestor is determined by what Fitch calls the rule of diminished ambiguity.[11] The precise formulation of this rule is encompassed in steps I and II of the backward pass Algorithm 3.2.3, to be presented further on.

Figure 3-4 (c) has three possible character-states for the ultimate ancestor and two states for the lower ancestral vertex. While the choice can be made in several ways to maximize parsimony, Figure 3-4 (d) presents an adequate solution which was not comprised in the possible alternatives in available in (c). This case is encompassed by
the rule of expanded ambiguity, which is described in steps III and IV of the Algorithm 3.2.3.

In Figure 3-4 (e) is shown a third reconstruction that has four leaves and two ancestors that need replacement. Figure 3-4 (f) represents a valid solution. In this type of case, two vertices, separated by a single vertex, both containing a character state not present in the set of the intermediary vertex can assign that character state to it. Hence, this is called the rule of encompassing ambiguity and is comprised in step V of Algorithm 3.2.3.

**Algorithm 3.2.3.** The preliminary character-state set for the root \( \psi(\rho) \) is made the final set for that vertex. Then, go to one of its descendent vertices and proceed according to the following six steps:

I. If the preliminary set \( \psi(v) \) contains all character-states present in the final set of its immediate ancestor, go to step II, otherwise, go to III.

II. Eliminate all character states from the preliminary set \( \psi(v) \) that are not present in the final set of its immediate ancestor and go to VI.

III. If the preliminary character-state \( \psi(v) \) was formed by a union of its descendent sets, go to IV, otherwise go to V.

IV. Add to the preliminary set \( \psi(v) \) any character states in the final set of its immediate ancestor that are not present in \( \psi(v) \) and go to VI.

V. Add to the preliminary \( \psi(v) \) any character states not already present provided that they are present in both the final set of the immediate ancestor and in at least one of the two immediately descendent preliminary sets and go to VI.

VI. The preliminary set \( \psi(v) \) is now final. Descend one vertex as long as any preliminary vertex sets remain and return to I above.

This algorithm applies to all interior vertices. When the first rule does not apply, then and only then does the second rule apply. The third rule can apply if and only if the first two do not. They mutually exclude one another.

Figure 3-5 illustrates an example of applying the forward and backward pass to the phylogenetic X-tree \( T \) presented in Figure 3-5 (a). It also shows the location where we
introduce the root $\rho$. Figure 3-5 (b) shows the rooted X-tree $T'$ after having performed the forward pass. Each interior vertex was assigned a preliminary character-state set and the parsimony score of that vertex. Based on Theorem 3.2.2, the parsimony score of the X-tree is $l(T') = l(\rho) = 4$. Figure 3-5 (c) presents $T'$ after the backward pass. The character states in parentheses are the ones overlooked by forward pass because of its deficiencies. The character-state sets of each vertex are the possible character states that the vertex can be assigned in order to get a maximum extension of the original X-tree $T$.

Figure 3-5. (a) The X-tree. (b) Forward pass. (c) Backward pass.

The fixed-tree problem for parsimony can be generalized in a number of directions. In certain situations, the character state assigned by some particular character(s) and species may be unspecified or ambiguous. The problem can be also extended to graphs. Generalized parsimony assigns different non-negative weights to the transitions.
among character states. Its premise is to penalize rare or unfrequent transitions. These directions are discussed in [32] but were not pursued in this thesis.

### 3.3 Tree Rearrangement Operations

When calculating maximum parsimony, one needs to consider multiple trees since the ancestral species represented by the interior vertices and the relationships among them are unknown. However, even in the setting of classical parsimony, the problem of finding a maximum parsimony tree for a sequence $C$ of binary characters is NP-hard. [12] For these tasks, branch-and-bound algorithms have been effective when dealing with modest number of species. For larger numbers, researchers use heuristic methods based on tree rearrangement operations. [32]

The premise for the rearrangement operations is that $T = (T; \phi)$ is a binary phylogenetic X-tree and $e = \{u, v\}$ is an edge of $T$. We introduce three types of tree rearrangement operations, beginning with the least restrictive operations. Not that all operations are reversible.

**Tree Bisection and Reconnection (TBR):** Let $T'$ be the binary tree obtained from $T$ by deleting $e$, adding an edge between a vertex that subdivides an edge of one component of $T \setminus e$ and a vertex that subdivides and edge of the other component of $T \setminus e$, and then suppressing any resulting degree-two vertices. If a component of $T \setminus e$ consists of a single vertex, then the added edge is attached to this vertex. The binary phylogenetic X-tree $(T'; \phi')$, where $\phi'(x) = \phi(x)$ for all $x \in X$, is obtained from $T$ by a single tree bisection and reconnection operation. Figure 3-6 illustrates the generic form of this operation, where $e$ is the deleted edge and $f$ is the added edge.

![Figure 3-6. A schematic representation on the generic TBR operation.](image)
Subtree Prune and Regraft (SPR): Let $T'$ be the binary tree obtained from $T$ by deleting $e$ (pruning a subtree), adding an edge between an end vertex $u$ of $e$ and a vertex that subdivides an edge in the component of $T \setminus e$ that does not contain $u$ (regrafting the subtree) and then suppressing any resulting degree-two vertices. The binary phylogenetic X-tree $(T'; \phi')$, where $\phi'(x) = \phi(x)$ for all $x \in X$, is said to be obtained from $T$ by a single subtree prune and regrafting operation. The generic form of the SPR operation is illustrated in Figure 3-7, where $e$ is the deleted edge and $f$ is the added edge.

Nearest Neighbor Interchange (NNI): Suppose that $e' = \{v, v'\}$ is an interior edge of $T$ adjacent to $e$. Let $T'$ be the binary tree obtained from $T$ by deleting $e$, adding an edge between $u$ and a vertex that subdivides an edge with the end vertex $v'$ in $T \setminus e$ (interchanging two subtrees across $e'$), and then suppressing any resulting degree-two vertices. The binary phylogenetic X-tree $(T'; \phi')$, where $\phi'(x) = \phi(x)$ for all $x \in X$, is said to be obtained from $T$ by a single nearest neighbor interchange operation. Figure 3-8 illustrates the generic form of the NNI operation, where $e$ is the deleted edge and $f$ is the added edge.
Clearly, any NNI operation is a restriction on an SPR operation and any SPR operation is a particular type of TBR operation. Also, every TBR operation is either a single SPR operation or the composition of two SPR operations.

**Proposition 3.3.1.** Let $T$ and $T'$ be elements of $B(n)$. Then, $T'$ can be obtained from $T$ by a sequence of NNI operations and, therefore, by a sequence of SPR (or TBR) operations. [29]

**Theorem 3.3.2.** Let $T$ be an element of $B(n)$. Then, the number of elements of $B(n)$, apart from $T$, that can be obtained by

(i) a single NNI operation on $T$ is $2(n - 3)$ and

(ii) a single SPR operation on $T$ is $2(n - 3)(2n - 7)$. [2]

**Proof.** To prove (i), let $e$ be an interior edge of $T$. Then, the number of distinct elements of $B(n)$ that can be obtained by a single NNI operation across $e$ is two. It is easily seen that neither of these elements can be obtained by a single NNI operation across any other interior edge of $T$. Since $T$ has $n - 3$ interior edges, the number of elements of $B(n)$ that can be obtained by a single NNI on $T$ is $2(n - 3)$.

For the proof of (ii), we note that any SPR operation is precisely one of three types:

(a) the pruned edge is adjacent to the edge on which the subtree is regrafted;

(b) the pruned edge is separated by exactly one edge from the edge on which the subtree is regrafted; and

(c) the pruned edge is separated by at least two edges from the edge on which the subtree is regrafted.

For a type (a) operation, the resulting element is $T$. For type (b), the single SPR operation corresponds to a single NNI operation on $T$. Thus, by part (i) of this theorem, there are $2(n - 3)$ elements of $B(n)$ resulting from a single type (b) SPR operation on $T$.

Now consider type (c). Every SPR operation on $T$ corresponds to an ordered pair of distinct edges of $T$, where the first component is the pruned edge and the second component is the edge on which the subtree is regrafted. One can easily check that two
distinct type (c) SPR operations on $T$ result in two distinct elements of $B(n)$. Therefore, we have $(2n - 3)(2n - 4)$ such ordered pairs. We need to subtract the number of such ordered pairs from from the sum of the ordered pairs from type (a) and type (b) SPR operations on $T$.

The number of ordered pairs corresponding to a type (a) of SPR operation on $T$ is $6(n - 2)$, since there are six such pairs for each of the $n - 2$ interior vertices. Now a type (b) SPR operation corresponds to an NNI operation across an interior edge. Therefore, the number of ordered pairs corresponding to a type (b) SPR operation on $T$ is $8(n - 3)$, since there are eight such pairs for each of the $n - 3$ interior edges. Therefore, the number of distinct elements on $B(n)$ resulting from a type (c) SPR operation on $T$ is

$$(2n - 3)(2n - 4) - [6(n - 2) + 8(n - 3)] = 4(n - 3)(n - 4).$$

Combining the results of (a)–(c), we deduce that the total number of distinct elements of $B(n)$, apart from $T$, resulting from a single SPR operation on $T$ is

$$0 + 2(n - 3) + 4(n - 3)(n - 4) = 2(n - 3)(2n - 7).$$

Observe from the statement of Theorem 3.3.2 that the number of elements of $B(n)$ that can be obtained by a single NNI (or SPR) operation on $T$ is independent of the shape of $T$. However, in the case of a single TBR operation on $T$ the number of such elements is dependent on the shape of $T$. Nevertheless, this number is $O(n^3)$ regardless of the shape of $T$. [32]

### 3.4 Relevance

Even though parsimony methods are based on a specific optimality criteria, they do not require explicit models of evolutionary change. There is considerable disagreement over whether being 'model free' is an advantage or disadvantage. Either way, not having an explicit model does not make parsimony more reliable. This method does make
assumptions, but they are implied and it is difficult to trying to define them. One example
is that the acceptance of an optimal tree under parsimony requires the assumption
that it is unlikely for parsimony methods to estimate an incorrect tree. The ability of an
estimation method to converge to a true value, in this case the correct tree, is known as
consistency. It has been shown that, in some cases, parsimony can favor incorrect trees
more if the number characters increases. This situation is known as the Felsenstein
zone (of inconsistency). [34] Nonetheless, maximum parsimony presents itself to
biologists as a reasonable estimator for the number of changes and the structure of the
tree, even if it is imperfect under some conditions.

Maximum parsimony method was used in the undergraduate research to determine
the branch lengths of the best-fit model because parsimony yields integer values. The
best fit model for the molecular data was established by the Akaike Information Criterion
on a preliminary tree according to expert evaluation [22]. Even though the tree was
calculated with maximum likelihood, maximum parsimony branch lengths were used so
molecular change would be represented by numbers of events (numbers of mutations)
that would be comparable to the morphological change.[20]
CHAPTER 4
MAXIMUM LIKELIHOOD

Maximum likelihood is another popular method for inferring phylogeny that has been gaining recognition especially in recent years. This method is a specific implementation of the popular statistical method used for fitting a statistical model to data, and providing estimates for the model's parameters called maximum likelihood estimation. For a fixed set of data and an underlying probability model, maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameters would make them. In contrast to parsimony methods, which are based mainly on graph theory and combinatorics, maximum likelihood methods of phylogenetic inference evaluate a hypothesis about evolutionary history in terms of probability. The assumption behind it is that a history with a higher probability of reaching the current set of species being observed is preferable to other hypotheses with lower probabilities of giving rise to the observed state. Maximum likelihood methods attempt to account for unobserved as well as observed substitutions. They choose the hypothesis that maximizes the probability of observing the current data. [34]

Maximum likelihood frequently yields estimates with lower variance than other methods, which means it is often the method least affected by sampling error. It also takes into account the fact that substitution processes taking place at different sites have much in common and that the major components determining the evolution of sequences can be described by just a few parameters, a fact overlooked by most methods. For this reason, likelihood tends to outperform other estimation methods even with very short sequences of characters. However, the perceived and actual complexities of obtaining a solution to problems with numerous alternative hypotheses have hindered the widespread use of this method. [34] As a rule, the examples provided in this chapter will use molecular character states since most of the time maximum likelihood is applied to molecular sequences of data and since the parameters of the
models presented are based on genetic data. Also, unless otherwise stated, the set of character states will be \( C = \{A, C, T, G\} \).

### 4.1 Basic Principles

A concrete evolutionary model is needed to perform maximum likelihood. This model may be fully defined or may contain parameters to be estimated from the data. A maximum likelihood approach evaluates the probability that the chosen model, a phylogenetic X-tree, will have generated the observed data. Phylogenetic trees are then inferred by finding the trees that yield the highest likelihoods. The actual process is complex because different tree topologies require different mathematical treatments. [31]

Let \( T = (T; \phi) \) be the input X-tree with \( T = (V, E) \) on which we will perform the maximum likelihood approach. Let \( \chi : X \rightarrow C \) be the full character on \( X \). Let \( C \) be a collection of characters \( \chi \) on \( X \). We want to calculate the probability that \( T \) could have generated the character set \( C \) under the chosen model. Most models are time-reversible, which means that the probability of character-state \( \alpha \) changing into character-state \( \beta \) is the same as the probability of \( \beta \) changing into \( \alpha \). If an unrooted tree is evaluated using likelihood, it is convenient to root the tree arbitrarily at an interior vertex since, if the model is time-reversible, the likelihood of a tree is generally independent of the location of the root. We define \( L(\chi) \) as the likelihood value of character \( \chi \) and \( L(C) \) as the likelihood of the set of characters \( C \). Under the assumption that character states evolve independently, we calculate the likelihood for each character \( \chi \in C \) separately and combine the likelihoods into \( L(C) \). Since the likelihood is a probability value, we have

\[
L(C) = L(\chi_1) \cdot L(\chi_2) \cdot \ldots \cdot L(\chi_N) = \prod_{j=1}^{N} L(\chi_j),
\]
where \( N = |C| \). Because the probability of any single observation is extremely small, we almost always evaluate the log of the likelihood instead.

\[
\ln L(C) = \ln L(\chi_1) + \ln L(\chi_2) + \ldots + \ln L(\chi_N) = \sum_{j=1}^{N} \ln L(\chi_j). \tag{34}
\]

To calculate the likelihood for character \( \chi \in C \), we must consider all possible scenarios by which we could get the set of leaves of \( T \). Some of these scenarios are more plausible than others, but every case has at least some probability of generating any pattern of observed leaves. Hence, there are \( |C|^{\tilde{V}(T)} \) possibilities to consider, where \( \tilde{V}(T) \) is the set of interior vertices of \( T \). Define a scenario as the set \((\tilde{\chi}(v_1), \tilde{\chi}(v_2), \ldots, \tilde{\chi}(v_k))\) where \( k = |\tilde{V}(T)| \) and \( \tilde{\chi} : \tilde{V}(T) \rightarrow C \). Since any of these scenarios could have led to the configuration at the leaves of the tree, we must calculate the probability of each and sum them to obtain the probability \( L(\chi) \) for each character \( \chi \in C \). \[34\]

To illustrate the process, consider the aligned set of nucleotide sequences introduced in Figure 4-1 (a). Suppose we want to evaluate the likelihood of the unrooted tree shown in Figure 4-1 (b), which we root as shown in Figure 4-1 (c). Each interior vertex of the tree might possess any of the character states A, C, T, or G. Since the tree has two interior vertices, namely vertex (5) and vertex (6), there are \( 4 \cdot 4 = 16 \) possibilities to consider. The calculation of the likelihood for character \( \chi \) is illustrated schematically in Figure 4-1 (d).

Maximum likelihood calculates the probabilities based on the branch length of the model tree. In likelihood methods, branch lengths represent the expected number of character-state changes along a branch, or an edge of the tree. If a branch is short, there is a relatively low probability of a single change occurring along that particular branch, and an almost negligible probability of more than one change. We assume that changes along different branches are independent. Thus, the probability of any single scenario \((\tilde{\chi}(v_1), \tilde{\chi}(v_2), \ldots, \tilde{\chi}(v_k))\) is equal to the product of the probabilities of the
changes required by that scenario.\[34]\] For example, the probability of the scenario represented by the first term of Figure 4-1 (d) is equal to the probability that the character state at node (6) is an A (typically $1/4$ or the average frequency of A in the original sequence, depending on the type of model) times the probability of retaining an A along the edge $\{5,6\}$, times the probability of an A $\rightarrow$ C change along the exterior edge leading to leaf (1), and so on.

4.2 Models of Sequence Evolution

We now examine how the probabilities of the various changes are calculated. These probabilities depend on several assumptions about the process of nucleotide substitution, which define a substitution model. The models explored in this thesis are restricted to Markov models, in which the probability of change from state $i$ to state
at a given site does not depend on the history of the site prior to it having state \(i\).

We also assume that the substitution probabilities do not change in different parts of the tree, in other words, that the evolutionary mechanisms constitute a homogeneous Markov process. These assumptions are not necessarily biologically plausible; they are consequences of modeling substitutions as stochastic Markovian processes. [34]

The mathematical expression of a model is a table of rates of substitutions per site per unit of evolutionary distance. For DNA sequences, these rates are expressed as a \(4 \times 4\) instantaneous rate matrix \(Q\). Each element \(Q_{ij}\) represents the rate of change from character state \(i\) to state \(j\) during some time period \(dt\). The rows and columns of \(Q\) correspond to the bases A, C, G, and T in this order. The most general form of this matrix is

\[
\begin{pmatrix}
  - & \mu a \pi_C & \mu b \pi_G & \mu c \pi_T \\
  \mu g \pi_A & - & \mu d \pi_G & \mu e \pi_T \\
  \mu h \pi_A & \mu j \pi_C & - & \mu f \pi_T \\
  \mu i \pi_A & \mu k \pi_C & \mu l \pi_G & -
\end{pmatrix}
\]

where the diagonal elements are set to the negative of the sum of the off-diagonal elements in the corresponding row. The factor \(\mu\) represents the mean instantaneous substitution rate, and it is modified by the relative rate parameters \(a, b, c, \ldots, l\), which correspond to each possible transformation from one base to another. The product of the mean instantaneous substitution rate and a relative rate parameter constitutes a rate parameter. The remaining parameters \(\pi_A, \pi_C, \pi_G\) and \(\pi_T\) are called frequency parameters that correspond to the frequencies of the bases A, C, G, and T in the known set of leaves. [37] We assume these frequencies remain constant over time and that the rate of change to each base is proportional to the equilibrium frequency but independent of the identity of the starting base. The diagonal elements of \(Q\) are always chosen so that the elements in the corresponding row sum to zero. Almost all DNA substitution
models proposed are special cases of this matrix. [34] Analogous matrices can be
defined for protein sequence data, except they would have 20 states instead of 4.

Time-reversible models have the following rate parameter restrictions: \( g = a, h = b, i = c, j = d, k = e, \) and \( l = f \). Hence the matrix \( Q \) will be symmetric for this restriction. The most general time-reversible model, GTR, is then represented by

\[
Q = \begin{pmatrix}
- & \mu a \pi_C & \mu b \pi_G & \mu c \pi_T \\
\mu a \pi_A & - & \mu d \pi_G & \mu e \pi_T \\
\mu b \pi_A & \mu d \pi_C & - & \mu f \pi_T \\
\mu c \pi_A & \mu e \pi_C & \mu f \pi_G & -
\end{pmatrix}
\]

with diagonal elements set to the negative of the sum of the off-diagonal elements
in the corresponding row.[25] Most of the remaining models used for maximum likelihood

tree inference can be obtained by further restricting the parameters of matrix \( Q \) of

GTR. It is often desirable to reduce the number of free parameters, especially when
they are unknown and need to be estimated from the data. This can be achieved
by introducing constraints based on some appropriate symmetries. For example,
nucleotide substitutions fall into two major groups. Substitutions in which a purine is
exchanged for a pyrimidine or vice versa are called transversions. Pyrimidines are
the single-ringed nucleobases C (cytosine) and T (thymine), whereas purines are the
double-ringed nucleobases A (adenine) and G (guanine). The possible transversions are
\( A \leftrightarrow C, A \leftrightarrow T, C \leftrightarrow G, \) and \( G \leftrightarrow T \). All other substitutions fall under transitions. These

can also be separated into substitutions between purines, called purine transitions
\( (A \leftrightarrow G) \), and substitutions between pyrimidines, called pyrimidine transitions \( (C \leftrightarrow T) \).

For instance, the model of TrN [35] separates substitutions into transversions, purine

transitions, and pyrimidine transitions by requiring that \( a = c = d = f \). Similarly, we
can obtain Kimura’s three-substitution-type (K3ST) [24] model by requiring that all bases

occur in equal frequency \((\pi_A = \pi_A = \pi_G = \pi_T = 1/4)\) and dividing the substitution types
into transitions \((b = e)\), \(A \leftrightarrow T\) or \(C \leftrightarrow G\) transversions \((c = d)\), and \(A \leftrightarrow C\) or \(G \leftrightarrow T\) transversions \((a = f)\). Zharkikh [38] described a model (SYM) almost identical to GTR, except it assumes equal base frequencies.

Further restrictions on the parameters lead to additional popular models. The simplest model is the one proposed by Jukes and Cantor (JC) [21] in which all base frequencies are equal \((\pi_A = \pi_A = \pi_G = \pi_T = 1/4)\) and all substitutions occur at the same rate \((a = b = c = d = e = f = 1)\).

\[
Q = \begin{pmatrix}
-\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\
\frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\
\frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu \\
\frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu \\
\end{pmatrix}
\]

The base frequency and substitution rate are typically combined into a single parameter \(\alpha = \mu/4\) for simplicity to form

\[
Q = \begin{pmatrix}
-3\alpha & \alpha & \alpha & \alpha \\
\alpha & -3\alpha & \alpha & \alpha \\
\alpha & \alpha & -3\alpha & \alpha \\
\alpha & \alpha & \alpha & -3\alpha \\
\end{pmatrix}
\]

Kimura’s two-parameter model (K2P) [23] is called this way because it assigns different rates to transitions and transversions but keeps base frequencies equal. Thus, we set \(a = c = d = f = 1\) and \(b = e = k\). Letting the transition rate be \(\alpha = \mu k/4\) and the transversion rate be \(\beta = \mu/4\), the matrix \(Q\) becomes

\[
Q = \begin{pmatrix}
-\alpha - 2\beta & \beta & \alpha & \beta \\
\beta & -\alpha - 2\beta & \beta & \alpha \\
\alpha & \beta & -\alpha - 2\beta & \beta \\
\beta & \alpha & \beta & -\alpha - 2\beta \\
\end{pmatrix}
\]
Note that $k$ represents the transition bias, or the transition to transversion ratio. If $k = 1$, there is no preference between the two rates and the model becomes the JC model. Since there are twice as many kinds of transversions as transitions, the expected ratio is $k = 1/2$. Similarly, if $k = 4$ one would expect twice as many transitions as transversions. Other models, HKY85 ([18]) and Felsenstein’s F81 [9], are generalizations of K2P and JC models respectively by allowing for unequal base frequencies. Felsenstein used a different method (F84) [10] to accommodate unequal base frequencies in a two-parameter model. It has a general substitution rate for all types of substitutions and a within-group substitution rate only for transitions. This can be achieved by setting $a = c = d = f$, $b = (1 + K/\pi_R)$, and $e = (1 + K/\pi_Y)$, where $K$ is the parameter determining the transition to transversion ratio, $\pi_R = \pi_a + \pi_G$, $\pi_Y = \pi_C + \pi_T$, and the diagonal elements are set to the negative of the sum of the off-diagonal elements in the corresponding row. This model has two components for each transition, because transitions can occur due to either the general substitution rate or to the within-group rate. [34] The matrix is of the form

$$Q = \begin{pmatrix}
\mu \pi_C & \mu \pi_G (1 + K/\pi_R) & \mu \pi_T \\
\mu \pi_A & \mu \pi_G & \mu \pi_T (1 + K/\pi_Y) \\
\mu \pi_A (1 + K/\pi_R) & \mu \pi_C & \mu \pi_T \\
\mu \pi_A & \mu \pi_C (1 + K/\pi_Y) & \mu \pi_G & \mu \pi_T
\end{pmatrix}$$

### 4.3 Calculating Change Probabilities

The instantaneous rate matrix $Q$ specifies the rates of change between pairs of nucleotides per instant of time $dt$. Once $Q$ is calculated based on the chosen model, we need the probabilities of change from any state to any other along a branch of length $t$. The substitution probability matrix is calculated by

$$P(t) = e^{Qt}.$$
Its components are \( P_{ij}(t) \) satisfy the following conditions:

\[
\sum_{j=1}^{n} P_{ij}(t) = 1 \quad \text{and} \quad P_{ij}(t) > 0 \quad \text{for} \quad t > 0.
\]

Moreover, it also fulfills the requirement that

\[
P(t + s) = P(t) \cdot P(s) \tag{4-1}
\]

known as the Chapman-Kolmogorov equation, and the initial condition

\[
P_{ij} = \begin{cases} 
1, & \text{for} \quad i = j \\
0, & \text{for} \quad i \neq j
\end{cases}
\]

From the Chapman-Kolmogorov Equation (4–1), the forward and backward differential equations are obtained:

\[
\frac{d}{dt} P(t) = P(t)Q = QP(t).
\]

The exponential can be evaluated with the use of matrix algebra by decomposing \( Q \) into its eigenvalues and eigenvectors. [31]

Several methods allow for simple expressions for the eigenvalues, providing an analytic calculation of the substitution probability matrix \( P(t) \). For example,

\[
\text{JC:} \quad P_{ij}(t) = \begin{cases} 
\frac{1}{4} + \frac{3}{4}e^{-\mu t} & (i = j) \\
\frac{1}{4} - \frac{1}{4}e^{-\mu t} & (i \neq j)
\end{cases}
\]

\[
\text{K2P:} \quad P_{ij}(t) = \begin{cases} 
\frac{1}{4} + \frac{1}{4}e^{-\mu t} + \frac{1}{2}e^{-\mu t(\frac{i+j}{2})} (i = j) \\
\frac{1}{4} + \frac{1}{4}e^{-\mu t} - \frac{1}{2}e^{-\mu t(\frac{i+j}{2})} (i \neq j, \text{ transition}) \\
\frac{1}{4} - \frac{1}{4}e^{-\mu t} & (i \neq j, \text{ transversion})
\end{cases}
\]

[34]

The \( P_{ii}(t) \) entry of the substitution probability matrix is the probability of no substitution over an edge of length \( t \), whereas \( P_{ij}(t) \) is the probability of substitution
from character state $i$ to $j$ along an edge of length $t$. This is the missing element to calculate the likelihood of a tree as presented in section 4.1.

### 4.4 Differences in Perspective between Parsimony and Likelihood

The parsimony and likelihood approaches to inferring a phylogeny have some common characteristics. They are both criterion-based methods and are able to rank and evaluate the observed trees. Also, the cost of a given change under parsimony is analogous to the likelihood of the given change from the substitution matrix $P(t)$. In parsimony, the cost of placing a given state at an internal node is the sum of the costs of deriving both of the daughter trees from that state, whereas the likelihood of an ancestral state is the product of the likelihoods of the state giving rise to the daughter trees. In parsimony, the total cost of the tree is the sum of the costs at each position. Similarly, the net log-likelihood of a tree is the sum of the log-likelihoods of the evolution at each sequence position. [34]

However, there are some essential differences between the two methods. The cost of a change in parsimony is not a function of branch length, unlike maximum likelihood. For this reason, even though several minimum extensions of a character may be equally ranked based on their parsimony score, some may have scored higher than others under likelihood because of the branch lengths. Maximum parsimony looks only at the single, lowest cost solution, whereas maximum likelihood looks at the combined likelihood for all solutions (ancestral states) consistent with the tree and the branch lengths. Another obvious difference maximum parsimony does not depend on a precise model as maximum likelihood does. The initial tree model needed for likelihood is based on a set of stated assumptions, while methods like maximum parsimony do not require one, which makes their assumptions implicit. Sometimes, "model-free" assumptions may be more likely to violate their implicit assumptions since they are not as obvious as the methods with more explicit assumptions. Having some idea of the phylogeny is relevant
to the development of good models, but ever-improving models can also lead to better phylogenetic inferences. Thus, both types of methods are useful and important. [34]

Even though the concept of maximum parsimony is based on Ockham's Razor, it is questionable whether the method, as implemented in phylogenetics, actually fulfills Ockham's razor. Since there are simple stochastic models that assume all characters have similar probabilities of change, arguably a more parsimonious situation, where maximum parsimony is inconsistent, it could be argued that maximum likelihood is actually more parsimonious.

In the undergraduate research, maximum likelihood methods were used to determine the best-fit model of evolution as well as to describe the phylogenetic trees estimated for all loci. Next, a computer program was used to estimate the expected branch lengths for the morphological and molecular data using the estimated branch lengths for all loci from PAUP*. The likelihood of the observed data was calculated using the Poisson model (in which case there is no variance inflation parameter) and the negative binomial model (where a variance inflation parameter is taken into account). The expectation given a universal rate of evolution for both molecules and morphology would be that the maximum likelihood estimates of the variance inflation parameter for the morphological data would fall within the diversity of variance inflation parameter estimates for the molecular data. The results of the undergraduate research suggested that, although the estimate for variance inflation parameter for morphology is high, it does fall within the range of estimates of the variance inflation parameter for molecular data. In conclusion, it does appear that the rates of molecular and morphological rates of evolution are correlated, at least within the galiforms.[20]
REFERENCES


BIOGRAPHICAL SKETCH

Naomi Iuhasz was born and grew up in Resita, Romania. She graduated from "Traian Lalescu" Theoretical High School in 2002 with a major in Computer Science. She then moved to Florida where she earned her Bachelor of Arts degree in mathematics from the University of Florida (UF) in 2008. Upon graduating in August 2008, she continued her graduate studies in the Department of Mathematics of the University of Florida earning a Master of Science degree in applied mathematics in May 2010.