Phylogenetic Tree Inferences Using Quartet Splits

By

Kevin Michael Hathcock

Bachelor of Science
Lenoir-Rhyne University 2010

Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in

Mathematics

College of Arts and Sciences

University of South Carolina

2012

Accepted by:

Éva Czabarka, Director of Thesis

Anton Schep, Reader

Lincoln Lu, Reader

Lacy Ford, Vice Provost and Dean of Graduate Studies

© Copyright by Kevin Michael Hathcock, 2012

All Rights Reserved.

Dedication

This work is dedicated to my beautiful fiancé Amanda. Your support has helped make this thesis possible. You inherited much more than your fair share of the stress and pressure that came along with writing it. I truly could not have done this without you. I love you.

Acknowledgments

Firstly, I would like to thank my advisor Dr. Éva Czabarka for her guidance and advice. Your help during the writing of this thesis cannot be overstated. This thesis would not have been possible if not for you. I would also like to thank my committee members Dr. Anton Schep and Dr. Lincoln Lu for your thoughtful suggestions, as well and your time and attention. I express my appreciation to Dr. George McNulty for making the process of writing a thesis much simpler than I ever could have.

I would also like to thank the members of the math department at the University of South Carolina. Special thanks to Kenny Brown, Virginia Johnson, Travis Johnston and Michael Lane for your help with formating and LaTeX. Your insight saved me many hours of extra work.

Last, but certainly not least, I would like to express my immense gratitude to my family, especially my mother, Kimberly Hathcock, my brother, Kane Hathcock, and my wonderful fiancé, and soon-to-be-wife, Amanda Huber. I could not have done this without your love and support. Your encouragement is directly responsible for my success. Thank you all.

Abstract

Phylogenetic trees are used by biologists and geneticists as a way of classifying the relationships between different taxonomic units. The branching of a tree represents one unit evolving into two or more different units. The leaves of phylogenetic trees are labeled with the units that are to be studied.

For any tree, and phylogenetic trees in particular, deleting an edge results in a bipartion of the leaf label set. We call such deletions splits and represent the split that results in bipartion sets A and B by $A|B$. A split σ is realized by a tree T if there is an edge deletion of T that results in the bipartition expressed by σ. The set of all splits realized by T is denoted $\Sigma(T)$. We discuss what it means for two splits to be compatible and give a proof of the Splits-Equivalence Theorem which states that for a collection Σ of splits, there is a tree T such that $\Sigma = \Sigma(T)$ if and only if the splits of Σ are pairwise compatible.

Of interest are splits in which both $|A|$ and $|B| = 2$. These are referred to as quartets. We say that a set $\mathbf{A}$ of quartets infers a quartet s if every tree that displays $\mathbf{A}$ must also display s. Such an inference is called a k-ary inference when $|A| = k$. A k-ary inference is called primitive, if it can not be derived from lower-order inferences. In his Master's Thesis *Reconstruction Methods for Derivation Trees*, Dekker characterized all primitive binary and ternary inferences. Steele and Bryant showed that for any positive integer k there are primitive k-ary inferences. We present an independent recreation of a portion of Dekker's work by listing all primitive binary quartet split inferences, as well as show that the only primitive ternary inferences involve < 8 leaves. We also provide an example of a valid ternary inference.

Contents

List of Figures

Chapter 1

Preliminaries

1.1 Introduction

Biologists use phylogenetic (evolutionary) trees as a visual representation of evolutionary events and the relationships between different taxonomic units (species, genus, family, etc). When one taxonomic unit splits and forms two or more new units, it is referred to as a speciation event. Biologists represent speciation events using internal vertices of a phylogenetic tree. The idea to represent speciation events using trees began with Charles Darwin as early as 1837 [6, 7]. Originally, the evolutionary "closeness" of different species was determined using physical comparisons but it was discovered that the appearance of similar traits did not always correlate with an evolutionary ancestor as well as might be expected. For example, based solely on appearance, it would seem that whales are more closely related to sharks than to hippopotamuses but this is not the case [3]. It is also difficult to distinguish between species that are extremely similar. In some cases, the only way to accurately distinguish between members of different species is by using DNA testing [13]. With the scientific advancements that came in the second half of the twentieth century, including the use of DNA and the field of genomics, biologists were able to get a much better picture of how different species are related and where to place them on phylogenetic trees [1]. With time, these abilities will only improve.

A natural method for creating a large phylogenetic tree is to construct smaller trees using only subsets of the taxonomic units and then attempt to merge these smaller trees together [4, 14]. In some cases however, there can be several different ways to merge these small trees together. When this is the case, it is a necessary endeavor to determine when these separate representations are compatible and can be combined into one comprehensive tree.

While the use of phylogenetic trees is by far the dominant method for representing evolutionary history, the methods used to place the different taxa on the trees are not without complications. The majority of biologic relationships are determined using genetic material. The chief means of transferring genetic information is referred to as vertical gene transfer. Vertical transfer occurs when genetic material is passed from parent to offspring. In multicellular organisms this is, with very few exceptions, the only means of transfer. However, single-celled organisms, in particular bacteria, have many methods of what is referred to as horizontal gene transfer [11]. Horizontal transfer occurs when genetic material is spread in a manner that is not from parent to offspring [9]. Bacteria can absorb and express foreign genetic materials. This is the primary cause of antibiotic resistance in bacteria [5]. The process of transduction involves bacteria and viruses [15]. Some viruses, called bacteriophages, have the ability to absorb bacterial DNA from a host organism and transfer that DNA to another bacterium, that may not be a close relative of the host [5]. Some bacteria also have the ability to transfer genetic material directly by means of cell-to-cell contact. In these ways, essentially unrelated organisms can display the same genes.

Among more complex organisms, there are also several difficulties that result from using genetic materials to infer evolutionary relationships. Hybridization can occur throughout several different taxonomic levels. Interbreeding between different subspecies within the same species is fairly common in nature: an example being mating between a Bengal tiger and a Siberian tiger [18]. Hybrids can be formed between different species within the same genus. Mules are the result of mating between a donkey and a horse. Interspecies hybridization has also been found among different species of yaks and squirrels [16]. Although rare, hybrids may also form between members of two different genera within the same family. While in general, the farther apart the two organisms are genetically, the lower the chances of a successful hybrid, such intergeneric hybrids do occur. The Leyland cypress is a result of the crossing of a Monterey cypress with an Alaska cypress [19]. In many cases, especially with organisms that are not closely related, these hybrids tend to be unable to reproduce, but this is not always the case [7].

Convergent evolution is another problem when considering the evolution of complex organisms. Plants, animals, fungi have so many traits, it is not out of the realm of possibility that some similar traits may evolve independently in rather unrelated taxa. When lineages that are not closely related develop similar traits independently, it is called convergent evolution [12]. A prime example involves birds and bats. The most recent common ancestor of birds and bats did not have wings, yet both have developed them over time. Convergent evolution can also occur at the DNA level with different organisms developing similar enzymes and proteins that display as similar traits independently of each other [10].

While genetic materials and information are readily available for living organisms, biologists must gather information about extinct species using the fossil record. This clearly results in a large amount of lost information. It is evident that some organisms that have become extinct, either were not fossilized or have yet to be discovered. Even among the organisms that have been discovered, only hard tissues are abundantly fossilized; although advancements are allowing for the inverstigation of some soft tissues [2]. The farther back in time an organism lived, the less information can be gained from the remains. Longer-extinct species tend to have more ambiguous evolutionary histories.

Even with these difficulties, the field of phylogenetics has the ability to straighten the tangled evolutionary history of the organisms on Earth. In this thesis, we intend to demonstrate how the idea of phylogenetic trees can be used to help visualize the relationships that biologists discover through the use of DNA sequencing and genomics. We begin by defining some fundamental ideas in graph theory. We then explore the idea of splits in a tree and ways in which these splits can be used to reconstruct the tree. Two important types of splits (quartets and rooted triples) are expounded upon separately in Chapter 3. Chapters 4 and 5 demonstrate how one can use an incomplete set of quartet splits to infer other compatible quartets.

1.2 Fundamental Definitions

As with any subject, in order to understand the complex, one must have a grasp of the basic. We give several essential definitions in this section. For readers who desire a more in depth introduction, see Introduction to Graph Theory by Trudeau [17].

Definition 1.1. A tree is a connected graph that contains no cycles.

Definition 1.2. A tree in which some vertices (including every vertex of degree one or two) are labeled with disjoint subsets of a set X is called an X-tree. Note that vertices of degree greater than two may or may not be labeled.

Example 1.3. Figure 1.1 below displays an X-tree where $X = \{1, \ldots, 25\}$. Note that there are some vertices of degree greater than 2 that are labeled and that some vertices are labeled by single elements of X.

Definition 1.4. A tree that has no vertices of degree two and has each leaf uniquely labelled by a singleton subset of the set X is called a phylogenetic tree or a phylogenetic X-tree.

Definition 1.5. A rooted phylogenetic tree is a pylogenetic tree that has an internal vertex that is distinguished and is called the root. The root is denoted by ρ and may have degree two. All internal vertices other than the root must have degree greater

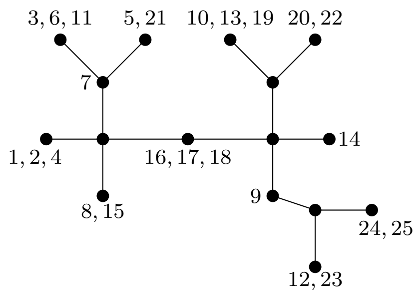


Figure 1.1: An X-tree with $X = \{1, \ldots, 25\}$

than or equal to three. In this thesis the term "phylogenetic tree" will refer to the unrooted case.

Definition 1.6. A tree in which every internal vertex has degree three is called a binary tree. A binary phylogenetic tree and a rooted binary phylogenetic tree, shown in Figure 1.2, are defined similarly. In a rooted binary phylogenetic tree however, the root still has degree two.

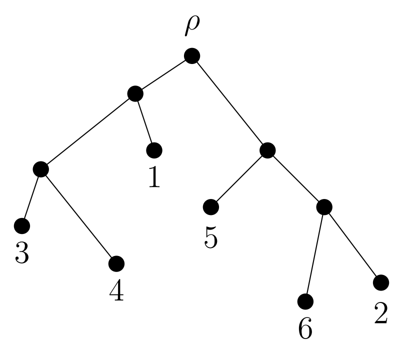


Figure 1.2: A rooted binary phylogenetic tree with label set $\{1, \ldots, 6, \rho\}$

Definition 1.7. Let the label set of an X-tree T be denoted $\mathcal{L}(T)$. The deletion of any edge of T results in exactly two smaller subtrees T_1 and T_2 of T. This deletion also partitions $\mathcal{L}(T)$ into two subsets $\mathcal{L}(T_1)$ and $\mathcal{L}(T_2)$ where $\mathcal{L}(T_1) \neq \emptyset$, $\mathcal{L}(T_2) \neq \emptyset$ and $\mathcal{L}(T_1) \dot{\cup} \mathcal{L}(T_2) = \mathcal{L}(T)$. We call this partition of $\mathcal{L}(T)$ a split of T. Splits are sometimes called X-splits. We will denote by $\Sigma(T)$, the collection of all the splits of T formed in this fashion. We call $\Sigma(T)$ the split set of T. The split that results in the partition of $\mathcal{L}(T)$ into sets A and B will be denoted $A|B$. Note that the split $A|B$ is the same as the split $B|A$.

Example 1.8. Consider the tree below in Figure 1.3. The split corresponding to deleting edge e_1 is $\{1,2\}|\{3,4,5,6\}$ while the split corresponding to edge e_2 is $\{1,2,3,4\}|\{5,6\}$.

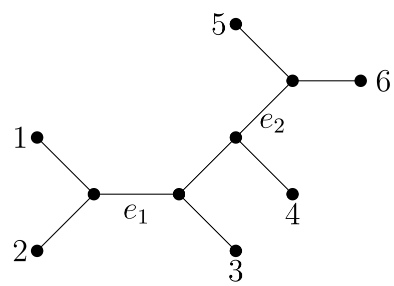


Figure 1.3: A phylogenetic tree

Definition 1.9. Let T be a tree and e be an edge of T with end vertices u and v. The new tree $T' = T/e$ is the tree formed by the contraction of e. The tree is formed by replacing vertices u and v with the single vertex ν such that $N(\nu) = (N(u) \setminus \{v\}) \cup (N(v) \setminus \{u\})$.

Example 1.10. The trees in Figure 1.4 demonstrate an edge contraction. The tree on the right has had the edge e contracted. It is important to notice that contracting an edge, removes a split from the tree. When considering phylogenetic trees, contracting an edge results in a loss of resolution. It takes two separate speciation events and joins them into one. It is then impossible to distinguish which event happened first.

For biologists, the best case scenario for representing evolutionary data would be a rooted binary phylogenetic tree. This tree would have a root that corresponds to the common ancestor from which all of the leaves evolved. Each internal vertex of the

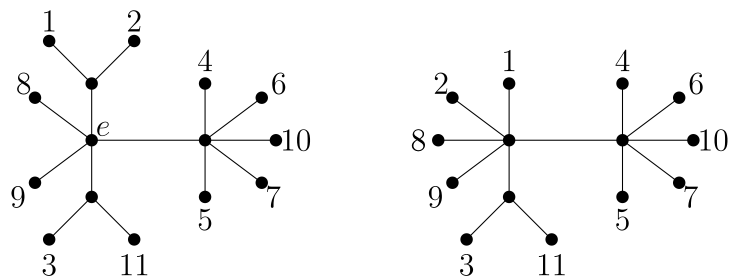


Figure 1.4: A tree T and the tree T/e after edge contraction

tree would have degree three and would correspond to the creation of exactly two new taxanomic units. A tree that displays all the information possible for its set of leaf labels is said to be "fully resolved." Since normally the information collected about the evolutionary relationships among a given set of taxa is not perfect, it cannot be expected that all data sets result in fully resolved trees. When a tree is not fully resolved, there are vertices of degree 4 or higher. This corresponds to a speciation event that resulted in 3 or more new taxa. Since this is rarely the case in nature, it is highly likely that this vertex of degree 4 represents two or more seperate speciation events whose order cannot be determined.

Chapter 2

Splits

This chapter will discuss splits on the label sets of phylogenetic trees. Definitions will include: split, split set, compatibility of splits, compatibility of trees, edge contraction and induced subtree. Theorems will include: Splits-Equivalence Theorem.

Definition 2.1. Let T be an unrooted phylogenetic tree and A be a subset of $\mathcal{L}(T)$. Let $T(A)$ be the minimal subtree of T that includes each element of A. Suppressing all vertices of $T(A)$ with degree two results in the subtree of T induced by A which is denoted $T_{|A}$. If T is a rooted phylogenetic tree, then distinguish the vertex of $T(A)$ that was closest to the root and suppress any other vertices of degree two in order the form $T|_A$. Induced subtrees are called restricted trees in Semple-Steele.

Definition 2.2. Let T and S be phylogenteic trees. Then we say that T is compatible with S if S can be formed by contractions of an induced subtree of T or S is an induced subtree of a contraction of T. We denote that T is compatible with S by $S \trianglelefteq T$.

Proposition 2.3. The relation $\trianglelefteq$ is a partial order

Proof. Let T be a tree. Clearly $T \trianglelefteq T$ since the induced subtree $T|_{\mathcal{L}(T)} = T$ along with the empty set of contractions again results in T.

Now suppose that $S \trianglelefteq T$ and $T \trianglelefteq S$. Then $\mathcal{L}(S) \subset \mathcal{L}(T)$ and $\mathcal{L}(T) \subset \mathcal{L}(S)$, hence $\mathcal{L}(S) = \mathcal{L}(T)$. Also, we know that S is T with all degree two vertices suppressed. But, since T was a phylogenetic tree, it had no degree two vertices from the beginning. So we see that $S = T$.

Now suppose that for phylogenetic trees S, T, U we know that $S \trianglelefteq T$ and $T \trianglelefteq U$.

Then $\mathfrak{L}(S) \subset \mathfrak{L}(T) \subset \mathfrak{L}(U)$. Now, T is formed from contractions of edges from U and S is formed from contractions of T. Hence S is formed from contractions of U and we see that $S \trianglelefteq U$. So we have shown that $\trianglelefteq$ is in fact a partial order.

□

The principle result of this section will be a proof of the Splits-Equivalence Theorem first proved by Buneman in 1971. We will first discuss several definitions and lemmata to aid in the proof of the theorem.

Definition 2.4. The pair of splits $A_1|B_1$ and $A_2|B_2$ are said to be compatible if any of the following intersections is empty

$$A_1 \cap A_2, \quad A_1 \cap B_2, \quad A_2 \cap B_1, \quad B_1 \cap B_2.$$

If a set Σ of splits is such that each pair of splits in Σ are compatible, we say that Σ is consistent.

Definition 2.5. A split of the form $A|B$ where $\min\{|A|, |B|\} = 1$ is called a trivial split. For a finite set X, a trivial split is of the form $\{x\}|\{X \setminus \{x\}\}$ where x is an element of X.

Proposition 2.6. A trivial split of a set X is compatible with every X-split.

Proof. Let $x \in X$. Then $\{x\}|\{X \setminus \{x\}\}$ is the trivial split corresponding to x. Let $A|B$ be another X-split. Notice that either $x \in A$ or $x \in B$ but not both. Without loss of generality, say $x \in A$. Then for the two splits $\{x\}|\{X \setminus \{x\}\}$ and $A|B$ we see that $\{x\} \cap B = \emptyset$ and from Definition 2.4 we know that these two splits are compatible. □

Lemma 2.7. *T is a phylogenetic $X-$tree if and only if T displays the set $\Sigma_{triv}(X)$ of trivial splits of X.*

Proof. ($\Rightarrow$) Let T be a phylogenetic $X-$tree for some finite set X. From the definition of a phylogenetic tree (Definition 1.4) we know that every leaf is labeled uniquely by a single element of a set X. The splits corresponding to deleting the pendant edges of T are precisely the trivial splits $\Sigma_{triv}(X)$.

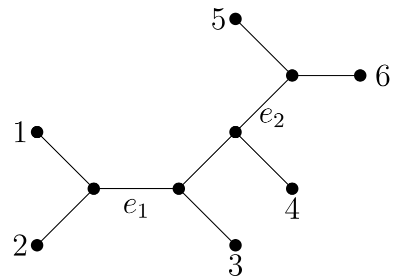
($\Leftarrow$) Let T be an $X-$tree that displays $\Sigma_{triv}(X)$. Then for each $x_i \in X$ there is an edge e_i such that $T - e_i$ results in the split $x_i|\{X \setminus \{x_i\}\}$. Since x_i is split from the rest of X it is clear that some vertex of T is labeled by the singleton set $\{x_i\}$. We need to show that this vertex is a leaf. Suppose not. Then this is an internal vertex and it corresponds to a branching of T. This must then lead to one or more leaves of T. But since T is an $X-$tree, all vertices of degree one (leaves) must be labeled. Suppose L is the set of labels from these leaves. Then $T - e_i$ would result in the split $\{\{x_i\} \cup L\}|\{X \setminus \{\{\{x_i\} \cup L\}\}$ which is a contradiction. So we see that there is a leaf that is labeled by the singleton set $\{x_i\}$ for every $x_i \in X$ and therefore T is a phylogenetic $X-$tree. □

Lemma 2.8. *(Robinson and Foulds 1981) Let T be an X-tree and let σ_1 and σ_2 be elements of $\Sigma(T)$ such that $\sigma_1 \neq \sigma_2$. Then X can be partitioned into three sets X_1, X_2, X_3 such that $\sigma_1 = X_1|(X_2 \cup X_3)$ and $\sigma_2 = (X_1 \cup X_2)|X_3$. Furthermore, $X_1 \cap X_3 = \emptyset$.*

Proof. Let T be an X-tree and $\sigma_1, \sigma_2 \in \Sigma(T)$. By definition, σ_1 corresponds to $T \setminus e_1$ where $e_1 = \{u_1, v_1\}$ is an edge of T connecting vertices u_1 and v_1. Similarly, σ_2 corresponds to $T \setminus e_2$ with $e_2 = \{u_2, v_2\}$. Since $\sigma_1 \neq \sigma_2$ we see that e_1 and e_2 are also distinct. Since T is a tree, there exists a unique path from u_1 to u_2 in T. Note that $u_1 \neq u_2$ but it may be the case that $v_1 = v_2$. Either way, we can see that X can be divided into three subsets, X_1, X_2, X_3 as follows. Consider the components of $T \setminus \{e_1, e_2\}$. Let C_1 be the component that includes u_1, C_2 be the component that includes v_1 and C_3 be the component that includes u_2. Setting $\mathcal{L}(C_i) = X_i$, we have found the desired sets.

□

Example 2.9. Consider the tree from Figure 1.3, that is reproduced below. Let σ_1 bet the split $T \setminus e_1$ and σ_2 be $T \setminus e_2$. Defining $X_1 = \{1, 2\}$, $X_2 = \{3, 4\}$, and $X_3 = \{5, 6\}$ results in the partition described above. We see that $\sigma_1 = X_1|(X_2 \cup X_3)$ and $\sigma_2 = (X_1 \cup X_2)|X_3$. Also, $X_1 \cap X_3 = \emptyset$.



Now consider any tree T. Let f be a function from a finite set X into the vertex set $V(T)$. In other words, each vertex of T is labeled by one or more elements from X and thus T is an X-tree. Color the elements of X either red or blue. We now color the vertices of the tree based on this coloring of X in the following way. Let $v \in V(T)$ be an element of $f(X)$. If every element of $f^{-1}(v)$ is the same color, then assign this color to v. If there are elements in $f^{-1}(v)$ of both colors, then color v red and blue. This coloring of the vertex set is referred to as the *coloring of V induced by f*. We say that a subgraph T' of T is *monochromatic* if all of the vertices in $V(T')$ have the same color.

Lemma 2.10. *Let T be a tree and f be a function from a finite set X into the vertices of T. Consider the coloring of $V(T)$ induced by f as described above. Now, suppose that for each edge $e \in E(T)$ that precisely one component of $T \setminus e$ is monochromatic in the induced coloring. Then there exists a unique vertex $v \in V$ such that every component of $T \setminus v$ is monochromatic.*

Proof. First we prove the existence of such a vertex. Begin by assigning an orientation to each edge of T away from the monochromatic component of $T \setminus e$. Call this oriented graph $\vec{T}$. With this orientation, there must be a vertex $v \in V(T)$ such that the out-degree of v is 0. If this were not the case then every vertex would have at least one edge directed away from it and we could construct an infinite directed path in $\vec{T}$. Now, if we consider $T \setminus v$ we see that every component is monochromatic. Thus we have found a vertex that satisfies the proposition. Now we will show that this vertex in unique. Suppose this is not the case. Then there are two distinct vertices $v, v' \in V(T)$ such that v and v' both have the desired property. Since T is a tree, there is a unique path P that connects v to v'. Since $v \neq v'$ this path P is at least one edge long. Select any edge e of P. Then from the statement of the proposition, exactly one component of $T \setminus e$ is not monochromatic. Assume without loss of generality that this is the component containing vertex v. This leads to a contradiction. We assumed that every component of $T \setminus v'$ was monochromatic but we have just shown that the component containing vertex v is not. So we have shown that such a vertex exists and is unique. □

Lemma 2.11. *Let $A|B$ be an X-split and let T be an X-tree so that $A|B$ is not a split of T but $A|B$ is compatible with with every split in $\Sigma(T)$. Then there is a unique vertex v of T such that, for each component T_i of $T \setminus v$ either $\mathcal{L}(T_i) \subset A$ or $\mathcal{L}(T_i) \subset B$.*

Proof. Let $f : X \to V(T)$ be the function that assigns the labels to the vertices of T. Color the elements of A red and the elements of B blue and consider the coloring of $V(T)$ induced by f. Now suppose that $A_1|B_1$ is a split of T. Then since $A_1|B_1$ is compatible with $A|B$ we know that one of the following intersections is empty

$$A \cap A_1, \quad A \cap B_1, \quad B \cap B_1, \quad B \cap A_1.$$

Without loss of generality, suppose that $B \cap B_1 = \emptyset$. Then, since A and B partition $V(T)$ we can see that $B_1 \subset A$ and hence B_1 is red. Since the split $A_1|B_1$ corresponds to deleting some edge e_1 of T, we know that $T \setminus e_1$ has a monochromatic component. If A_1 were a subset of B and B_1 were a subset of A, then it would be the case that $A|B = A_1|B_1$ which would contradict the assumption that $A|B \notin \Sigma(T)$. Now we see that $T \setminus e_1$ has exactly one monochromatic component for each $e \in E(T)$ and we can appeal to Lemma 2.10. So we know that there exists a unique vertex $v \in V(T)$ such that every component of $T \setminus v$ is monochromatic. In other words, each component T_i of $T \setminus v$ is either completely red or completely blue and hence $\mathcal{L}(T_i) \subset A$ or $\mathcal{L}(T_i) \subset B$. □

We will now prove the Splits-Equivalence Theorem as stated below.

Theorem 2.12. ***(Splits-Equivalence Theorem)*** *Let Σ be a collection of X-splits. Then, there is an X-tree T such that $\Sigma = \Sigma(T)$ if and only if the splits of Σ are pairwise compatible. Furthermore, if such an X-tree T exists, then up to isomporphism, T is unique.*

Proof. Suppose that Σ is a collection of X-splits induced by the edges of some X-tree. Let σ_1 and σ_2 be distinct elements of the split set Σ. We know from Lemma 2.8 that there must be a partition of X into sets X_1, X_2, X_3 such that $\sigma_1 = X_1|(X_2 \cup X_3)$ and $\sigma_2 = (X_1 \cup X_2)|X_3$. Since $X_1 \cap X_3 = \emptyset$, we can see from Definition 2.4 that the splits σ_1 and σ_2 are compatible. Therefore all the splits of Σ must be pairwise compatible. Now, suppose that Σ is a collection of pairwise compatible X-splits. We will show by induction on the cardinality of Σ to show that $\Sigma = \Sigma(T)$ for some tree T and also that, up to isomorphism, this tree T is unique. **Base Case:** Suppose that $|\Sigma| = 0$. Then, up to isomorphism, there is a unique X-tree T for which $\Sigma = \Sigma(T)$. In this case, we see that T the tree consisting of one vertex labeled by all of X. **Induction Step:** Now suppose that $|\Sigma| = k + 1$ where $k \geq 0$ and that the theorem is true for

$|\Sigma| = k$. Now let $A|B$ be an element of Σ. Since Σ is a set of pairwise compatible X-splits we see that $\Sigma - A|B$ must be also. So by the induction hypothesis, there exists, up to isomorphism, a unique tree T' where $\Sigma - A|B = \Sigma(T')$. Let the function $f' : X \to V(T')$ label the vertices of T'. Using Lemma 2.11, we see that there is a vertex $v' \in V(T')$ such that for every component T'_i of $T' \setminus v'$, either $\mathcal{L}(T'_i) \subset A$ or $\mathcal{L}(T'_i) \subset B$. We now create a tree T in the following manner. Replace v' in T' by two new vertices v_A and v_B so that $\{v_A, v_B\} \in E(T)$. Now attach the vertices of $f'(A)$ to the new vertex v_A and attach the vertices of $f'(B)$ to v_B. Now we define a new function f as follows

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

This new function f labels the vertices of our new tree T. It is clear that T is an X-tree. Any split of T' is also a split of T and also, by construction $T \setminus \{v_A, v_B\} = A|B$. So we have constructed a tree T so that $\Sigma = \Sigma(T)$. By the induction hypothesis, we know that T' is the only tree, up to isomporphism, that displays $\Sigma - A|B$. It follows then, that, up to isomorphism, T is the only tree that has split set Σ. $\square$

Chapter 3

Quartets and Rooted Triples

3.1 Quartets

This chapter will discuss quartet splits and rooted triples. Definitions will include: quartet, rooted triple, span, quartet set and rooted triple set. Proofs of several lemmas on tree compatibility will be given.

Definition 3.1. An unrooted binary tree with four leaves is called a quartet. We will denote the quartet with leaf pairs $\{a, b\}$ and $\{c, d\}$ joined by a single internal edge as $ab|cd$. We use this notation because deleting the single non-pendant edge results in the splitting of the two vertex pairs. For this reason, quartets are sometimes referred to as quartet splits.

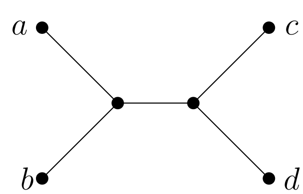


Figure 3.1: The quartet $ab|cd$

Definition 3.2. For a set Q of quartets, the span of Q, which we denote $\langle Q \rangle$, is the set of all unrooted trees that are compatible with each quartet in Q and have leaves labelled by $\mathcal{L}(Q)$.

Definition 3.3. For an unrooted tree T the quartet set $q(T)$ is the set of all quartets that are induced subtrees of T. That is, the set of quartets that can be formed by contracting edges of induced subtrees of T.

3.2 Rooted Triples

This section will discuss rooted triples

Definition 3.4. A rooted binary tree with exactly three leaves is called a rooted triple. We will denote the rooted triple with leaf set $\{a, b\}$ connected by the root the leaf c by $ab|c$.

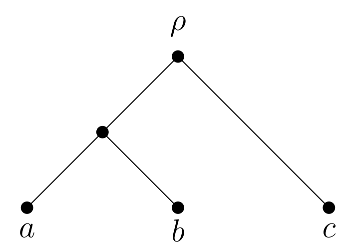


Figure 3.2: The rooted triple $ab|c$

Definition 3.5. For a set R of rooted triples, the span of R, which we denote $\langle R \rangle$ is the set of all rooted trees that are compatible with each rooted triple in R and have leaves labeled by $\mathcal{L}(R)$.

Definition 3.6. For a rooted tree T the rooted triple set $r(T)$ is the set of all rooted triples that are induced subtrees of T. That is, the set of rooted triples that can be formed by contracting edges of induced subtrees of T.

Recall from definition 2.2 that for two trees S and T, we say $S \trianglelefteq T$ if S is formed by contractions of an induced subtree of T or if S is an induced subtree of contractions of T. We showed in proposition 2.3 that $\trianglelefteq$ is a partial order. Below, in Theorem 3.2 we give a way to determine, for trees S and T, if T is compatible with S, i.e. $S \trianglelefteq T$.

Lemma 3.7. *For a phylogenetic tree S, if $\lambda|\bar{\lambda}$ is a split of S, with* $\min\{|\lambda|, |\bar{\lambda}|\} \geq 2$ *then $ab|cd$ is a quartet split of S for all pairs $a, b \in \lambda$ and $c, d \in \bar{\lambda}$.*

Proof. Since $\lambda|\bar{\lambda}$ is a split of S, there is an edge $e \in S$ so that $S - e$ produces the split $\lambda|\bar{\lambda}$ where $\lambda \cap \bar{\lambda} = \emptyset$ and $\lambda \cup \bar{\lambda} = \mathcal{L}(S)$. Let $a, b \in \lambda$ and $c, d \in \bar{\lambda}$, then clearly, $S - e$ results in $ab|\bar{\lambda}$ and $\lambda|cd$. Therefore $S - e$ results in the quartet split $ab|cd$ and we see that $ab|cd \in q(S)$. □

Lemma 3.8. *If T is a phylogenetic tree and $ab|cd \in q(T)$ for all pairs $\{a, b\} \in \lambda$ and $\{c, d\} \in \bar{\lambda}$, then $\lambda|\bar{\lambda}$ is a split of T.*

Proof. Let T be a phylogenetic tree. Let a, a' be elements of λ and let u_0 be the vertex at which the tree splits into the separate branches leading to a and a'. Here we use induction on $|\bar{\lambda}|$. **Base Case:** If $|\bar{\lambda}| = 2$ then clearly $\lambda|\bar{\lambda} \in \Sigma(T)$. **Induction Step:** Assume that $|\bar{\lambda}| = n$ and the statement is true for all sets of size $< n$. Let $x_0 \in \bar{\lambda}$. Then, by the induction hypothesis, the split $aa'|\{\bar{\lambda} \setminus \{x_0\}\}$ is in $\Sigma(T)$ since this set is of size one less than $\bar{\lambda}$. This means that there is an edge $e_0 = \{u_0, v_0\}$ so that $T \setminus e_0$ produces the aforementioned split. Also, for each $x_i \in \bar{\lambda}$ there is an edge e_i whose removal results in $aa'|x_0x_i$. Consider one such edge e_1. Suppose that $e_0 \neq e_1$. Then $e_1 = \{u_0, v_1\}$. But now, u_0—v_1—x_1—v_0—u_0 is a cycle inside T (shown in Figure 3.3) which is a contradiction since T is a tree. So we see that for all i, it must be that $e_i = e_0$ and that removing e_0 from T results in $aa'|\bar{\lambda}$. There must be such an edge for each pair of vertices in λ and a similar argument shows that $\lambda|\bar{\lambda}$ must be a split of T. □

Lemma 3.9. *For phylogenetic trees S and T on the same label set $\mathcal{L}$, if $\Sigma(S) \subset \Sigma(T)$, then $S \trianglelefteq T$.*

Proof. Let T and S be phylogenetic trees on the same label set $\mathcal{L}$ such that $\Sigma(S) \subset \Sigma(T)$. For each edge $e_i \in E(S)$, $S - e_i$ results in some split $\alpha|\beta$. Since $\Sigma(S) \subset \Sigma(T)$, the split $\alpha|\beta$ must correspond to $T - e_i'$ for some edge $e_i' \in E(T)$. However, there may be splits in $\Sigma(T)$ that have no corresponding edge in $E(S)$. So we can see that S is formed by contracting exactly these edges of T. So S is formed by contractiosn on the edges of the induced tree $T|_{\mathcal{L}}$ and therefore $S \trianglelefteq T$. $\square$

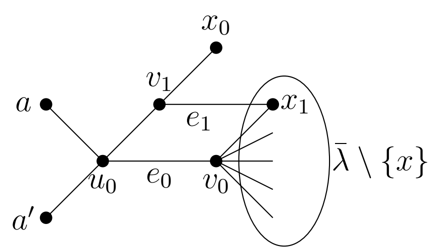


Figure 3.3: Contradictory cycle showing that $aa'|\bar{\lambda} \in \Sigma(T)$

Theorem 3.10. *Let S and T be unrooted phylogenetic trees. T is compatible with S (i.e. $S \trianglelefteq T$) if and only if $q(S) \subset q(T)$ and $\mathcal{L}(S) \subset \mathcal{L}(T)$.*

Proof. ($\Rightarrow$) Assume that $S \trianglelefteq T$. If the split $ab|cd$ is in $q(S)$, then surely S is compatible with the quartet with leaf sets $\{a, b\}$ and $\{c, d\}$ since the quartet can be formed by contracting edges of $S|_{\{a,b,c,d\}}$. We showed in Proposition 2.3 that $\trianglelefteq$ is transitive and so the quartet $ab|cd \trianglelefteq S$ and $S \trianglelefteq T$ implies that $ab|cd \trianglelefteq T$. Therefore $ab|cd \in q(T)$. Since $S \trianglelefteq T$ by assumption, it is clear that $\mathcal{L}(S) \subset \mathcal{L}(T)$.
($\Leftarrow$). Suppose that $q(S) \subset q(T)$ and $\mathcal{L}(S) \subset \mathcal{L}(T)$. Clearly, if $T|_{\mathcal{L}(S)}$ is compatible with S then T is also compatible with S. So we can restrict ourselves the case when $\mathcal{L}(S) = \mathcal{L}(T)$. Let $\lambda|\bar{\lambda}$ be a split of S. By Lemma 3.7, we see that $ab|cd \in q(S)$ for all $a, b \in \lambda$ and $c, d \in \bar{\lambda}$. From the assumption, we see that this means $ab|cd \in q(T)$ for all such pairs and therefore $\lambda|\bar{\lambda} \in \Sigma(T)$ by Lemma 3.8. This accounts for every split $\lambda|\bar{\lambda}$ where $\min\{|\lambda|, |\bar{\lambda}|\} \geq 2$. If, without loss of generality, $|\lambda| = 1$, then the split $\lambda|\bar{\lambda}$ is a trivial splits. We know from Lemma 2.7 that since S and T are both phylogenetic trees, they both display all the trivial splits. So we see that $\Sigma(S) \subset \Sigma(T)$. Since we restricted T to $\mathcal{L}(S)$ we can appeal to Lemma 3.9 to show that $S \trianglelefteq T$. $\square$

Chapter 4

Binary Inferences

Definition 4.1. We say that a set of quartet splits $\mathbf{Q}$ infers the quartet split s when every tree T that displays every split in $\mathbf{Q}$ also displays s. If $\mathbf{Q}$ infers s it is denoted $\mathbf{Q} \rightarrow s$. When $|\mathbf{Q}| = k$ these are called k-ary inferences. If $\mathbf{Q} \rightarrow s$ for some $\mathbf{Q}$ of size k and there is no proper subset of $\mathbf{Q}$ that infers s, this is called a primitive k-ary inference.

Bryant and Steel showed that there are primitive k-ary inferences for all k [4]. In his Master's Thesis from 1986, Reconstruction Methods for Derivation Trees, M.C.H. Dekker enumerated all binary and ternary quartet split inferences[8]. His results were not available to me and I have therefore not seen any of Dekker's work. What follows is an independent re-creation of a portion of his Master's Thesis.

Let $\mathbf{A}$ be a set of two quartet splits A_1 and A_2. Let s be quartet split such that $s \notin \mathbf{A}$. We wish to know under what circumstances does $\mathbf{A} \rightarrow s$. We let the label set $\mathcal{L}(\mathbf{A})$ of $\mathbf{A}$ be $\{\mathcal{L}(A_1) \cup \mathcal{L}(A_2)\}$.

Lemma 4.2. *If* $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| < 2$ *then* $\mathbf{A} \not\rightarrow s$.

Proof.

Case 1. *Let* $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| = 0$ *and let* T *be a tree that displays both splits in* $\mathbf{A}$*. Let* s *be the split* $ab|cd$ *where* $\{a, b, c, d\} \cap \mathcal{L}(\mathbf{A}) = \emptyset$*. Now choose some internal vertex* v *of* T *and attach four new edges to* v *with one edge leading to each of the four new vertices* $\{a, b, c, d\}$*. Call this new tree* T'*. Since* a, b, c *and* d *are all adjacent to the vertex* v *there is no way to split* a, b *from* c, d *in* T'*. It is clear that* $q(T) \subset q(T')$ *and* *therefore T' displays $\mathbf{A}$ but T' does not display s. Since we have created a tree that displays $\mathbf{A}$ and not s we see that $\mathbf{A} \not\to s$.*

Case 2. *Let $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| = 1$ and let T be a tree that displays both splits in $\mathbf{A}$. Let s be the split $ab|cd$ where $\{a, b, c, d\} \cap \mathcal{L}(\mathbf{A}) = \{a\}$ without loss of generality. Choose an internal vertex v of T and attach three new edges to v with one edge leading to each of the new vertices $\{b, c, d\}$. Again, there is no split that separates $\{a, b\}$ from $\{c, d\}$. T' is an example of a tree that displays $\mathbf{A}$ but not s. So we see that $\mathbf{A} \not\to s$.*

□

Lemma 4.3. *If $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| = 2$ then $\mathbf{A} \not\to s$.*

Proof. Let $\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s) = \{a, b\}$ without loss of generality. So we have that $\mathcal{L}(s) = \{a, b, s_1, s_2\}$. Let T be a tree that displays both splits in $\mathbf{A}$. If we add s_1 and s_2 to T in such a way that s_1 and s_2 are adjacent to the same internal vertex of T, then there can be no splits of the form $v_1s_1|v_2s_2$ where $v_1, v_2 \in \mathcal{L}(\mathbf{A})$ as seen in Figure 4.1.

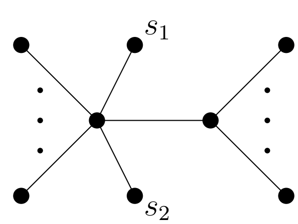


Figure 4.1: The tree T along with new vertices s_1, s_2

So suppose that s is of the form $v_1v_2|s_1s_2$. Add s_1 to T so that it is adjacent to v_1 and add s_2 so that it is adjacent to v_2. We see that $T \cup \{s_1, s_2\}$ cannot display the split s. There are no more possibilities for the arrangement of s and so we see that when $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| = 2$, $\mathbf{A} \not\to q$. □

Lemma 4.4. *If* $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| = 3$ *then* $\mathbf{A} \not\to s$.

Proof. Suppose that $\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s) = \{a, b, c\}$. Then without loss of generality, let s be the split $ab|cs_1$. Let T be a tree that displays both splits in $\mathbf{A}$. Since $s_1 \notin \mathcal{L}(\mathbf{A})$ we see that s_1 can be added so that it is adjacent to any internal vertex of T. In particular, s_1 can be added so that it is adjacent to vertex a. Since a is adjacent to s_1 there is no split that will separate a from s_1. Hence for any split s we see that $\mathbf{A} \not\to s$. □

Note that we have now shown that if $\mathbf{A} \to s$ then it must be the case that $|\mathcal{L}(\mathbf{A}) \cap \mathcal{L}(s)| \geq 4$. Since $|\mathcal{L}(s)| = 4$, we see that when $\mathbf{A} \to s$ it must be that $\mathcal{L}(q) \subset \mathcal{L}(\mathbf{A})$. In order for an inference to be made, we see from definition 4.1 that *all* trees displaying $\mathbf{A}$ must also display s. That is, if $\mathcal{T}$ is the set of all trees T that display $\mathbf{A}$, and $\mathbf{A} \to s$, then $s \in q(\mathcal{T})$. It is clear then that if some number of trees in $\mathcal{T}$ are shown to have no common quartet splits other than $\mathbf{A}$ then it has been shown that there are no inferences to be made from $\mathbf{A}$.

Theorem 4.5. *The following is an exhaustive list of binary inferences involving quartet splits. The set* $\{a, b, c, d, e\}$ *represents the leaves of each tree* T *that displays the quartet splits.*

$$ab|cd, ab|ce \to ab|de$$

$$ab|cd, ac|be \to ae|cd$$

$$ab|cd, ac|be \to ad|be$$

$$ab|cd, ac|be \to be|cd$$

Proof. We can see that $|\mathfrak{L}(\mathbf{A})| \in \{5, 6, 7, 8\}$. We will prove the theorem by cases.

Case 1. *Suppose that* $|\mathfrak{L}(\mathbf{A})| = 8$ *i.e.* $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| = 0$. *Without loss of generality let* A_1 *be* $ab|cd$ *and* A_2 *be* $ef|gh$.

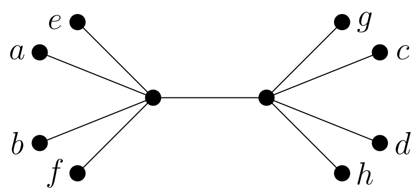


Figure 4.2: Trees T_1 and T_2

We can see from Figure 4.2 that the trees T_1 *and* T_2 *both display the splits of* $\mathbf{A}$ *but have no other splits in common. From this we can see that the splits of* $\mathbf{A}$ *do not infer any other quartet splits.*

Case 2. *Suppose that* $|\mathfrak{L}(\mathbf{A})| = 7$ *i.e.* $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| = 1$. *Without loss of generality let* A_1 *be* $ab|cd$ *and* A_2 *be* $ae|fg$.

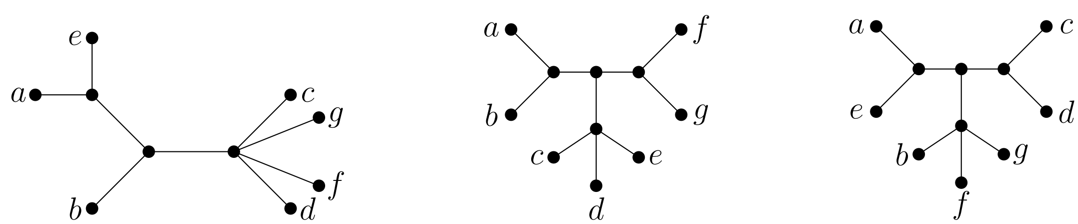


Figure 4.3: Trees T_1, T_2, and T_3 share no common splits other than $\mathbf{A}$.

In Figure 4.3, we have three trees that all display both A_1 *and* A_2. *If it were the case that* $\mathbf{A}$ *infered another quartet split* s, *then* T_1, T_2, *and* T_3 *would all display* s. *Since there are not splits that are common to all three trees, other than those in* $\mathbf{A}$, *we can see than no inferences can be made when* $|\mathfrak{L}(\mathbf{A})| = 7$.

Case 3. *Suppose that* $|\mathcal{L}(\mathbf{A})| = 6$ *i.e.* $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| = 2$. *Here we must look at several subcases for there are four possible arrangements of the leaves in which* $|\mathcal{L}(\mathbf{A})| = 6$.

Subcase 3.1. *Suppose that* A_1 *is the quartet split* $ab|cd$ *and* A_2 *is the quartet split* $ab|ef$. *It is clear from Figure 4.4 that these two splits do not infer a third.*

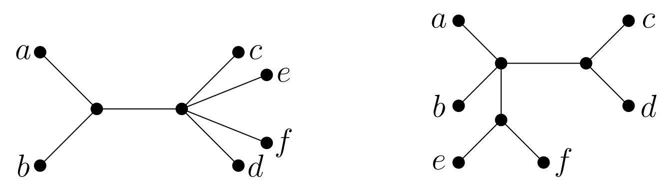


Figure 4.4: Trees T_1 and T_2 share no common splits other than $\mathbf{A}$.

Subcase 3.2. ***Subcase 3.2*** *Suppose that* A_1 *is the quartet split* $ab|cd$ *and* A_2 *is the quartet split* $ac|ef$. *As seen in Figure 4.5, there are no splits that can be infered from* A_1 *and* A_2.

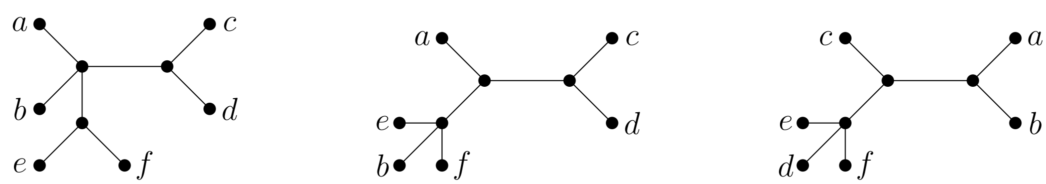


Figure 4.5: Trees T_1, T_2, and T_3 share no common splits other than $\mathbf{A}$.

Subcase 3.3. *Suppose that* A_1 *is the quartet split* $ab|cd$ *and* A_2 *is the quartet split* $ae|cf$. *As evidenced by the trees in Figure 4.6, the splits in* $\mathbf{A}$ *do not infer a third quartet split.*

Subcase 3.4. *Suppose that* A_1 *is the quartet split* $ab|cd$ *and* A_2 *is the quartet split* $ae|bf$. *It is shown in Figure 4.7 that no inferences can be made from these two quartet splits.*

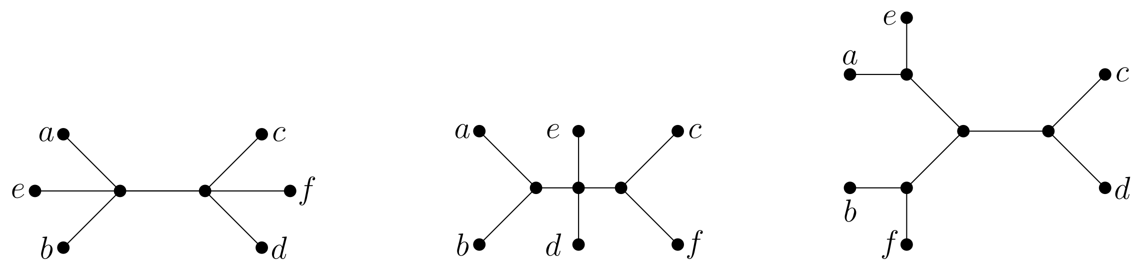


Figure 4.6: Trees T_1, T_2, and T_3 share no common splits other than $\mathbf{A}$.

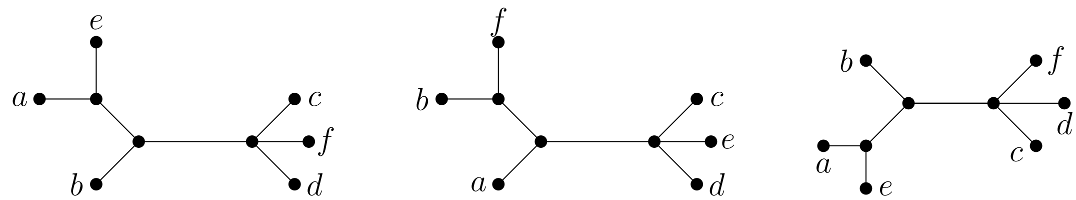


Figure 4.7: Trees T_1, T_2, and T_3 share no common splits other than $\mathbf{A}$.

We have now exhausted all possibilities for two quartet splits A_1 and A_2, for which $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| \leq 6$. The only remaining case involves splits between which only one leaf in different.

Case 4. *Now, assume that $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| = 3$ i.e. $|\mathcal{L}(\mathbf{A})| = 5$. Without loss of generality, let A_1 be the split $ab|cd$. In this scenario, we can consider a 4-leaf tree T that displays quartet split A_1 and explore the possible locations for our fifth leaf. Consider Figure 4.8 below.*

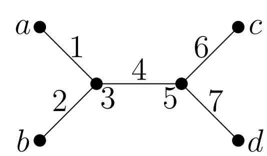


Figure 4.8: Tree showing possible locations for additional leaf

There are 7 possible locations to which the fifth leaf can be added to T. We will consider a second quartet split A_2, with $\mathcal{L}(A_2) = \{a, b, c, e\}$, and determine which of these positions would be valid locations for the new leaf f. We will then determine if there are inferences to be made from A_1 and A_2.

Subcase 4.1. *Suppose that A_1 is the quartet split $ab|cd$ as above, and A_2 is the quartet split $ab|ce$. To form a new tree $T' = T + \{e\}$ that will display A_1 as well as A_2, we see that e can be attached at positions 4, 5, 6, or 7. If e were added to positions 1, 2 or 3, the resulting tree T' would not display A_2. Now let us consider the trees in Figure 4.9. By construction, each of these trees displays splits A_1 and A_2 and we determined using Figure 4.8 that these are the only trees that do so. Notice, however, that each of these trees also displays the quartet split $ab|de$.*

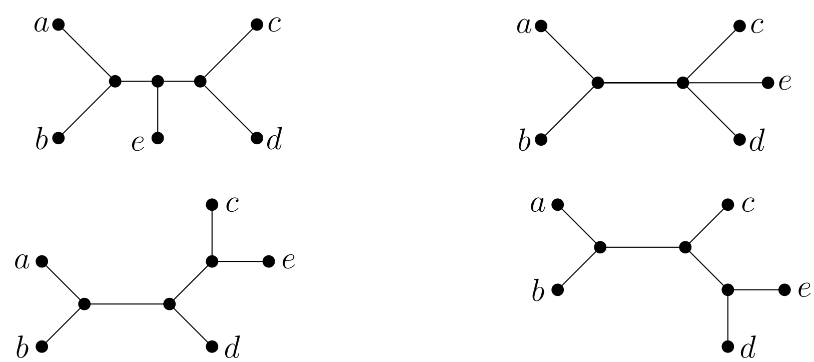


Figure 4.9: Trees with leaf e added in positions 4, 5, 6, and 7

Since every tree that dislays $\mathbf{A}$ also displays $ab|de$, we have shown that

$$ab|cd,\ ab|ce \rightarrow ab|de.$$

This is the first of the four binary inferences.

Subcase 4.2. *Again, we let A_1 be the split $ab|cd$. If A_2 is the split $ac|be$ then, from Figure 4.8, we see that the only possible location for leaf e is in position 2.*

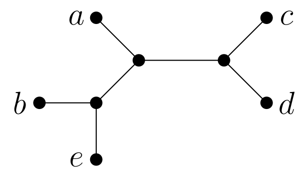


Figure 4.10: Tree with leaf e attached to position 2

In Figure 4.10 we see that the only tree that displays A_1 and A_2 also displays three other splits. So from this tree we can determine the following binary inferences

$$ab|cd,\ ac|be \to ae|cd$$

$$ab|cd,\ ac|be \to ad|be$$

$$ab|cd,\ ac|be \to be|cd$$

As there are no other possibilities for A_1 and A_2 we have shown that these four binary inferences are precisely the inferences that can be drawn from two quartet splits.

$\square$

Chapter 5

Ternary Inferences

We continue the work of the previous chapter, this time letting $\mathbf{A} = \{A_1, A_2, A_3\}$. Since the size of $\mathbf{A}$ is now 3, we refer to any inferences found from such a set **ternary inferences**.

Lemma 5.1. *If* $|\mathcal{L}(\mathbf{A})| = 12$ *then there is no quartet split* s *such that* $\mathbf{A} \rightarrow s$*.*

Proof. Without loss of generality, let $A_1 = ab|cd$, $A_2 = ef|gh$, and $A_3 = ij|kl$. The proof of the above lemma is clear from Figure 5.1 below. We can see that since the pairs $\{ef\}$ and $\{gh\}$ can be interchanged and the pairs $\{ij\}$ and $\{kl\}$ can be interchanged that there are no inferences that can be drawn from the tree below.

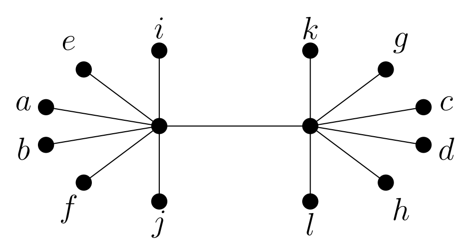


Figure 5.1: A tree with quartet splits A_1,A_2, and A_3

□

Lemma 5.2. *If* $|\mathcal{L}(\mathbf{A})| = 11$ *then there are no inferences that can be made.*

Proof. Without loss of generality we can let $A_1 = ab|cd$, $A_2 = ef|gh$, and $A_3 = ai|jk$. If we can find a collection of trees that display each of these three splits but have no others in common, then we have proved the lemma. Consider Figure 5.2 below. These five trees each display A_1, A_2, A_3 but have no other splits common to all five.

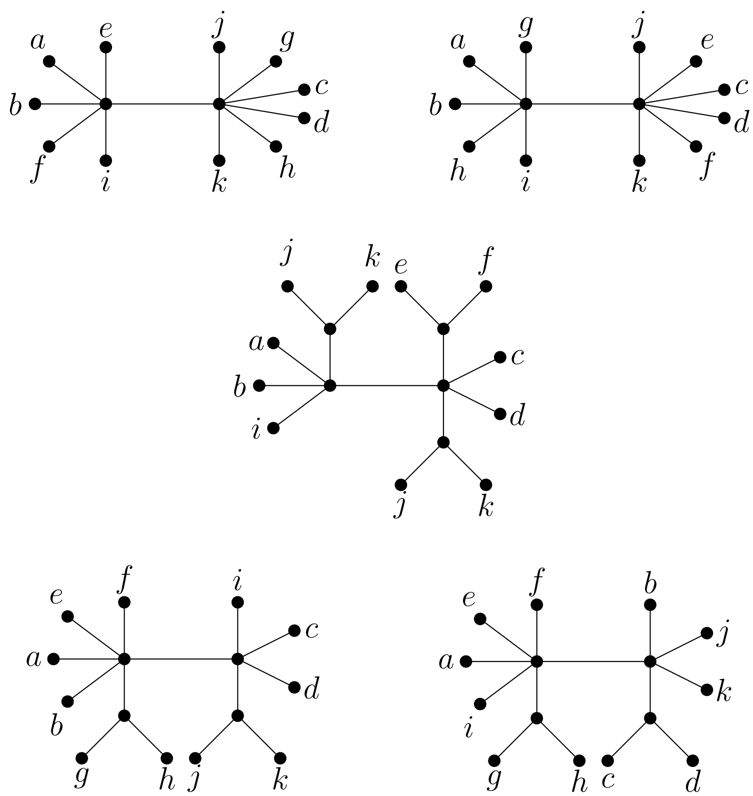


Figure 5.2: These five trees share only the splits A_1, A_2, A_3

$\square$

Lemma 5.3. *If* $|\mathcal{L}(\mathbf{A})| = 10$ *then there are no inferences that can be made.*

Proof. When $|\mathcal{L}(\mathbf{A})| = 10$ we are forced to consider two cases, each of which lead to several subcases. We can see that with a label set of size 10, that there are two different ways to choose our first two splits. Without loss of generality, let $A_1 = ab|cd$. Now, A_2 must be such that $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| \in \{0, 1, 2\}$ in order to get 10 distinctly labeled leaves. Note however, that if $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| = 2$, then since there are 10 total leaves, it must be that $|\mathcal{L}(A_1) \cap \mathcal{L}(A_3)| = 0$. Since we can arbitratily designate which split is A_2 and which is A_3 this is essentially the same as the case when

$|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| = 0$. This realization allows us to only consider the two cases when $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| \in \{0, 1\}$.

Case 5. *Suppose that* $|\mathcal{L}(A_1) \cap \mathcal{L}(A_2)| = 0$*. Then, without loss of generality,* $A_1 = ab|cd$ *and* $A_2 = ef|gh$*. Here we consider each possible arrangement of* A_3*.*

Subcase 5.1. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ef|gh \quad A_3 = ac|ij$$

We can see from Figure 5.3 below that these three splits do not infer a fourth.

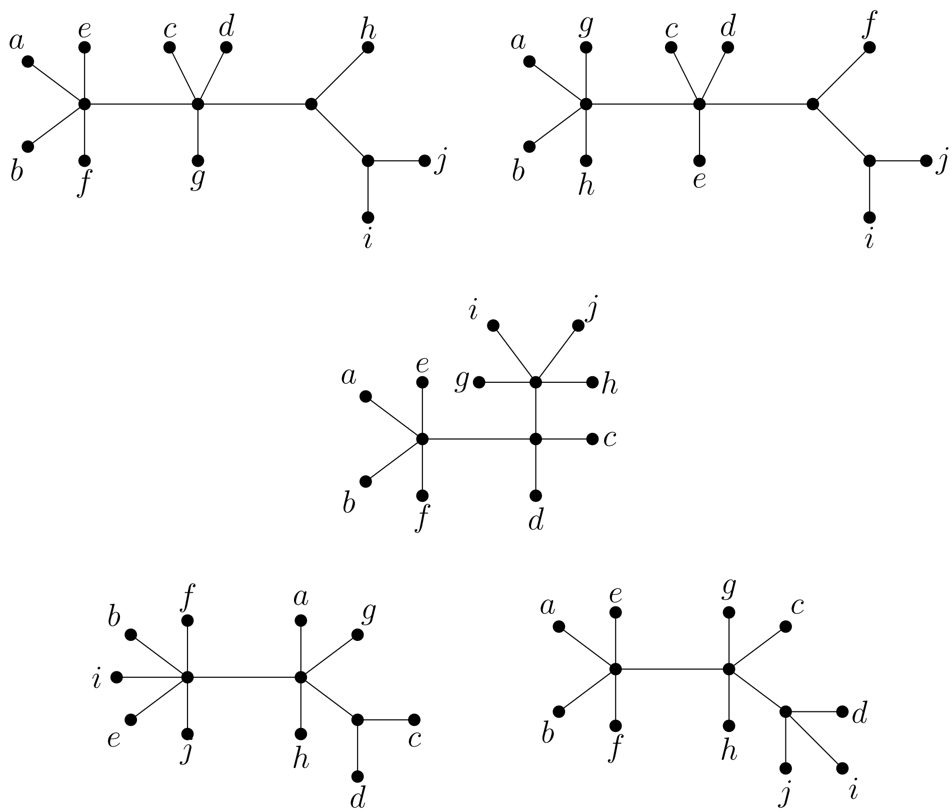


Figure 5.3: The five trees above have only three common splits

Subcase 5.2. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ef|gh \quad A_3 = ae|ij$$

We can see from Figure 5.4 below that these three splits do not infer a fourth.

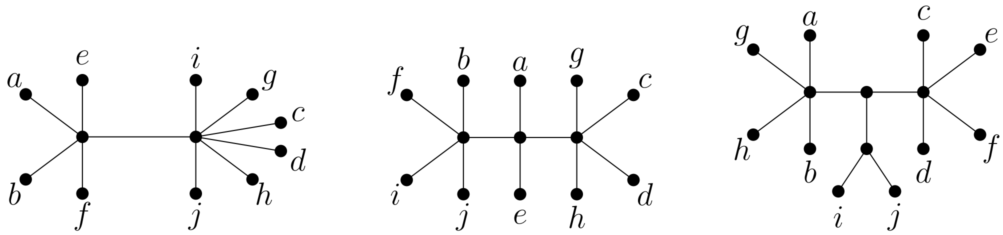


Figure 5.4: The three trees above share only splits A_1, A_2, and A_3

Subcase 5.3. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ef|gh \quad A_3 = ai|cj$$

We can see from Figure 5.5 below that these three splits do not infer a fourth.

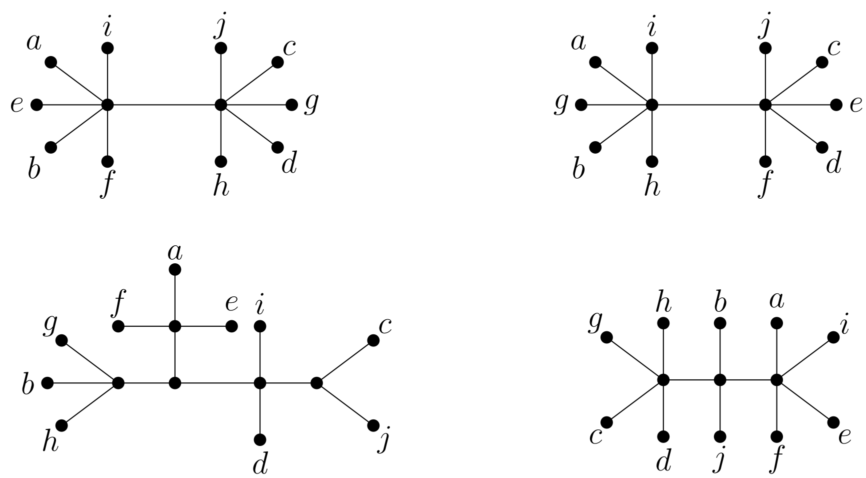


Figure 5.5: The four trees above share only splits A_1 A_2 and A_3.

Subcase 5.4. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ef|gh \quad A_3 = ab|ij$$

We can see from Figure 5.6 below that these three splits do not infer a fourth.

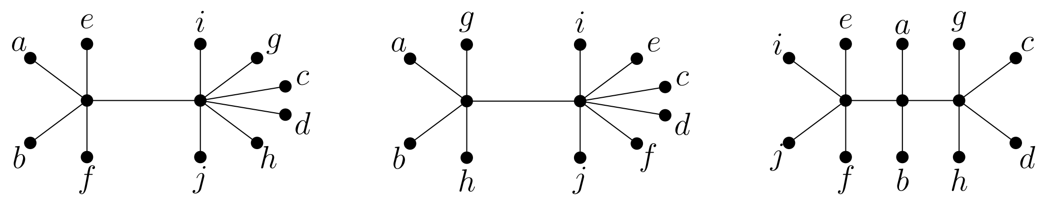


Figure 5.6: The three trees above share only splits A_1, A_2, and A_3

Subcase 5.5. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ef|gh \quad A_3 = ai|ej$$

We can see from Figure 5.7 below that these three splits do not infer a fourth.

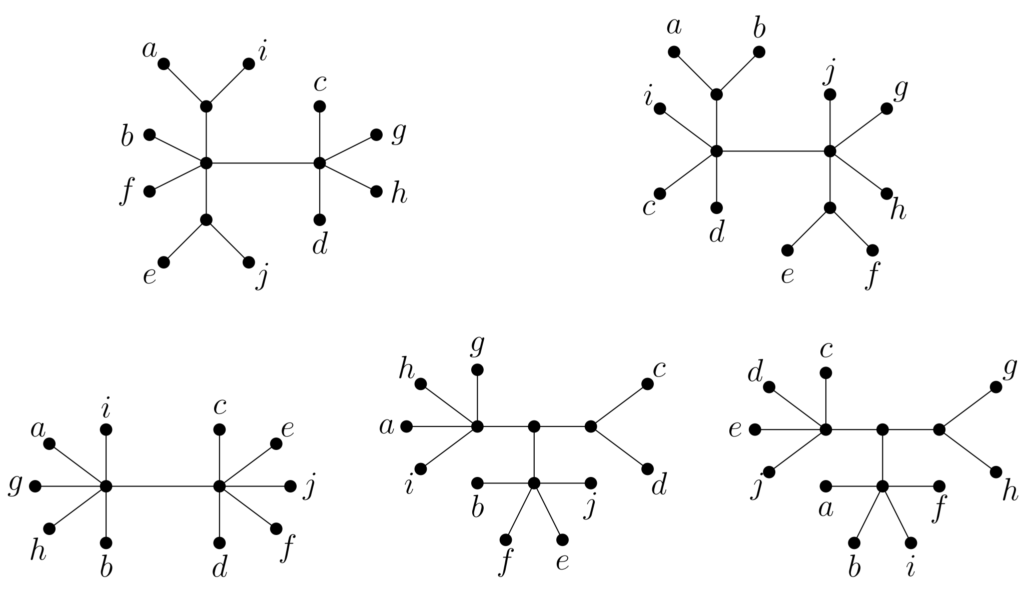


Figure 5.7: The five trees above share only splits A_1, A_2, and A_3

Subcase 5.6. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ef|gh \quad A_3 = ai|bj$$

We can see from Figure 5.8 below that these three splits do not infer a fourth.

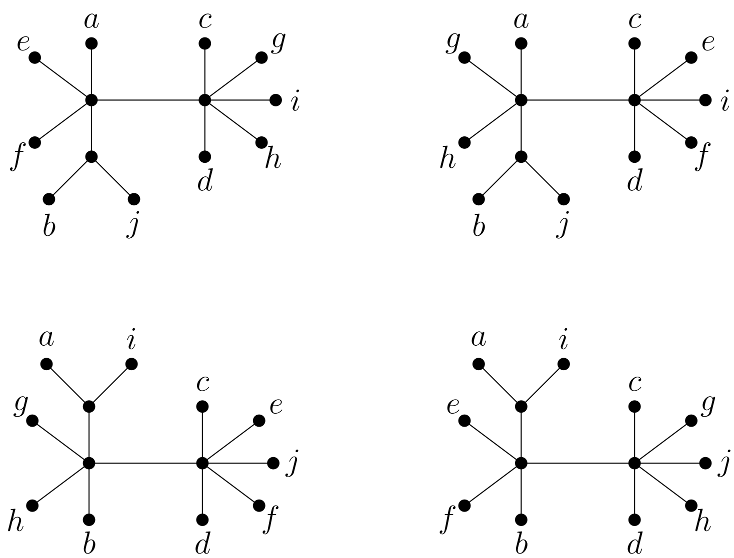


Figure 5.8: The four trees above share only splits A_1, A_2, and A_3

Case 6. *Now suppose that $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| = 1$. Then, without loss of generality, $A_1 = ab|cd$ and $A_2 = ae|fg$. Here we consider each possible arrangement of A_3.*

Subcase 6.1. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = ah|ij$$

We can see from Figure 5.9 below that these three splits do not infer a fourth.

Subcase 6.2. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = bh|ij$$

We can see from Figure 5.10 below that these three splits do not infer a fourth.

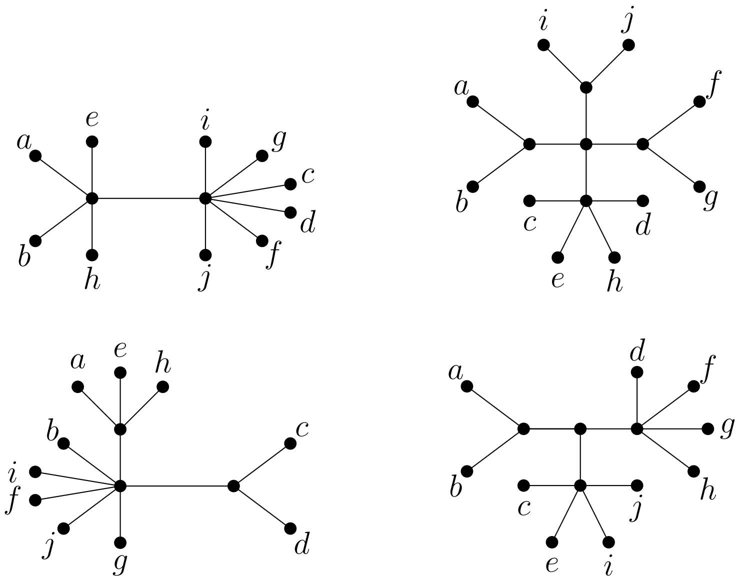


Figure 5.9: The four trees above share only splits A_1, A_2, and A_3

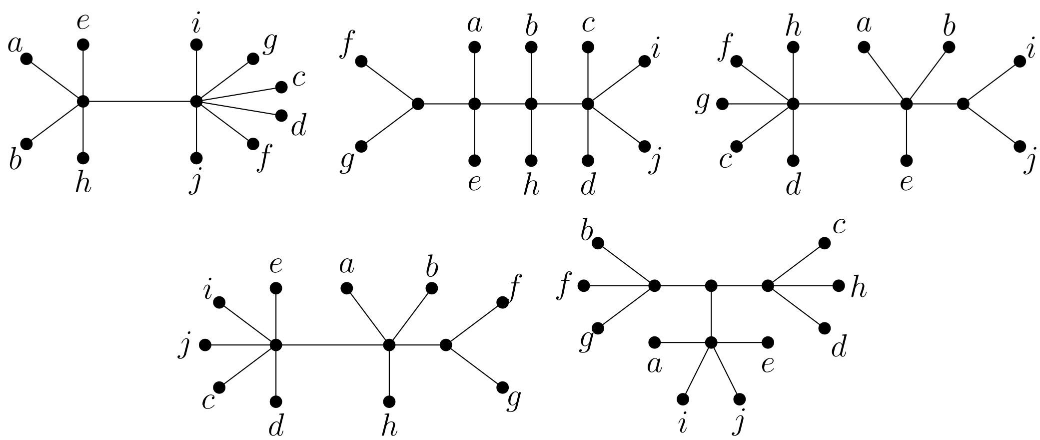


Figure 5.10: The five trees above share only splits A_1, A_2, and A_3

Subcase 6.3. *We have the following three splits:*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = ch|ij$$

We can see from Figure 5.11 below that these three splits do not infer a fourth.

□

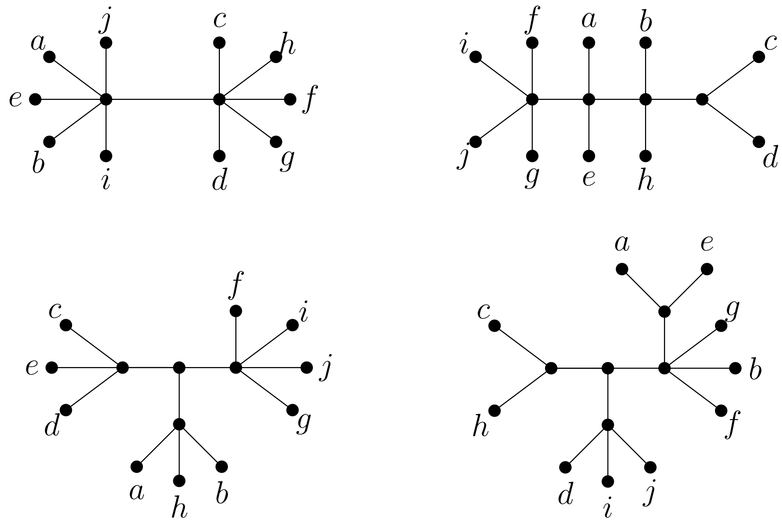


Figure 5.11: The four trees above share only splits A_1, A_2, and A_3

Lemma 5.4. *If $|\mathfrak{L}(A)| = 9$, there are no inferences that can be made.*

Proof. When we consider three quartet splits with a total of 9 distinct leaf labels, we see that the possibilities are divided into 5 cases depending on the intersections of the label sets of each of the three quartet pairs. We have splits A_1, A_2, and A_3 and we therefore have $\binom{3}{2} = 3$ intersections between them. For convenience, we always let $A_1 = ab|cd$. We can assume without loss of generality that $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| \geq |\mathfrak{L}(A_1) \cap \mathfrak{L}(A_3)|$. If this is not the case, we can simply relabel A_2 and A_3 so that this is true. We also notice that $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| \in \{0, 1, 2, 3\}$ since two quartet splits on the same 4 leaves are either the same split or they are imcompatible. The four cases of our proof will be based on intersection triples which will be ordered triples where the coordinates are the sizes of the intersections between the three quartet splits. The first coordinate of the intersection triple will be $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)|$, the second $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_3)|$, and the third coordinate will be $|\mathfrak{L}(A_2) \cap \mathfrak{L}(A_3)|$.

Case 1. *First assume that $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)|$ is as large as possible, namely 3. Since there must be exactly 9 distinct leaves, and only 5 are coming from splits A_1 and A_2, we see A_3 must consist of 4 previously unused labels. Therefore, we have the* *intersection triple (3,0,0). Notice that by relabeling and using our above assumption that* $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)|$ *is a largest intersection that (3,0,0) is equivalent to (0,3,0) and (0,0,3). This equivalence greatly reduces the number of subcases we must check. So we are considering sets where splits* A_1 *and* A_2 *share three leaves which, without loss of generality, we will name* a, b, c*. We now have two subcases to consider.*

Subcase 1.1. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ab|ce \quad A_3 = fg|hi$$

Notice that splits A_1 *and* A_2 *imply the third split* $ab|de$ *from Theorem 4.5. Since this inference is made by only two of our splits it is not considered a ternary inference. We can see from Figure 5.12 below that these three splits yield no ternary inferences.*

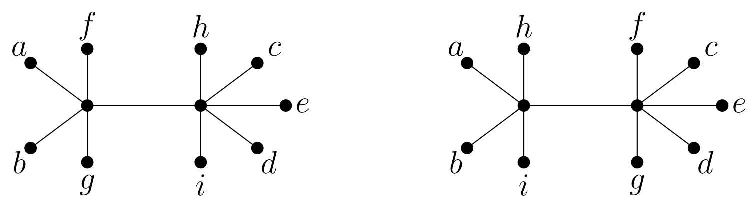


Figure 5.12: The two trees above share only splits A_1, A_2, A_3 and the aforementioned binary inference

Subcase 1.2. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ac|be \quad A_3 = fg|hi$$

Notice that splits A_1 *and* A_2 *imply the splits* $ae|cd$*,* $ad|be$*, and* $be|cd$ *from Theorem 4.5. Since these inferences are made by only two of our splits the are not considered ternary inferences. The tree in Figure 5.13 below is the only way to display both* A_1 *and* A_2 *as found during the proof of the* Binary Inferences Theorem. *Since the leaves* f, g, h, i *can be attached arbitrarily to this tree so long as split* A_3 *is realized, we see that there can be no inferences drawn other than the three mentioned previously.*

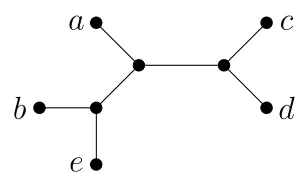


Figure 5.13: This tree shows that A_1, A_2, A_3 can yield no ternary inferences

Case 2. *Now we consider the case when the intersection triple is (2,1,1). From the second coordinate we can see that* $|\mathcal{L}(A_1) \cap \mathcal{L}(A_3)| = 1$. *Since we assume that* A_1 *is always* $ab|cd$, *we can let this single element of the intersection be* a. *We now have four leaves from* A_1 *plus two additional leaves from* A_2, *which means that we must get three unused leaves from the remainder of* A_3. *From this we see that it must be that* $a \in \mathcal{L}(A_2)$. *We know that* A_2 *must have two leaves in common with* A_1 *and that one of these has to be* a. *We can see that this points us toward two possible combinations,* a, b *and* a, c. *From this information, we arrive at the four subcases below.*

Subcase 2.1. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ab|ef \quad A_3 = ag|hi$$

Notice that splits A_1 *and* A_2 *imply the third split* $ab|de$ *from Theorem 4.5. Since this inference is made by only two of our splits it is not considered a ternary inference. We can see from Figure 5.14 below that these three splits yield no ternary inferences.*

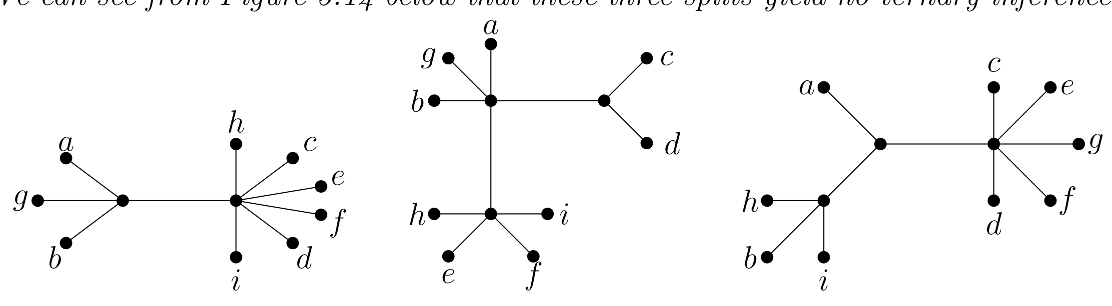


Figure 5.14: The trees above share only splits A_1, A_2, A_3

Subcase 2.2. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|bf \quad A_3 = ag|hi$$

We can see from Figure 5.15 below that these three splits yield no ternary inferences.

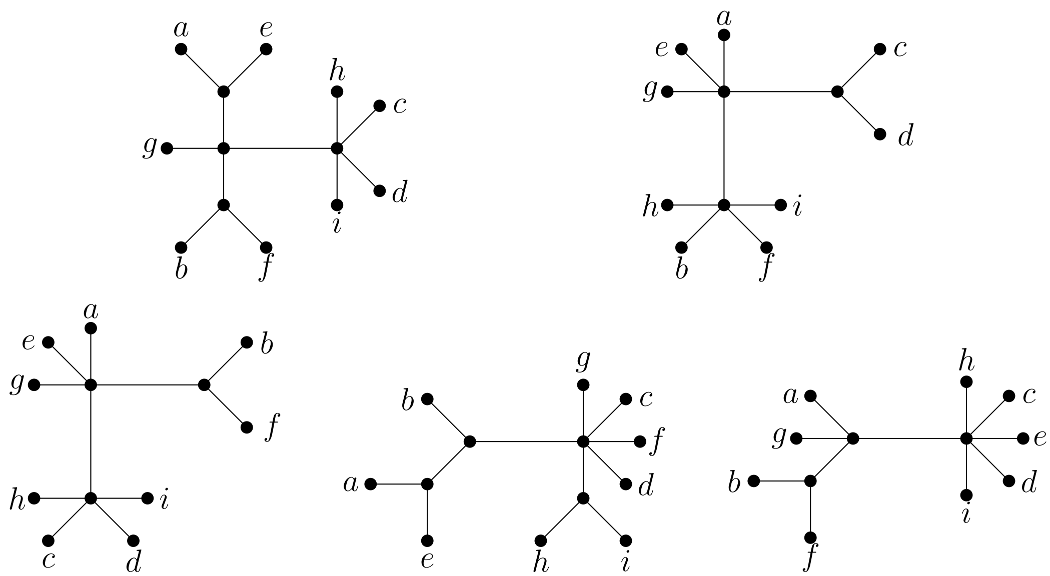


Figure 5.15: The trees above share only splits A_1, A_2, A_3

Subcase 2.3. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ac|ef \quad A_3 = ag|hi$$

We can see from Figure 5.16 below that these three splits yield no ternary inferences.

Subcase 2.4. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|cf \quad A_3 = ag|hi$$

We can see from Figure 5.17 below that these three splits yield no ternary inferences.

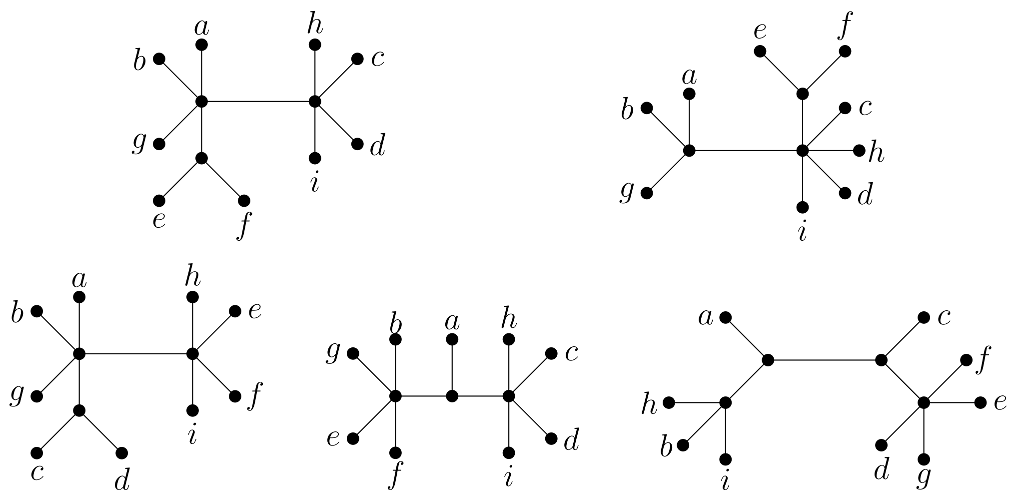


Figure 5.16: The trees above share only splits A_1, A_2, A_3

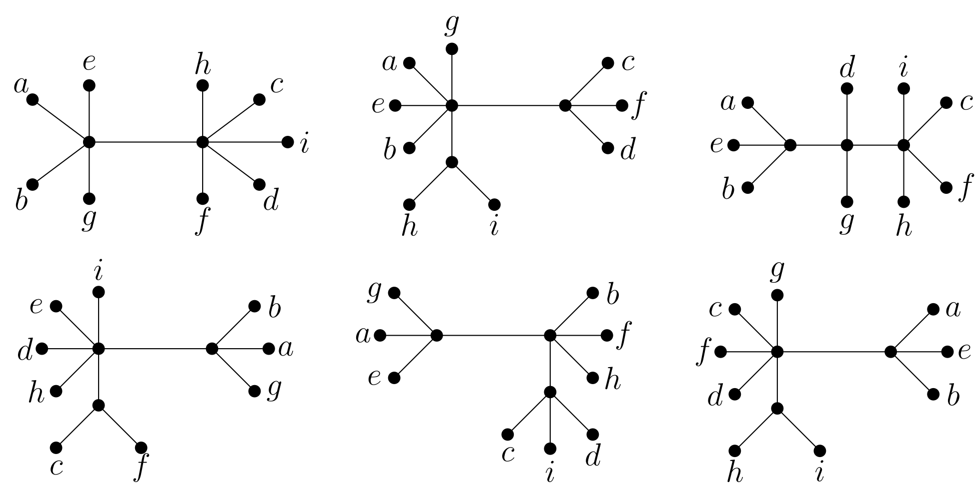


Figure 5.17: The trees above share only splits A_1, A_2, A_3

Case 3. *Now we consider the case when the intersection triple is (2,1,0). Since the first coordinate tells us that* $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| = 2$ *we can say that this intersection must either be* $\{a, b\}$ *or* $\{a, c\}$*. Suppose that* $A_2 = ab|ef$*. Then in order to conform to the intersection triple,* A_3 *must be either* $cg|hi$ *or* $dg|hi$*. Looking at the symmetry of* A_1 *(*$ab|cd = ab|dc$*) we see that these two splits are essentially the same. Using these principles, we arrive at the following four subcases.*

Subcase 3.1. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ab|ef \quad A_3 = cg|hi$$

We can see from Figure 5.18 below that these three splits yield no ternary inferences.

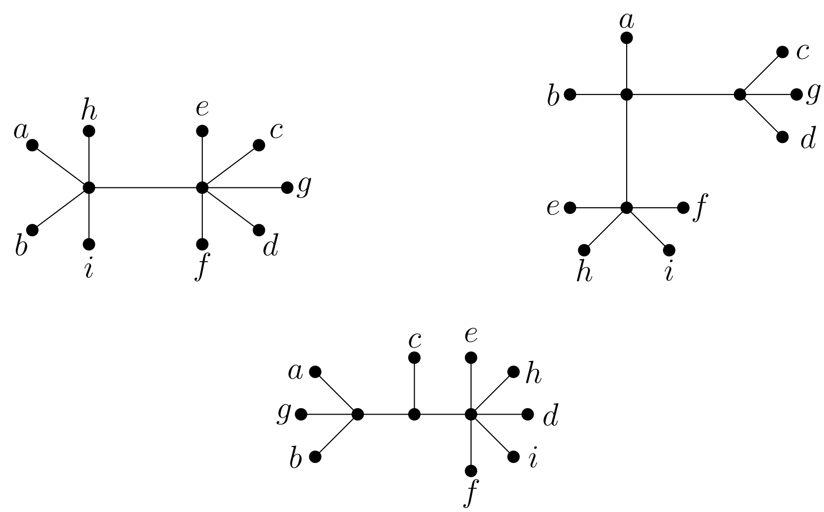


Figure 5.18: The trees above share only splits A_1, A_2, A_3

Subcase 3.2. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|bf \quad A_3 = cg|hi$$

We can see from Figure 5.19 below that these three splits yield no ternary inferences.

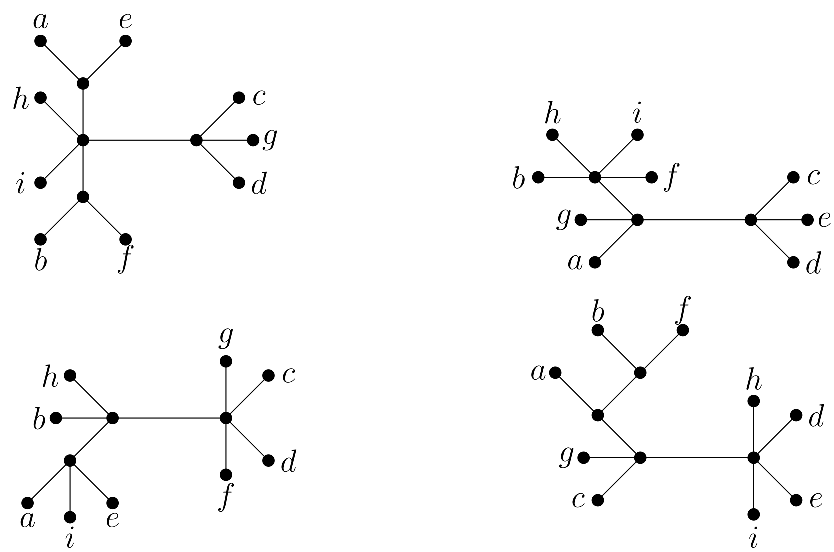


Figure 5.19: The trees above share only A_1, A_2, A_3

Subcase 3.3. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ac|ef \quad A_3 = bg|hi$$

We can see from Figure 5.20 below that these three splits yield no ternary inferences.

Subcase 3.4. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|cf \quad A_3 = bg|hi$$

We can see from Figure 5.21 below that these three splits yield no ternary inferences.

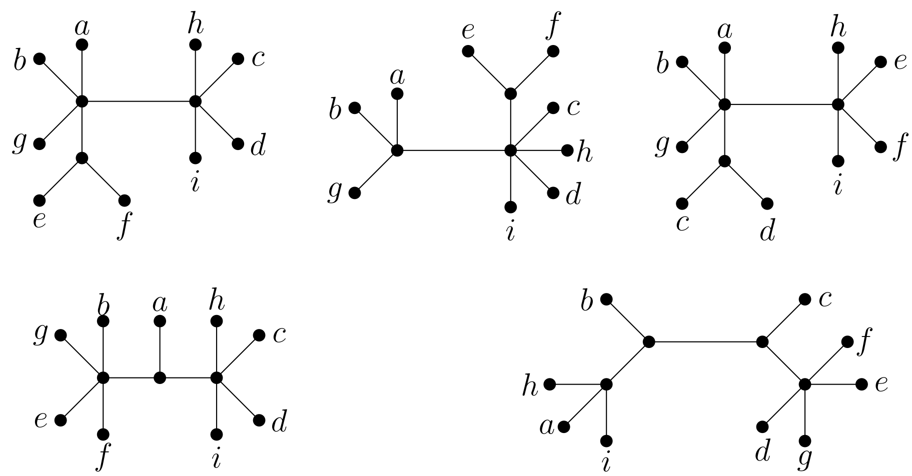


Figure 5.20: The trees above share only splits A_1, A_2, A_3

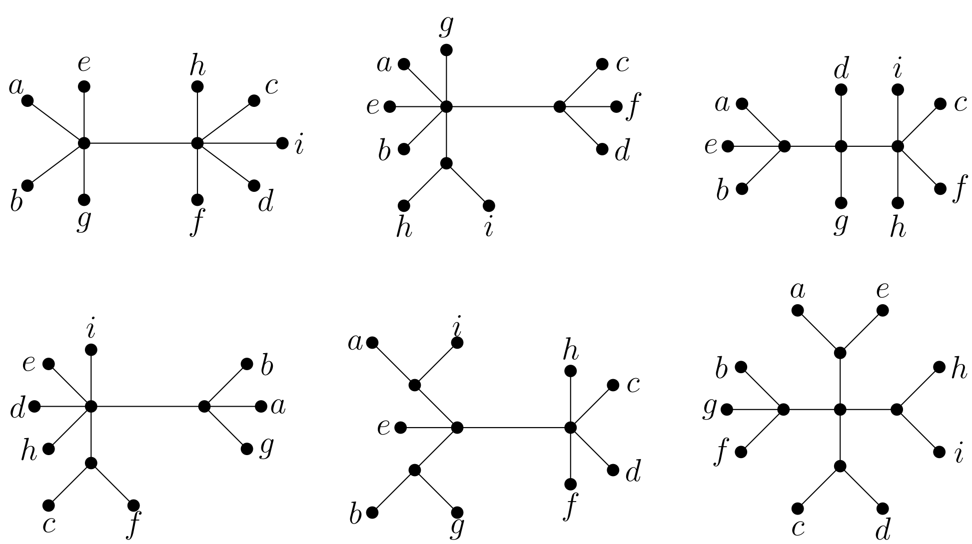


Figure 5.21: The trees above share only splits A_1, A_2, A_3

Case 4. *Now we consider the case when the intersection triple is (1,1,1). For the first coordinate, we can set* $A_1 \cap A_2 = \{a\}$ *without loss of generality. So in each case we have* $A_1 = ab|cd$ *and* $A_2 = ae|fg$*. The only variety comes from* A_3*. It must have one leaf in common with each of* A_1 *and* A_2 *but this leaf cannot be* a*. If* a *were an* *element of* $\mathfrak{L}(A_3)$*, then it would have the form* $ah|ij$*, but this gives us more than 9 leaves. So we see that* $\mathfrak{L}(A_3)$ *must contain exactly one of the following pairs in each case:*

$$\{b, e\} \qquad \{b, f\} \qquad \{c, e\} \qquad \{c, f\}$$

However, if we consider these four possibilities closely we will notice that there essentially is no difference between the pair $\{b.f\}$ *and the pair* $\{c, e\}$*. Thus we have three choices that lead to 6 subcases in the proof of our lemma.*

Subcase 4.1. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = be|hi$$

We can see from Figure 5.22 below that these three splits yield no ternary inferences.

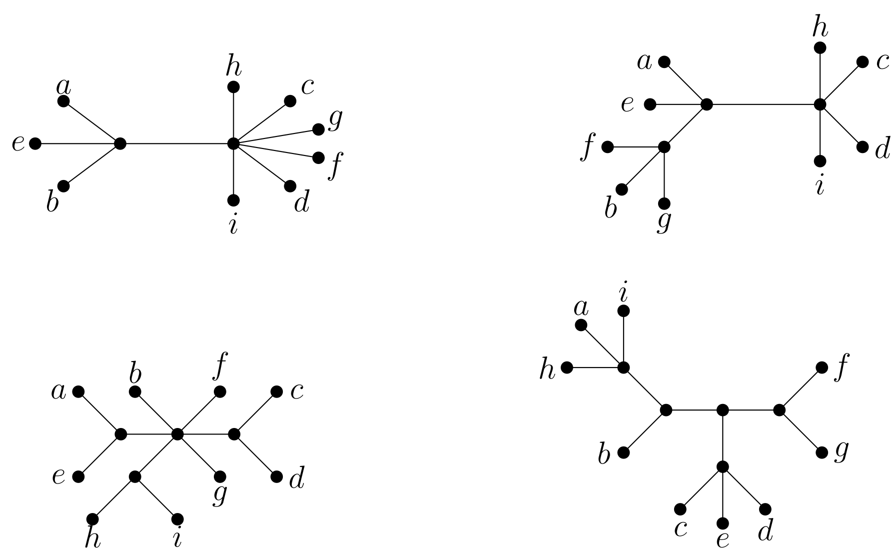


Figure 5.22: The trees above share only splits A_1, A_2, A_3

Subcase 4.2. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = bh|ei$$

We can see from Figure 5.23 below that these three splits yield no ternary inferences.

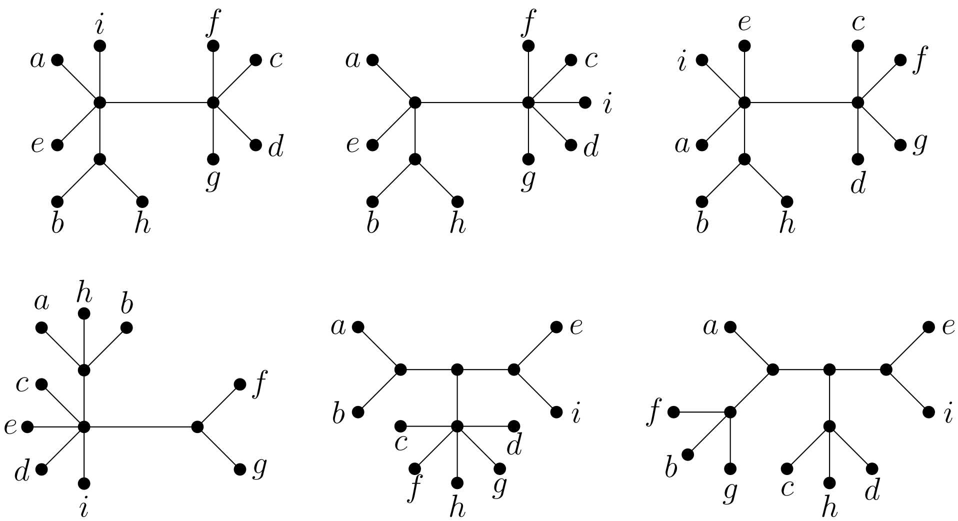


Figure 5.23: The trees above share only splits A_1, A_2, A_3

Subcase 4.3. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = bf|hi$$

We can see from Figure 5.24 below that these three splits yield no ternary inferences.

Subcase 4.4. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = bh|fi$$

We can see from Figure 5.25 below that these three splits yield no ternary inferences.

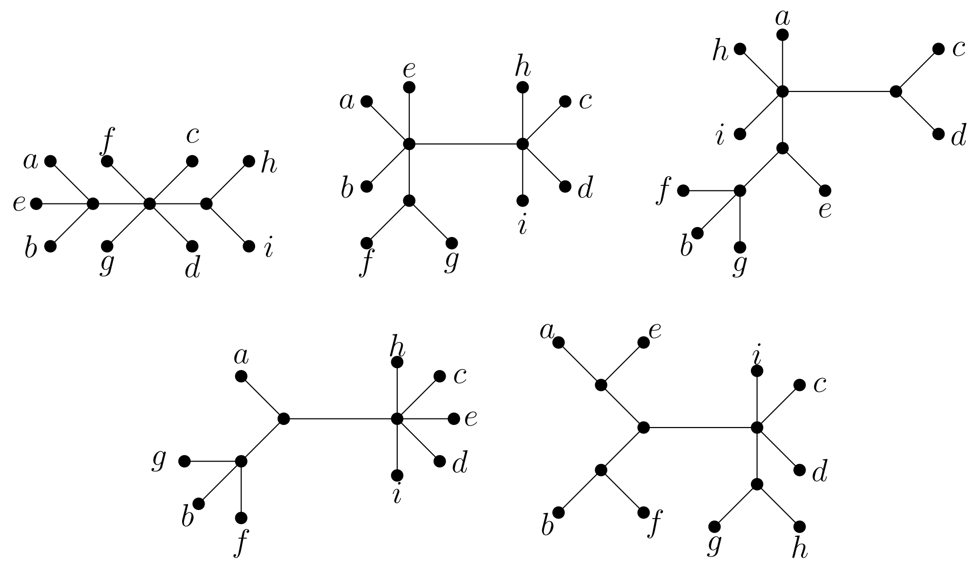


Figure 5.24: The trees above share only splits $A_1, A_2.A_3$

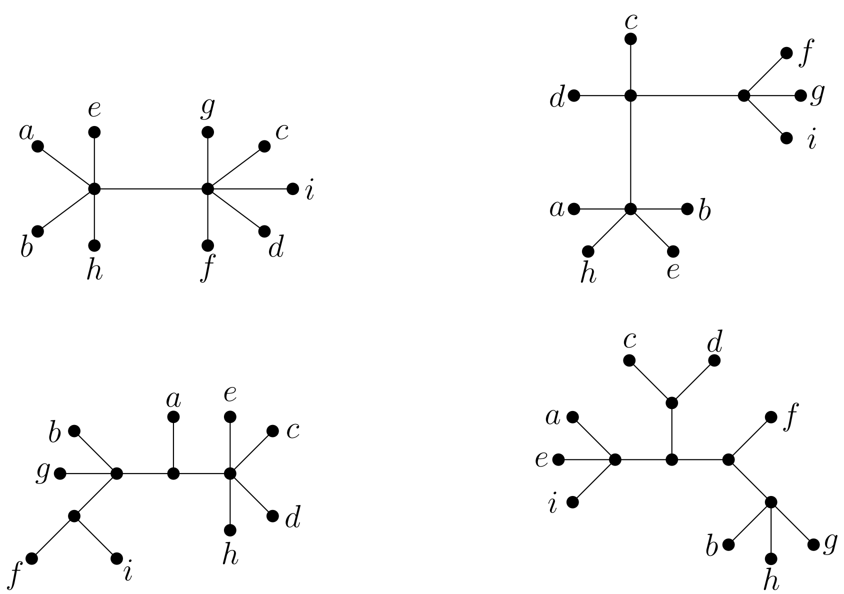


Figure 5.25: The trees above share only splits A_1, A_2, A_3

Subcase 4.5. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = cf|hi$$

We can see from Figure 5.26 below that these three splits yield no ternary inferences.

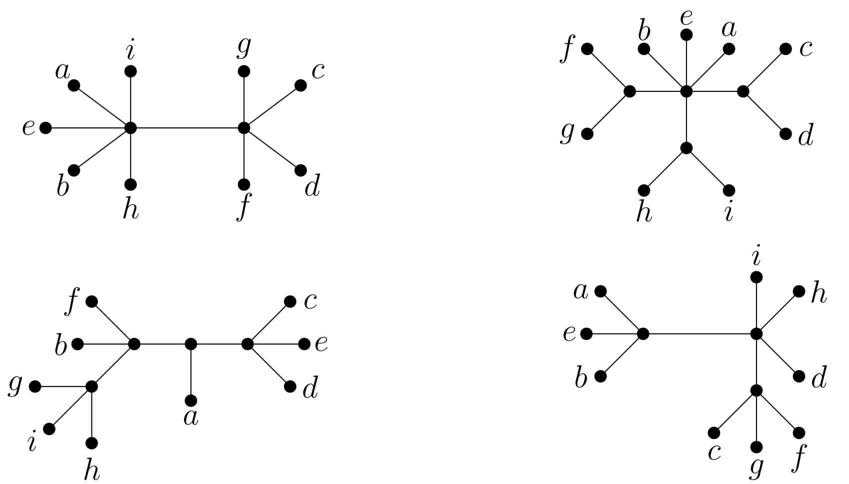


Figure 5.26: The trees above share only splits A_1, A_2, A_3

Subcase 4.6. *We have the following three splits*

$$A_1 = ab|cd \quad A_2 = ae|fg \quad A_3 = ch|fi$$

We can see from Figure 5.27 below that these three splits yield no ternary inferences.

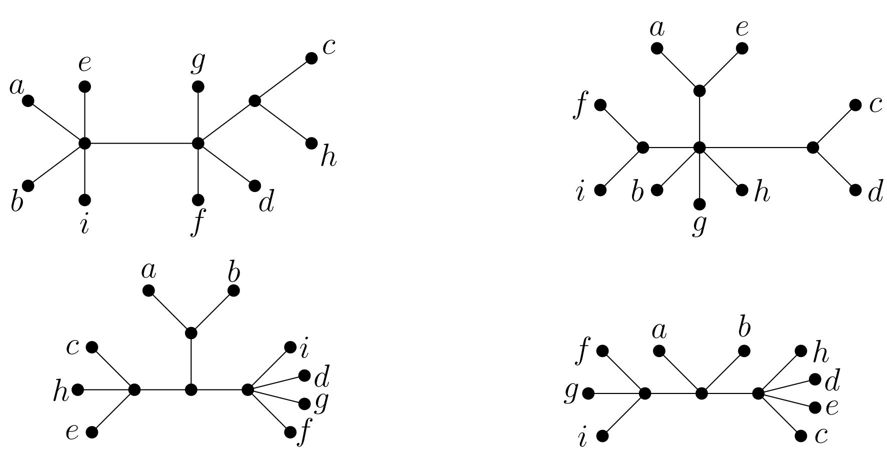


Figure 5.27: The trees above share only splits A_1, A_2, A_3

$\square$

Lemma 5.5. *If $|\mathfrak{L}(A)| = 8$, there are no inferences that can be made.*

Proof. When we consider three quartet splits with a total of 8 distinct leaf labels, we see that the possibilities are divided into 6 cases depending on the intersections of the label sets of each of the three quartet pairs. Again for convenience, we always let $A_1 = ab|cd$. We can assume without loss of generality that $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| \geq |\mathfrak{L}(A_1) \cap \mathfrak{L}(A_3)|$. If this is not the case, we can simply relabel A_2 and A_3 so that this is true. We also notice that $|\mathfrak{L}(A_1) \cap \mathfrak{L}(A_2)| \in \{0, 1, 2, 3\}$ since two quartet splits on the same 4 leaves are either the same split or they are imcompatible. The six cases of our proof will again be based on intersection triples.

Case 1. *First we consider the intersection triple (3,1,1). This scenario results in two possible subcases.*

Subcase 1.1. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ce \qquad A_3 = af|gh.$$

First, notice that A_1 and A_2 imply the split $ab|de$. We can see from Figure 5.28 below however, that these splits do not yield any ternary inferences.

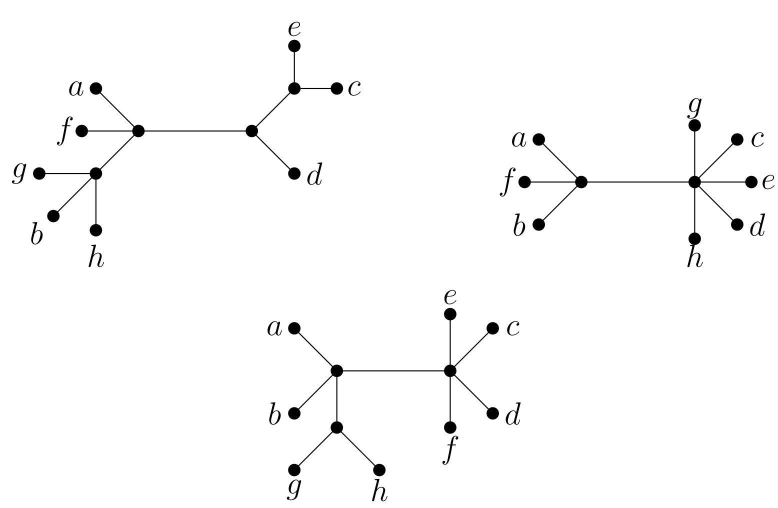


Figure 5.28: The trees above share only splits A_1, A_2, A_3

Subcase 1.2. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|be \qquad A_3 = af|gh.$$

First, notice that A_1 *and* A_2 *imply the splits* $ae|cd$ $ad|be$ *and* $be|cd$. *We can see from Figure 5.29 below however, that these splits do not yield any ternary inferences.*

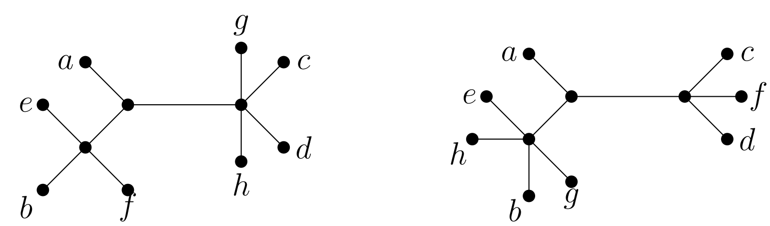


Figure 5.29: The trees above share only splits A_1, A_2, A_3

Case 2. *Now we consider the intersection triple (3,1,0). This scenario results in two possible subcases.*

Subcase 2.1. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ce \qquad A_3 = df|gh.$$

First, notice that A_1 *and* A_2 *imply the split* $ab|de$. *We can see from Figure 5.30 below however, that these splits do not yield any ternary inferences.*

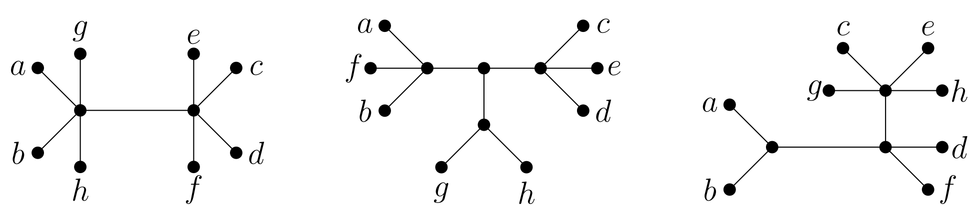


Figure 5.30: The trees above share only splits A_1, A_2, A_3

Subcase 2.2. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|be \qquad A_3 = df|gh.$$

First, notice that A_1 and A_2 imply the splits $ae|cd$, $ad|be$ and $be|cd$. We can see from Figure 5.31 below however, that these splits do not yield any ternary inferences.

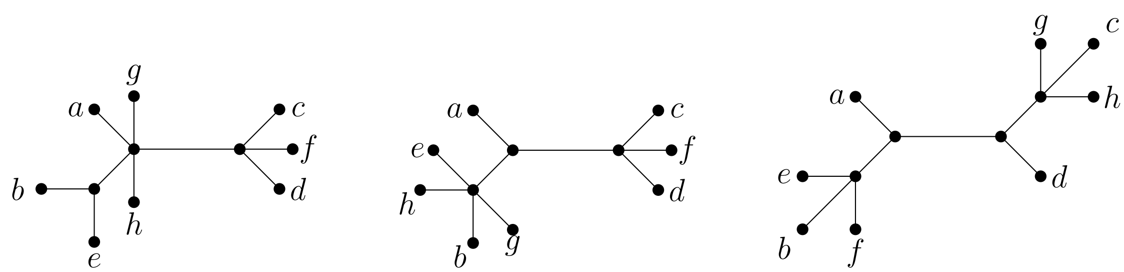


Figure 5.31: The trees above share only splits A_1, A_2, A_3

Case 3. *Now we consider the intersection triple (2,2,2). This scenario results in six possible subcases.*

Subcase 3.1. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = ab|gh.$$

We can see from Figure 5.32 below however, that these splits do not yield any ternary inferences.

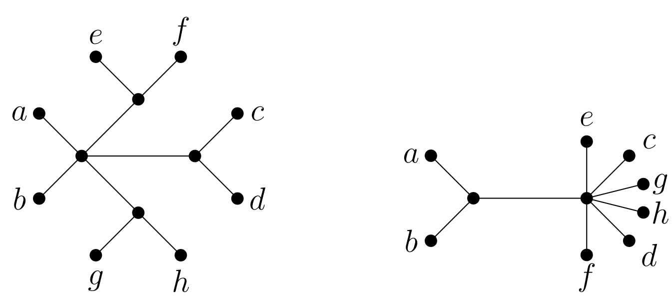


Figure 5.32: The trees above share only splits A_1, A_2, A_3

Subcase 3.2. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = ag|bh.$$

We can see from Figure 5.33 below however, that these splits do not yield any ternary inferences.

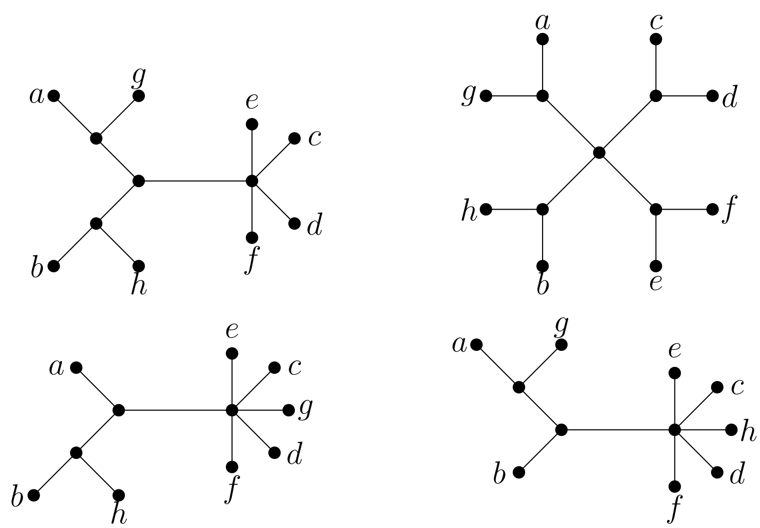


Figure 5.33: The trees above share only splits A_1, A_2, A_3

Subcase 3.3. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|bf \qquad A_3 = ag|bh.$$

We can see from Figure 5.34 below however, that these splits do not yield any ternary inferences.

Subcase 3.4. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = ac|gh.$$

We can see from Figure 5.35 below however, that these splits do not yield any ternary inferences.

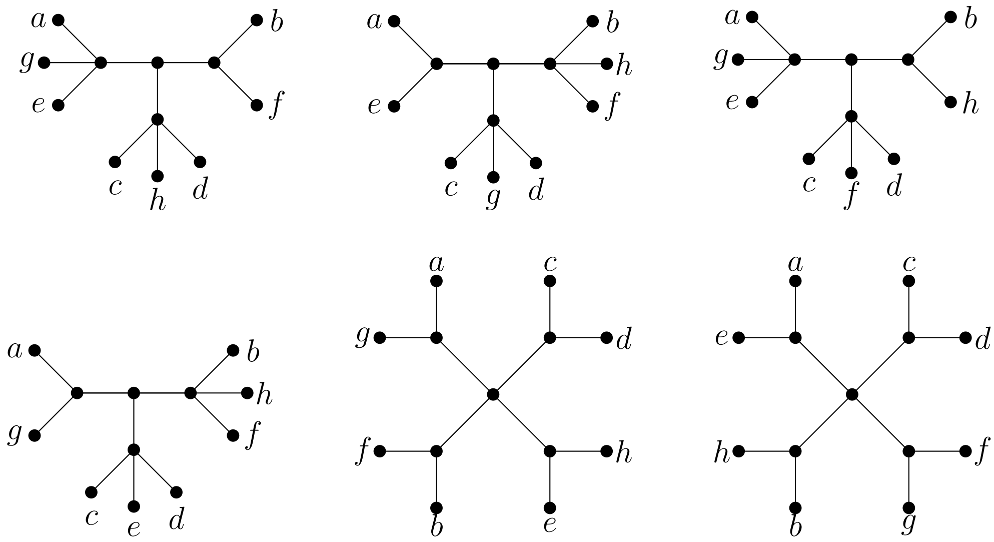


Figure 5.34: The trees above share only splits A_1, A_2, A_3

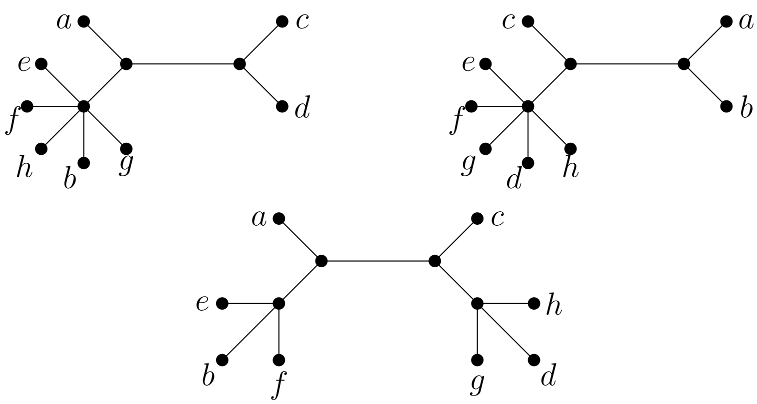


Figure 5.35: The trees above share only splits A_1, A_2, A_3

Subcase 3.5. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = ag|ch.$$

We can see from Figure 5.36 below however, that these splits do not yield any ternary inferences.

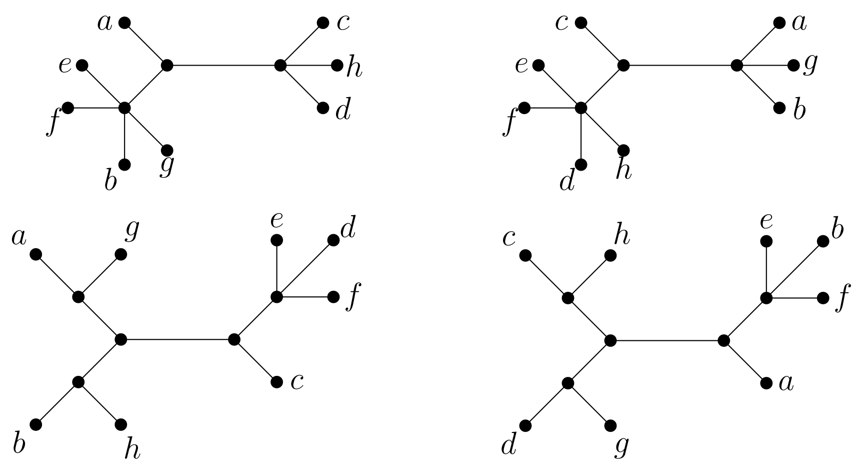


Figure 5.36: The trees above share only splits A_1, A_2, A_3

Subcase 3.6. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|cf \qquad A_3 = ag|ch.$$

We can see from Figure 5.37 below however, that these splits do not yield any ternary inferences.

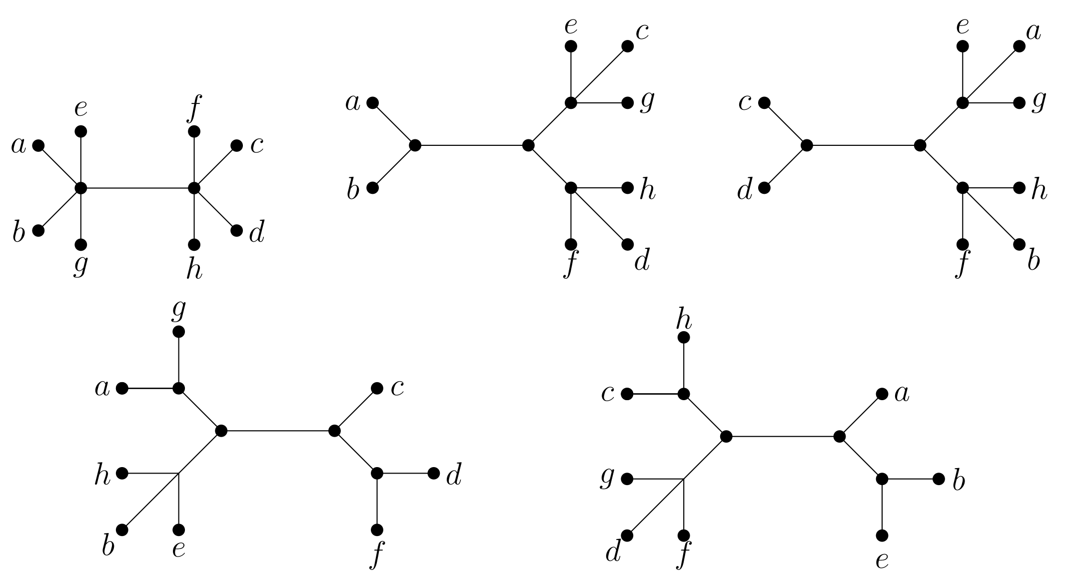


Figure 5.37: The trees above share only splits A_1, A_2, A_3

Case 4. *Now we consider the intersection triple (2,2,1). This scenario results in seven possible subcases.*

Subcase 4.1. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = ac|gh.$$

We can see from Figure 5.38 below however, that these splits do not yield any ternary inferences.

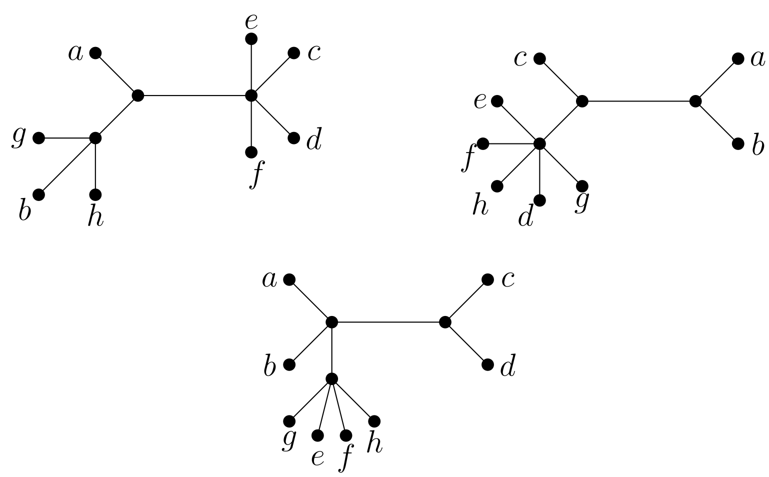


Figure 5.38: The trees above share only splits A_1, A_2, A_3

Subcase 4.2. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = ag|ch.$$

We can see from Figure 5.39 below however, that these splits do not yield any ternary inferences.

Subcase 4.3. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|bf \qquad A_3 = ac|gh.$$

We can see from Figure 5.40 below however, that these splits do not yield any ternary inferences.

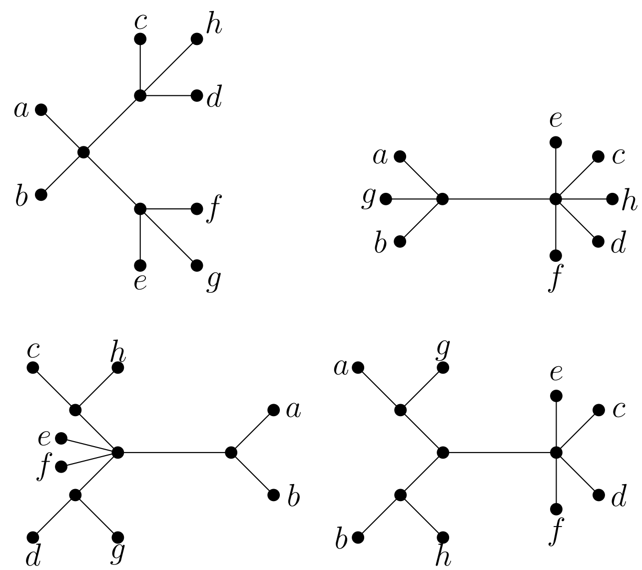


Figure 5.39: The trees above share only splits A_1, A_2, A_3

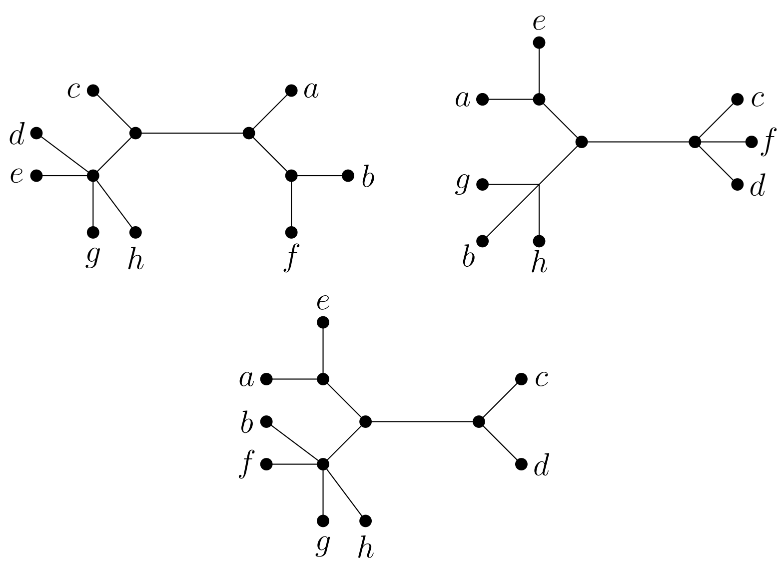


Figure 5.40: The trees above share only splits A_1, A_2, A_3

Subcase 4.4. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|bf \qquad A_3 = ag|ch.$$

We can see from Figure 5.41 below however, that these splits do not yield any ternary inferences.

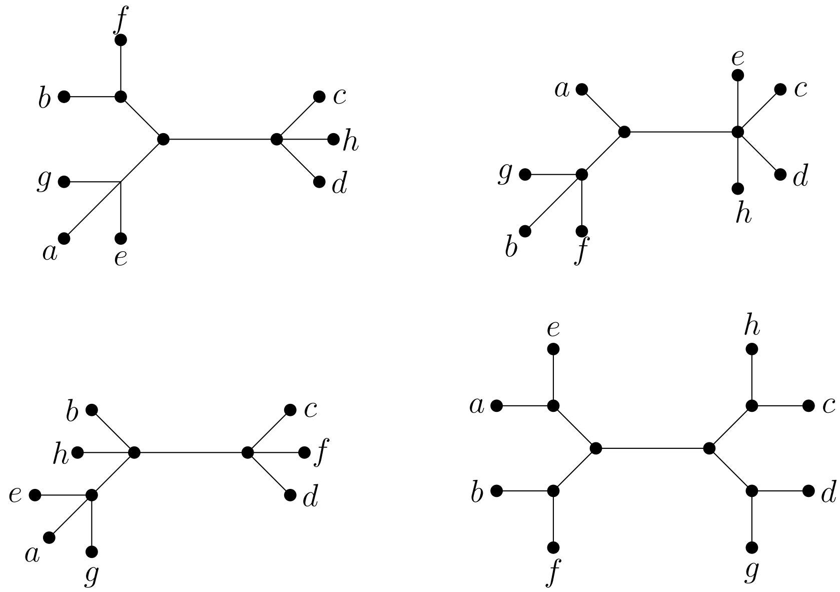


Figure 5.41: The trees above share only splits A_1, A_2, A_3

Subcase 4.5. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = ad|gh.$$

We can see from Figure 5.42 below however, that these splits do not yield any ternary inferences.

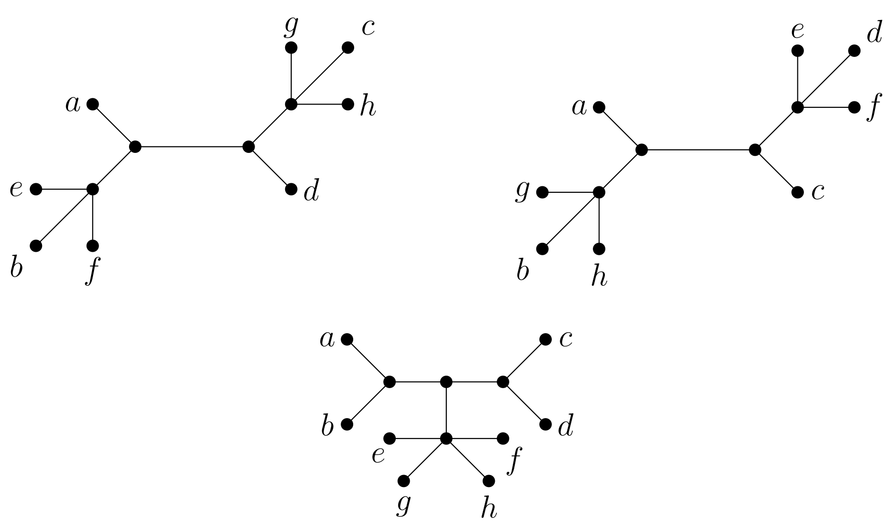


Figure 5.42: The trees above share only splits A_1, A_2, A_3

Subcase 4.6. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = ag|dh.$$

We can see from Figure 5.43 below however, that these splits do not yield any ternary inferences.

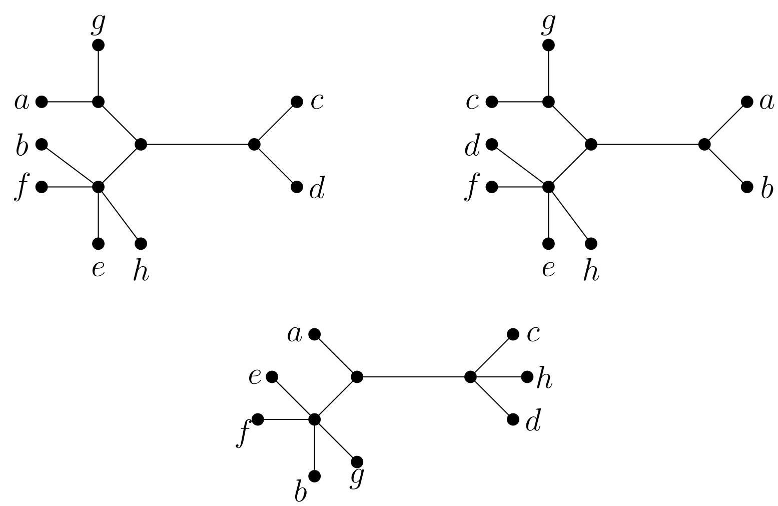


Figure 5.43: The trees above share only splits A_1, A_2, A_3

Subcase 4.7. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|cf \qquad A_3 = ag|dh.$$

We can see from Figure 5.44 below however, that these splits do not yield any ternary inferences.

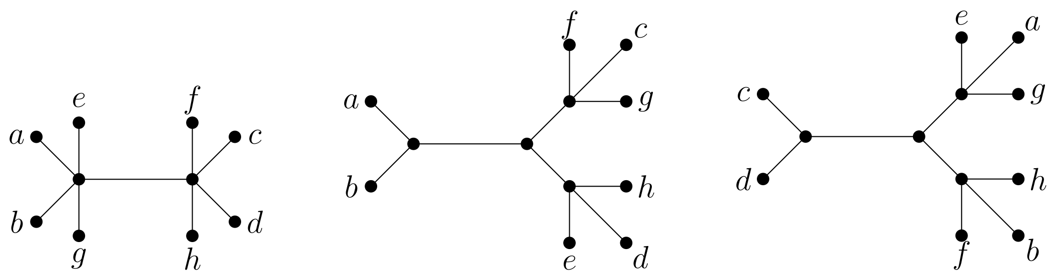


Figure 5.44: The trees above share only splits A_1, A_2, A_3

Case 5. *Now we consider the intersection triple (2,2,0). This scenario results in six possible subcases.*

Subcase 5.1. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = cd|gh.$$

We can see from Figure 5.45 below however, that these splits do not yield any ternary inferences.

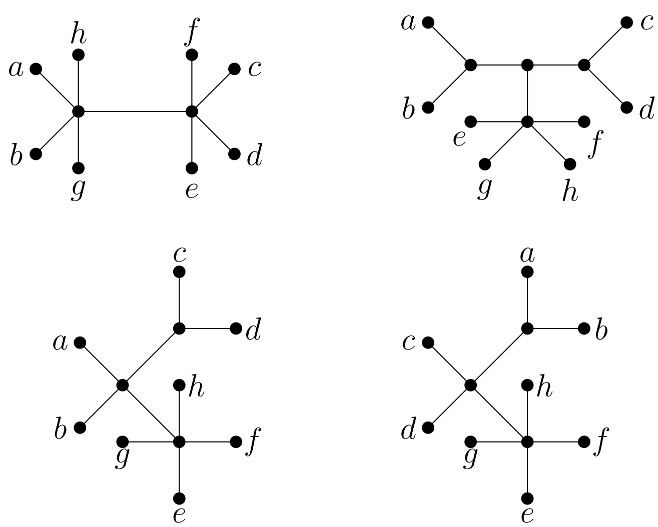


Figure 5.45: The trees above share only splits A_1, A_2, A_3

Subcase 5.2. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = cg|dh.$$

We can see from Figure 5.46 below however, that these splits do not yield any ternary inferences.

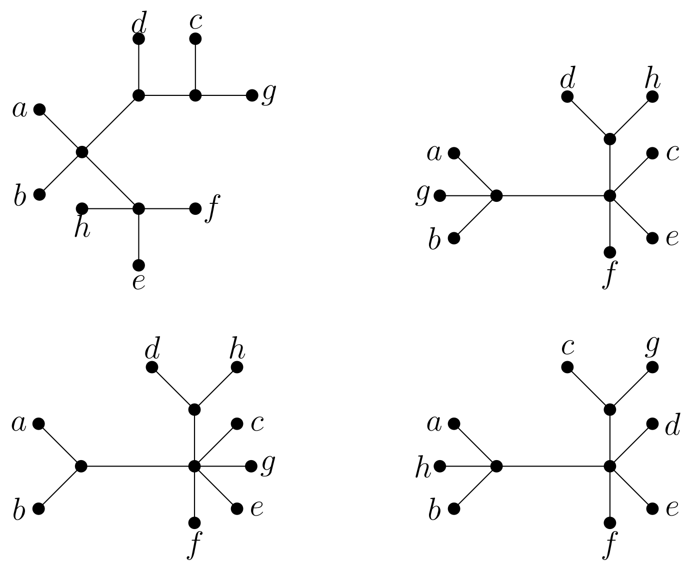


Figure 5.46: The trees above share only splits A_1, A_2, A_3

Subcase 5.3. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|bf \qquad A_3 = cg|dh.$$

We can see from Figure 5.47 below however, that these splits do not yield any ternary inferences.

Subcase 5.4. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = bd|gh.$$

We can see from Figure 5.48 below however, that these splits do not yield any ternary inferences.

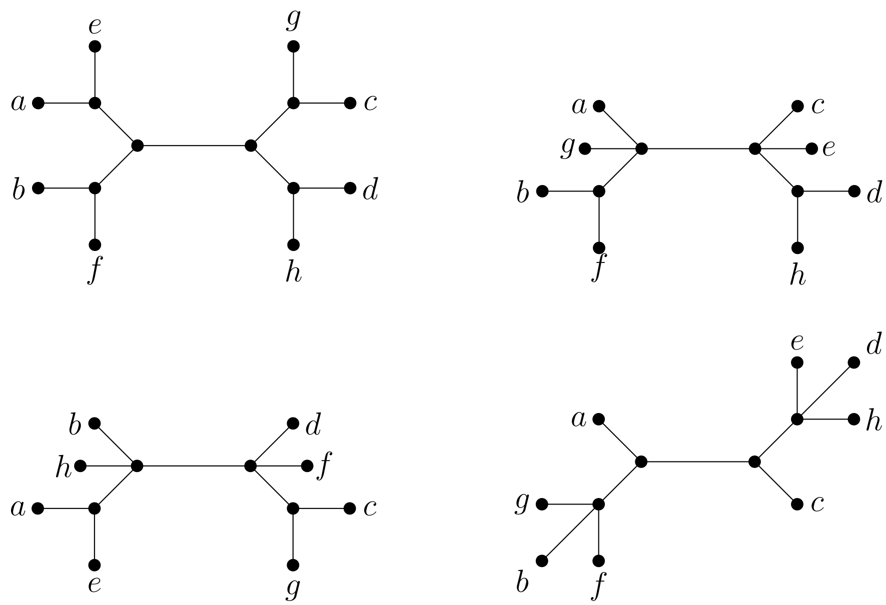


Figure 5.47: The trees above share only splits A_1, A_2, A_3

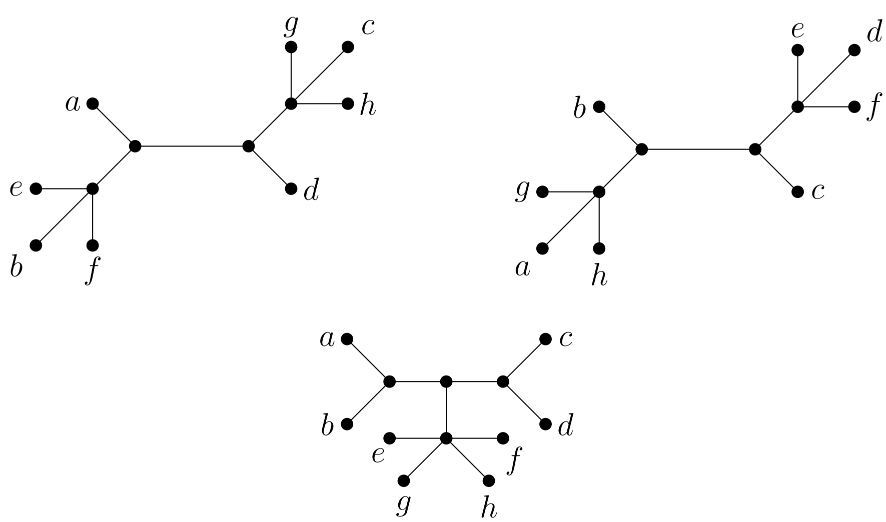


Figure 5.48: The trees above share only splits A_1, A_2, A_3

Subcase 5.5. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = bg|dh.$$

We can see from Figure 5.49 below however, that these splits do not yield any ternary inferences.

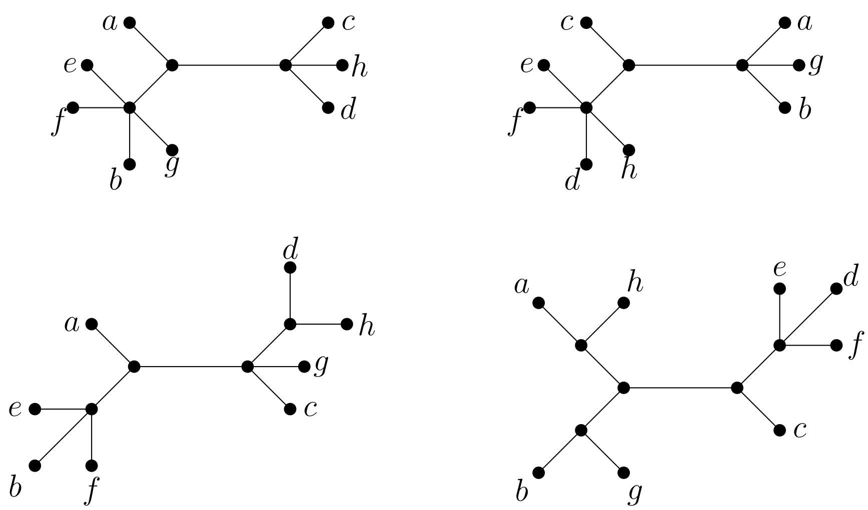


Figure 5.49: The trees above share only splits A_1, A_2, A_3

Subcase 5.6. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|cf \qquad A_3 = bg|dh.$$

We can see from Figure 5.50 below however, that these splits do not yield any ternary inferences.

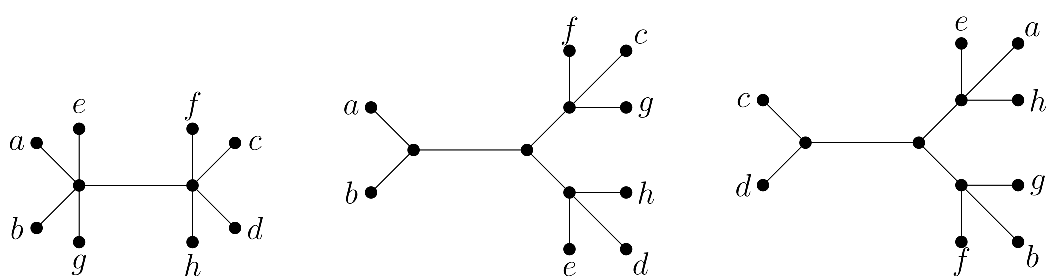


Figure 5.50: The trees above share only splits A_1, A_2, A_3

Case 6. *Now we consider the intersection triple (2,1,1). This scenario results in seven possible subcases.*

Subcase 6.1. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = ce|gh.$$

We can see from Figure 5.51 below however, that these splits do not yield any ternary inferences.

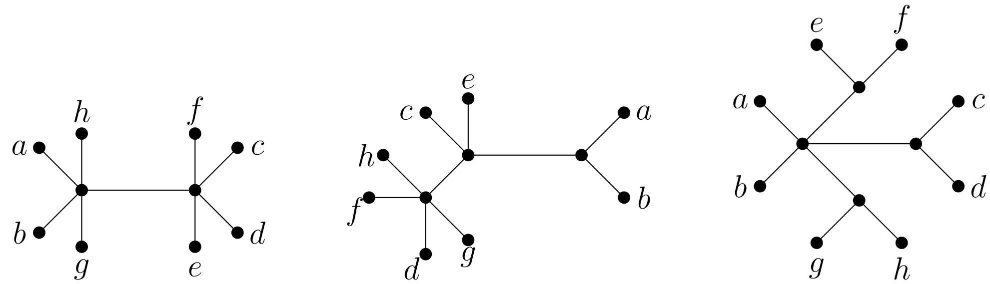


Figure 5.51: The trees above share only splits A_1, A_2, A_3

Subcase 6.2. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = cg|eh.$$

We can see from Figure 5.52 below however, that these splits do not yield any ternary inferences.

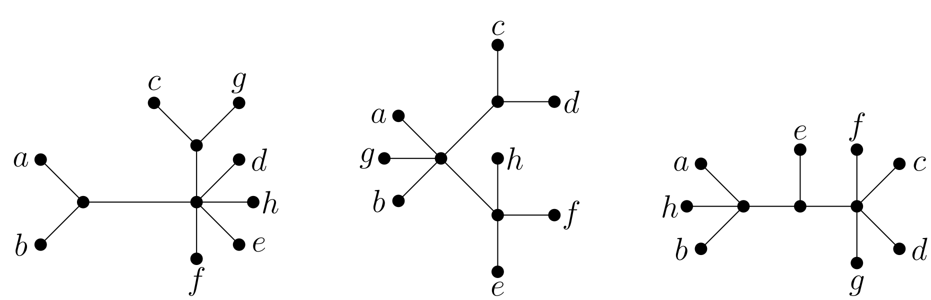


Figure 5.52: The trees above share only splits A_1, A_2, A_3

Subcase 6.3. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|bf \qquad A_3 = ce|gh.$$

We can see from Figure 5.53 below however, that these splits do not yield any ternary inferences.

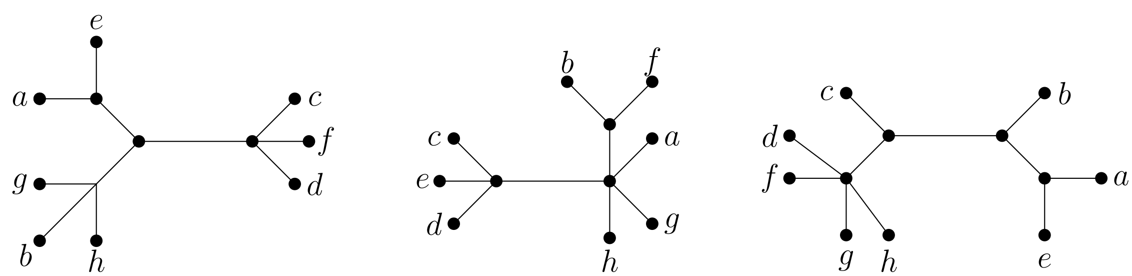


Figure 5.53: The trees above share only splits A_1, A_2, A_3

Subcase 6.4. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|bf \qquad A_3 = cg|eh.$$

We can see from Figure 5.54 below however, that these splits do not yield any ternary inferences.

Subcase 6.5. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = be|gh.$$

We can see from Figure 5.55 below however, that these splits do not yield any ternary inferences.

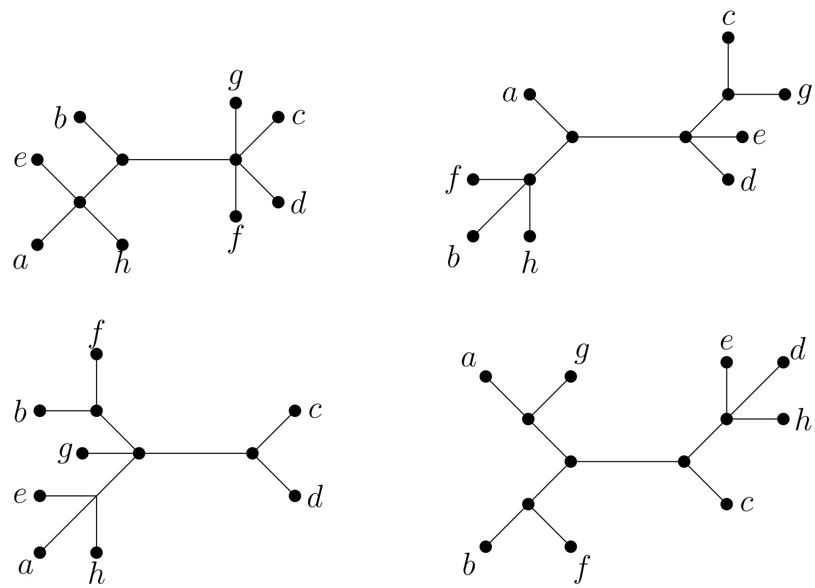


Figure 5.54: The trees above share only splits A_1, A_2, A_3

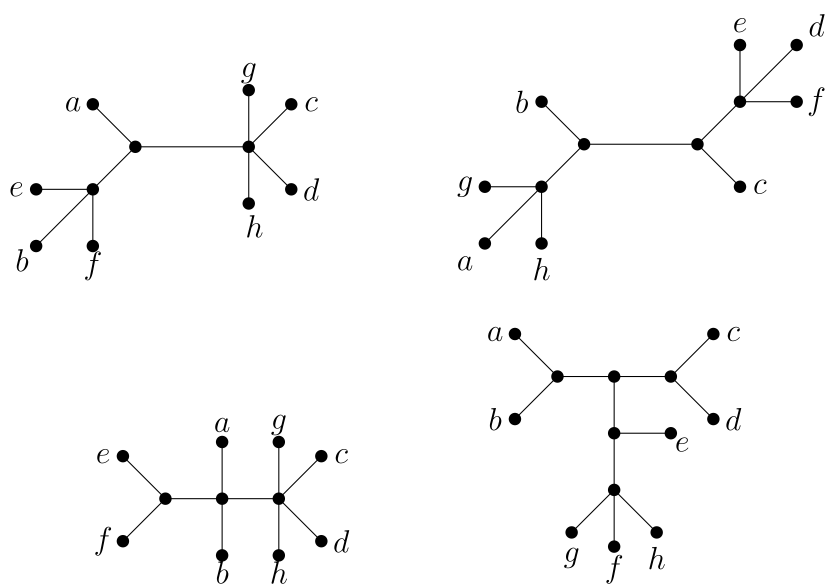


Figure 5.55: The trees above share only splits A_1, A_2, A_3

Subcase 6.6. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ac|ef \qquad A_3 = bg|eh.$$

We can see from Figure 5.56 below, that these splits do not yield any inferences.

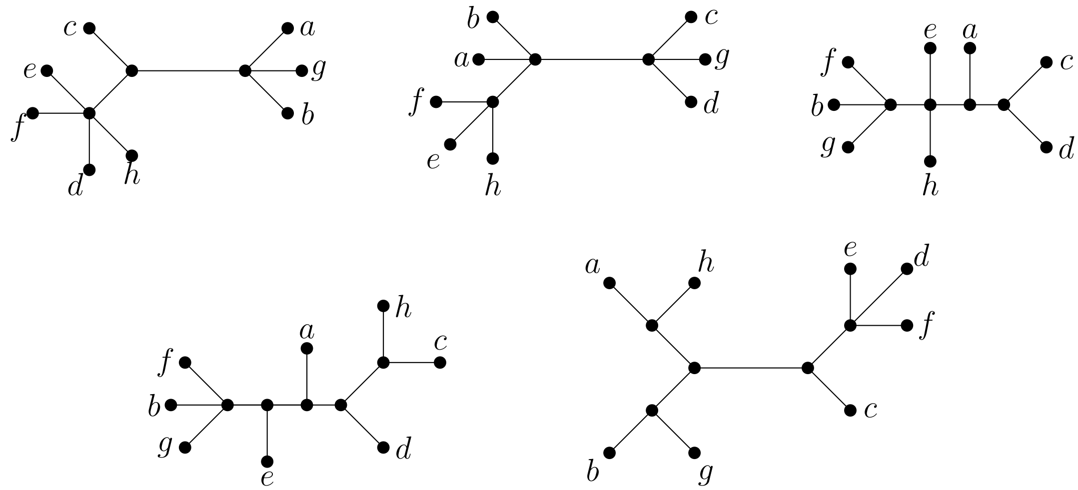


Figure 5.56: The trees above share only splits A_1, A_2, A_3

Subcase 6.7. *We have the following three splits*

$$A_1 = ab|cd \qquad A_2 = ae|cf \qquad A_3 = be|gh.$$

We can see from Figure 5.57 below however, that these splits do not yield any ternary inferences.

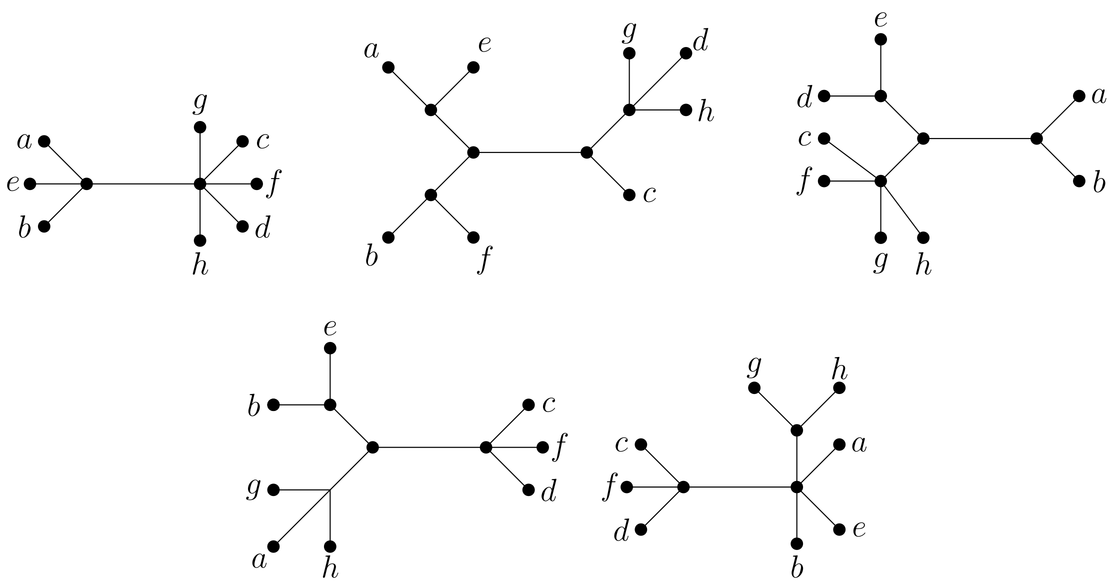


Figure 5.57: The trees above share only splits A_1, A_2, A_3

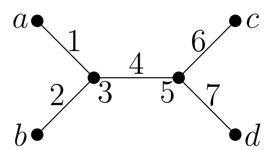
□

We can see that there do exist ternary influences from the example below.

Example 5.6. Consider the quartet splits

$$A_1 = ab|cd \qquad A_2 = ab|ef \qquad A_3 = ce|df$$

We refer back to Figure 4.8 from Chapter 4, reproduced below.



We need to adjoin the two new leaves e and f to this tree so that splits A_2 and A_3 are realized. We see from A_2 that neither e nor f can be placed at positions 1,2, or 3. Also, from A_3 we see that e and f cannot be adjacent to the same vertex. It is clear however that there are positions that will allow all three splits to be displayed by the same tree: for example, e at position 6 and f at postion 7. We can see that any valid positioning of e and f at positions 4,5,6, or 7 will indeed infer the new quartet $ab|ce$. This new split, along with the original three, then implies other splits using the binary inference rules from Theorem 4.5.

Bibliography

[1] Joshi Amitabh, *Genomics and evolution*, Curr. Sci. **82** (2002), 122–124.

[2] U. Bergmann et al, *Archaeopteryx feathers and bone chemistry fully revealed via synchrotron imaging*, Proc. Nat. Acad. Sci. USA **107** (2010), 9060–9065.

[3] Jean-Renauld Boisserie, Fabrice Lihoreau, and Michel Brunet, *Origins of hippopotamidae (mammalia, cetartiodactyla): towards resolution*, Zoologica Scripta **34** (2005), 119–143.

[4] David Bryant and Mike Steel, *Extension operations on sets of leaf-labelled trees*, Adv. in Appl. Math. **16** (1995), no. 4, 425–453. 1362849 (96j:05035)

[5] Marta Colomer-Lluch, Joan Jofre, and Maria Teresa Muniesa Perez, *Antibiotic resistance genes in the bacteriophage dna fraction of environmental samples*, PLoS ONE **6** (2011).

[6] Charles Darwin, *First notebook on transmutation of species*, 1837-1838.

[7] ———, *The origin of species*, Barnes & Noble Books, 1859.

[8] M.C.H. Dekker, *Reconstruction methods for derivation trees*, Master's thesis, University of Amsterdam, 1986.

[9] J. Peter Gogarten, W. Ford Doolittle, and Jeffrey G. Lawrence, *Prokaryotic evolution in light of gene transfer*, Mol. Biol. Evol. **19** (2002).

[10] Marcus R. Kronforst et al, *Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation*, Pigment Cell Melanoma Res. **25** (2012), 411–433.

[11] Jeffrey G. Lawrence, *Horizontal and vertical gene transfer: the life history of pathogens*, Contrib. Microbiol. **12** (2005), 255–271.

[12] Jonathan B Losos, *Convergence, adaptation, and constraint*, Evolution **65** (2011), 1827–1840.

[13] Gabriela Parra-Olea and David B. Wake, *Extreme morphological and ecological homoplasy in tropical salamanders*, Proc. Nat. Acad. Sc. USA **98** (2001), 7888–7891.

[14] Charles Semple and Mike Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, vol. 24, Oxford University Press, Oxford, 2003.

[15] R.Y. Stanier, J.L. Ingraham, M.L. Wheelis, and P.R. Painter, *General microbiology*, Macmillan, 1987.

[16] S. Titov, A. Shmyrov, and A Kuz'min, *Biotope principles of sympatry and interspecies hybridization in mammals*, Biol. bull. Russ. Acad. Sci. **39** (2012), 36–44.

[17] Richard J. Trudeau, *Introduction to graph theory*, Dover Publications Inc., New York, 1993, Corrected reprint of the 1976 original. 1262129 (95e:05031)

[18] E.O. Wilson and W.L. Brown, *The subspecies concept and its taxonomic application*, Syst. Biol. **2** (1953), 97–111.

[19] Naoko Yamaguchi, Shitaishi Susumu, Yoshinori Sasaki, and Chiaki Yamamoto, *Molecular evidence of the hybrid origins of leyland cypress*, J. of Forest Res. **5** (2000), 35–38.