A.1 The use of molecular and morphological characters in phylogenetic studies

Phylogenetic systematics is the study of the historical relationships among lineages, the recognition and the understanding of biodiversity. It attempts to determine the way in which characters change within groups over time (the direction in which characters change), and the relative frequency with which they change. It is useful in many different contexts as it makes it possible to compare the descendants of a single ancestor to evaluate the patterns of origin and extinction, or to determine the relative size and diversity of the groups (Richardson, 1999). Phylogenetics can use either morphological or molecular data in order to determine the pattern of relationships among taxa. These have both advantages and disadvantages, and each of them addresses questions that cannot be addressed by the other (Hillis & Wiens 2000; Hillis, 1987).

Morphological data can be extremely valuable in exploring evolutionary hypotheses for any given group. In general, morphological studies tend to be less costly, allow for greater taxonomic sampling, can be scored for many characters without damage to the specimens, and are the only method for analysing extinct taxa known only from the fossil record (Hillis & Wiens, 2000). However, inclusion of genetic information obtained by molecular techniques increases the number of characters and results in production of more robust phylogenetic hypotheses (Hillis & Wiens, 2000).

Molecular data, particularly DNA sequence data, are more versatile for phylogenetic studies than morphological data. The assumption of this methodology is that the similarities between genomes of individuals will help to develop an understanding of the taxonomic relationship among species. Molecular data have fewer problems with adaptive convergence and homology, can include a larger number of characters, provides more objective character scoring, and can identify evolutionary patterns among otherwise indistinguishable taxa, as well as distantly related lineages (Hillis & Wiens, 2000). Nonetheless, different portions of the genomes evolve at different rates, and as a result, a wide range of possibilities exists for resolving relationships.

A.2 Choosing an appropriate gene in the molecular phylogenetic study of *Passerina*

The field of molecular systematic uses data from DNA sequences to analyse relationships among species. An understanding of relationships among species allows us to produce useful taxonomic classification systems, to study the evolution of morphological characters, the biogeography, and speciation mechanisms. Plant cells have three different genomes that contain information that could be exploited to analyse species relationships (Judd *et al*., 1999), but for this molecular study, I have relied primarily on data from the chloroplast genome and the

ribosomal DNA sequences of the nuclear genome. These different genomes differ greatly in size, structure, and evolutionary rate (Bennetzena, 2002).

The chloroplast genome of land plants is highly conserved with most groups having the same general structure and gene order. It is a circular molecule that possesses two long inverted repeats (IR) which separate a large single copy region (LSC) from a small single copy region (SSC; Soltis & Soltis, 1998). The advantages of the use of the chloroplast genome for phylogeny reconstruction include the fact that it is small (120 and 200kb) and most genes are single-copy that makes it relatively easy to examine the entire genome (Soltis & Soltis, 1998). The genome is considered conservative in its evolution compared to the nucleur genome and evolves slowly at the nucleotide sequence level (Palmer, 1991; Downie & Palmer, 1992). However, different portions of the chloroplast genome evolve at different rates (Soltis & Soltis, 1998). In some instances changes can also occurred in the genome that include inversions, large deletions and insertions, gene losses and/or transfers, and intron losses (Jansen, 1999). These changes not only provide interesting insights into the evolution of chloroplast genomes but they also provide a wide range of possibilities for resolving relationships, from the level of species and genus to family and even higher levels (Soltis & Soltis, 1998; Jansen, 1999).

The nuclear genome is the DNA complement located in the cell nucleus. It is enormous and has more gene diversity, but despite this, most attempts to infer phylogeny with nuclear gene sequences have involved the nuclear ribosomal DNA (rDNA). Nuclear ribosomal genes are in transcription units, that occur in tandem repeats and each such repeat is reiterated thousands of times within most plant genomes (Appels & Dvorak, 1982; Appels & Honeycutt, 1986). These numbers of rDNA repeats units are highly variable (Riven *et al*., 1986; Bobola *et al*., 1992; Govindaraju & Culli, 1992). The rDNA gene family comprises many conserved regions (18S and 26S genes) that can be used to infer phylogeny at higher taxonomic levels, whereas rapidly evolving regions such as ITS is often best suited for comparing species and closely related genera (Soltis & Soltis, 1998). The fact that this gene family undergoes rapid concerted evolution permits the use of the gene for inferring phylogeny (Appels & Dvoral, 1982; Soltis & Soltis, 1998).

Genes used for resolving relationships within *Passerina* included three plastid genes (*rbcL, trnL-F* and *rps16*) and one nuclear gene (ITS):

i. The gene *rbcL* was one of the first plant genes to be sequenced (Zurawski *et al*., 1981; 1986). It is located in the large single-copy region of the genome, and encodes the large subunit of ribulose RUBISCO (1.5-bisphosphate carboxylase/oxygenase; Klein *et al*., 1994; Soltis & Soltis, 1998). The gene is typically 1434bp in length and has a broad range of applicability in terms of phylogenetic studies; from phylum all the way to genus level.

Analysis of *rbcL* sequences have been used to resolve generic relationships within several families of angiosperms (Xiang *et al*., 1993; Morgan *et al*., 1994; Kim & Jansen, 1996; Soltis & Soltis, 1998).

ii. Other noncoding sequences of the chloroplast genome included the supposedly faster evolving regions, *rps16* and *trnL-trnF* (*trnL* (UAA) intron and the intergenetic spacer between the *trnL* (UAA) 3' exon and the *trnF* (GAA) gene). The *trnL-F* region is relatively small, with the *trnL* intron ranging from 350-360bp and the *trnL-trnF* spacer from roughly 120-350bp (Gielly & Taberlet, 1994; 1996).These noncoding plastid DNA regions are length-variable and depending on the study group, these regions may evolve at rates similar to that of *rbcL*, to as much as three times faster than *rbcL* (Soltis & Soltis, 1998).

iii. Rapidly evolving regions such as the internal transcribed spacer (ITS) regions of 18S-26S nuclear rDNA are use in comparative sequencing at the specific and generic levels in angiosperms (Baldwin *et al*., 1995). The internal transcribed spacers are part of the nuclear rDNA transcript but not incorporated into ribosomes. According to Hershkovitz and Zimmer (1996), ITS-1 and ITS-2 sequences are inherently G+C rich and that portions of these regions are quite conserved among flowering plants. ITS regions not only possess high information content at lower taxonomic levels, but also exhibit conserved sequence patterns and are highly alignability over a broad taxonomic scale (Hershkovitz & Zimmer, 1996; Soltis & Soltis, 1998).

To be able to take advantage of molecular data, however, one must understand the methodology of molecular phylogenetics, and the aim of the next section is to provide some basic principles of phylogenetic inference.

A.3 Analyses of phylogenetic relationships

There are two major groups of analyses to examine phylogenetic relationships between sequences namely distant and discrete methods. Distance methods first convert aligned sequences into pairwise distance matrix, and then that matrix is put into a tree building method, whereas discrete methods consider each nucleotide site (Page & Holmes, 1998).

A.3.1 Analysis based on distance methods

Distance methods involve the use of various measures of overall similarity for the ranking of species. In this approach, a tree is constructed by considering the similarities of sequences that does not necessarily reflect evolutionary relationships. Starting with a multiple alignment, pairwise distances are calculated between DNA sequences as the sum of all pair differences between two sequences. This creates a distance matrix (Hall, 2001; Felsenstein, 2004). From

the obtained distance matrix, a phylogenetic tree is calculated with clustering algorithms (will be discussed later). These cluster methods construct a tree by linking the least distant pair of taxa, followed by successively more distant taxa. The two most popular distance methods in current use are UPGMA clustering (Unweighted Pair Group Method using Arithmetic Average) and Neighbour Joining (NJ) where NJ attempts to correct UPGMA for its assumption that the rate of evolution is the same in all taxa (Hall, 2001). These methods have the advantage of being easy to implement, resulting in very fast computer programs. However, the taxonomic clusters that result from such an analysis do not necessarily reflect genetic similarity or evolutionary relatedness. The lack of evolutionary significance has meant that this system has had little impact on classification and has been declining in recent years (Hall, 2001).

Unweighted Pair Group Method using Arithmetic Average (UPGMA)

This method has gained popularity mostly because of its simplicity and because of its speed. It is one of the few tree-building methods that construct an ultrametric tree (Swofford, *et al*., 1996). In an ultrametric tree all the tips are equidistant from the root of the tree, which is equivalent to assuming a molecular clock, an assumption that is very unlikely. The method does not consider rate heterogeneity and can therefore produce an incorrect topology if some lineages have evolved faster than others. For that and other reasons, UPGMA is rarely used today (Page & Holmes, 1998; Richardson, 1999; Hall, 2001).

Neighbour Joining (NJ)

Neighbour Joining (Saitou & Nei, 1987) is similar to UPGMA in that it manipulates a distance matrix, reducing it in size at each step, and then reconstructs an unrooted tree without the assumption of a clock. It differs from UPGMA in that it does not construct clusters but directly calculates distances to internal nodes. The least distant pairs of nodes are linked and their common ancestral node is added to the tree, and their terminal nodes with their respective branches are pruned from the tree. This continues until only two nodes remain, separated by a single branch (Swofford *et al*., 1996). It is a widely used method for tree building which combines computational speed with uniqueness of results with most implementations gives a single tree. The branch lengths are not optimised by the least squares criterion but the methods are very fast and thus can handle much larger data sets. Neighbour joining is a clustering method rather than an optimality method, and hence suffers from the limitation that it does not guarantee finding the tree with the smallest overall distance (Page & Holmes, 1998).

A.3.2 Analysis based on discrete methods

In contrast to distance methods, discrete methods operate directly on the sequences, or on functions derived from the sequences, rather than on pairwise distance (Page & Holmes, 1998). This method relies on knowledge of ancestral relationships and the position of each alignment. Via this method, a tree is reconstructed by considering the various possible pathways of evolution and choosing from amongst these the best possible tree. In order to use this software with sequence data, certain sequences must be designated as ancestral and others as derived. As a result, changes at certain positions will have a larger effect than others on the location of each sequence in the predicted tree.

The basic assumption behind character distinction is dependent on two assumptions (Swofford *et al*., 1996). The first is that any group of species is related by descent from a common ancestor, by either separation or bifurcation. The second is the assumption that change in characteristics occurs in lineages over time. It is only when characteristics change that we are able to recognise different lineages or groups. Therefore, in contrast to distance methods, discrete groupings do not depend on whether species share morphological traits but depend on their evolutionary relationships. For character data and for higher levels of taxonomy, the discrete approach is almost certainly superior. Discrete methods are now accepted as the best method available for phylogenetic analysis, for it provides an explicit and testable hypothesis of relationships (Page & Holmes, 1998; Richardson, 1999). However, discrete methods are often difficult to implement with assumptions that are not always satisfied with molecular data.

Computer algorithms based on the discrete model included Maximum Likelihood (ML), Bayesian methods and Maximum Parsimony (MP). All methods use the multiple alignments directly by comparing characters within each site in the alignment (Hall, 2001).

Maximum Likelihood (ML)

The method of ML attempts to reconstruct a phylogeny by finding that tree that maximises the probability of the observed data (Page & Holmes, 1998; Hall 2001). This method also uses each position in the alignment, evaluates all possible trees, and calculates the likelihood for each individual tree using an explicit model of evolution. Since each nucleotide site evolves independently, the product of the likelihood is for each site provides the overall likelihood of the observed data (Richardson, 1999). The advantage of the ML methods is that it allows users to specify the evolutionary model they want to use, and seems to give the best result and most information about the tree (Hall, 2001). Unfortunately, this is computationally difficult to do and even with simple models of evolutionary change, the computational task is enormous and this is

thus the slowest of all methods. This method works best when it is used to test (or improve on) an existing tree (Siddall & Kluge, 1997; Hall, 2001).

Bayesian analysis

Bayesian inference of phylogenies (Rannala & Yang 1996; Mau & Newton 1997; Mau *et al.,* 1999) is closely related to ML methods in that the user postulates a model of evolution and the program searches for the best trees that are consistent with both the model and the data (Felsenstein, 2004). Instead of seeking the tree that maximises the probability of observing the data, Bayesian methods seeks those trees that have the greatest probability, given the data and the model for evolution (Hillis *et al.,* 1996). Unlike ML, that seeks the single most likely tree, Bayesian analysis searches for the best set of trees (Hall, 2001). The results of a Bayesian analysis are easy to interpret, because the frequency of a given clade in that set of trees is virtually identical to the probability of that clade, so no bootstrapping is necessary to assess the confidence in the structure of the tree. The method is easy to use, fast and capable of dealing with very large phylogenies (Hall, 2001).

Maximum Parsimony (MP)

Maximum parsimony is based on the assumption that the most likely tree is the one that requires the fewest evolutionary changes for all sequences to derive from a common ancestor (Hall, 2001), thus minimising homoplasy and maximising congruence between the characters. Finding a maximally parsimonious cladogram is usually computationally intensive task requiring computer analysis and is a more time-consuming method than the distance methods (Page & Holmes, 1998; Weston & Crisp, 1998; Hall, 2001). There are three basic steps to find the most parsimonious tree/trees, which are outlined in the next section.

A.3.3 Variants of parsimony (optimally criterion)

The parsimony principle helps to optimise complex trees (Kitching, 1992) by minimising the number of changes in character states. There are a number of parsimony methods, all of which choose the most parsimonious tree as that tree that minimizes the total number of character state changes across the entire tree (Holsinger & Jansen, 1993; Liu, 1998). The criterion used to set the number of changes between each state is called the optimality criterion and will be discussed.

1. Wagner Parsimony (Wagner, 1961)

Wagner parsimony assumes that ancestral states are unknown, and that roughly equal rates of substitutions can occur in either direction, among all states (Holsinger & Jansen, 1993; Liu, 1998). The criterion is to find the tree, which requires the minimum number of changes. For a binary character, a change from state zero to state one is given equal weight to that of a change from state one to state zero. This means that an unrooted tree can be rooted at any point without changing its length (Richardson, 1999). Since the characters are measured on an interval scale, this method is appropiate for binary, ordered multistate and continuous characters (Kitching, 1992).

2. Fitch Parsimony (Fitch, 1971)

Characters with three or more states are unordered, which means that they transformed directly into any other state. This criterion was formulated for DNA sequences, which have four character states. Wagner and Fitch parsimony criteria are appropriate whenever those character state changes are symmetrical that change from 0 to 1 and that it has the same probability as a change from 1 to 0 (Richardson, 1999).

3. Dolly Parsimony (Dolly, 1893)

In Dolly parsimony, each character state is allowed to be gained only once and if the distribution of character states does not fit, it is explained by extra reversals (losses). This is appropriate when the probability of a reverse change (1 to 0) is zero. It has been proposed as a way to analyse restriction site data, where the probability of a loss is much higher than the probability of a gain (Holsinger & Jansen, 1993; Liu, 1998) It requires specification of polarity (to know what states are gained), and is generally considered unnecessarily strict and unrealistic (Felsenstein, 2004).

4. Camin-Sokal Parsimony (Camin & Sokal, 1965)

This was the first parsimony methods proposed, and perhaps the simplest parsimony method (Felsenstein, 2004). Here character evolution is assumed irreversible and no reversals are allowed, but only multiple gains (Swofford *et. al*., 1996). This method explains the data by allowing the changes 0 to 1 but not changes 1 to 0 (Felsenstein, 2004). It is therefore probably more appropriate for use with morphologic traits rather than molecular data (Holsinger & Jansen, 1993; Liu, 1998).

A.3.4 Additive and ultrametric trees

The output from a phylogenetic analysis is a hypothesis of relationships of different taxa. A variety of trees is used to depict different aspects of evolutionary history (Page & Holmes, 1998). Distance trees can be either additive or ultrametric as shown in Figure A.1. In additive trees, the distance between pairs of taxa equals the sum of the lengths of the branches connecting them. These numbers associated with each branch correspond to the amount of evolutionary change (Swofford *et al*., 1996; Page & Holmes, 1998), and do not assume a molecular clock as in the case of ultrametric trees.

Ultrametric trees are a special kind of additive tree in which the tips of the tree are all equidistant from the root and can be used to depict evolutionary time, expressed directly as years or indirectly as the amount of sequence divergence using a molecular clock (Swofford *et al*., 1996; Page & Holmes, 1998).

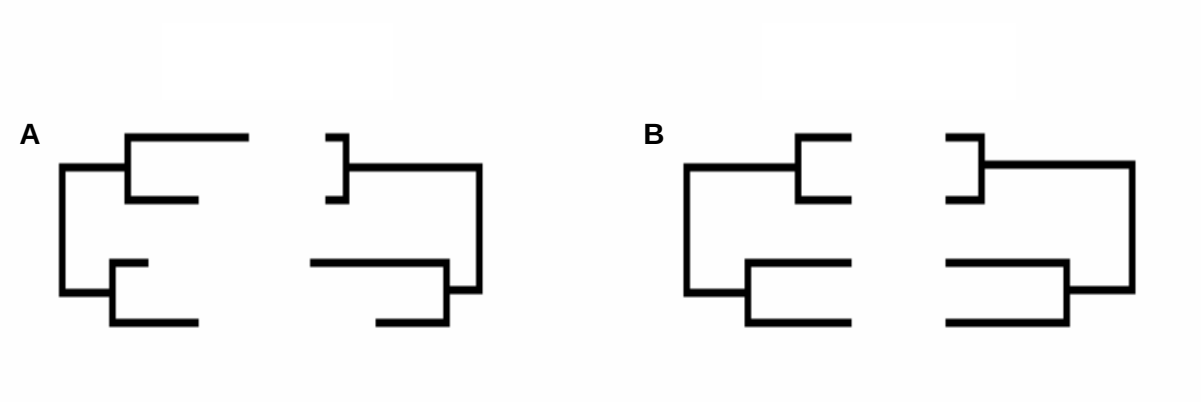

Figure A.1 Example of an additive (A) and ultrametric tree (B; Page, 1995).

A.3.5 Tree building methods

Tree-building methods use optimal criteria to choose among the set of all possible trees. Optimality methods have the great advantage of requiring an explicit function that relates data and tree (Page & Holmes, 1998). PAUP provides two basic classes of methods for searching for optimal trees that include exact and heuristic methods. Exact methods (exhaustive search and branch-bond search) guarantee to find the optimal tree, but it requires a lot of computer time and can handle only small matrices, typically less than 20 taxa or sequences. For larger problems, fast heuristic algorithms must be employed, and although they cannot guarantee to find the optimal tree, it generally requires far less computer time (Weston & Crisp, 1998).

A.3.5.1 Exact methods

Exhaustive search

When the number of taxa is small, an exhaustive search (Figure A.2) guarantees to find the best tree by evaluating the individual trees. If every possible tree is evaluated, the best tree will be found. In this search the first three taxa are connected with only one possible unrooted tree. The fourth and fifth taxa are added and evaluated in every topology. This is repeated until all taxa are added. Difficulty with this search is that the number of possible trees increases rapidly with the addition of further taxa and the search become very large (Kitching, 1992; Swofford *et al*., 1996).

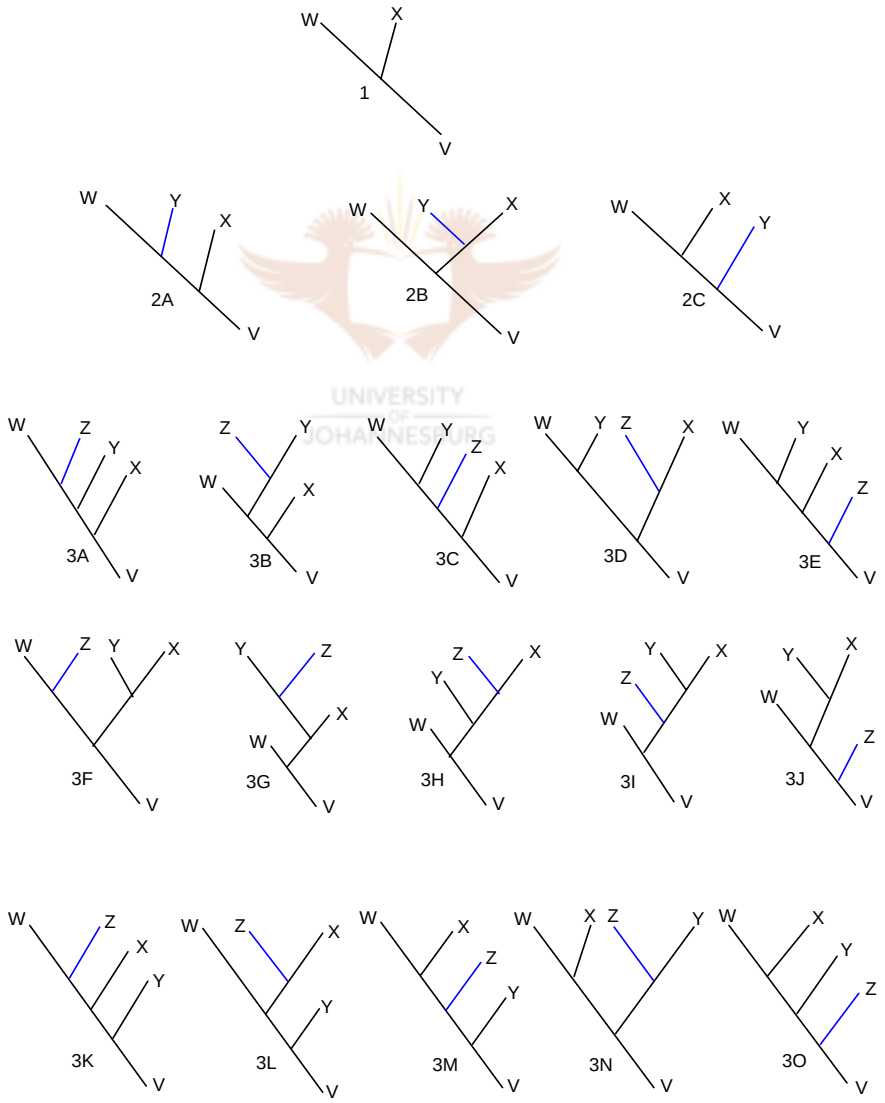

Figure A.2 Illustration of an exhaustive search (Swofford *et al*., 1996).

Branch-and-bound search

This algorithm also guarantees to find the best tree, but has a provision for discarding some of the trees and evaluates only the subset of all possible trees. This can be achieved by assembling a tree by adding one taxon at a time (Figure A.3). Three taxa are used to obtain the first tree. Another taxon is then added to this tree which results in three possible tree topologies. The fifth taxon may then be placed into five different places within each of the three trees yielding 15 trees altogether. If an upper bound is set with regard to the tree length, all trees which exceed this need not to be considered further (Swofford *et al*., 1996). The success of such a method depends on how messy (homoplasy) the data are, that can cause a decrease in efficiency. This algorithm may be useful for data sets of up to about 15-20 taxa.

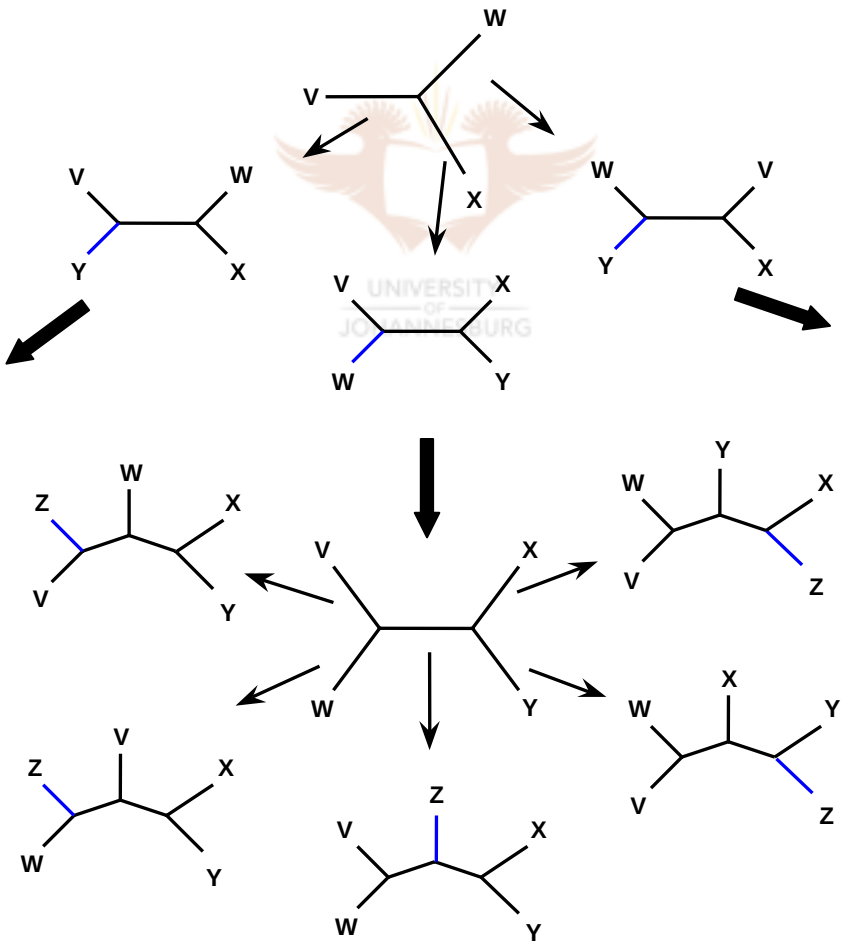

Figure A.3 A diagrammatic representation of the branch-and-bound approach (Felsenstein, 2004).

A.3.5.2 Heuristic search

When the data set is too large and evaluating each tree would be too slow to permit using the exact methods, a heuristic strategy are used (Hall, 2001). However, these methods do not guaranteed to find the best tree(s), but are useful when working with large numbers of taxa greater than 20 (Hillis *et al*., 1996). A heuristic approach is essentially a hill-climbing algorithm (Felsenstein, 2004) that usually proceeds in two steps. First, the initial tree is constructed, by adding one taxon at a time, placing the added taxon on the branch which gives the best tree for the subset. When a tree contains all taxa at hand, the second step tries to find a better tree by moving subtrees to other branches, keeping any rearrangement that produce a better tree (Page & Holmes, 1998; Hall, 2001). There are variations on the procedures in both steps, and their performance will depend on the data at hand.

1. Obtaining the initial tree(s)

There are several algorithms to obtain the initial tree(s), and all are correct but some are more efficient depending on the data.

Stepwise addition

One common approach is the stepwise addition method. It starts with a three-taxon tree, than ads branches to make each of the three possible four-taxon trees. At this point, each of the trees is evaluated and the one with the best score is selected to make the five possible five-taxon trees that can be derived from it. At each level, only the best of the trees are used to add the next taxon (Hall, 2001). The process is termined when all taxa have been joined to the growing tree. The problem with taxon addition is to decide which three taxa should be used for a starting tree and which of the unplaced taxa should be attached to the tree next (Hillis *et al.,* 1996; Swofford *et al*., 1996).

Star decomposition

An alternative alternative to stepwise addition is star decomposition. The algorithm start with all taxa connected in a star topology (connected to a single internal node). Then all the trees that can be constructed by joining two of the terminal nodes in a new group are evaluated and the tree with the best score is used for the next step. In each step when a new group is form, the number of branches connected to the central node is reduced by one. This continues until a dichotomous tree is found (Hall, 2001).

2. Improving the initial tree(s)

This step is commonly referred to as branch swapping. Branch swapping involves making predefined rearrangements of trees by a variety of means (Hall, 2001). These arrangements are performed in the hope that a better tree will be found. Three methods of branch swapping are used, namely Neighbour Interchange (NNI), Subtree Pruning and Regrafting (SPR) and Tree Bisection and Reconnection (TBR; Swofford *et al*., 1996).

Nearest Neighbor Interchange (NNI)

In this process (Figure A.4), the interior branches on the tree are identified and erased and the two branches connected to it at each end. The four subtrees disconnected from each other are then reformed in one of the two possible alternative ways (Felsenstein, 2004).

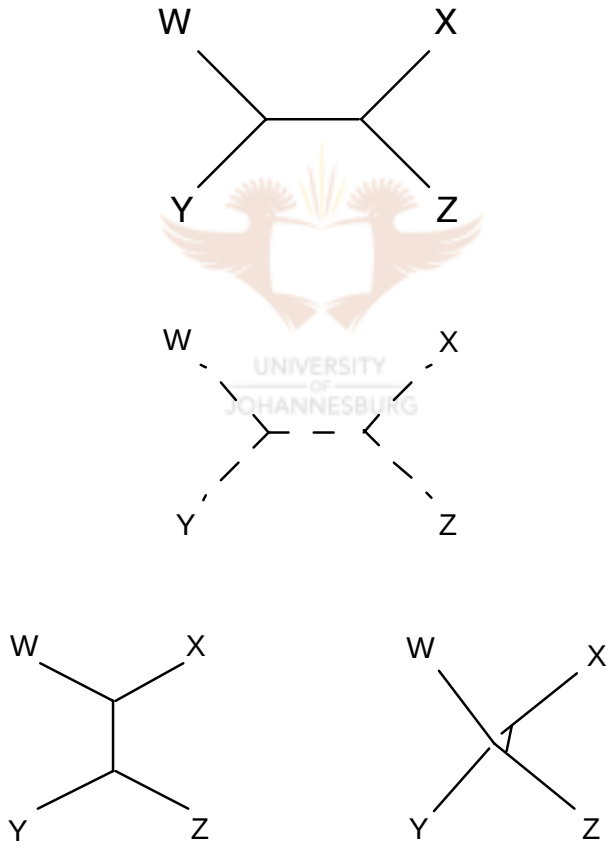

Figure A.4 The process of Nearest-Neighbour Interchange (NNI; Felsenstein, 2004).

Subtree Pruning and Regrafting (SPR)

This algorithm is very similar to the NNI process, but with a more elaborate branch swapping scheme (Figure A.5). It consists of identify and removing a branch from the tree with a subtree attached to it. The subtree is then reattached to each possible branch of the remaining tree, in

order to find a shorter tree (Kitching *et al*., 1998; Felsenstein, 2004). All the possible subtree removals and reattachments are then evaluated. While this method is slower than NNI, it will often find shorter trees (Kitching *et al*., 1998).

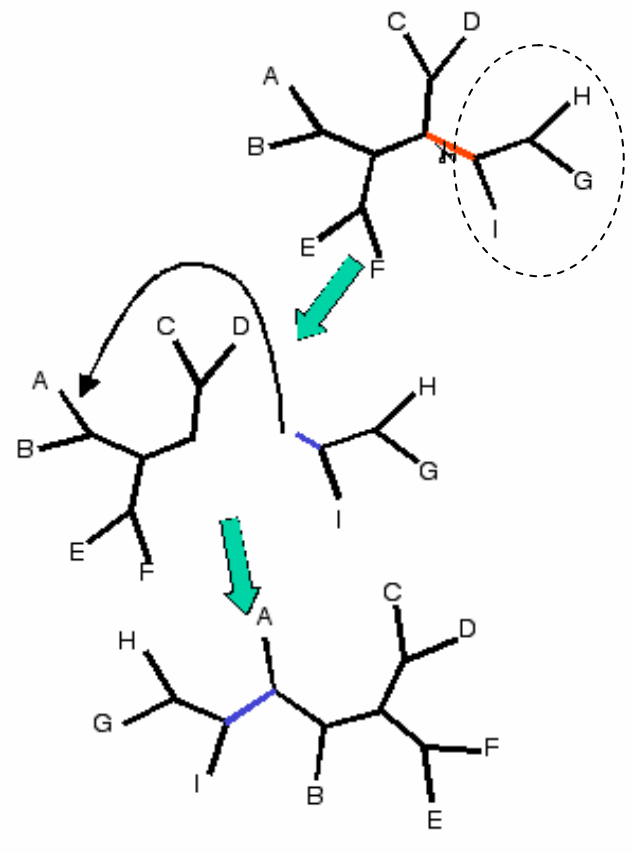

Figure A.5 Subtree Pruning and Regrafting (SPR) rearrangement (Swell & Thollesson, 2001).

Tree Bisection and Reconnection (TBR)

In this algorithm (Figure A.6), an interior branch is broken and the two resulting fragments are considered as separate trees. Again all possible bisections and reconnections are evaluated to find a shorter tree (Felsenstein, 2004). TBR will often find shorter trees than SPR and NNI, at the cost of longer computation time (Kitching *et al*., 1998).

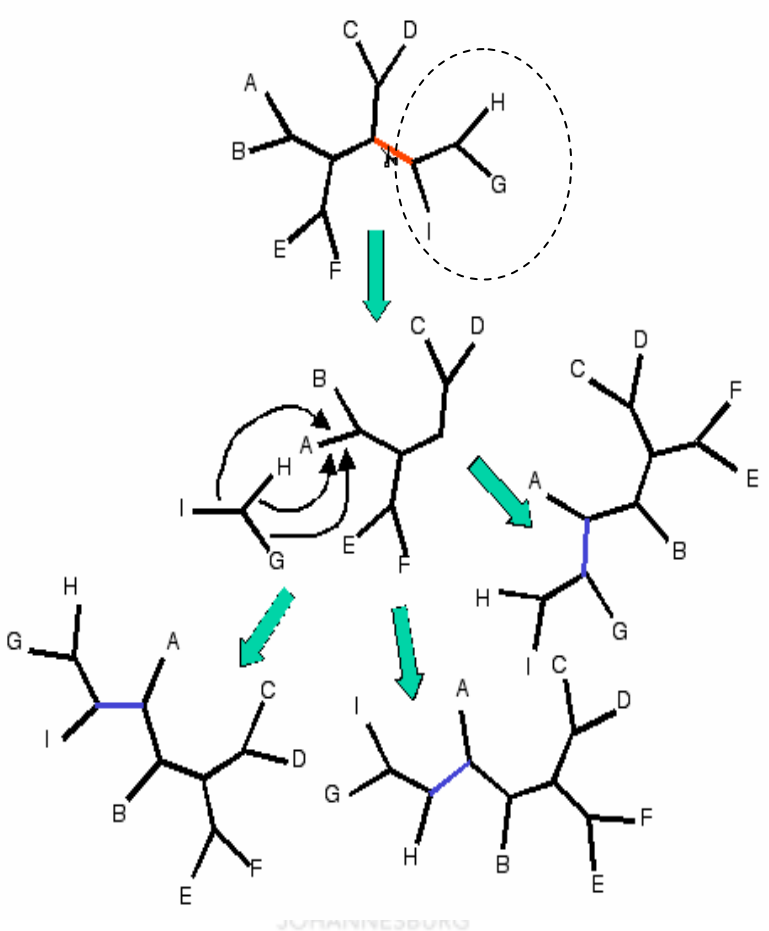

Figure A.6 Tree Bisection and Reconnection (TBR; Swell & Thollesson, 2001).

A.3.6 Credibility of the hypotheses

These methods allow us to evaluate the quality of any tree, thus we can compare how well competing hypotheses of evolutionary relationship fit the data (Page & Holmes, 1998).

1. Measures of the robustness of a cladogram

The Consistency Index (CI) serves to measure the relative amount of homoplasy in a cladogram and to explain all distributions of character states over taxa (Kluge & Farris, 1969). If the characters in the data set are perfectly congruent with each other, without homoplasy, the observed number of steps will equal the minimum, and the CI will be one (one is the highest possible value). The more chaotic the data on the tree, and the greater the amount of homoplasy, the greater the observed number of steps, and the more the CI shrinks. The CI value is very limited and highly correlated within the datasets and therefore only used as a comparative metric within the datasets (Kitching *et al*., 1998).

A second index used is the Retention Index (RI) that measures the amount of synapomorphy in a data set, by examining the actual amount of homoplasy, as a function of maximum possible homoplasy (Farris, 1989; Kitching *et al*., 1998). Farris (1989) introduced RI as a substitute for CI, since CI do not remove autapomorphies (which have an automatic CI of 1.0) and is highly correlated with the number of taxa in a data set (Leseure, 1998). For example, in a binary data matrix, the maximum number of steps on the tree is the total number of taxa with state one or state zero, summed for all characters. The closer the RI is to one the better the tree is considered to be.

2. Estimating the reliability of phylogenetic trees

Several methods have been developed for quantifying the level of congruence, and in particular, for providing a measure of the level of support for groupings of taxa (Weston & Crisp, 1998). I will only discuss bootstrap and Jack-knife methods.

Bootstrap method

Bootstrap analysis is the most commonly used method. It involves creating a new data set by resampling 100 to a 1000 characters randomly with replacement, so that the resulting data set has the same size as the original, but some characters are deleted and others duplicated (Felsenstein, 1985; Weston & Crisp, 1998). These bootstrap data sets are each analysed in the same manner as for the real data, and the number of times that each grouping of species appears in the resulting profile of cladograms, it can be consider as an index of relative support for that grouping (Siebert, 1992; Weston & Crisp, 1998). The method assumes that the characters evolve independently, an assumption that may not be realistic for many kinds of data (Felsenstein, 1985; Hall, 2001).

Jack-knife method

Whereas bootstrapping applies resampling with replacement, the Jack-knife technique applies resampling without replacement. Another way to view sampling without replacement is that a certain proportions of the different sites in a sequence is randomly omitted from the analysis (Siebert, 1992; Felsenstein, 2004).

A.3.7 Consensus trees

The consensus trees represent the commonality among a set of trees produced by a search. The tree summarises information common to two or more trees, where all (or most) of the branch points are in common. Different types of consensus trees are described below and each is calculated differently to answer different kinds of questions.

The strict consensus tree (Figure A.7) is the simplest approach for constructing a consensus tree. It includes only those monophyletic branches occurring in all the original trees, so it is the most conservative consensus. This constraint is particularly restrictive and usually results in a tree with unresolved polytomies (Koning, 1994; Kitching *et al*., 1998; Page & Holmes, 1998; Felsenstein, 2004).

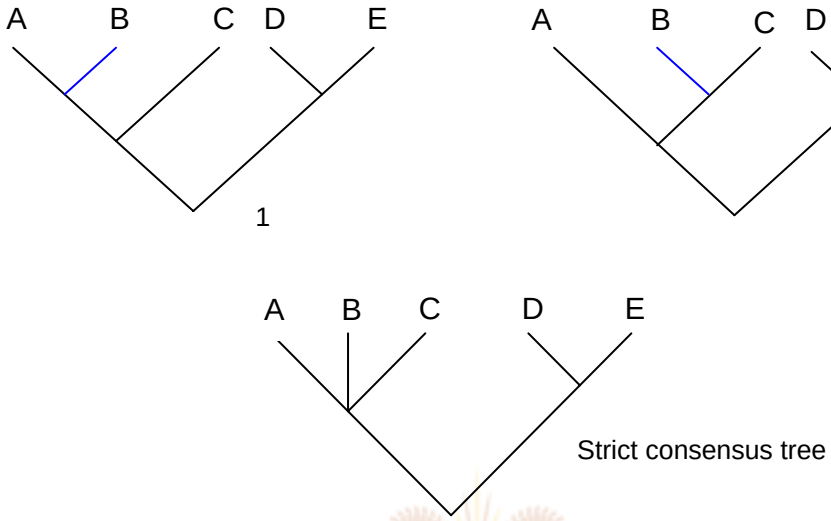

Figure A.7 Two trees (1 and 2) and their strict consensus tree (after Felsenstein, 2004).

A common alternative is the 50% majority-rule consensus tree (Figure A.8), where any split in more than half the trees is included in the consensus tree, so two splits shared by two of the trees are also included. However, the trees that have no split in common may still have points of similarity (Kitching *et al*., 1998; Page & Holmes, 1998).

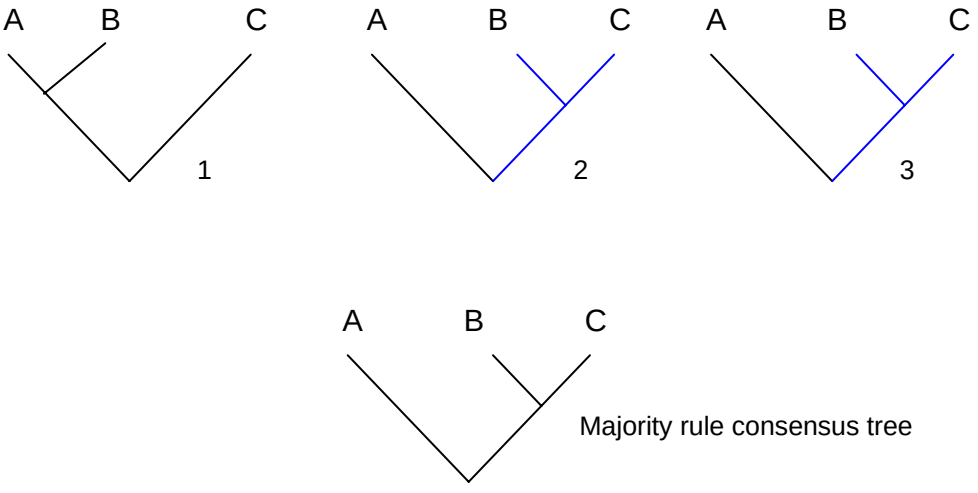

Figure A.8 Three trees (1, 2 and 3) and their majority rule consensus tree (after Felsenstein, 2004).

The Adams consensus tree (Figure A.9; Adams, 1972), can sometimes be difficult to inter, but it is very useful for summarise similarities in topology among fundamental trees when they contain one or more taxa that have very different positions on different trees (Page & Holmes, 1998). If some trees are not logically consistent, conflicting branches relocated to the nearest common node. Therefore, this consensus does not necessarily reflect the same monophyletic groups supported by the original data.

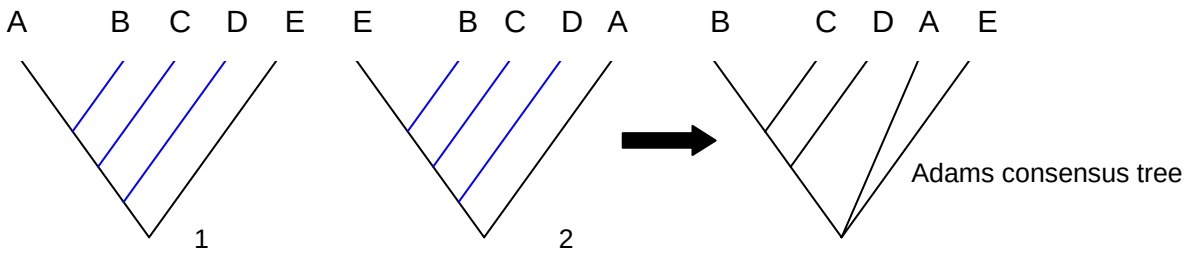

Figure A.9 Two trees (1 and 2) and their adams consensus tree (after Felsenstein, 2004).

A.3.8 Choice of outgroups

An important characteristic of molecular phylogenetic reconstructions is the inclusion of outgroups. An outgroup is a group of several species used in phylogenetic analysis that are less related to any of the study taxa than any pair of ingroup taxa are related to each other. Outgroups are included in the analysis, to provide a root for the tree and to help resolve the polarity of characters. The most ideal outgroup, is the sister group that is most closely related to the study group. The outgroup taxa are treated exactly as the ingroup analysis. When the analysis is done and a resulting tree is achieved, the root from the perspective of the ingroup will be on the edge between the ingroup and outgroup. If outgroup taxa are found scattered among ingroup taxa, the outgroup was not really an outgroup and another one must be selected (Swofford *et al*., 1996).

A.3.9 Choice of methods used to analyse sequence data

A variety of methods is used to construct phylogenetic trees. For any given set of data, these models may be violated because of various occurrences. Therefore, when interpreting a given analysis, it is important to understand most of the methods to select the best method for constructing phylogenetic trees. No method is ideal for all performance criteria. Some of the desirable properties that need to be considered when choosing a method include efficiency, power, consistency, robustness and falsifiability.

Efficiency is a measure of how quickly the method converges on the correct tree as the amount of data increase. An efficient method will take shorter time to produce a result than a not-so-efficient method (Hall, 2001). The power of a method is a measure of how much data we need to collect before we can be reasonable sure of arriving at the correct result. A powerful method will need less data to reach a conclusion with the same level of confidence as a less powerful method. Another consideration is whether the method will converge towards the correct solution as more data is added. The desirable property is consistency while an inconsistent method can give a positively wrong result, given enough data (Page & Holmes, 1998). The tree-building methods make assumptions about the data and/or the evolutionary process, assumptions that is not necessarily met. Robustness measures how well the method can tolerate deviations from its assumptions and still recover the correct tree (Page & Holmes, 1998; Hall, 2001). Ideally, we would like to know whether these violations are sufficient to rule out a particular model and such method is falsifiable when data reject the appropriateness of a method (Page & Holmes, 1998).

The ideal tree-building method would meet all five criteria, but such a method does not exist. There are often tradeoffs among these criteria in that methods that increase one measure decrease another (Hillis *et al*., 1996). Algorithmic methods (NJ; UPGMA) are usually very efficient, but have no falsifiability and are not robust to its implicit assumption of a molecular clock. Maximum likelihood methods on the other hand are falsifiable and consistent, but with low efficiency (Kitching *et al*., 1998; Page & Holmes, 1998). Therefore, distant methods, MP and ML (Bayesian) were selected to analyse the datasets, as parsimony have a high power and makes only few assumptions about evolutionary processes.