

"WAGNER TREES"

Wagner Trees (and Networks) are estimates of evolutionary trees. The construction of a Wagner Tree begins with a character x taxon matrix. In the matrix the states of characters are coded numerically, and the initial direction of evolution of the states within each character is specified. The Wagner Tree for a given data set is the cladistic topology that requires the minimum number of evolutionary steps, i.e., it makes the fewest number of assumptions about the number of evolutionary steps and it is the "most parsimonious" hypothesis overall.

The chief references on Wagner Trees are the papers by Wagner (1969) and Farris et al. listed on the references handout.

I. Initial treatment of characters

- A. Discover characters in the taxa under study. This involves estimating homologous relations of features in different organisms using the criteria noted in class. Verbalize or, if possible, enumerate the states of each character.
- B. Construct a morphocline for the states of each character. A morphocline is an arrangement of states in which the states are ordered on the basis of their overall similarity. For example:

(a) 5 toes - 4 toes - 3 toes - 2 toes - 1 toe

(b) 5 mm - 20 mm - 21 mm

(c) blue - green - yellow

(d) eyes present - eyes absent

(e) horizontal

stripes



vertical

bars



stripes

&

bars



stripes & bars

& dots in the middle



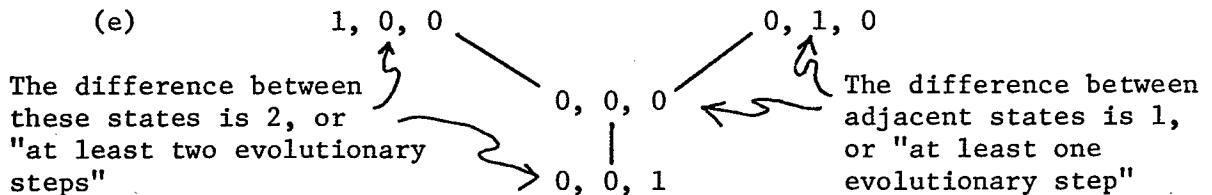
Morphoclines are estimates of the pattern of evolution within a character. Thus, if a 5 toed ancestor gave rise to a 1 toed descendant we would suppose a series of intermediate ancestors with 4, 3 and 2 toes respectively.

- C. Estimate an ancestral state for each character using the criteria for primitiveness discussed in class. The selection of an ancestral state yields a hypothesis on the direction of evolution for a character. For example:

blue → green → yellow,
 or,
 blue ← green → yellow,
 or,
 blue ← green ← yellow.

- D. Using integers in an additive fashion, and a similar scale for each character (by convention a unit difference between adjacent states in a morphocline) code numerically to preserve the estimated form and direction of character state evolution. The coding schemes below correspond to the morphoclines verbalized on page 2.

- (a) 0 - 1 - 2 - 3 - 4 or 4 - 3 - 2 - 1 - 0
- (b) 0 - 1 - 2 or 2 - 1 - 0
- (c) 0 - 1 - 2 or 2 - 1 - 0
- (d) 0 - 1 or 1 - 0
- (e)



II. The ANCESTOR

The most recent common ancestor for the group of taxa under study is taken to be the collection of all estimated ancestral states. Add the ancestor to the coded character x taxon data matrix.

III. The Steps in Tree Construction & a Hypothetical Example

The steps in the construction of a Wagner Tree are illustrated here with a hypothetical data set of 4 taxa and 12 characters (see attached sheet). The characters are already coded and an ancestor for the four taxa has been estimated and added to the initial data matrix.

Step 1. Calculate the difference (similarity) matrix according to the formula:

$$d(J,K) = \sum_{i=1}^n |X(J,i) - X(K,i)|$$

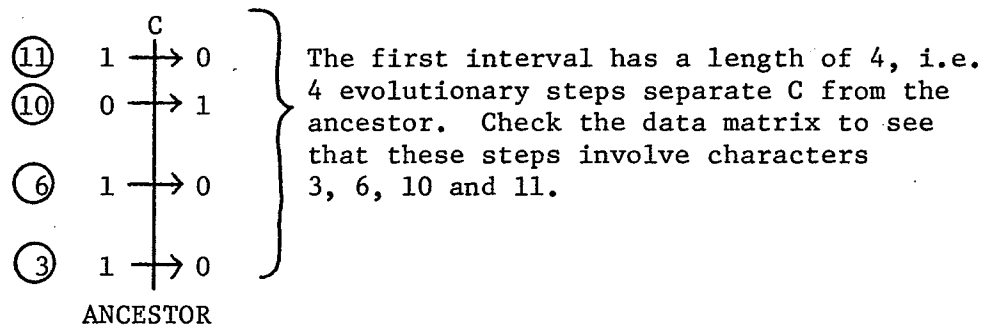
where $d(J,K)$ is the difference between two taxa J & K ,

X is the character state value for a character i ,

n is the number of characters

[the difference matrix based on the hypothetical data set is below the data set]

Step 2. Form the first interval of the tree by connecting the ancestor to the taxon to which it is most similar.



Step 3. Select the next taxon to be placed on the tree. Here we must find the closest "unplaced taxon-interval pair," i.e. the next taxon to be placed is that one which is most similar to any interval of the tree.

Calculate the difference between each unplaced taxon and every interval according to the formula:

$$d(L, INT(J,K)) = (d(L,J) + d(L,K) - d(J,K)) \left(\frac{1}{2}\right)$$

where $d(L, INT(J,K))$ is the difference between an unplaced taxon L

and an interval composed of the taxa J and K.

$d(L,J)$, $d(L,K)$ and $d(J,K)$ are the distances between these taxa taken directly from the difference matrix.

Thus in the example, there are three remaining unplaced taxa A, B and D.

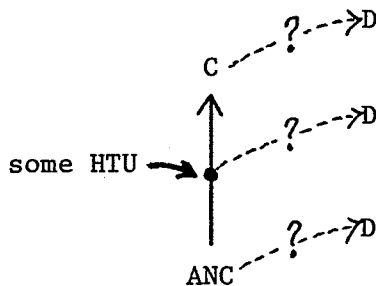
$$\begin{aligned} d(A, \text{INT}(ANC, C)) & \\ &= d(A, ANC) + d(A, C) - d(ANC-C) \\ &= 7 + 11 - 4 \\ &= \textcircled{14} \times \frac{1}{2} = 7 \end{aligned}$$

$$\begin{aligned} d(B, \text{INT}(ANC, C)) & \\ &= d(B, ANC) + d(B, C) - d(ANC, C) \\ &= 9 + 11 - 4 \\ &= \textcircled{16} \times \frac{1}{2} = 8 \end{aligned}$$

$$\begin{aligned} d(D, \text{INT}(ANC, C)) & \\ &= 6 + 8 - 4 \\ &= \textcircled{10} \times \frac{1}{2} = 5 \end{aligned}$$

Since D is the closest to the interval it shall be placed next.

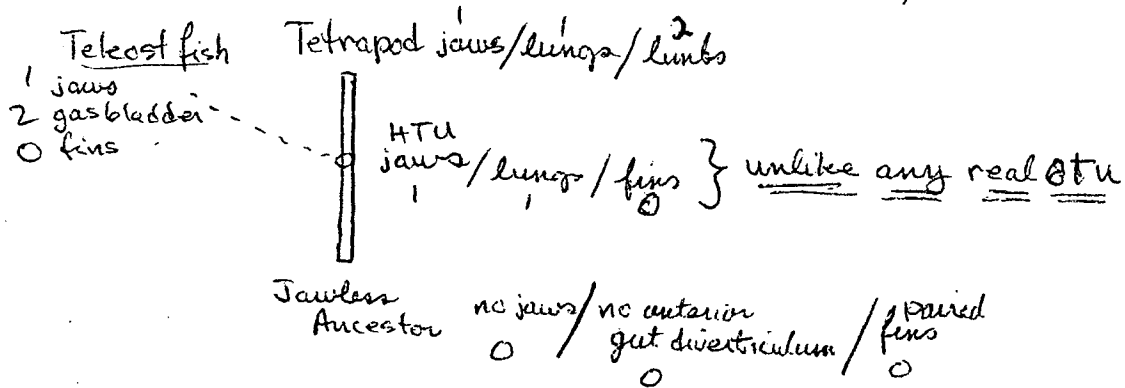
Step 4. Construct a Hypothetical Intermediate Ancestor (= Hypothetical Taxonomic Unit, = HTU) for the unplaced taxon selected in Step 3 and the two members of the interval to which that unplaced taxon is closest. HTU's must be considered because it is possible (in most cases probable) that the unplaced taxon will diverge from the lineage leading from ANCESTOR to C at an intermediate level instead of directly from the ANCESTOR or C. Thus, where does D diverge from?



* Insert this ~~the~~ ^{after} explanation of STEP 4.

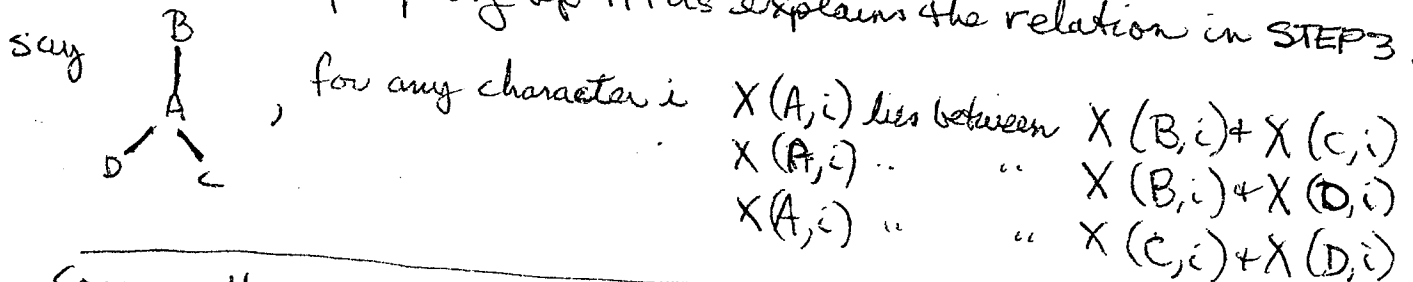
5a

Concept of HTUs - reconstructions of intermediate ancestors in order to achieve minimum length trees (max parsimony)



Examine step 4 and median state property of HTUs on handout

The median state property of HTUs explains the relation in STEP 3:



Consequently,

$$|X(B, i) - X(C, i)| = |X(A, i) - X(B, i)| + |X(A, i) - X(C, i)|$$

$$|X(B, i) - X(D, i)| = |X(A, i) - X(B, i)| + |X(A, i) - X(D, i)|$$

$$|X(C, i) - X(D, i)| = |X(A, i) - X(C, i)| + |X(A, i) - X(D, i)|$$

Summing over all characters using Manhattan metric:

$$d(B, C) = d(A, B) + d(A, C) \quad \textcircled{1}$$

$$d(B, D) = d(A, B) + d(A, D) \quad \textcircled{2}$$

$$d(C, D) = d(A, C) + d(A, D) \quad \textcircled{3}$$

Now add ① + ②:

$$d(B, C) + d(B, D) = 2d(A, B) + d(A, C) + d(A, D) \quad \textcircled{4}$$

By ③, ④ becomes:

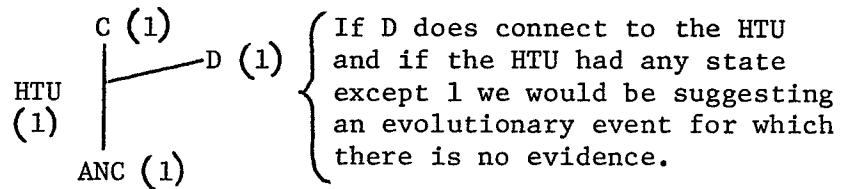
$$d(B, C) + d(B, D) = 2d(A, B) + d(C, D)$$

Rearranging: $d(A, B) = (d(B, C) + d(B, D) - d(C, D)) \times \frac{1}{2}$

The states of an HTU are always based on the states of 3 other taxonomic units (real or hypothetical). HTU's are constructed in such a way as to reduce the number of evolutionary steps implied in the final tree. The rules for HTU construction are as follows:

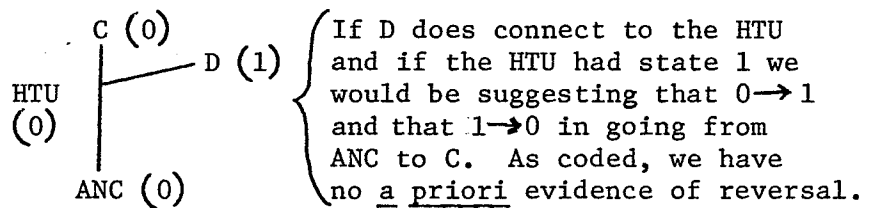
- (a) If all three taxa have the same state for a character, the HTU will take that state.

e.g. Character 1



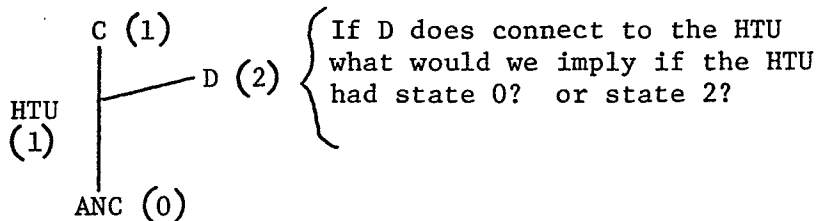
- (b) If two of the three taxa have the same state, the HTU will take that state.

e.g. Character 4



- (c) If all three taxa have different states, the hypothetical taxonomic unit will take the intermediate state.

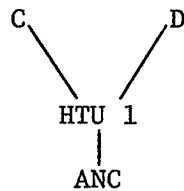
e.g.



The HTU for D, C and ANC, called HTU 1, is different from any other taxonomic unit so far in the study and it is added to both the data and difference matrices. If the HTU turns out to be identical to C or ANC we could ignore it.

Step 5. Connect the unplaced taxon to the real or hypothetical taxon from which it differs least.

In our hypothetical case,

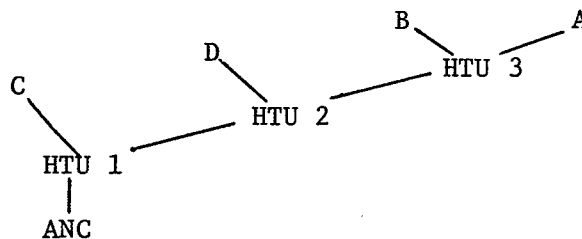


} How long is each interval?
Identify the characters that change state in each interval.

Step 6. If any real taxon remains unplaced return to Step 3, otherwise stop.

(Note in our hypothetical case we have now 2 remaining taxa. In the next Step 3, these will have to be compared to each of the three intervals that currently exist in the tree. Our old first interval, ANC-C, was destroyed with the intercalation of HTU 1.)

Complete the construction of the tree based on the data at hand. The answer is



Plot the positions of all character state changes. Which characters undergo homoplasious evolution? List the derived states that are shared by groups of taxa and the derived states that are unique to single taxa.

What are the consequences, in terms of homoplasy, of shifting the cladistic positions of taxa? Or, can you make a shorter tree in terms of total number of evolutionary steps by shifting the positions of taxa or groups of taxa?

Wagner Networks. Wagner Networks are an extension of Wagner Trees. The difference is that no ancestral states are estimated a priori and, therefore, there is no hypothetical common ancestor to serve as a starting point for construction. The first interval is made by connecting (by convention) the two most different real taxa. After this one proceeds as in tree construction. Wagner Networks yield hypotheses on cladistic pattern but not sequence. But, sometimes the pattern may help in inferring sequence. If you are interested see me for references.