Parsimony
(along with a little bit of the history of phylogenetics)
1950: Willi Hennig published *Grundzüge einer Theorie der Phylogenetischen Systematik*, which was later (1966) translated into English as *Phylogenetic Systematics*.

Emphasized that phylogenetic groups should be defined on the basis of shared, derived characters (synapomorphies), not on the basis of shared, ancestral characters (symplesiomorphies).

Although widely considered to be the father of cladistics, Hennig never described an algorithm based on the parsimony criterion, nor did he suggest what to do in the case of character conflict.

Changing terminology

Transformation series:
leaf pubescence

Characters:
glabrous
tomentose

Character:
leaf pubescence

Character states:
glabrous
tomentose

Photo (tomentose): [Link](http://www.sbs.utexas.edu/bio406d/images/pics/eup/croton_capitatus.htm)
First computed phylogeny

1957: Charles Michener and Robert Sokal published the first phylogeny estimated using a computer (but not using parsimony).

“...the computation of a 97 species X 97 species matrix of correlation coefficients presents serious technical difficulties. An operation of this magnitude cannot be reasonably undertaken without punched card or electronic computing equipment.”


Paul O. Lewis ~ Phylogenetics, Spring 2022

Photo: Fig. 10.1, p. 124, in Felsenstein, J. 2004. Inferring phylogenies. Sinauer, Sunderland, MA. Photo courtesy of Robert Sokal.
First parsimony analysis

1964: Anthony Edwards and Luca Cavalli-Sforza first used parsimony and likelihood for human blood group frequency data

“...we do not anticipate any major difficulties in the corresponding computer programmes, although these are not yet working.”

This quote refers to the maximum likelihood approach: it would not be until 1973 that Joe Felsenstein identified the central problem (trying to estimate too many parameters) and produced a workable maximum likelihood method for continuous data.


Anthony Edwards, 1963

Paul O. Lewis ~ Phylogenetics, Spring 2022
Irreversible Parsimony

1965: Joseph Camin and Robert Sokal published the first parsimony method using discrete character data.

Character states ordered from most ancestral to most derived, and reversals were not allowed.

“Comparison by Camin of these various schemes with the “truth” led him to the observation that those trees which most closely resembled the true cladistics invariably required for their construction the least number of postulated evolutionary steps for the characters studied. Subsequently we examined the possibility of reconstructing cladistics by the principle of evolutionary parsimony.”


Photo: Fig. 10.6, p. 129, in Felsenstein, J. 2004. Inferring phylogenies. Sinauer, Sunderland, MA.

Photo courtesy of the Snow Entomological Division, Natural History Museum, University of Kansas.
First molecular phylogeny

1966: Richard Eck and Margaret Dayhoff published the first molecular phylogeny using a numerical algorithm

“In principle, for a family of closely related proteins, we seek the topology that has the minimum number of mutations.”

Eck and Dayhoff method:
- **characters unordered** (any amino acid can change to any other amino acid with 1 step)
- **starting tree by stepwise addition** (parsimony score approximate)
- **branch swapping** (not explicitly defined)


Photo: [http://www.dayhoff.cc/MODayhoff_Contrib_Science.html](http://www.dayhoff.cc/MODayhoff_Contrib_Science.html)
Wagner’s Ground Plan Divergence

Warren “Herb” Wagner is well-known for his work on ferns as well as his “ground plan divergence” method for constructing phylogenies.

Wagner ground plan divergence diagram (Fig. 5.30, p. 177, in Wiley, 1981)

State 0 = presumed ancestral state, State 1 = presumed derived state

<table>
<thead>
<tr>
<th>Character</th>
<th>Neocteniza</th>
<th>Actinopus</th>
<th>Missulena</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>total</td>
<td><strong>7</strong></td>
<td><strong>10</strong></td>
<td><strong>8</strong></td>
</tr>
</tbody>
</table>
Ordered (Wagner) parsimony

1969: Arnold Kluge and Steve Farris produce software for computing Wagner parsimony trees (and introduce CI)

Photo of Farris: http://systbio.org.whsites.net/?q=taxonomy/term/2&from=30
Photo of Kluge: http://um2017.org/faculty-history/faculty/arnold-g-kluge

Phylogeny of anuran (frogs and toads) families (Fig. 5, p. 14, in Kluge and Farris, 1969)
Ordered states

- alternate
- opposite
- whorled

1 step
2 steps
1 step
1971: Walter Fitch described his algorithm for computing tree length for parsimony with four unordered states (e.g., DNA or RNA data)

Nearly all parsimony analyses of nucleotide data have used this algorithm

Photo: Fig. 10.7, p. 131, in Felsenstein, J. 2004. Inferring phylogenies. Sinauer, Sunderland, MA. Photo courtesy of Walter Fitch.
## Sequence Data

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Characters (=sites in this case)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agave</td>
<td>GGGCTGGGCG-CGGCCCGGTCGCGC-CTCT...</td>
</tr>
<tr>
<td>Bougainvillia</td>
<td>GGGTTGGGCT-CGACCGGTCGCGC-CTTG...</td>
</tr>
<tr>
<td>Ceratophyllum</td>
<td>GGGTTGGGCG-CGGTCGGTCGTCGACTCTCT...</td>
</tr>
<tr>
<td>Digitalis</td>
<td>GGGTTGGGCT-CGGCCCGGTCGCGC-CTTT...</td>
</tr>
<tr>
<td>Epilobium</td>
<td>GGGTTGGGCT-CGACCGGTCGCGC-CTTA...</td>
</tr>
<tr>
<td>Fagus</td>
<td>GGGTTGGGCG-AGAGCGGTCGCGC-CCCT...</td>
</tr>
<tr>
<td>Galax</td>
<td>GGGTTGGCGC-CGCCCGGGTCGCGC-CTAG...</td>
</tr>
</tbody>
</table>

**Taxa**

- **Agave**
- **Bougainvillia**
- **Ceratophyllum**
- **Digitalis**
- **Epilobium**
- **Fagus**
- **Galax**

**Characters (sites in this case)**

1. **Constant site**
2. **Parsimony informative site**
3. **Parsimony uninformative site**
4. **Variable site**

---

Paul O. Lewis ~ Phylogenetics, Spring 2022
## Discrete Morphology

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>P._fimbriata</td>
<td>0000000001101010110?01...</td>
</tr>
<tr>
<td>P._robusta</td>
<td>000000001101010110001...</td>
</tr>
<tr>
<td>P._articulata</td>
<td>100111010110000001100...</td>
</tr>
<tr>
<td>P._parksii</td>
<td>000111012110000001000...</td>
</tr>
<tr>
<td>P._americana</td>
<td>100111010111100001000...</td>
</tr>
<tr>
<td>P._myriophylla</td>
<td>100111010111000001000...</td>
</tr>
<tr>
<td>P._macrophylla</td>
<td>11011101210000000000000...</td>
</tr>
<tr>
<td>P._polygama</td>
<td>1101110131001000000000...</td>
</tr>
<tr>
<td>P._gracilis</td>
<td>101111013000001000110...</td>
</tr>
<tr>
<td>P._ciliata</td>
<td>001110112000001000110...</td>
</tr>
<tr>
<td>P._basiramia</td>
<td>001110112000001000110...</td>
</tr>
</tbody>
</table>

*Paul O. Lewis ~ Phylogenetics, Spring 2022*
Parsimony

123456789...
Taxon1  CGACCA\textbf{AGGT}...
Taxon2  CGACCA\textbf{AGGT}...
Taxon3  CGGTCC\textbf{GGT}...
Taxon4  CGGCC\textbf{TGGT}...

One of the three possible unrooted trees

Same tree but with data from site 6 inserted in place of taxon names
"Standard" Parsimony

2 steps

parsimony score?  number of most parsimonious reconstructions?

Paul O. Lewis ~ Phylogenetics, Spring 2022
Parsimony Steps

123456789...
Taxon1: CGACCAGGT...
Taxon2: CGACCAGGT...
Taxon3: CGGCTGGT...
Taxon4: CGGCCCTGGT...
Steps: 001102000...

Let's call this tree 1: (1,2,(3,4))

Tree 1's length for first 9 sites = 4
Parsimony Steps

\begin{itemize}
  \item Taxon 1: CGAACCAGGT...
  \item Taxon 2: CGAACCAGGT...
  \item Taxon 3: CGGTCCGGT...
  \item Taxon 4: CGGCCTGGT...
  \item Steps: 002102000...
\end{itemize}

Tree 2: \((1,3,(2,4))\)

Tree 2's length for first 9 sites = 5
Parsimony Steps

123456789...
Taxon1  CGA\text{ACC}AGGT...
Taxon2  CGA\text{ACC}AGGT...
Taxon3  CG\text{GTC}CGGT...
Taxon4  CG\text{GCC}TGGT...
Steps  002102000...

Tree 3: (1,4,(2,3))

Tree 3's length for first 9 sites = 5
Parsimony (using only 9 sites)

This is the simplest explanation of the data for the first 9 sites according to the parsimony criterion.

most parsimonious
Ockham's Razor

"Essentia non sunt multiplicanda praeter necessitatem"
(Entities should not be multiplied unnecessarily)
— William of Ockham, 14th. century

In **science**, this is the **principle of parsimony**: all other things equal, go with the simplest, least complicated hypothesis

In **phylogenetics**, it has become the **parsimony criterion**: all other things equal, go with the tree that requires the fewest inferred character state changes
Counting steps with a minimum of effort

4 steps

1971 Fitch
Parsimony variants

\[
\begin{align*}
\text{transversions} & \quad \text{transitions} \\
\text{transversions more reliable} & \quad \text{transversion parsimony} \\
\text{format equate=}& \quad "A=G \ C=T" \\
\end{align*}
\]
Transversion Parsimony

$R = A, G$

$Y = C, T$

Paul O. Lewis ~ Phylogenetics, Spring 2022
MINIMAL MUTATION TREES OF SEQUENCES*

DAVID SANKOFF†

Abstract. Given a finite tree, some of whose vertices are identified with given finite sequences, we show how to construct sequences for all the remaining vertices simultaneously, so as to minimize the total edge-length of the tree. Edge-length is calculated by a metric whose biological significance is the mutational distance between two sequences.

1. Introduction. The problem solved in this paper may be summarized as follows. Given a finite tree \( T \), some of whose vertices are identified with points in a certain metric space \( (S, d) \), locate the remaining vertices in \( S \) so as to minimize the total length of the edges of \( T \). For \( S = \mathbb{R}^n \), this is a well-known generalization

1975: David Sankoff describes an algorithm for determining tree length for generalized parsimony

David Sankoff

Step Matrices

<table>
<thead>
<tr>
<th>From</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Step matrix for Fitch parsimony
Step Matrices

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Step matrix for transversion parsimony
Step Matrices

Step matrix for analysis in which all changes are allowed, but transversions are weighted 5 times more than transitions.
Generalized Parsimony

A | A | A | T
10 steps

A | A | C | T
6 steps

A | A | G | T
11 steps

A | A | T | T
6 steps

A | C | A | T
25 steps

A | C | C | T
11 steps

A | C | G | T
25 steps

A | C | T | T
12 steps

A | G | A | T
13 steps

A | G | C | T
8 steps

A | G | G | T
12 steps

A | G | T | T
8 steps

A | T | A | T
25 steps

A | T | C | T
12 steps

A | T | G | T
25 steps

A | T | T | T
11 steps
Terminology of cladistics

apomorphy, synapomorphy, autapomorphy

plesiomorphy, symplesiomorphy

homoplasy

consistency index: $CI = \frac{m}{s}$

retention index: $RI = \frac{(g-s)}{(g-m)}$