SVDQuartets Tutorial

Laura Kubatko and Dave Swofford

Slide credit: Dave Swofford

June 4, 2018
What SVDQuartets is ....

- **SVDQuartets** is a method for species tree inference based on the multispecies coalescent that can be applied to multilocus, SNP, or coalescent independent sites data.

- The theory underlying the method is valid for data arising from very general models:
  - GTR+I+G model and all submodels
  - with or without the molecular clock
  - variation in rates and effective population sizes along branches
  - gene flow between sister taxa
What SDVQuartets is not ....

- **SVDQuartets is not a concatenation method!**

- It is NOT a summary statistics method – at least not in the traditional way.

- It does NOT try to approximate maximum likelihood (or anything else).

- It is NOT model-free.
Why is it called SVDQuartets?

Basic idea:

- Break the problem into quartets = sets of 4 taxa
- Infer the unrooted four-taxon tree for each quartet – this is done using a mathematical technique called singular value decomposition (SVD)
- Reassemble the quartets to form an overall species tree estimate

Lots of small details for each step – we’ll review the major ideas in the first half of today’s tutorial
Site pattern frequencies

- Example:

A tree for 4 taxa, which may be a subtree of a larger tree
Site pattern frequencies
Site pattern frequencies
Site pattern frequencies

A

C
C
A
A

A
G
C

G
G
G

.
Site pattern frequencies

- For each set of 4 sequences (quartet), we can count the relative frequencies of the 256 possible site patterns

<table>
<thead>
<tr>
<th>$p_{ijkl}$</th>
<th>Taxon A</th>
<th>Taxon B</th>
<th>Taxon C</th>
<th>Taxon D</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>$p_{AAAA}$</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>$p_{AAAC}$</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>$p_{AAAG}$</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>$p_{AAAT}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>129</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>$p_{GGGA}$</td>
</tr>
<tr>
<td>130</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>$p_{GGGC}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>255</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>$p_{TTTG}$</td>
</tr>
<tr>
<td>256</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>$p_{TTTT}$</td>
</tr>
</tbody>
</table>
**Flattening matrix**

- For each set of 4 sequences (quartet), we represent the pattern frequencies by a flattening matrix:

\[
\text{Flat}_{\{1,3\},\{2,4\}}(P) = \begin{pmatrix}
AA & AC & AG & AT & CA & CC & \ldots \\
AA & p_{AAAA} & p_{AAAC} & p_{AAAG} & p_{AAAT} & p_{ACAA} & p_{ACAC} & \ldots \\
AC & p_{AACA} & p_{AACC} & p_{AAGG} & p_{AAGT} & p_{ACGA} & p_{ACCC} & \ldots \\
AG & p_{AAGA} & p_{AAGC} & p_{AAGG} & p_{AAGT} & p_{ACGA} & p_{ACGC} & \ldots \\
AT & p_{AATA} & p_{AATC} & p_{AATG} & p_{AATT} & p_{ACTA} & p_{ACTC} & \ldots \\
CA & p_{CAAA} & p_{CAAC} & p_{CAAG} & p_{CAAT} & p_{CCAA} & p_{CCAC} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
\]

- **Matrix rank**: number of linearly independent rows and columns

- **Main result**: when the flattening corresponds to the tree that generated the data, the matrix rank will be *fewer* than the number of rows/columns.
Intuition on reduced rank/linear dependencies

True tree

1 2 3 4
A C A C
A C C A
C A A C
C A C A

E.g., all 4 of these site patterns have the same expected frequency

\[ f(AC|AC) = f(AC|CA) = f(CA|AC) = f(CA|CA) \]
Intuition on reduced rank/linear dependencies

E.g., all 4 of these site patterns have the same expected frequency

\[
f(AC|AC)=f(AC|CA)=f(CA|AC)=f(CA|CA)
\]

These patterns are not all expected to have the same expected frequency if they evolved on the other tree

\[
f(AA|CC) \neq f(AC|CA) \neq f(CA|AC) \neq f(CC|AA)
\]
Intuition on reduced rank/linear dependencies

\[ \begin{array}{c|cccc}
  & AA & AG & GA & GG \\
\hline
AA & a & & & \\
AG & & & & \\
GA & & & & \\
GG & & & & a \\
\end{array} \]

Flattening matrix for $1,2|3,4$
Intuition on reduced rank/linear dependencies

Flattening matrix for $1,2|3,4$

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>$a$</td>
<td>$b$</td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>$b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>$b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td>$a$</td>
</tr>
</tbody>
</table>
Intuition on reduced rank/linear dependencies

\[
\begin{align*}
A & \quad G \\
G & \quad G \\
G & \quad G \\
A & \quad G
\end{align*} =
\begin{align*}
G & \quad A \\
G & \quad G \\
G & \quad G \\
A & \quad G
\end{align*}
\]

Flattening matrix for $1,2|3,4$

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>$a$</td>
<td>$b$</td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>$b$</td>
<td></td>
<td></td>
<td>$b$</td>
</tr>
<tr>
<td>GA</td>
<td>$b$</td>
<td></td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td></td>
<td>$b$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
</tbody>
</table>
Intuition on reduced rank/linear dependencies

\[
\begin{array}{c}
A \\
A
\end{array} \\
\begin{array}{c}
G \\
G
\end{array} = \\
\begin{array}{c}
G \\
G
\end{array} \\
\begin{array}{c}
A \\
A
\end{array}
\]

Flattening matrix for 1,2|3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>AG</td>
<td>b</td>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>GA</td>
<td>b</td>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>GG</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>
Intuition on reduced rank/linear dependencies

\[
\begin{align*}
\text{A} & \quad \text{G} \\
\text{A} & \quad \text{G} \\
\text{G} & \quad \text{A} \\
\text{A} & \quad \text{G}
\end{align*}
= \begin{align*}
\text{G} & \quad \text{A} \\
\text{G} & \quad \text{A} \\
\text{G} & \quad \text{G} \\
\text{A} & \quad \text{A}
\end{align*}
\]

Flattening matrix for 1,2|3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>AG</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GA</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GG</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>
Numerical example

Expected flattening matrix for 1,2|3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.093008</td>
<td>0.061355</td>
<td>0.061355</td>
<td>0.068115</td>
</tr>
<tr>
<td>AG</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GA</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GG</td>
<td>0.068115</td>
<td>0.061355</td>
<td>0.061355</td>
<td>0.093008</td>
</tr>
</tbody>
</table>

Expected site-pattern frequencies

- $p_{AAAA} = 0.09300841$
- $p_{AAAG} = 0.06135527$
- $p_{AAGA} = 0.06135527$
- $p_{AAGG} = 0.06811487$
- $p_{AGAA} = 0.06135527$
- $p_{AGAG} = 0.04672782$
- $p_{AGGA} = 0.04672782$
- $p_{AGGG} = 0.06135527$
- $p_{GAAT} = 0.06135527$
- $p_{GAAG} = 0.04672782$
- $p_{GAGA} = 0.04672782$
- $p_{GAGG} = 0.06135527$
- $p_{GAAA} = 0.06811487$
- $p_{GAGG} = 0.06135527$
- $p_{GGAA} = 0.06135527$
- $p_{GGAG} = 0.06135527$
- $p_{GGGG} = 0.09300841$

etc.
Numerical example

Expected flattening matrix for 1,2|3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.093008</td>
<td>0.061355</td>
<td>0.061355</td>
<td>0.068115</td>
</tr>
<tr>
<td>AG</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GA</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GG</td>
<td>0.068115</td>
<td>0.061355</td>
<td>0.061355</td>
<td>0.093008</td>
</tr>
</tbody>
</table>

Delete redundant 3rd row and column...

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.093008</td>
<td>0.061355</td>
<td>0.068115</td>
</tr>
<tr>
<td>AG</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GG</td>
<td>0.068115</td>
<td>0.061355</td>
<td>0.093008</td>
</tr>
</tbody>
</table>
Numerical example

Expected flattening matrix for $1,2 | 3,4$

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.093008</td>
<td>0.061355</td>
<td>0.061355</td>
<td>0.068115</td>
</tr>
<tr>
<td>AG</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GA</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GG</td>
<td>0.068115</td>
<td>0.061355</td>
<td>0.061355</td>
<td>0.093008</td>
</tr>
</tbody>
</table>

Delete redundant 3rd row and column...

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.093008</td>
<td>0.061355</td>
<td>0.068115</td>
</tr>
<tr>
<td>AG</td>
<td>0.061355</td>
<td>0.046728</td>
<td>0.061355</td>
</tr>
<tr>
<td>GG</td>
<td>0.068115</td>
<td>0.061355</td>
<td>0.093008</td>
</tr>
</tbody>
</table>

Note that we can now obtain the last column of the above matrix as a linear combination of the first two columns:

\[
f_{AA,GG} = -f_{AA,AA} + 2.62617 f_{AA,AG} = 0.068115 \\
f_{AG,GG} = -f_{AG,AA} + 2.62617 f_{AG,AG} = 0.061355 \\
f_{GG,GG} = -f_{GG,AA} + 2.62617 f_{GG,AG} = 0.093008
\]
Numerical example

"True" branch lengths in expected substitutions/site

Expected site-pattern frequencies

\[ p_{AAAA} = 0.09300841 \]
\[ p_{AAAG} = 0.06135527 \]
\[ p_{AAGA} = 0.06135527 \]
\[ p_{AAGG} = 0.06811487 \]
\[ p_{AGAA} = 0.06135527 \]
\[ p_{AGAG} = 0.04672782 \]
\[ p_{AGGA} = 0.04672782 \]
\[ p_{AGGG} = 0.06135527 \]
\[ p_{GAAA} = 0.06135527 \]
\[ p_{GAAG} = 0.04672782 \]
\[ p_{GAGA} = 0.04672782 \]
\[ p_{GAGG} = 0.06135527 \]
\[ p_{GGAA} = 0.06811487 \]
\[ p_{GGA} = 0.06135527 \]
\[ p_{GGAG} = 0.06135527 \]
\[ p_{GGGA} = 0.06135527 \]
\[ p_{GGGG} = 0.09300841 \]

Expected flattening matrix for 1,2 | 3,4

\[
\begin{array}{cccc}
\text{AA} & \text{AG} & \text{GA} & \text{GG} \\
0.093008 & 0.061355 & 0.061355 & 0.068115 \\
0.061355 & 0.046728 & 0.046728 & 0.061355 \\
0.061355 & 0.046728 & 0.046728 & 0.061355 \\
0.068115 & 0.061355 & 0.061355 & 0.093008 \\
\end{array}
\]

Delete redundant 3rd row and column...

\[
\begin{array}{ccc}
\text{AA} & \text{AG} & \text{GG} \\
0.093008 & 0.061355 & 0.068115 \\
0.061355 & 0.046728 & 0.061355 \\
0.068115 & 0.061355 & 0.093008 \\
\end{array}
\]

Note that we can now obtain the last column of the above matrix as a linear combination of the first two columns:

\[ f_{AA,GG} = -f_{AA,AA} + 2.62617 f_{AA,AG} = 0.068115 \]
\[ f_{AG,GG} = -f_{AG,AA} + 2.62617 f_{AG,AG} = 0.061355 \]
\[ f_{GG,GG} = -f_{GG,AA} + 2.62617 f_{GG,AG} = 0.093008 \]

\[ \therefore \text{matrix has only two linearly independent rows and columns; rank is 2} \]
What if we construct a flattening matrix for a tree that did NOT generate the data?

Flattening matrix for 1,2 | 3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>AG</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GA</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GG</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

Flattening matrix for 1,3 | 2,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What if we construct a flattening matrix for a tree that did NOT generate the data?

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>AG</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GA</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GG</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

Flattening matrix for 1,2 | 3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>d</td>
</tr>
<tr>
<td>AG</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Flattening matrix for 1,3 | 2,4
What if we construct a flattening matrix for a tree that did NOT generate the data?

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>AG</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GA</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GG</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

Flattening matrix for 1,2|3,4

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AG</th>
<th>GA</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>d</td>
</tr>
<tr>
<td>AG</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>b</td>
</tr>
<tr>
<td>GA</td>
<td>b</td>
<td>d</td>
<td>c</td>
<td>b</td>
</tr>
<tr>
<td>GG</td>
<td>d</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

No redundant rows; matrix is full rank (=4)
How can we use this for species tree inference?

- **Fact:** Under the multispecies coalescent model for DNA sequence data:
  - the flattening matrix corresponding to the true tree has rank 10
  - the flattening matrix corresponding to each of the two alternative topologies has rank 16 (there are 16 rows and 16 columns)

- **Complication:** For empirical data, the site pattern frequencies approximate the true probabilities, but aren’t exact

- **Solution:** Find a way to measure how close we are to a reduced rank matrix 
  *Use singular value decomposition!*
Singular value decomposition

- **Basic idea:** Decompose an initial matrix into 3 new ones, such that multiplying the new matrices as shown below returns the original matrix exactly

\[ M = U \Sigma V' \]

- The matrix \( \Sigma \) contains the **singular values** (16 values here, since the flattening matrix is 16 \( \times \) 16)

- **Fact:** The number of non-zero singular values is equivalent to the matrix rank
The SVD Score

\[
\text{score} = \sqrt{\sum_{i=5}^{16} s_i^2}
\]

= “Frobenius distance” to nearest rank 4 matrix

Simulation conditions:

- tree = (((1:0.05,2:0.05):0.05,3:0.1):0.05,4:0.15)
- 1,000,000 sites
- HKY model: \( \kappa=4 \pi=(0.1, 0.2, 0.3, 0.4) \)
- all sites share same history (no incomplete lineage sorting, horizontal transfer, gene duplication and loss, etc.)

<table>
<thead>
<tr>
<th>SV (s)</th>
<th>1,2 ( \uparrow ) 3,4</th>
<th>1,3 ( \uparrow ) 2,4</th>
<th>1,4 ( \uparrow ) 2,3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.279686</td>
<td>0.278714</td>
<td>0.278716</td>
</tr>
<tr>
<td>2</td>
<td>0.21899</td>
<td>0.219191</td>
<td>0.219191</td>
</tr>
<tr>
<td>3</td>
<td>0.10902</td>
<td>0.110392</td>
<td>0.110389</td>
</tr>
<tr>
<td>4</td>
<td>0.066873</td>
<td>0.05709</td>
<td>0.05709</td>
</tr>
<tr>
<td>5</td>
<td>6E-05</td>
<td>0.006875</td>
<td>0.006886</td>
</tr>
<tr>
<td>6</td>
<td>6.14E-05</td>
<td>0.006315</td>
<td>0.006305</td>
</tr>
<tr>
<td>7</td>
<td>4.93E-05</td>
<td>0.003286</td>
<td>0.003286</td>
</tr>
<tr>
<td>8</td>
<td>3.8E-05</td>
<td>0.003244</td>
<td>0.003246</td>
</tr>
<tr>
<td>9</td>
<td>3.26E-05</td>
<td>0.002905</td>
<td>0.002903</td>
</tr>
<tr>
<td>10</td>
<td>3.09E-05</td>
<td>0.002499</td>
<td>0.002499</td>
</tr>
<tr>
<td>11</td>
<td>2.69E-05</td>
<td>0.001471</td>
<td>0.001472</td>
</tr>
<tr>
<td>12</td>
<td>2.23E-05</td>
<td>0.001182</td>
<td>0.001181</td>
</tr>
<tr>
<td>13</td>
<td>1.3E-05</td>
<td>0.001009</td>
<td>0.001008</td>
</tr>
<tr>
<td>14</td>
<td>1.03E-05</td>
<td>0.000937</td>
<td>0.000937</td>
</tr>
<tr>
<td>15</td>
<td>6.19E-06</td>
<td>0.000382</td>
<td>0.000384</td>
</tr>
<tr>
<td>16</td>
<td>1.58E-06</td>
<td>0.000377</td>
<td>0.000375</td>
</tr>
</tbody>
</table>

\( \text{score} \) 0.000133  0.011353  0.011354
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

\[
egin{align*}
12|34 & \quad \text{Suppose we infer} \\
12|35 & \quad \text{these quartet} \\
12|45 & \quad \text{relationships for} \\
14|35 & \quad 5 \text{ taxa} \\
23|45 & \\
\end{align*}
\]
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of inconsistent quartets (i.e., seek a solution to the Maximum Quartet Consistency problem).
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

```
12|34
12|35
12|45
14|35
23|45
```
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

```
12|34
12|35
12|45
14|35
23|45
```
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

12|34
12|35
12|45
14|35
23|45
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

12|34
12|35
12|45
14|35
23|45
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

\[
\begin{align*}
12|34 \\
12|35 \\
12|45 \\
14|35 \\
23|45
\end{align*}
\]

4 consistent quartets, 1 inconsistent quartet
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

```
12|34  
12|35  
12|45  
14|35  
23|45  
```
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

12|34
12|35
12|45
14|35
23|45
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

```
12|34
12|35
12|45
14|35
23|45
```
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

\[
\begin{align*}
12|34 \\
12|35 \\
12|45 \\
14|35 \\
23|45 \\
\end{align*}
\]
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of *inconsistent quartets* (i.e., seek a solution to the Maximum Quartet Consistency problem).

12|34
12|35
12|45
14|35
23|45

1  2
3----4
    \  
     5
More than 4 taxa

Compute invariant scores for all quartets, choosing the best resolution for each one.

Search for a tree that minimizes the number of inconsistent quartets (i.e., seek a solution to the Maximum Quartet Consistency problem).

2 consistent quartets, 3 inconsistent quartet

Now evaluate the remaining 13 trees and choose the one that maximizes the number of consistent quartets.
More than 4 taxa

While evaluation of each possible tree might work well for 5-tip trees, the number of possible trees for $n$ tips grows too quickly to make it a general strategy.

Must use a heuristic algorithm to search for the best tree:

- The default in PAUP* is a heavily modified version of “QFM” (Reaz et al., 2014)
- Other algorithms are available in PAUP* and elsewhere
- Unfortunately, the MQC problem is NP-hard (i.e., exact solution will be slow for large numbers of tips)
On to the tutorial!