Splits

aka Splits Tree

Developers:
Daniel Huson and David Bryant

downloadable from

http://www-ab.informatik.uni-tuebingen.de/software/jsplits/welcome.html

Alexandra Han, 10/29/2003
Demo

Algorithms

What is jSplits?
What is jSplits?
SplitsTreeReloaded

• Otherwise known as SplitsTree4.

• Software for analysing evolutionary data using phylogenetic methods including:
  (i) Split decomposition
  (ii) Parsimony splits
  (iii) Neighbor-net (Bryant and Moulton 2002)
  (iv) Neighbor-joining and Bio-Nj
  (v) Buneman Tree
  (vi) quartets
  (vii) networks

• Datatypes supported:
  (i) (un)aligned sequences
  (ii) distances
  (iii) trees
  (iv) splits
  (v) networks
  (vi) quartets

Alexandra Han, 10/29/2003
New methods are easily added to the program by providing plug-ins, and a number are in preparation, e.g., a faster version of the refined Buneman tree.

Consensus network construction (Holland and Moulton 2003),

SplitsTree now supports the nexus trees blocks and currently provides the new

SplitsTree supports analysis plugins, for example one providing quartet mapping and
BEGIN taxa;

DIMENSIONSntax=6;

TAXLABELS
[1] Bernd
[2] Seth
[3] Ian
[5] Rudy

END[ taxa ];
BEGIN distances!

DIMENSIONSntax=6;

FORMAT triangle=LOWER diagonal labels missing=?

MATRIX
Bernd 0 10 12 13 7 11
Seth 6 10 11 12 0 12
Ian 8 8 12 7 0 13
Me 0 7 7 12 6 0
Rudy 0 7 0 6 12 0
Josephine 7 13 12 7 11 0

END;

MATRIX

tables missing=

diamond triangle=LOWER

FORMATTABLES
DIMENSIONS ntax=6:
BEGIN distances:

Input File: us.nex

Alexandra Han, 10/29/2003
Figure 1: Our true relationship
BEGIN taxa;
DIMENSIONSntax=8;
TAXLABELS
Ottontobos
An-ntdul
Engteta
Chlorotta
Chlamy
Marchantia
Rice
Tobacco
TAILABLES
DIMENSIONSntax=8;
BEGIN taxa;
[; A comment starting with ";!" is printed when read
#NEXUS [Comments come in square brackets]
Input File: algae.nex
Alexandra Han, 10/29/2003

BEGIN characters;

DIMENSIONS nchar=920;
FORMAT datatype=RNA gap=- labels interleave;

MATRIX

TobaccoAAGAACCUGCCCUUGGGAGCUGGAAACGGCUGCUAAUACCCC
RiceAAGAACCUGCCCUUGGGAACUGGAAACGGUUGCUAAUACCCC
MarchantiaAAGAACCUGCCCUUGGGAGCUGGAAACGGUUGCUAAUACCCC
ChlamyAAGAACCUACCUAUCGGAUUGGGAAACUGUUGCUAAUACCCC
ChlorellaAAGAACCUACCUUUAGGAACUGGAAACGGUUGCUAAUACCCC
EuglenaAAGAAUCUGCGCUUGGGAGAUGGAAACGUUUGCUAAUGCCUC
An_nidulGAGAAUCUGCCUACAGGAGUUGGAAACGACUGCUAAUACCCG
OlithodiscusGAGAAUCUGCCUUUAGGAUUUGGAAACGAAUGCUAAUACCUU

END!

METHODS

Align mitochondrial English Chlamydomonas Marchantia Rice Tobacco
MATRIX

BEGIN characters;

DIMENSIONS nchar=920;

BEGIN characters;

END;

END;

Alexandra Han, 10/29/2003

Input File: algae.nex
BEGIN distances;

<table>
<thead>
<tr>
<th></th>
<th>Tobacco</th>
<th>Rice</th>
<th>Marchantia</th>
<th>Chlamydomonas</th>
<th>Olithodiscus</th>
<th>Anacystis</th>
<th>Euglena</th>
<th>Chlorella</th>
<th>Chlamy</th>
<th>Marchantia</th>
<th>Rice</th>
<th>Tobacco</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco</td>
<td>0</td>
<td>0</td>
<td>0.028150</td>
<td>0.032410</td>
<td>0.129040</td>
<td>0.152350</td>
<td>0.17907</td>
<td>0.16007</td>
<td>0.16569</td>
<td>0.13724</td>
<td>0.18069</td>
<td>0.15286</td>
</tr>
<tr>
<td>Rice</td>
<td>0.16007</td>
<td>0</td>
<td>0.06990</td>
<td>0.09909</td>
<td>0.08754</td>
<td>0.13856</td>
<td>0.11317</td>
<td>0.09909</td>
<td>0.11363</td>
<td>0.12710</td>
<td>0.11065</td>
<td>0.15208</td>
</tr>
<tr>
<td>Marchantia</td>
<td>0.16569</td>
<td>0.06990</td>
<td>0</td>
<td>0.102815</td>
<td>0.032410</td>
<td>0</td>
<td>0.14353</td>
<td>0.032410</td>
<td>0.03241</td>
<td>0</td>
<td>0.03241</td>
<td>0.04554</td>
</tr>
<tr>
<td>Chlamydomonas</td>
<td>0.13724</td>
<td>0.09909</td>
<td>0.102815</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.12904</td>
<td>0.04554</td>
</tr>
<tr>
<td>Olithodiscus</td>
<td>0.18069</td>
<td>0.11317</td>
<td>0.11317</td>
<td>0.12904</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.102815</td>
<td>0</td>
<td>0.08754</td>
<td>0.032410</td>
</tr>
<tr>
<td>Anacystis</td>
<td>0.15286</td>
<td>0.15208</td>
<td>0.15208</td>
<td>0.15235</td>
<td>0.15286</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.16007</td>
<td>0</td>
</tr>
<tr>
<td>Euglena</td>
<td>0.15208</td>
<td>0.15208</td>
<td>0.15208</td>
<td>0.17907</td>
<td>0.15208</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.16007</td>
<td>0</td>
</tr>
<tr>
<td>Chlorella</td>
<td>0.04554</td>
<td>0.04554</td>
<td>0.04554</td>
<td>0.12710</td>
<td>0.04554</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.102815</td>
<td>0</td>
<td>0.08754</td>
<td>0</td>
</tr>
<tr>
<td>Chlamy</td>
<td>0.11065</td>
<td>0.11065</td>
<td>0.11065</td>
<td>0.11363</td>
<td>0.11065</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.16007</td>
<td>0</td>
</tr>
<tr>
<td>Marchantia</td>
<td>0.13724</td>
<td>0.13724</td>
<td>0.13724</td>
<td>0.13856</td>
<td>0.13724</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.16007</td>
<td>0</td>
</tr>
<tr>
<td>Rice</td>
<td>0.03241</td>
<td>0.03241</td>
<td>0.03241</td>
<td>0.03241</td>
<td>0.03241</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.102815</td>
<td>0</td>
<td>0.08754</td>
<td>0</td>
</tr>
<tr>
<td>Tobacco</td>
<td>0.15286</td>
<td>0.15208</td>
<td>0.15208</td>
<td>0.15235</td>
<td>0.15286</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14353</td>
<td>0</td>
<td>0.16007</td>
<td>0</td>
</tr>
</tbody>
</table>

FORMAT triangle=LOWER diagonal labels;

MATRIX

END;

Alexandra Han, 10/29/2003
Figure 2: ntax=8

FIT=87.3  ntax=8  nchar=920  gaps=20  const=550  nonparsi=783 -dsplits -logdet 0.01
What is jSplits?
Given a finite set \( X \) and a (pseudo-)metric \( d \), a split \( S \) of set \( X \) is a bipartition of \( X \) into two non-empty sets, \( A \) and \( B \).

### Split Decomposition

**Split (pseudo)metric** \( \delta_S \) associated to this split \( S \) is defined by

\[
\delta_S(x, y) = \begin{cases} 
0 & \text{if } x, y \in A \text{ or } x, y \in B, \\
1 & \text{otherwise,}
\end{cases}
\]

The isolation index w.r.t. the (pseudo-)metric \( d \) is defined by

\[
\alpha_d(S) = \min_{a, a' \in A, b, b' \in B} \max \left\{ \frac{d(a, b)}{2} + \frac{d(a', b')}{2}, \frac{d(a', b) + d(a, b')}{2}, \frac{d(a, b') + d(a', b)}{2} \right\} - \frac{d(a, a')}{2} - \frac{d(b, b')}{2}
\]

Furthermore, \( S \) is a \( d \)-split if \( \alpha_d(S) > 0 \).

**Split Decomposition**

\[
\left\{ \frac{\|q - p\|}{2} - \frac{\|q + p\|}{2}, \frac{\|q + p, q + p\|}{2}, \frac{\|q + p, q + p\|}{2}, \frac{\|q + p, q + p\|}{2} \right\} = \sum \frac{2}{\|q - p\|}
\]
Let $S$ be the set of all $p$-splits of $X$.

The splitability index

$$\frac{\left\| \sum_{\gamma \in \mathcal{S}} \frac{\gamma}{\gamma - \sum_{\gamma \in \mathcal{S}} \alpha_{\gamma} \cdot \delta_{\gamma}} \right\|}{\left\| \sum_{\gamma \in \mathcal{S}} \frac{\gamma}{\gamma - \sum_{\gamma \in \mathcal{S}} \alpha_{\gamma} \cdot \delta_{\gamma}} \right\|} \cdot 100$$

where

$$p_0 - p =: \rho$$

is a split-prime pseudo-metric, a pseudo-metric with isolation index $\alpha_{\gamma} = 0$ for all splits $\gamma$.

The main theorem of [BD92a] gives us the split-prime residue of $p$.

Then the main theorem of [BD92a] gives us the split-prime residue of $p$.

It was also shown in [BD92a] that if $p$ satisfies the four-point condition, then

$$X.$$
Step 1. Find all $d$-splits for our (pseudo-)metric $d$ by finding the $d$-splits on subset \{1, ..., $i$\} by checking for each $p$-split $B$, \{1, ..., $i$\} \in $A$ whether $B \cap \{1, ..., $i$\} is a $p$-split, bounded by $B$ \{1, ..., $i$\} \cap $A$. By checking for each $p$-split $A$, $B$ \{1, ..., $i$\} \in $A$ by finding the $p$-splits on subset $B$.

Step 2. Produce a graph from the $N$-splits on our metric $p$ that is a subgraph of the $N$-dimensional hypercube.

Step 3. Make the length of the edges proportional to the weight.

Step 4. Done if graph is planar. Otherwise, remove "redundant" edges to make the graph planar.
Given $X = \{a,b,c,d,e,f,g\}$, metric $d$ with $d$-splits:

- $S_1 = \{a,b,f,g\}, \{c,d,e\}$
- $S_2 = \{a,f,g\}, \{b,c,d,e\}$
- $S_3 = \{a,b,c,d\}, \{e,f,g\}$
- $S_4 = \{a,b,c,g\}, \{d,e,f\}$

Given $X$ with $p$-metric with $p$-splits:

- $\{f,p,e\}, \{b,c,q,v\} = S$
- $\{b,f,e\}, \{p,c,q,v\} = S$
- $\{b,f,v\} = S$
- $\{b,f,q,v\} = S$

Figure 3: Subgraph of $N$-Hypercube
Figure 4: Before and after removing "redundant" edges.
Demo
Figure 5: ntax=10

Tree. Biometics, 43, 751-758.

D.H. Huson, R. Wetzal. Analyzing and visualizing distance data with SPLITSTREE1.0.

H.-J. Bandelt, A. Dress. Reconstructing the shape of a tree from observed dissimilarity data.

http://www-ab.informatik.uni-tuebingen.de/software/jsplits/welcome.html

Want to have a go at it yourself?