Phylogeny

Quartet tree puzzling

(Strimmer & von Haeseler, 1996)

Rolf Backofen

Lehrstuhl für Bioinformatik
Institut für Informatik
Albert-Ludwigs-Universität Freiburg

Course Bioinformatics I

built on January 31, 2019
Finally: Problem of find correct topology for unrooted tree

- very complex problem $\Rightarrow$ heuristics
- best so far: Quartet tree puzzling (Strimmer & von Haeseler, 1996)

Quartet tree of $\{a, b, c, d\} = \begin{array}{c}
  a \\
  b \\
  c \\
  d \\
\end{array}$

$\Rightarrow$ expresses neighbourship between four taxa:

\emph{a-b are more closely related than a-c or a-d etc.}
Observation 1

Why are we interested in quartet trees of taxa?

Theorem

Any topology of an unrooted binary tree for $n$ taxa can be uniquely decomposed into $\binom{n}{4}$ quartet trees.

---

Diagram showing five taxa $a$, $b$, $c$, $d$, and $e$, and their relationships through quartet trees.
Example: 5 taxa \( \{a, b, c, d, e\} \)

Consider given but unknown tree for 5 taxa

How many quartet trees?

\[ \Rightarrow 1 \text{ for every 4-subset of 5 taxa} \]

\[ \Rightarrow \binom{5}{4} = 5 \]

For each 4-subset, extract quartet tree..

\{a, b, c, d\} \rightarrow \quad \begin{array}{c}
\text{a} \\
\text{b} \\
\text{c} \\
\text{d}
\end{array}

\{a, c, d, e\} \rightarrow \quad \begin{array}{c}
\text{a} \\
\text{b} \\
\text{e} \\
\text{c}
\end{array}

\{b, c, d, e\} \rightarrow \quad \begin{array}{c}
\text{b} \\
\text{c} \\
\text{e} \\
\text{d}
\end{array}

\{a, b, c, e\} \rightarrow \quad \begin{array}{c}
\text{a} \\
\text{b} \\
\text{d} \\
\text{c}
\end{array}

\{a, b, d, e\} \rightarrow \quad \begin{array}{c}
\text{a} \\
\text{b} \\
\text{e} \\
\text{d}
\end{array}

Alternative quartet tree encodings:

\( N_1(a, b|c, d), N_2(a, e|c, d), N_3(b, e|c, d), N_4(a, e|b, c), N_5(a, e|b, d), \)
Observation 2

- Why are we interested in the $\binom{n}{4}$ many quartet trees for $n$ taxa?

**Theorem**

*Given the $\binom{n}{4}$ many quartet trees associated with an arbitrary topology for $n$ taxa. Then the set of all quartet trees uniquely determines the final topology.*

⇒ Idea of quartet tree puzzling (Strimmer & von Haeseler, 1996)

**Step 1** for each of the $\binom{n}{4}$ subsets of $n$ taxa, determine the optimal quartet tree using Felsenstein

⇒ 3 possibilities for each subset of 4 taxa to check

**Step 2** combine the found quartet trees into final topology

- Step 2 “simple” if all quartet trees are correct
- however, this is usually not the case (some are wrong...)

⇒ heuristic combination needed! = quartet tree puzzling
Idea Quartet tree Puzzling

Idea of heuristic in Step 2: minimize error!

- given reconstructed tree (topology) $T$ for $k$ taxa $\{t_1, \ldots, t_k\}$
  
  (e.g. start with the quartet tree of some $\{t_1, \ldots, t_4\}$)

⇒ Apply iterative procedure to extend tree by one taxon $t_{k+1}$ in every step:

- consider from Step 1 all (optimal) quartet trees (=neighborships)
  for $\{t_1, \ldots, t_{k+1}\}$ that contain $t_{k+1}$

- for each resp. $N(t_{k+1}, t|t', t'')$: identify path in $T$ connecting $t'$, $t''$
  
  = edges where $t_{k+1}$ can not be added according to $N(t_{k+1}, t|t', t'')$

⇒ increase “violation counter” for edges on $t' \Rightarrow t''$ path

- add $t_{k+1}$ to edge with lowest violation count

⇒ generates new topology $T'$ for $\{t_1, \ldots, t_{k+1}\}$

⇒ iterate until all taxa have been added . . .
Example: Quartet tree Puzzling adding taxon e

\[ T = \begin{array}{c}
N_1(a, b | c, d) \\
\text{init with without e}
\end{array}
\]

\[ N_2(a, e | c, d) \Rightarrow \]

\[ N_3(b, e | c, d) \]

\[ N_4(a, e | b, c) \leftarrow \]

\[ N_5(a, e | b, d) \]

\[ = T' \]
**Final Outcome of one Iteration**

- **Note:** there might be several edges with minimal error

- **Solution:**
  - randomize extension order of taxa
  - select minimal edge at random
  - run several times
  - build final **majority tree** (see literature)