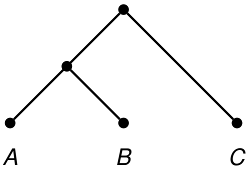
Rooted Triples

Rooted triplet= rooted binary phylogenetic tree with exactly three leaves.



For three leaves A, B, C in T we write $((A, B), C)$ if the path from A to B does not intersect the path from C to the root ρ.

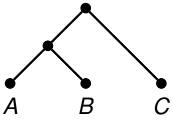
That is the unique rooted triplet with

$$lca(A, B) \prec lca(A, C) = lca(B, C)$$

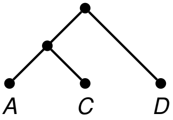
Any rooted phylogenetic tree can be represented by a set of rooted triplets.

Combining Rooted Triples

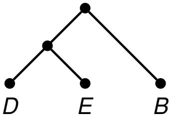
$((A, B)C)$



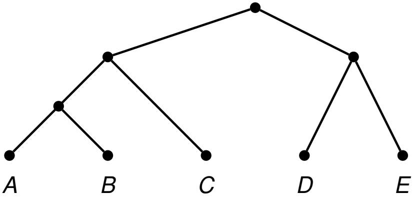
$((A, C)D)$



$((D, E)B)$

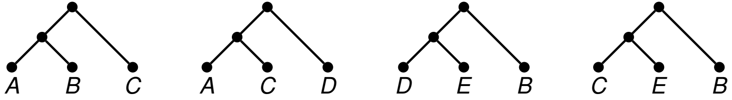


Consensus Tree “displays” all rooted triples:



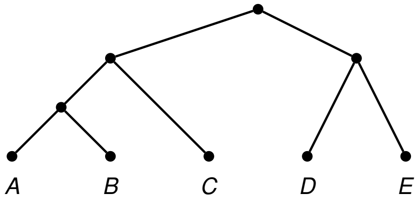
Combining Rooted Triples

$((A,B)C)$ $((A,C)D)$ $((D,E)B)$ $((C,E)B)$



Consensus Tree does not always exist!!

Consistence



For three leaves A, B, C in T we write $((A, B), C)$ if the path from A to B does not intersect the path from C to the root ρ.

That is the unique rooted triplet with

$$lca(A, B) \prec lca(A, C) = lca(B, C)$$

T and an arbitrary triple $((A, B), C)$ are consistent iff

$$lca(A, B) \prec lca(A, C) = lca(B, C)$$

T displays $((A, B), C)$.

BUILD

Theorem (Aho, Sagiv, Szymanski, Ullman - 1981; Semple & Steel - 2003)

Let $\mathscr{R}$ by a collection of rooted triples with leaf set $\mathscr{L}$. Then there is an $O(|\mathscr{R}||\mathscr{L}|)$ time algorithm – called `BUILD` *– that either*

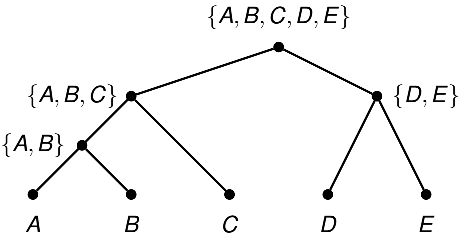
- *constructs a phylogenetic tree $T_{|\mathscr{R}}$ that displays each member of $\mathscr{R}$*

  or

- *recognizes $\mathscr{R}$ as inconsistent.*

BUILD

Idea of this recursive, top-down approach: Partition $\mathscr{L}$ into blocks according to $\mathscr{R}$. Output a tree consisting of a root whose children are roots of the trees obtained by recursing on each block.



BUILD

Let $\mathscr{R}$ be a set of triples defined on a leaf set $\mathscr{L}$.

For any $L \subseteq \mathscr{L}$ define $\mathscr{R}_{|L} = \{((x,y)z) \in \mathscr{R} \mid x,y,z \in L\}$.

To find blocks use auxiliary graph $G(\mathscr{R}_{|L}, L) = (L, E)$ with $(x,y) \in E$ iff there is a triple $((x,y)z) \in \mathscr{R}_{|L}$

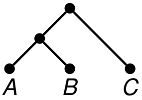
BUILD

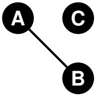
Let $\mathscr{R}$ be a set of triples defined on a leaf set $\mathscr{L}$.

For any $L \subseteq \mathscr{L}$ define $\mathscr{R}_{|L} = \{((x,y)z) \in \mathscr{R} \mid x,y,z \in L\}$.

To find blocks use auxiliary graph $G(\mathscr{R}_{|L}, L) = (L, E)$ with $(x,y) \in E$ iff there is a triple $((x,y)z) \in \mathscr{R}_{|L}$

Exmpl: $L = \{A,B,C\}$, $\mathscr{R} = ((A,B)C)$, $G(\mathscr{R}_{|L}, L)$





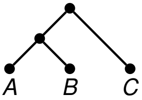
BUILD

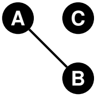
Let $\mathscr{R}$ be a set of triples defined on a leaf set $\mathscr{L}$.

For any $L \subseteq \mathscr{L}$ define $\mathscr{R}_{|L} = \{((x,y)z) \in \mathscr{R} \mid x,y,z \in L\}$.

To find blocks use auxiliary graph $G(\mathscr{R}_{|L}, L) = (L, E)$ with $(x,y) \in E$ iff there is a triple $((x,y)z) \in \mathscr{R}_{|L}$

Exmpl: $L = \{A,B,C\}$, $\mathscr{R} = ((A,B)C)$, $G(\mathscr{R}_{|L}, L)$





Crucial observation: If ((xy)z) is consistent with a tree T then the leaves labeled by x and y cannot descend from two different children of the root of T, i.e., x and y must belong to the same block.

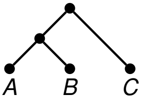
BUILD

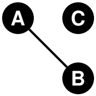
Let $\mathscr{R}$ be a set of triples defined on a leaf set $\mathscr{L}$.

For any $L \subseteq \mathscr{L}$ define $\mathscr{R}_{|L} = \{((x,y)z) \in \mathscr{R} \mid x,y,z \in L\}$.

To find blocks use auxiliary graph $G(\mathscr{R}_{|L}, L) = (L, E)$ with $(x,y) \in E$ iff there is a triple $((x,y)z) \in \mathscr{R}_{|L}$

Exmpl: $L = \{A,B,C\}$, $\mathscr{R} = ((A,B)C)$, $G(\mathscr{R}_{|L}, L)$





Crucial observation: If ((xy)z) is consistent with a tree T then the leaves labeled by x and y cannot descend from two different children of the root of T, i.e., x and y must belong to the same block.

Therefore, the algorithm defines the partition of $L \subseteq \mathscr{L}$ by:
Blocks of leaves iff connected components in $G(\mathscr{R}_{|L}, L)$

BUILD

Lemma (Aho, Sagiv, Szymanski, Ullman (1981), Bryant & Steel (1995))

A given triple set $\mathscr{R}$ on a leaf set $\mathscr{L}$ is consistent if and only if for all $L \subseteq \mathscr{L}$ with $|L| > 1$ the graph $G(\mathscr{R}_{|L}, L)$ is disconnected.

BUILD

INPUT: Set of triples in $\mathscr{R}$, leaf set $\mathscr{L}$.

OUTPUT: A rooted, phylog. tree distinctly leaf-labeled by $\mathscr{L}$ consistent with all rooted triplets in $\mathscr{R}$, if one exists; otherwise *null*.

compute $G(\mathscr{R},\mathscr{L})$

compute connected components $C_1,\dots,C_s$ of $G(\mathscr{R},\mathscr{L})$

if $s = 1$ and $|\mathscr{L}| = 1$ **then**

  return tree $\simeq K_1$

else if $s = 1$ and $|\mathscr{L}| > 1$ **then**

  return *null*

else

  for i = $1,\dots s$ **do**

    $T_i = \mathrm{BUILD}(\mathscr{R}_{|V(C_i)}, V(C_i))$

  end for

  if $T_i \neq null$ for all $i = 1,\dots s$ **then**

    attach all of these trees to a common parent node and let T be the resulting tree; else $T = null$.

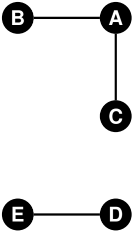
  end if

end if

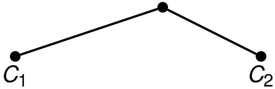
BUILD - Example

$\mathscr{R} = \{((AB)C), ((AC)D), ((DE)B)\}$

$G(\mathscr{R}, \mathscr{L})$:



$\mathrm{BUILD}(\mathscr{R}, \mathscr{L} = \{A, B, C, D, E\})$



$C_1 := \mathrm{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{A, B, C\})$
$C_2 := \mathrm{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{D, E\})$

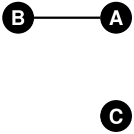
BUILD - Example

$\mathscr{R} = \{((AB)C), ((AC)D), ((DE)B)\}$

$C_1 := \texttt{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{A, B, C\})$
$\mathscr{R}_1 := \{((AB)C)\}$

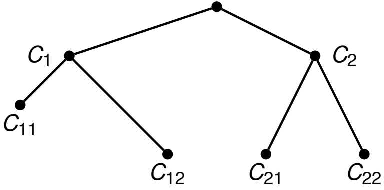
$C_2 := \texttt{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{D, E\})$
$\mathscr{R}_2 := \emptyset$

$G(\{A, B, C\})$:



$G(\{D, E\})$:

$\texttt{BUILD}(\mathscr{R}, \mathscr{L} = \{A, B, C, D, E\})$



BUILD - Example

$\mathscr{R} = \{((AB)C), ((AC)D), ((DE)B)\}$

$C_1 := \texttt{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{A, B, C\})$

$C_2 := \texttt{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{D, E\})$

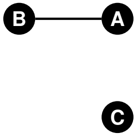
$C_{11} := \texttt{BUILD}(\mathscr{R}_{|\mathscr{L}}, \mathscr{L} = \{A, B\})$

$C_{12} := \texttt{BUILD}(\emptyset, \{C\})$

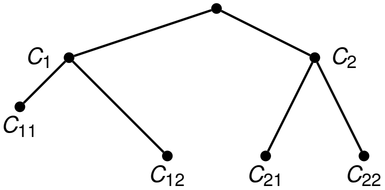
$C_{21} := \texttt{BUILD}(\emptyset, \{D\})$

$C_{22} := \texttt{BUILD}(\emptyset, \{E\})$

$G(\{A, B, C\})$:



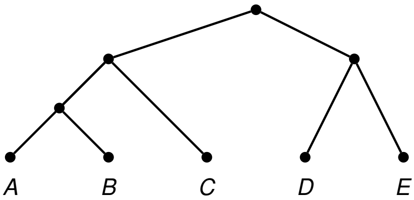
$\texttt{BUILD}(\mathscr{R}, \mathscr{L} = \{A, B, C, D, E\})$



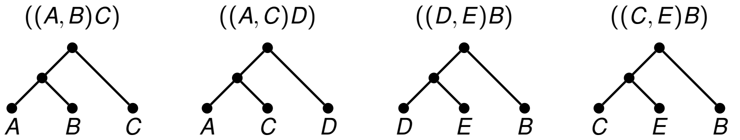
$G(\{D, E\})$:

BUILD - Example

$\texttt{BUILD}(\mathscr{R}, \mathscr{L} = \{A, B, C, D, E\})$



BUILD - Example



Consensus Tree does not always exist!!

$G(\mathscr{R}, \mathscr{L})$: