

From Gene Trees with HGT to Species Trees

Marc Hellmuth

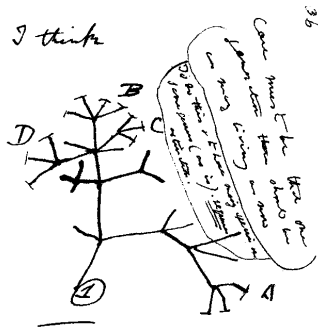
Dptm. Mathematics and Computer Science
University of Greifswald, Germany

+ Nikolai, Manuela, Nic, Sarah, Paul, John, Peter

TBI-Winterseminar 2017

Phylogenomics ...

... aims at finding plausible hypotheses of the evolutionary history of genes or species based on genomic sequence information

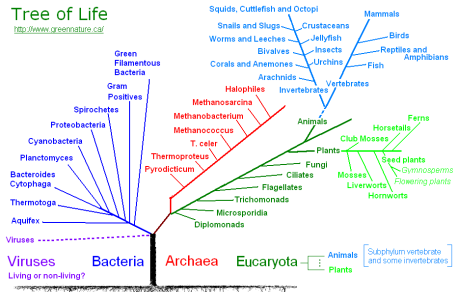


(1837)

"I think" by Charles Darwin

Tree of Life

<http://www.greennature.ca>

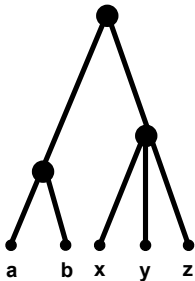


(2014)

<http://www.greennature.ca>

Central Objects: Rooted Trees and Triples

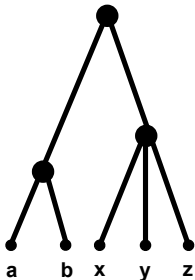
Rooted tree T:



connected, acyclic graph

Central Objects: Rooted Trees and Triples

Rooted tree T :

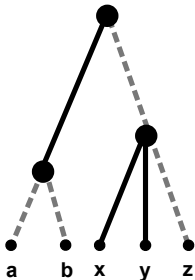


Triples:

T displays a triple $(ab|z)$ if the path from a to b does not intersect the path from z to the root.

Central Objects: Rooted Trees and Triples

Rooted tree T :

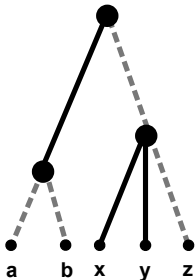


Triples:

T displays a triple $(ab|z)$ if the path from a to b does not intersect the path from z to the root.

Central Objects: Rooted Trees and Triples

Rooted tree T :



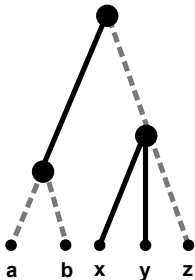
Triples:

T displays a triple $(ab|z)$ if the path from a to b does not intersect the path from z to the root.

$$\mathcal{R}(T) = \{(ab|x), (ab|y), (ab|z), (xy|a) \dots\}$$

Central Objects: Rooted Trees and Triples

Rooted tree T :



Triples:

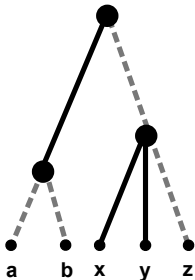
T **displays** a triple $(ab|z)$ if the path from a to b does not intersect the path from z to the root.

$$\mathcal{R}(T) = \{(ab|x), (ab|y), (ab|z), (xy|a), \dots\}$$

An arbitrary set of triples R is **consistent**, if there is a tree T with $R \subseteq \mathcal{R}(T)$

Central Objects: Rooted Trees and Triples

Rooted tree T :



Triples:

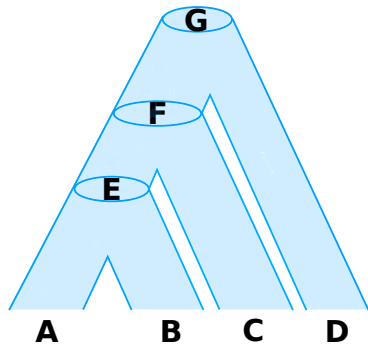
T **displays** a triple $(ab|z)$ if the path from a to b does not intersect the path from z to the root.

$$\mathcal{R}(T) = \{(ab|x), (ab|y), (ab|z), (xy|a), \dots\}$$

An arbitrary set of triples R is **consistent**, if there is a tree T with $R \subseteq \mathcal{R}(T)$

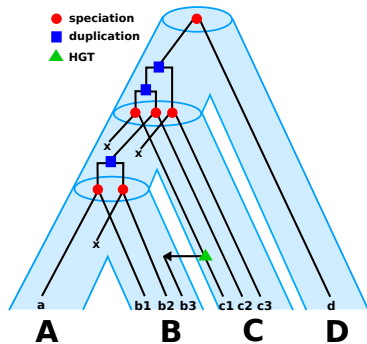
Consistency-check via `BUILD` in polynomial time.
`BUILD` returns also a tree T with $R \subseteq \mathcal{R}(T)$ if one exists.

The “true” evolutionary History



The “true” evolutionary History

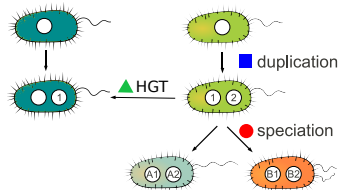
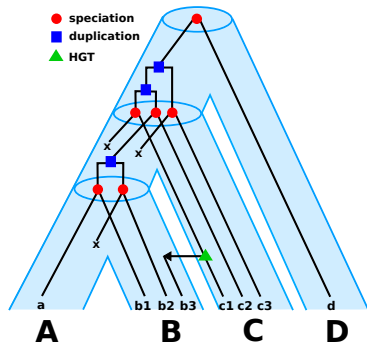
- species are characterized by its genome:
a “bag of genes”
- “Genes” evolve along a *rooted* tree with
unique *event labeling*
 $t: V^0 \rightarrow M = \{\bullet, \blacksquare, \blacktriangle\}$



The “true” evolutionary History

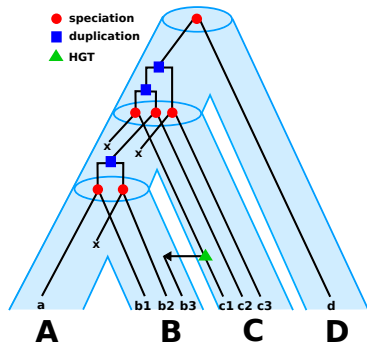
- species are characterized by its genome:
a “bag of genes”
- “Genes” evolve along a *rooted tree* with unique *event labeling*
 $t : V^0 \rightarrow M = \{\bullet, \blacksquare, \blacktriangle\}$

- **Gene duplication** : an offspring has two copies of a single gene of its ancestor
- **Speciation** : two offspring species inherit the entire genome of their common ancestor
- ▲ **HGT** : transfer of genes between organisms in a manner other than traditional reproduction and across different species

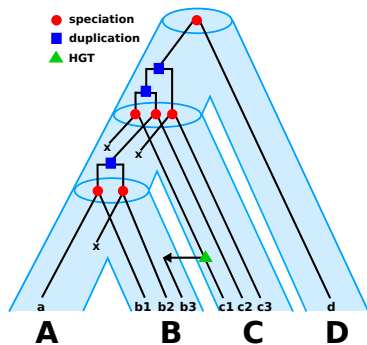
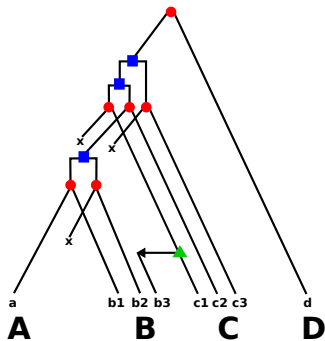


The “true” evolutionary History

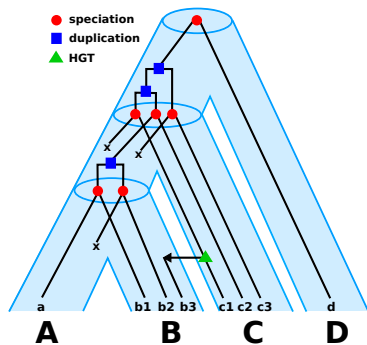
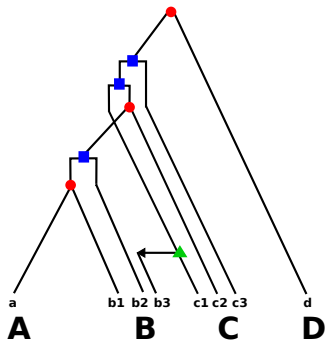
- species are characterized by its genome:
a “bag of genes”
- “Genes” evolve along a *rooted* tree with
unique *event labeling*
 $t: V^0 \rightarrow M = \{\bullet, \blacksquare, \blacktriangle\}$



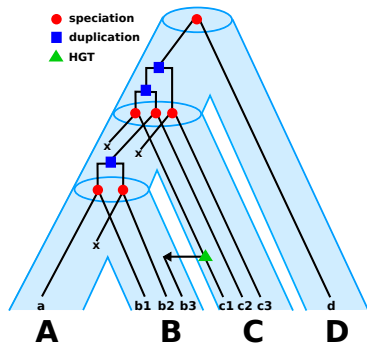
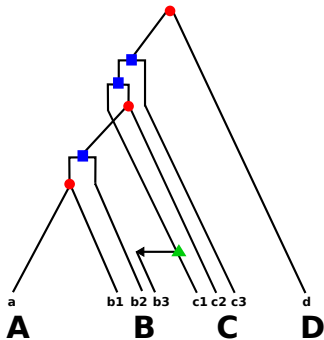
In practice, it is possible to compute (observable part of) event-labeled gene trees directly from sequence data (using orthologs, xenologs, paralogs) *without the need to know the species tree!*



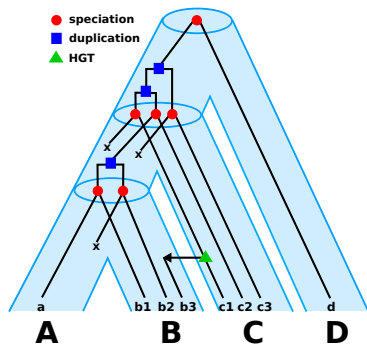
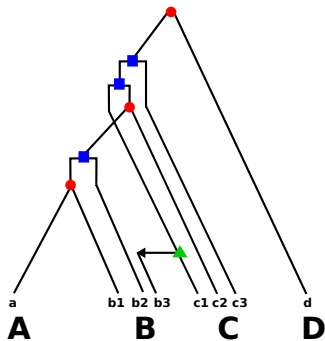
In practice, it is possible to compute (observable part of) event-labeled gene trees directly from sequence data (using orthologs, xenologs, paralogs) *without the need to know the species tree!*



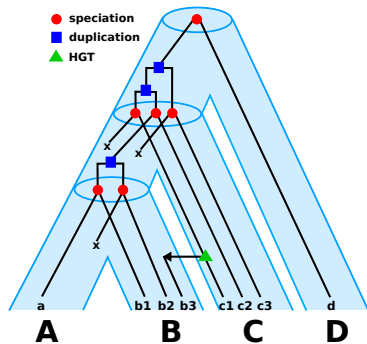
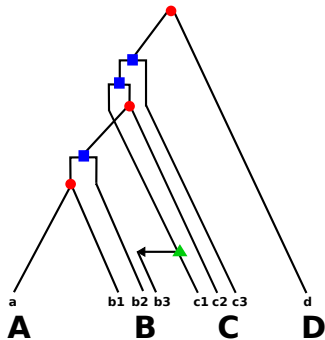
In practice, it is possible to compute (observable part of) event-labeled gene trees directly from sequence data (using orthologs, xenologs, paralogs) *without the need to know the species tree!*



Questions: Are the gene trees biologically feasible? \Leftrightarrow
Are there species trees for the gene trees?

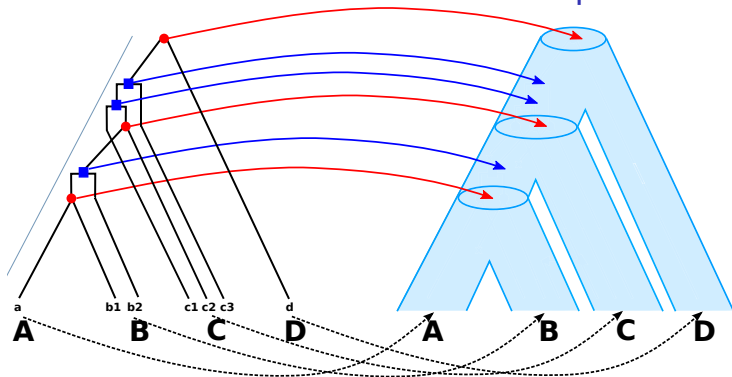


- Questions:** Are the gene trees biologically feasible? \iff
 Are there species trees for the gene trees?
 If so, how to construct species tree? \iff



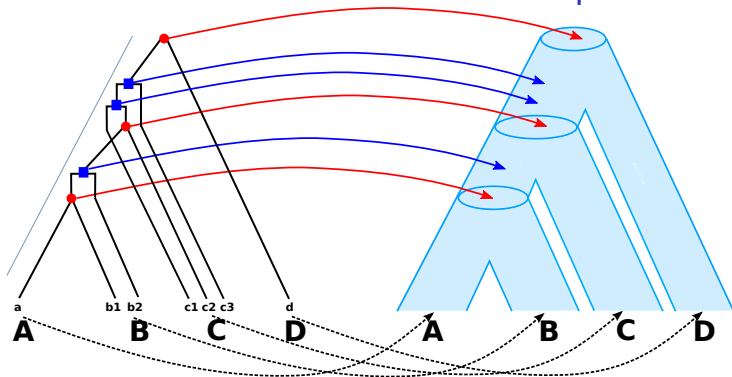
- Questions:** Are the gene trees biologically feasible? \iff
 Are there species trees for the gene trees?
 If so, how to construct species tree? \iff
 How much information about a (putative) species tree is contained in the gene tree?

From Gene Tree -without HGT- to Species Tree



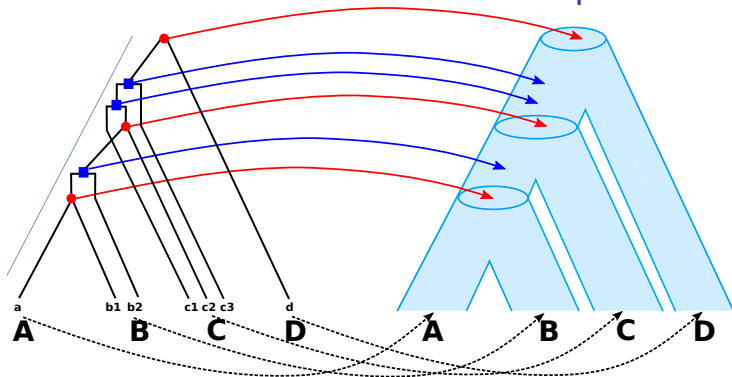
Question: When does there exist a species tree for a given gene tree

From Gene Tree -without HGT- to Species Tree



Question: When does there exist a species tree for a given gene tree and a reconciliation map μ between them?

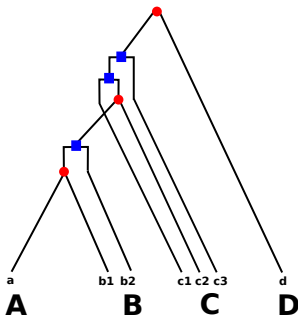
From Gene Tree -without HGT- to Species Tree



Question: When does there exist a species tree for a given gene tree and a reconciliation map μ between them?

$$\mu : V(T) \rightarrow V(S) \cup E(S) \text{ such that } \mu \text{ preserves } \preceq_T .$$

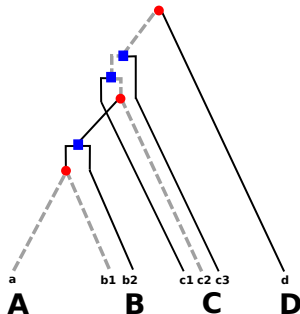
From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

Question: When does there exist a species tree for a given gene tree and a reconciliation map μ between them?

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

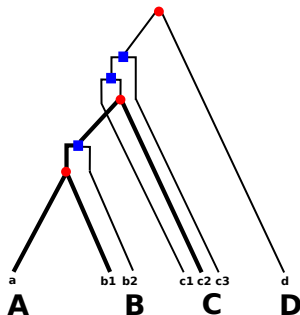
$$\mathcal{R}(T) = \{(ab_1|x) \text{ with } x = b_2, c_1, c_2, c_3, d;$$

$$(ab_2|x) \text{ with } x = c_1, c_2, c_3, d;$$

$$(b_1b_2|x) \text{ with } x = c_1, c_2, c_3, d;$$

$$\dots\}$$

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

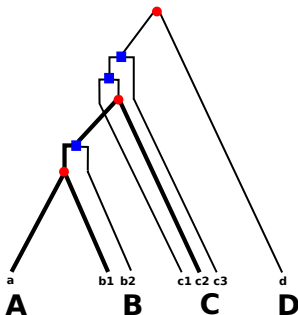
$$\mathcal{R}(T) = \{(ab_1|x) \text{ with } x = b_2, c_1, c_2, c_3, d;$$

$$(ab_2|x) \text{ with } x = c_1, c_2, c_3, d;$$

$$(b_1b_2|x) \text{ with } x = c_1, c_2, c_3, d;$$

$$\dots\}$$

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

We write $(ab|c)^\bullet$ if $(ab|c) \in \mathcal{R}(T)$

$\text{lca}(a, b, c) = \bullet = \text{"speciation"}$

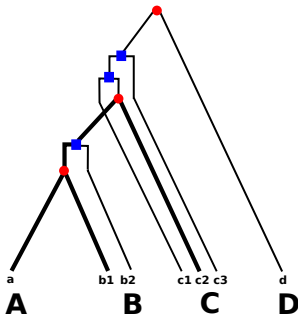
$$\mathcal{R}(T) = \{(ab_1|x) \text{ with } x = b_2, c_1, c_2, c_3, d;$$

$$(ab_2|x) \text{ with } x = c_1, c_2, c_3, d;$$

$$(b_1b_2|x) \text{ with } x = c_1, c_2, c_3, d;$$

$$\dots\}$$

From Gene Tree -without HGT- to Species Tree



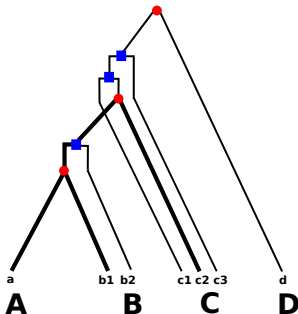
For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

We write $(ab|c)^\bullet$ if $(ab|c) \in \mathcal{R}(T)$

$\text{lca}(a, b, c) = \bullet = \text{"speciation"}$

Examples: $(ab_1|c_2)^\bullet$, $(ab_1|d)^\bullet$, $(b_2c_3|d)^\bullet$, $(ac_2|d)^\bullet$, ...

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

We write $(ab|c)^\bullet$ if $(ab|c) \in \mathcal{R}(T)$

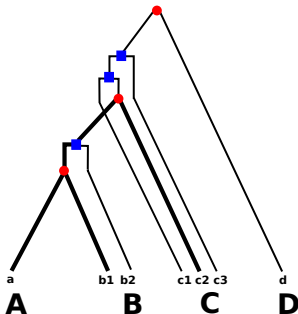
$$lca(a, b, c) = \bullet = \text{"speciation"}$$

We know the assignment of genes to the species in which they occur. This gives us the triple set:

$$\mathbb{S} = \{(AB|C) : \exists (ab|c)^\bullet \text{ with } a \in A, b \in B, c \in C\}$$

Examples: $(ab_1|c_2)^\bullet$, $(ab_1|d)^\bullet$, $(b_2c_3|d)^\bullet$, $(ac_2|d)^\bullet$, ...

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

We write $(ab|c)^\bullet$ if $(ab|c) \in \mathcal{R}(T)$

$$lca(a, b, c) = \bullet = \text{"speciation"}$$

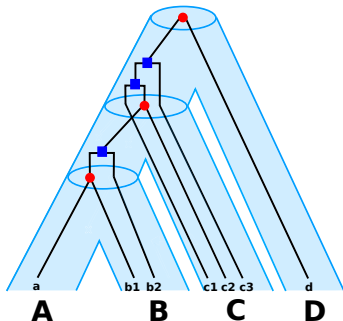
We know the assignment of genes to the species in which they occur. This gives us the triple set:

$$\mathcal{S} = \{(AB|C) : \exists (ab|c)^\bullet \text{ with } a \in A, b \in B, c \in C\}$$

Examples: $(ab_1|c_2)^\bullet$, $(ab_1|d)^\bullet$, $(b_2c_3|d)^\bullet$, $(ac_2|d)^\bullet$, ...

$$\mathcal{S} = \{(AB|C), (AB|D), (BC|D), (AC|D)\}$$

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

We write $(ab|c)^{\bullet}$ if $(ab|c) \in \mathcal{R}(T)$

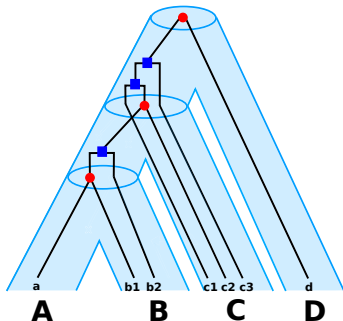
$$\text{lca}(a, b, c) = \bullet = \text{"speciation"}$$

We know the assignment of genes to the species in which they occur. This gives us the triple set:

$$\mathcal{S} = \{(AB|C) : \exists (ab|c)^{\bullet} \text{ with } a \in A, b \in B, c \in C\}$$

$$\mathcal{S} = \{(AB|C), (AB|D), (BC|D), (AC|D)\}$$

From Gene Tree -without HGT- to Species Tree



For three leaves a, b, c in T we write $(ab|c)$ if the path from a to b does not intersect the path from c to the root.

We write $(ab|c)^*$ if $(ab|c) \in \mathcal{R}(T)$

$\text{lca}(a, b, c) = \bullet = \text{"speciation"}$

We know the assignment of genes to the species in which they occur. This gives us the triple set:

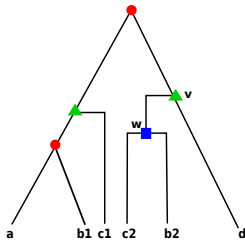
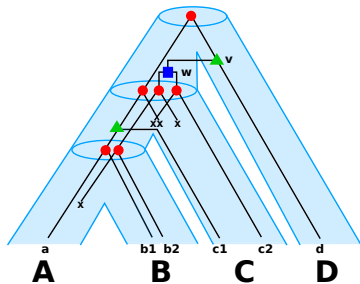
$\mathcal{S} = \{(AB|C) : \exists (ab|c)^* \text{ with } a \in A, b \in B, c \in C\}$

Theorem (2012)

There is a species tree S for the gene tree $T \iff$ the triple set \mathcal{S} is consistent (can be tested efficiently).

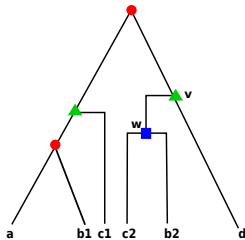
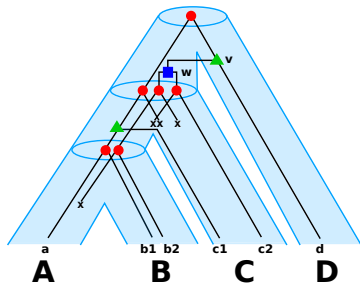
A reconciliation map μ from T to S can be constructed in polynomial time.

From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

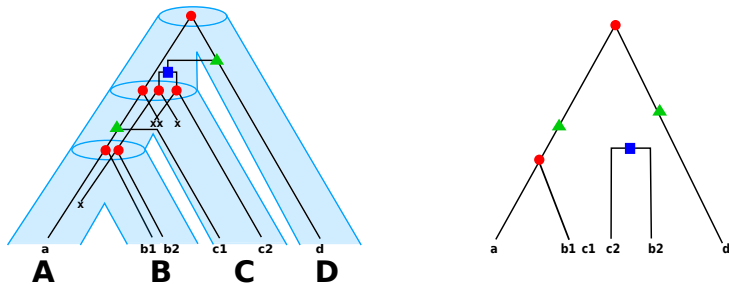
From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^* \implies (AC|D)$ and $(CD|A) \not\Leftarrow$

From Gene Tree -with HGT- to Species Tree

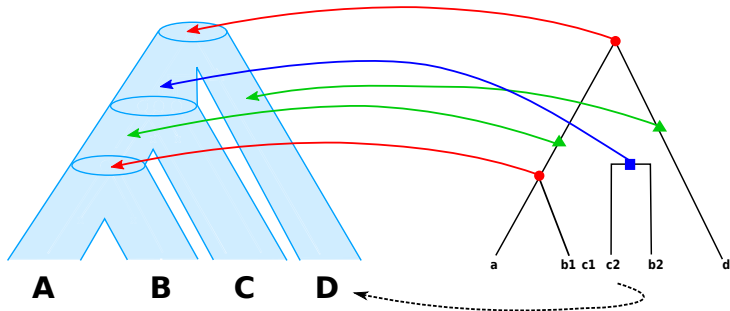


Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\in$

Idea: Remove Transfer edges from T , we get T^*

From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

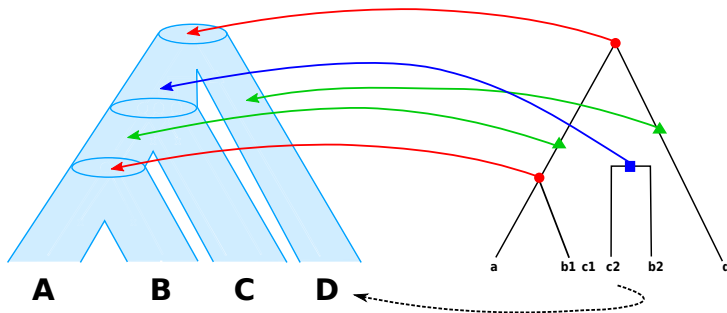
Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\prec$

Idea: Remove Transfer edges from T , we get T^*

Define μ on T^* and \preceq_{T^*} (μ preserves order on conn.components of T^*)

If (v, w) is transfer-edges, then $\mu(v)$ and $\mu(w)$ should be incomparable

From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\leq_S \mu(v)$

Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\equiv$

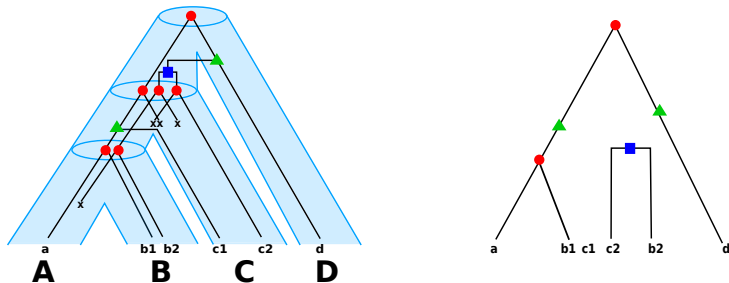
Idea: Remove Transfer edges from T , we get T^*

Define μ on T^* and \preceq_{T^*} (μ preserves order on conn.components of T^*)

If (v, w) is transfer-edges, then $\mu(v)$ and $\mu(w)$ should be incomparable

→ **Consistent with standard “DTL-scenarios”**

From Gene Tree -with HGT- to Species Tree

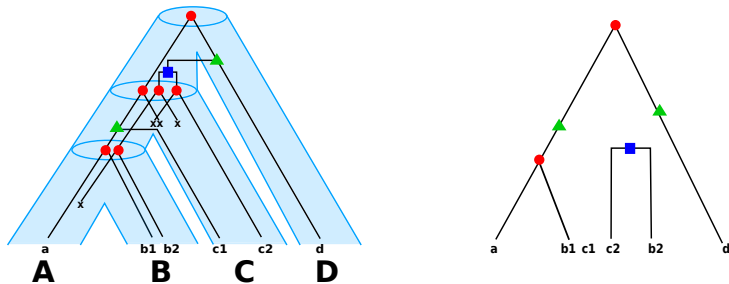


Observation: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\Leftarrow$

Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^*$ of conn.comp. of $T^* \rightarrow (ab_1|d)^* \rightarrow$ species triple $(AB|D)$

From Gene Tree -with HGT- to Species Tree

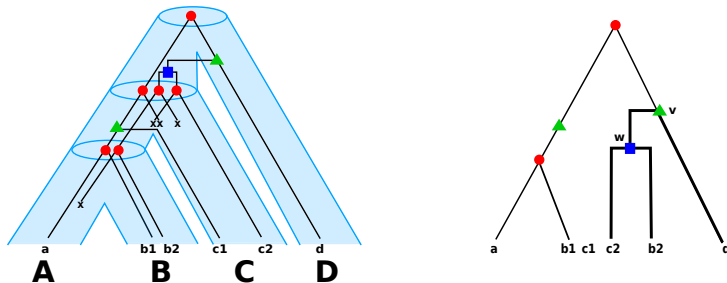


Observation: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\Leftarrow$

Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^*$ of conn.comp. of $T^* \rightarrow (ab_1|d)^* \rightarrow$ species triple $(AB|D)$ (*not sufficient*)

From Gene Tree -with HGT- to Species Tree



Observation: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\Leftarrow$

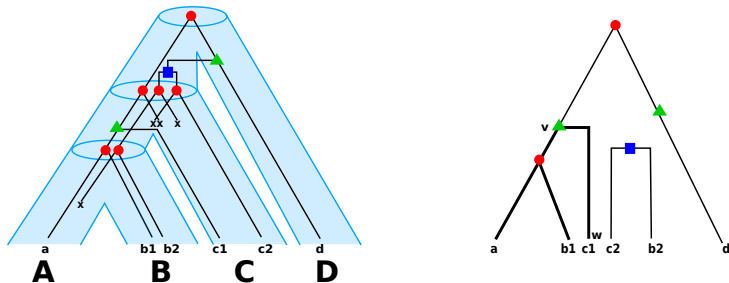
Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^*$ of conn.comp. of $T^* \rightarrow (ab_1|d)^* \rightarrow$ species triple $(AB|D)$ (*not sufficient*)

For each transfer edge (v, w) add all triples $(xy|z)^\Delta$ with x, y descend. of w , z descend. of v in T^* (or vice versa):

$$(c_2b_2|d)^\Delta \rightarrow (CB|D)$$

From Gene Tree -with HGT- to Species Tree



Observation: We see $(ac_1|d)^\bullet$ and $(c_2d|a)^\bullet \implies (AC|D)$ and $(CD|A) \not\Leftarrow$

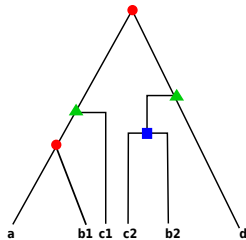
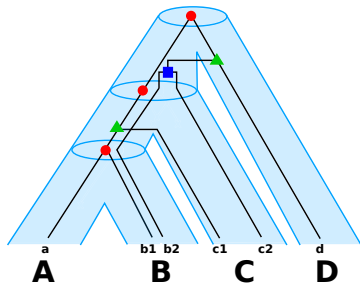
Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^\bullet$ of conn.comp. of $T^* \rightarrow (ab_1|d)^\bullet \rightarrow$ species triple $(AB|D)$ (*not sufficient*)

For each transfer edge (v, w) add all triples $(xy|z)^\blacktriangle$ with x, y descend. of w , z descend. of v in T^* (or vice versa):

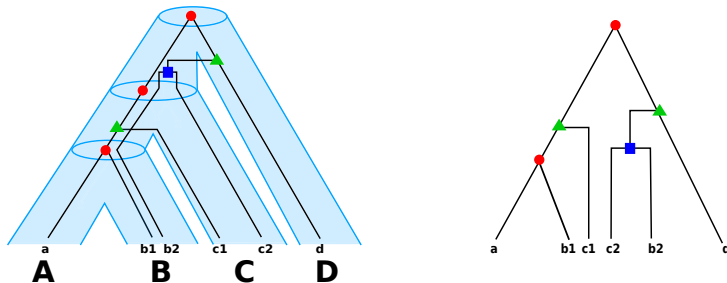
$$(c_2b_2|d)^\blacktriangle \rightarrow (CB|D) \text{ and } (ab_1|c_1)^\blacktriangle \rightarrow (AB|C)$$

From Gene Tree -with HGT- to Species Tree



$$\mathbb{S} = \{(AB|D), (CB|D), (AB|C)\}$$

From Gene Tree -with HGT- to Species Tree



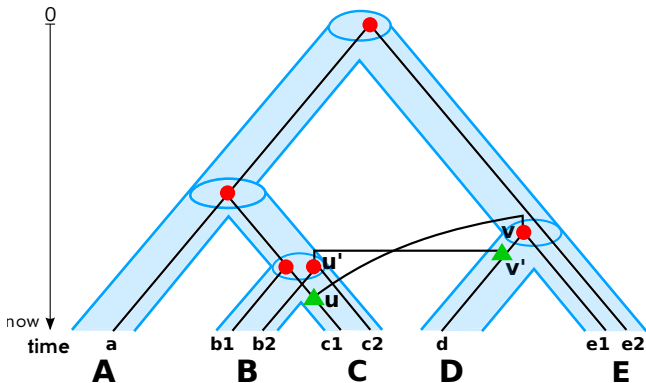
$$\mathbb{S} = \{(AB|D), (CB|D), (AB|C)\}$$

Theorem (2017)

There is a species tree S for the gene tree T with HGT \iff the triple set \mathbb{S} is consistent (can be tested efficiently).

A reconciliation map μ from T to S can be constructed in polynomial time.

Time Travel of Genes



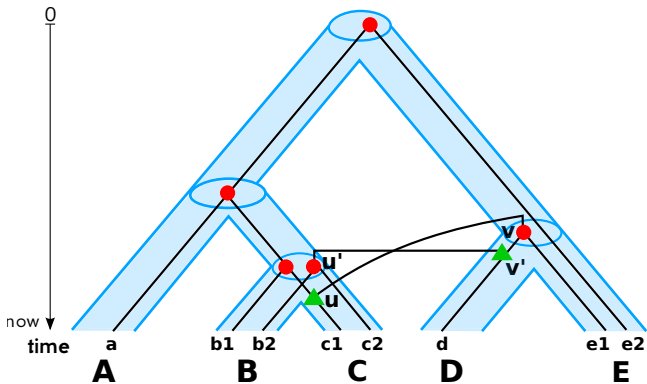
An issue that is not covered, so-far:

$$\mu(u) \prec_S \mu(u') \text{ and } \mu(v) \succ_S \mu(v')$$

If we associate a “time” τ to vertices of T :

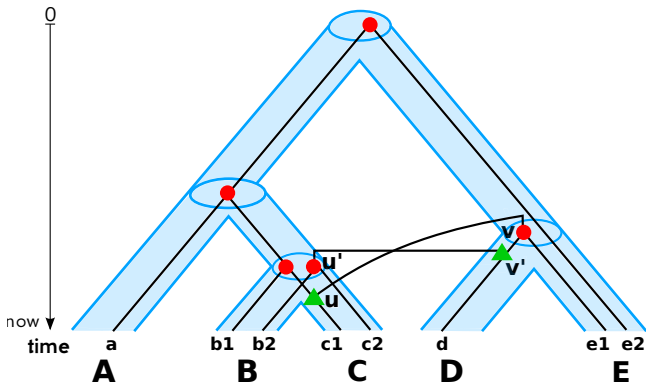
$$\tau_{u'} = \tau_{v'}, \text{ but } \tau_{u'} < \tau_u = \tau_v < \tau_{v'} \not\leq$$

Time Travel of Genes



Each species tree S must display \mathbb{S} .

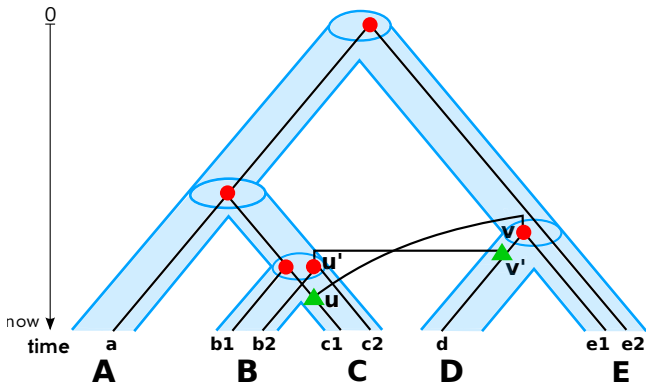
Time Travel of Genes



Each species tree S must display \mathbb{S} .

Here, S is the only tree that displays \mathbb{S} and μ is the only reconciliation map from T to S

Time Travel of Genes

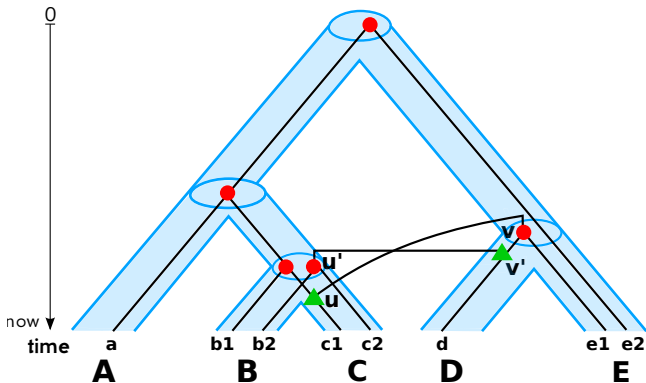


Each species tree S must display \mathbb{S} .

Here, S is the only tree that displays \mathbb{S} and μ is the only reconciliation map from T to S

⇒ There is no TIME-CONSISTENT (TC) scenario for T !

Time Travel of Genes



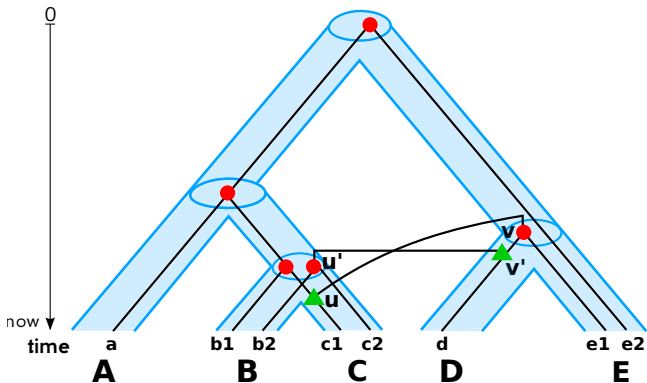
Each species tree S must display \mathbb{S} .

Here, S is the only tree that displays \mathbb{S} and μ is the only reconciliation map from T to S

⇒ There is no TIME-CONSISTENT (TC) scenario for T !

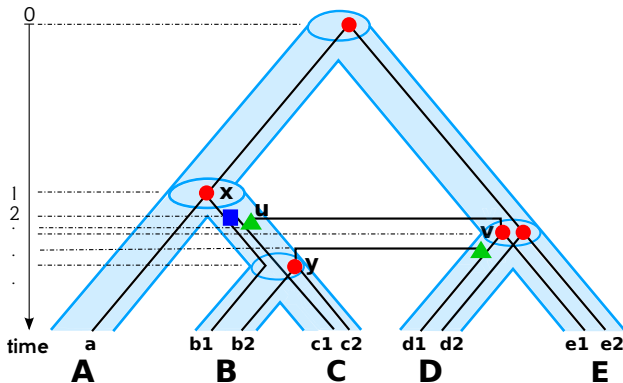
⇒ Consistency of \mathbb{S} is still necessary for TC scenario, but not sufficient!

Time Travel of Genes



There is no sufficient TC-condition provided in the literature.
Hence, we need to establish one!

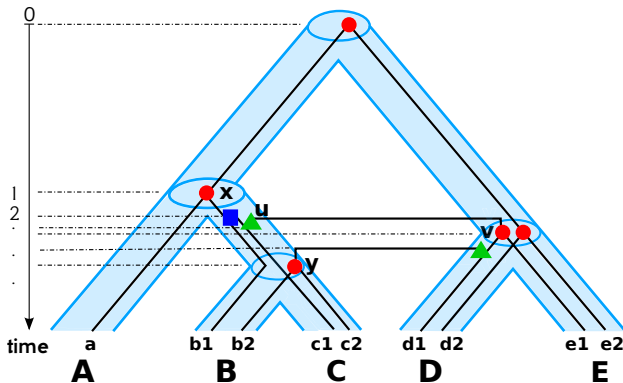
Time Travel of Genes



A reconciliation map μ is TC if there is a gene-tree time map $\tau_T : V(T) \rightarrow \mathbb{R}$ and species tree time map $\tau_S : V(S) \rightarrow \mathbb{R}$ such that

- C0 $u \prec_T v \implies \tau_T(u) > \tau_T(v)$
- C1 $u \prec_S v \implies \tau_S(u) > \tau_S(v)$
- C2 If u speciation, then $\tau_T(u) = \tau_S(\mu(u))$
- C3 If u is no speciation, then $\tau_S(x) < \tau_T(u) < \tau_S(y)$ where $\mu(u) = (x, y) \in E(S)$

Time Travel of Genes

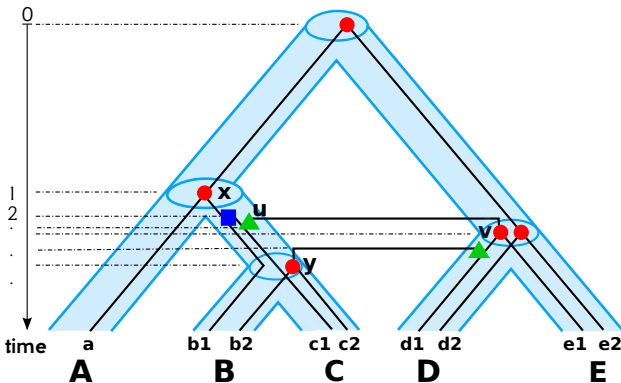


C0-C3 are satisfied \iff there is a time map $\tau : V(T) \rightarrow \mathbb{R}$ such that for any $x, y \in V(T)$:

T0 If $y \prec_T x$ then $\tau(y) > \tau(x)$.

T1 If x and y are speciation vertices and $\mu(x) = \mu(y)$, then $\tau(x) = \tau(y)$.

Time Travel of Genes

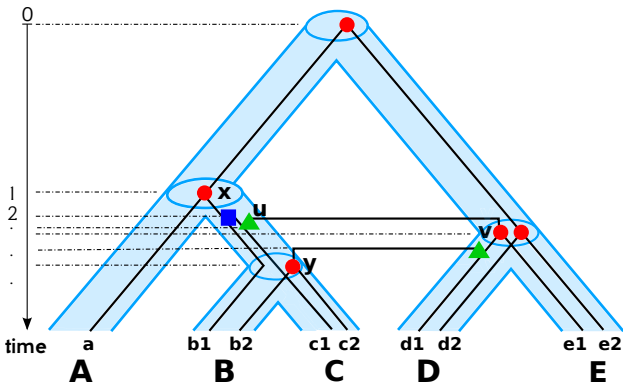


C0-C3 are satisfied \iff there is a time map $\tau : V(T) \rightarrow \mathbb{R}$ such that for any $x, y \in V(T)$:

T0 If $y \prec_T x$ then $\tau(y) > \tau(x)$.

T1 If x and y are speciation vertices and $\mu(x) = \mu(y)$, then $\tau(x) = \tau(y)$.

Time Travel of Genes



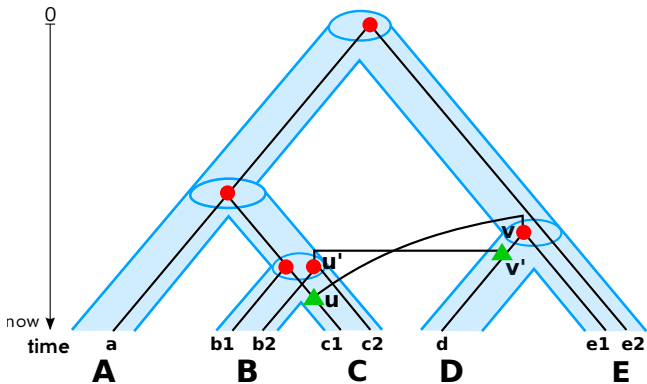
C0-C3 are satisfied \iff there is a time map $\tau : V(T) \rightarrow \mathbb{R}$ such that for any $x, y \in V(T)$:

T0 If $y \prec_T x$ then $\tau(y) > \tau(x)$.

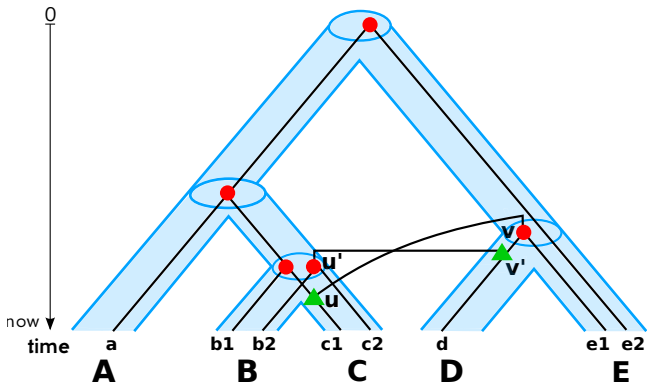
T1 If x and y are speciation vertices and $\mu(x) = \mu(y)$, then $\tau(x) = \tau(y)$.

For given T and S one can determine in poly-time whether there is a $\tau\text{-}\mu$ from T to S and construct it. (proof by Nikolai Nøjgaard in Bled 2017)

Time Travel of Genes - Open Problems



Time Travel of Genes - Open Problems



Showing that for given S there is no TC $\mu \not\Rightarrow$ **NO** species tree exists for T .

Maybe, there is another species tree that displays all triples in \mathbb{S} with valid μ ?!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display S

A *biol. meaningful* gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A *biol. meaningful* gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A *biol. meaningful* gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A biol. meaningful gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A biol. meaningful gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A biol. meaningful gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A biol. meaningful gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau\mathcal{C}\text{-}\mu$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau\mathcal{C}\text{-}\mu$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A biol. meaningful gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau_{C-\mu}$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau_{C-\mu}$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

The general problem statement

A gene tree T is “*biol. meaningful*” IFF \mathcal{S} is consistent.

Let \mathcal{S} be the set of species tree that display \mathcal{S}

A biol. meaningful gene tree T is “*biologically feasible*” \iff
there is a species tree $S \in \mathcal{S}$ with $\tau_{C-\mu}$ from T to S .

Observation:

$|\mathcal{S}|$ “small” \implies only few restrictions on the species tree $S \implies$
number of possible species trees $|\mathcal{S}|$ “high”

Example: If $\mathcal{S} = \emptyset$ and n species, then $|\mathcal{S}| = (2n - 3)!!$

Tasks:

Check if T is “*biol. meaningful*” \rightarrow DONE

Check if for T exists $\tau_{C-\mu}$ w.r.t. a given $S \rightarrow$ DONE

Check if T is “*biologically feasible*” \rightarrow Poly-time? NP-hard?

Great Task for a PhD-student (Manuela, Sarah, Nikolai)!

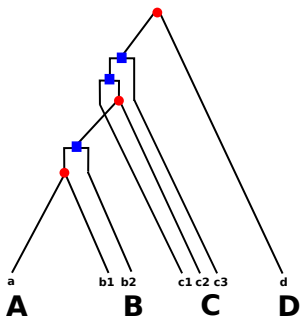
**THANKS TO
Nikolai, Nic, Peter, John, Paul, Sarah, Manuela ..**

**THANKS TO
Nikolai, Nic, Peter, John, Paul, Sarah, Manuela ..
.. and YOU!**



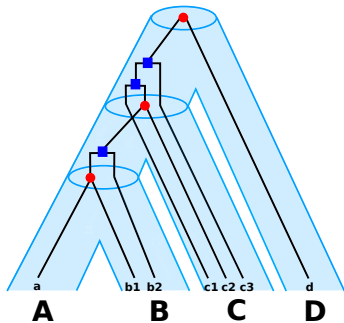
– It's time for an isotonic sports drink! –

From Gene Tree -without HGT- to Species Tree



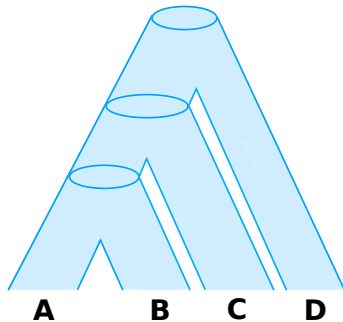
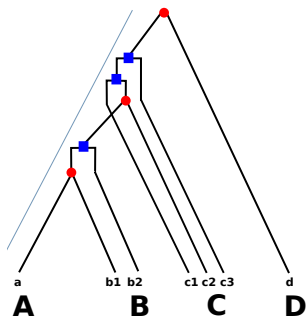
Question: When does there exist a species tree for a given gene tree

From Gene Tree -without HGT- to Species Tree



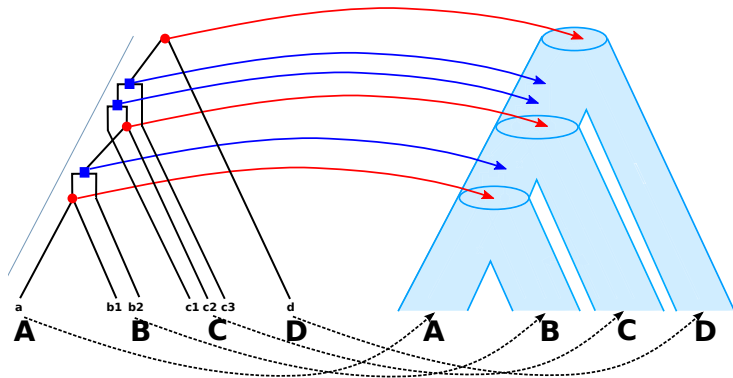
Question: When does there exist a species tree for a given gene tree

From Gene Tree - *without HGT* - to Species Tree



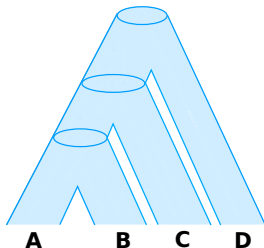
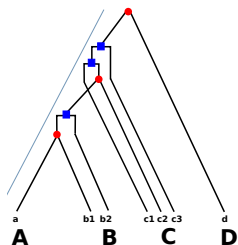
Question: When does there exist a species tree for a given gene tree

From Gene Tree -without HGT- to Species Tree



Question: When does there exist a species tree for a given gene tree and a reconciliation map μ between them?

From Gene Tree -without HGT- to Species Tree



Given:

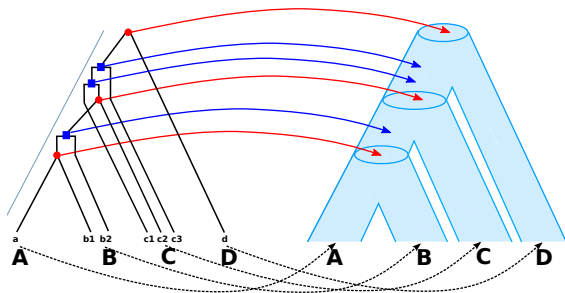
Gene tree T on \mathbb{G}

Species tree S on \mathbb{S}

Gene-Species map

$\sigma : \mathbb{G} \rightarrow \mathbb{S}$

From Gene Tree -without HGT- to Species Tree



Given:

Gene tree T on \mathbb{G}

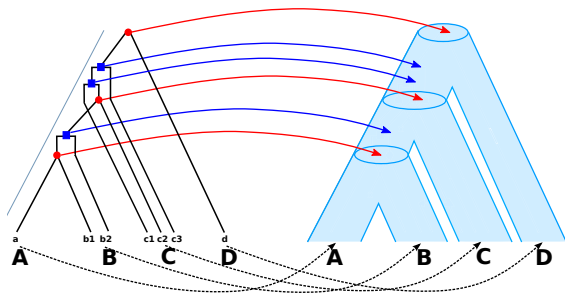
Species tree S on \mathbb{S}

Gene-Species map

$\sigma : \mathbb{G} \rightarrow \mathbb{S}$

Map $\mu : V(T) \rightarrow V(S) \cup E(S)$ is a *reconciliation map* if for all $x \in V(T)$:

From Gene Tree -without HGT- to Species Tree



Given:

Gene tree T on \mathbb{G}

Species tree S on \mathbb{S}

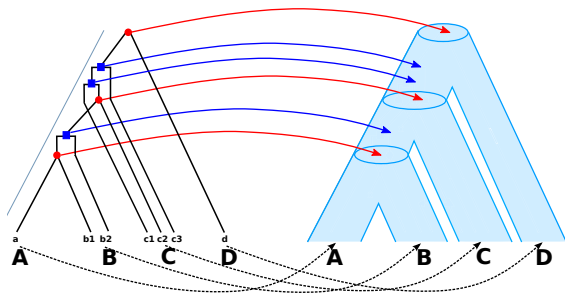
Gene-Species map

$\sigma : \mathbb{G} \rightarrow \mathbb{S}$

Map $\mu : V(T) \rightarrow V(S) \cup E(S)$ is a *reconciliation map* if for all $x \in V(T)$:

Leaf Constraint. If $x \in \mathbb{G}$, then $\mu(x) = \sigma(x)$.

From Gene Tree -without HGT- to Species Tree



Given:

Gene tree T on \mathbb{G}

Species tree S on \mathbb{S}

Gene-Species map

$\sigma : \mathbb{G} \rightarrow \mathbb{S}$

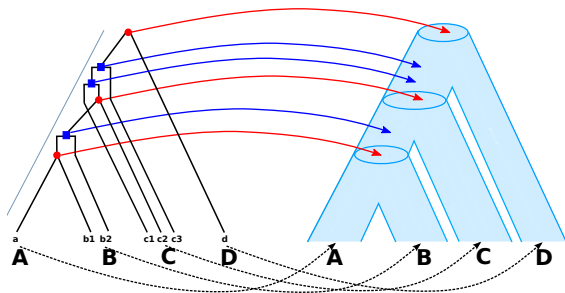
Map $\mu : V(T) \rightarrow V(S) \cup E(S)$ is a *reconciliation map* if for all $x \in V(T)$:

Leaf Constraint. If $x \in \mathbb{G}$, then $\mu(x) = \sigma(x)$.

Event Constraint.

- (i) If $t(x) = \bullet$, then $\mu(x) = \text{lca}_S(\sigma(L_T(x)))$.
- (ii) If $t(x) \in \{\square\}$, then $\mu(x) \in F$.

From Gene Tree -without HGT- to Species Tree



Given:

Gene tree T on \mathbb{G}

Species tree S on \mathbb{S}

Gene-Species map

$\sigma : \mathbb{G} \rightarrow \mathbb{S}$

Map $\mu : V(T) \rightarrow V(S) \cup E(S)$ is a *reconciliation map* if for all $x \in V(T)$:

Leaf Constraint. If $x \in \mathbb{G}$, then $\mu(x) = \sigma(x)$.

Ancestor Constraint.

Event Constraint.

Let $x, y \in V$ with $x \prec_T y$.

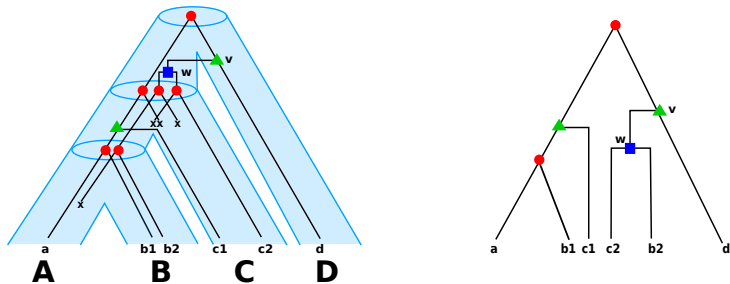
- (i) If $t(x) = \bullet$, then $\mu(x) = \text{lca}_S(\sigma(L_T(x)))$.

- (i) If $t(x), t(y) \in \{\square\}$, then $\mu(x) \preceq_S \mu(y)$,

- (ii) If $t(x) \in \{\square\}$, then $\mu(x) \in F$.

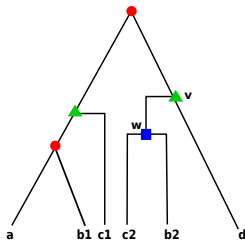
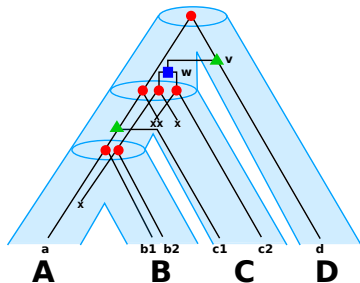
- (ii) otherwise, i.e., at least one of $t(x)$ and $t(y)$ is a speciation \bullet , $\mu(x) \prec_S \mu(y)$.

From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

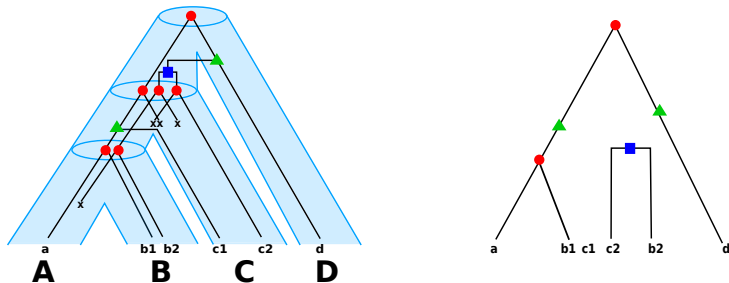
From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\vdash$

From Gene Tree -with HGT- to Species Tree

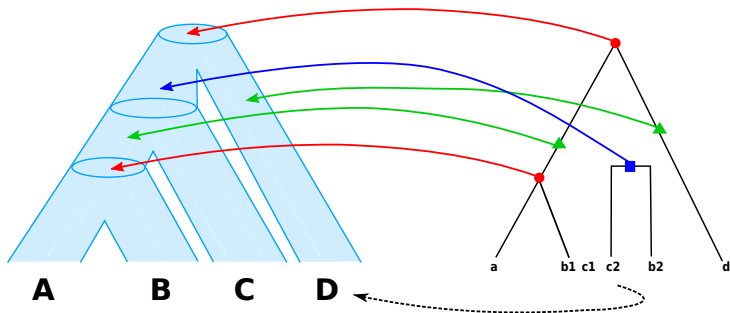


Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\in$

Idea: Remove Transfer edges from T , we get T^*

From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\prec_S \mu(v)$

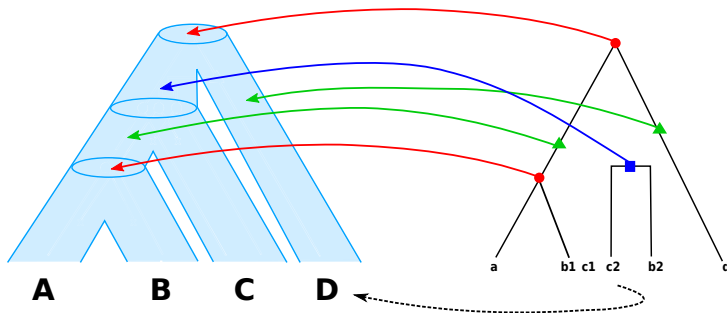
Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\prec$

Idea: Remove Transfer edges from T , we get T^*

Define μ on T^* and \preceq_{T^*} (μ preserves order on conn.components of T^*)

If (v, w) is transfer-edges, then $\mu(v)$ and $\mu(w)$ should be incomparable

From Gene Tree -with HGT- to Species Tree



Observation 1: Partial Order in Gene tree T and Species tree S : $w \prec_T v$ but $\mu(w) \not\leq_S \mu(v)$

Observation 2: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\vdash$

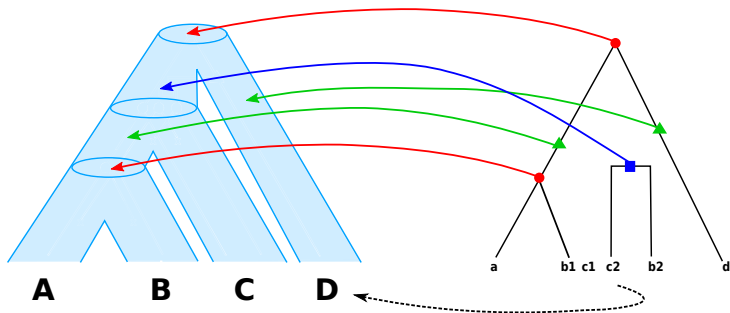
Idea: Remove Transfer edges from T , we get T^*

Define μ on T^* and \preceq_{T^*} (μ preserves order on conn.components of T^*)

If (v, w) is transfer-edges, then $\mu(v)$ and $\mu(w)$ should be incomparable

→ **Consistent with standard “DTL-scenarios”**

From Gene Tree -with HGT- to Species Tree



Leaf Constraint. If $x \in \mathbb{G}$, then $\mu(x) = \sigma(x)$.

Event Constraint.

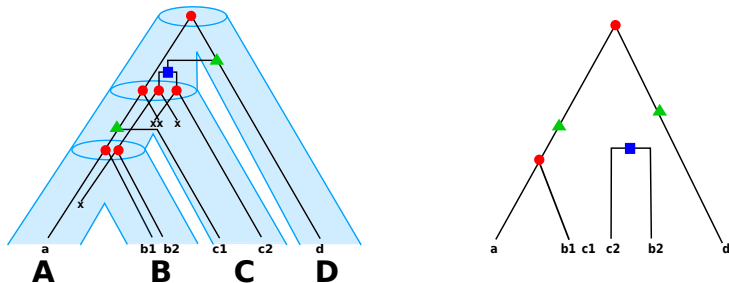
- (i) If $t(x) = \bullet$, then $\mu(x) = \text{lca}_S(\sigma(L_{T^*}(x)))$.
- (ii) If $t(x) \in \{\square, \triangle\}$, then $\mu(x) \in F$.
- (iii) If $t(x) = \triangle$ and $(x, y) \in \mathcal{E}$, then $\mu(x)$ and $\mu(y)$ are incomparable in S .

Ancestor Constraint.

Let $x, y \in V$ with $x \prec_{T^*} y$.

- (i) If $t(x), t(y) \in \{\square, \triangle\}$, then $\mu(x) \preceq_S \mu(y)$,
- (ii) otherwise, i.e., at least one of $t(x)$ and $t(y)$ is a speciation \bullet , $\mu(x) \prec_S \mu(y)$.

From Gene Tree -with HGT- to Species Tree

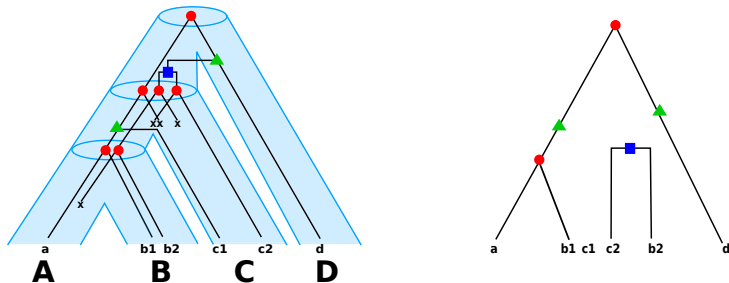


Observation: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\Leftarrow$

Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^*$ of conn.comp. of $T^* \rightarrow (ab_1|d)^* \rightarrow$ species triple $(AB|D)$

From Gene Tree -with HGT- to Species Tree

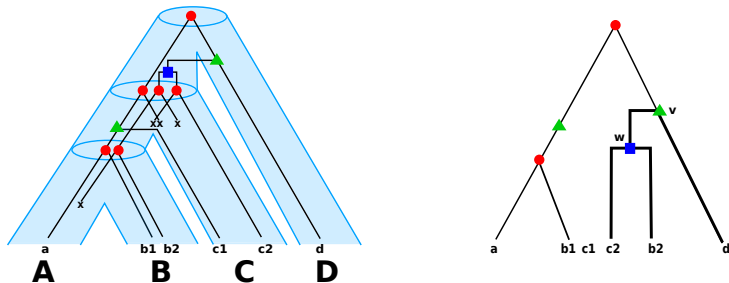


Observation: We see $(ac_1|d)^*$ and $(c_2d|a)^*$ $\implies (AC|D)$ and $(CD|A) \not\Leftarrow$

Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^*$ of conn.comp. of $T^* \rightarrow (ab_1|d)^* \rightarrow$ species triple $(AB|D)$ (*not sufficient*)

From Gene Tree -with HGT- to Species Tree



Observation: We see $(ac_1|d)^\bullet$ and $(c_2d|a)^\bullet \implies (AC|D)$ and $(CD|A) \not\Leftarrow$

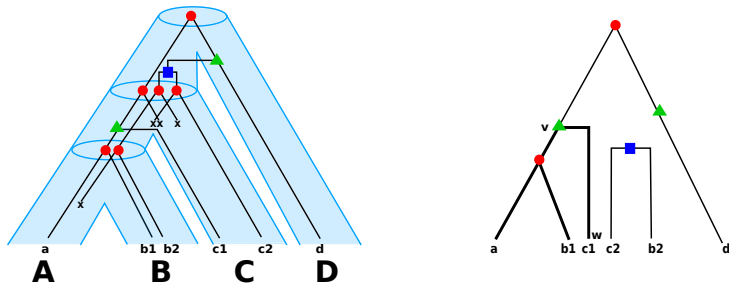
Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^\bullet$ of conn.comp. of $T^* \rightarrow (ab_1|d)^\bullet \rightarrow$ species triple $(AB|D)$ (*not sufficient*)

For each transfer edge (v, w) add all triples $(xy|z)^\blacktriangle$ with x, y descend. of w , z descend. of v in T^* (or vice versa):

$$(c_2b_2|d)^\blacktriangle \rightarrow (CB|D)$$

From Gene Tree -with HGT- to Species Tree



Observation: We see $(ac_1|d)^\bullet$ and $(c_2d|a)^\bullet \implies (AC|D)$ and $(CD|A) \not\Leftarrow$

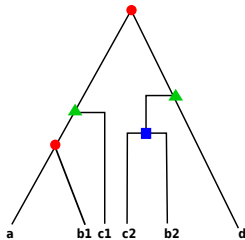
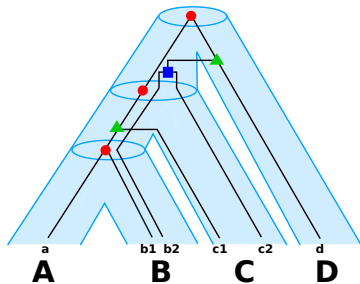
Idea: Remove Transfer edges from T , we get T^*

Use only $(xy|z)^\bullet$ of conn.comp. of $T^* \rightarrow (ab_1|d)^\bullet \rightarrow$ species triple $(AB|D)$ (*not sufficient*)

For each transfer edge (v, w) add all triples $(xy|z)^\blacktriangle$ with x, y descend. of w , z descend. of v in T^* (or vice versa):

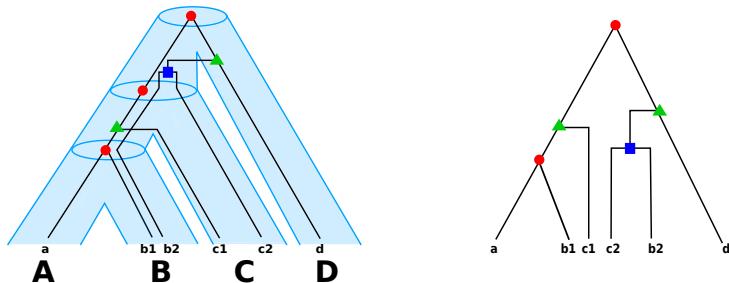
$$(c_2b_2|d)^\blacktriangle \rightarrow (CB|D) \text{ and } (ab_1|c_1)^\blacktriangle \rightarrow (AB|C)$$

From Gene Tree -with HGT- to Species Tree



$$\mathbb{S} = \{(AB|D), (CB|D), (AB|C)\}$$

From Gene Tree -with HGT- to Species Tree

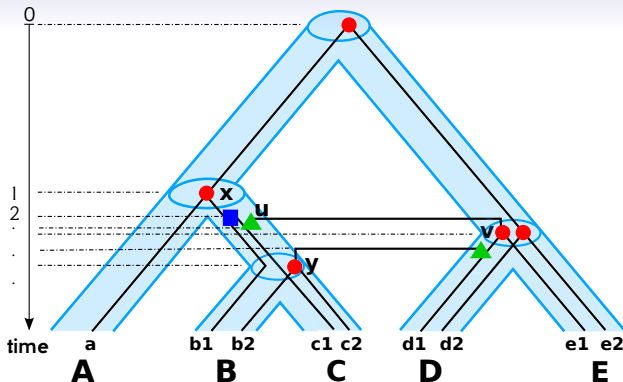


$$\mathbb{S} = \{(AB|D), (CB|D), (AB|C)\}$$

Theorem (2017)

There is a species tree S for the gene tree T with HGT \iff the triple set \mathbb{S} is consistent (can be tested efficiently).

A reconciliation map μ from T to S can be constructed in polynomial time.



A reconciliation map μ is TC if there is a gene-tree time map $\tau_T : V(T) \rightarrow \mathbb{R}$ and species tree time map $\tau_S : V(S) \rightarrow \mathbb{R}$ such that

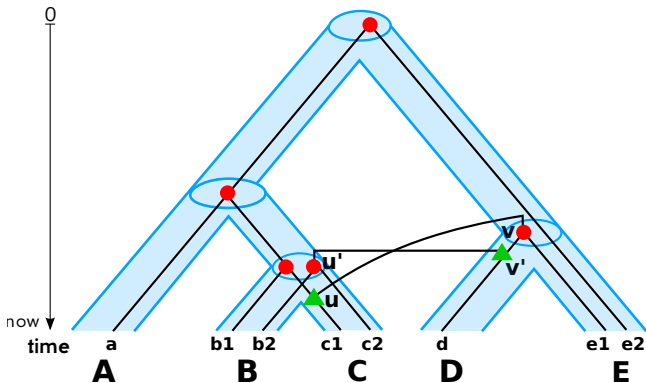
C0 $u \prec_T v \implies \tau_T(u) > \tau_T(v)$

C1 $u \prec_S v \implies \tau_S(u) > \tau_S(v)$

C2 If u speciation, then $\tau_T(u) = \tau_S(\mu(u))$

C3 If u is no speciation, then $\tau_S(x) < \tau_T(u) < \tau_S(y)$ where $\mu(u) = (x, y) \in E(S)$

Time Travel of Genes



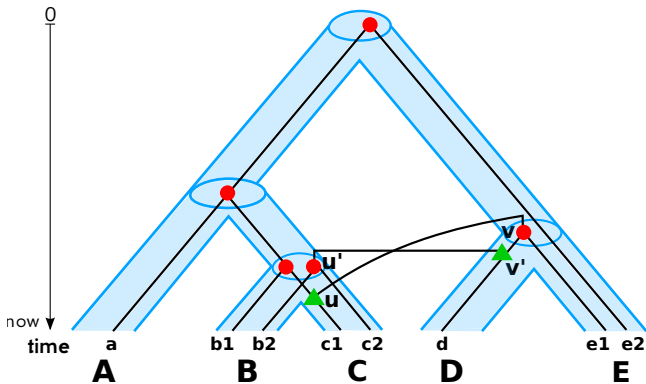
An issue that is not covered, so-far:

$$\mu(u) \prec_S \mu(u') \text{ and } \mu(v) \succ_S \mu(v')$$

If we associate a “time” τ to vertices of T :

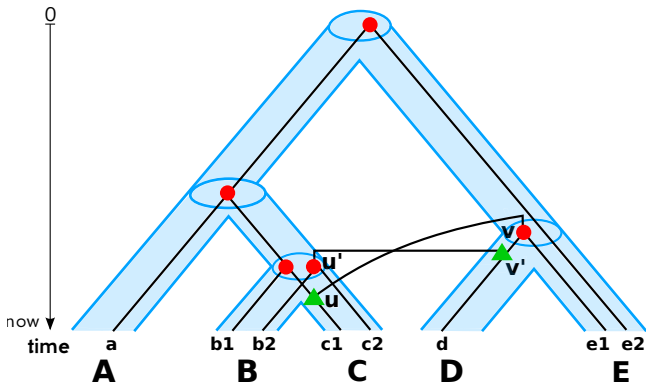
$$\tau_{u'} = \tau_{v'}, \text{ but } \tau_{u'} < \tau_u = \tau_v < \tau_{v'} \not\Leftarrow$$

Time Travel of Genes



Each species tree S must display \mathcal{S} .

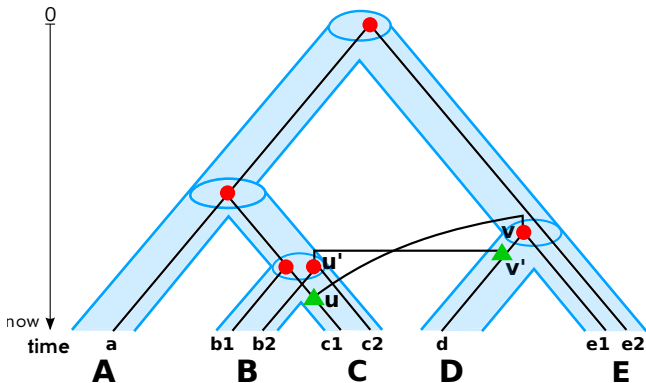
Time Travel of Genes



Each species tree S must display \mathbb{S} .

Here, S is the only tree that displays \mathbb{S} and μ is the only reconciliation map from T to S

Time Travel of Genes

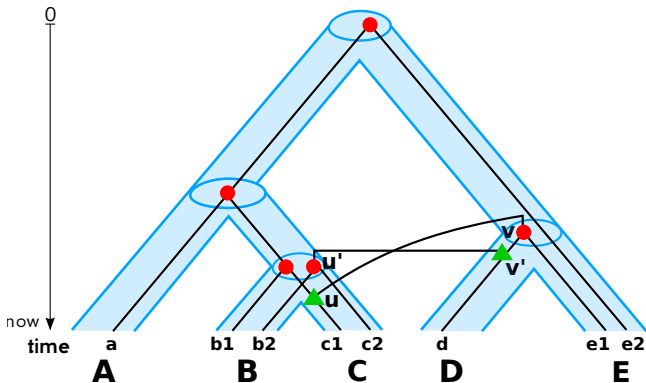


Each species tree S must display \mathcal{S} .

Here, S is the only tree that displays \mathcal{S} and μ is the only reconciliation map from T to S

⇒ There is no TIME-CONSISTENT (TC) scenario for T !

Time Travel of Genes



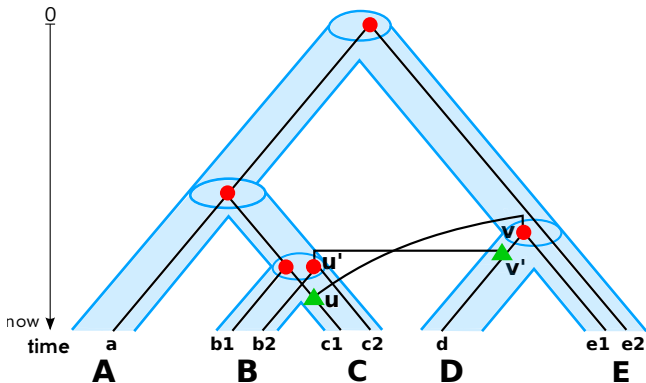
Each species tree S must display \mathbb{S} .

Here, S is the only tree that displays \mathbb{S} and μ is the only reconciliation map from T to S

⇒ There is no TIME-CONSISTENT (TC) scenario for T !

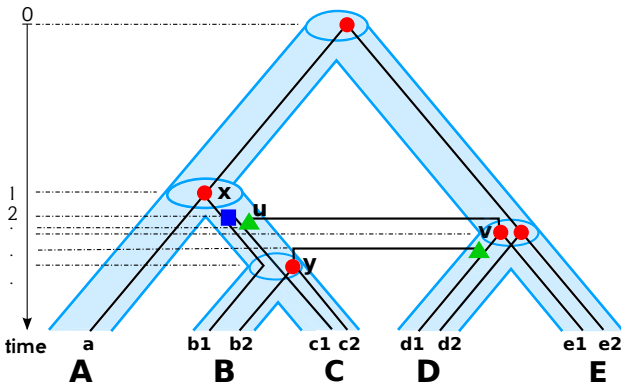
⇒ Consistency of \mathbb{S} is still necessary for TC scenario, but not sufficient!

Time Travel of Genes



There is no sufficient TC-condition provided in the literature.
Hence, we need to establish one!

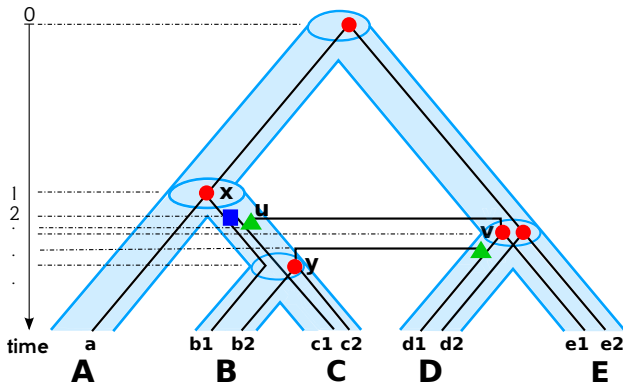
Time Travel of Genes



A reconciliation map μ is TC if there is a gene-tree time map $\tau_T : V(T) \rightarrow \mathbb{R}$ and species tree time map $\tau_S : V(S) \rightarrow \mathbb{R}$ such that

- C0** $u \prec_T v \implies \tau_T(u) > \tau_T(v)$
- C1** $u \prec_S v \implies \tau_S(u) > \tau_S(v)$
- C2** If u is a speciation node, then $\tau_T(u) = \tau_S(\mu(u))$
- C3** If u is not a speciation node, then $\tau_S(x) < \tau_T(u) < \tau_S(y)$ where $\mu(u) = (x, y) \in E(S)$

Time Travel of Genes

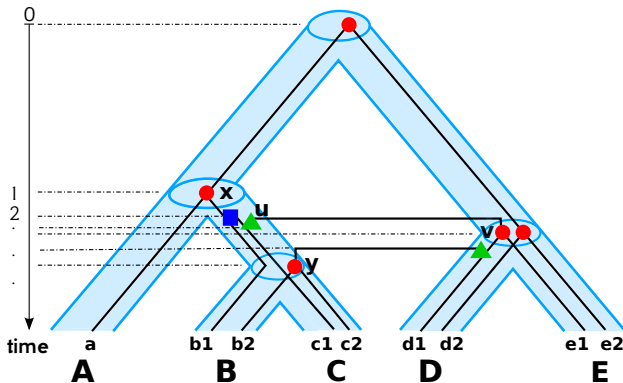


C0-C3 are satisfied \iff there is a time map $\tau : V(T) \rightarrow \mathbb{R}$ such that for any $x, y \in V(T)$:

T0 If $y \prec_T x$ then $\tau(y) > \tau(x)$.

T1 If x and y are speciation vertices and $\mu(x) = \mu(y)$, then $\tau(x) = \tau(y)$.

Time Travel of Genes

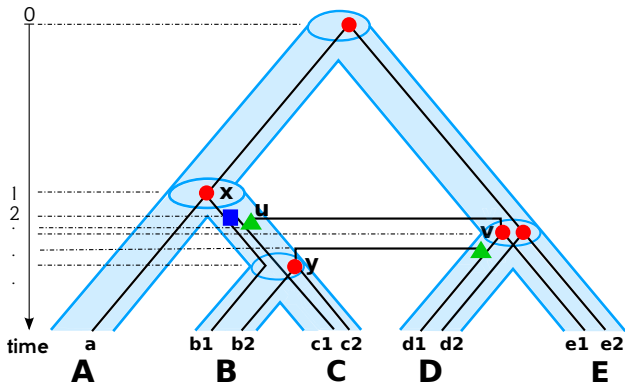


C0-C3 are satisfied \iff there is a time map $\tau : V(T) \rightarrow \mathbb{R}$ such that for any $x, y \in V(T)$:

T0 If $y \prec_T x$ then $\tau(y) > \tau(x)$.

T1 If x and y are speciation vertices and $\mu(x) = \mu(y)$, then $\tau(x) = \tau(y)$.

Time Travel of Genes



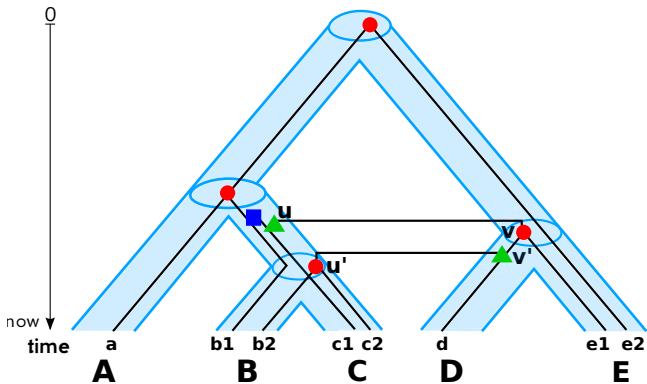
C0-C3 are satisfied \iff there is a time map $\tau : V(T) \rightarrow \mathbb{R}$ such that for any $x, y \in V(T)$:

T0 If $y \prec_T x$ then $\tau(y) > \tau(x)$.

T1 If x and y are speciation vertices and $\mu(x) = \mu(y)$, then $\tau(x) = \tau(y)$.

For given T and S one can determine in poly-time whether there is a $\tau\text{-}\mu$ from T to S and construct it. (proof by Nikolai Nøjgaard in Bled 2017)

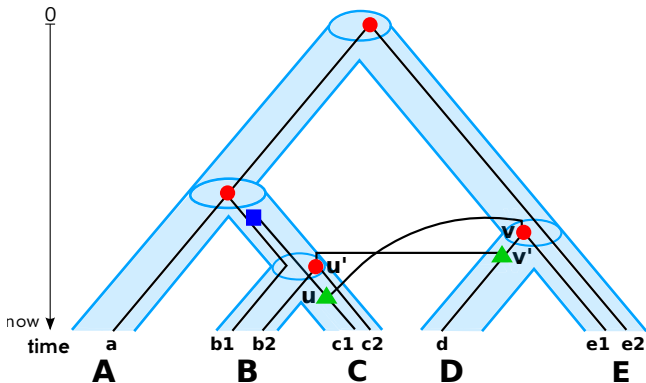
Time Travel of Genes - Open Problems



In general, neither μ nor species tree S must be unique!

If μ is not TC, then there might be another μ' that is TC.

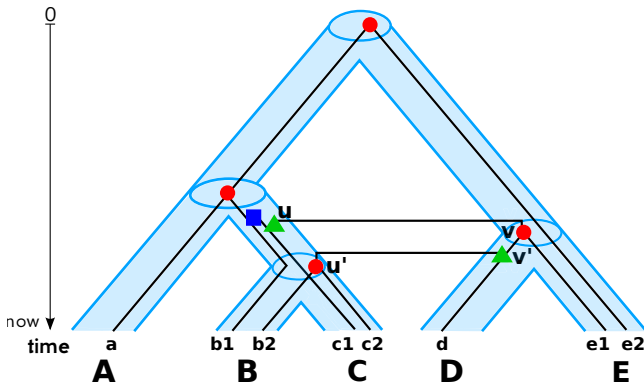
Time Travel of Genes - Open Problems



In general, neither μ nor species tree S must be unique!

If μ is not TC, then there might be another μ' that is TC.

Time Travel of Genes - Open Problems

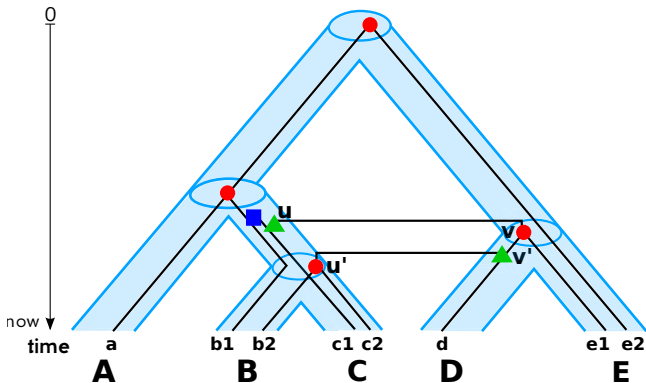


In general, neither μ nor species tree S must be unique!

If μ is not TC, then there might be another μ' that is TC.

Showing that for given S there is no TC $\mu \not\Rightarrow$ **NO** species tree exists for T .

Time Travel of Genes - Open Problems



In general, neither μ nor species tree S must be unique!

If μ is not TC, then there might be another μ' that is TC.

Showing that for given S there is no TC $\mu \not\Rightarrow$ **NO** species tree exists for T .

Maybe, there is another species tree that displays all triples in \mathcal{S} with valid μ !?