A Formal Definition of Orthology Using the Relation θ

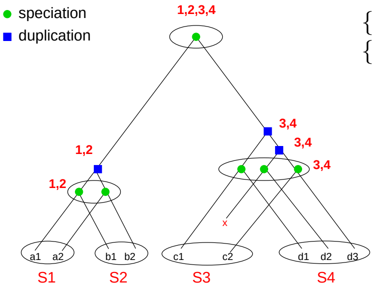
Marc Hellmuth

Bioinformatics Group
Department of Computer Science
University Leipzig

joint work with: Nicolas Wieseke, Maribel Hernandez-Rosales and Peter F. Stadler

February 16, 2011

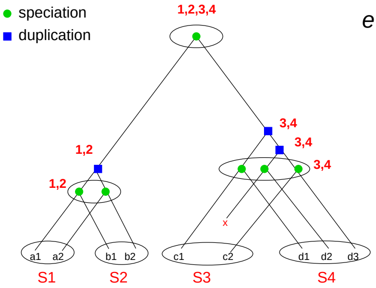
Reconciliation Tree Π



A *reconciliation tree* Π is a rooted, binary tree (V,E) with vertex labeling $e(v) \in \{\text{duplication, speciation, loss, none}\}$ and $s(v) \subseteq \{1,\ldots,n\}$ such that for all vertices $v,w \in V$ hold:

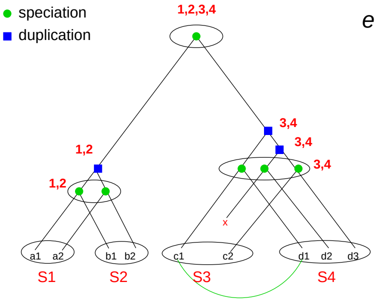
1. For the root vertex r holds $s(r) = \{1,\ldots,n\}$
2. $v \notin L(\Pi) \Leftrightarrow e(v) \in \{\text{dupl., spec.}\}$ $v \in L(\Pi) \Leftrightarrow e(v) \in \{\text{loss, none}\}$
3. $e(v) = \text{duplication} \Leftrightarrow$ $s(v.1) = s(v.2) = s(v)$
4. $e(v) = \text{speciation} \Leftrightarrow$ there exists a bipartition $s(v) = s(v.1) \cup s(v.2)$.
5. $e(v) = \text{none} \Rightarrow |s(v)| = 1$

Orthology and Relation θ



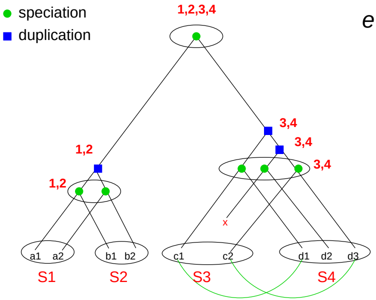
Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Orthology and Relation θ



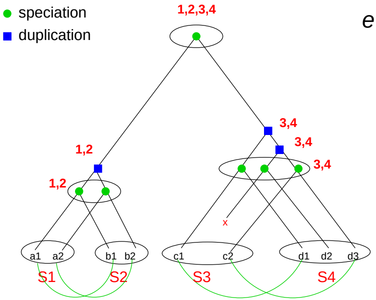
Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Orthology and Relation θ



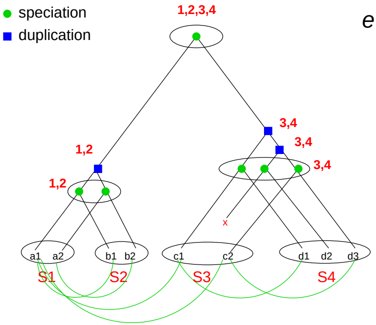
Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Orthology and Relation θ



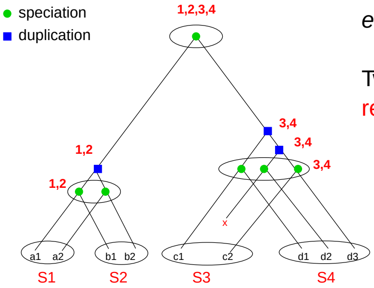
Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Orthology and Relation θ



Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Orthology and Relation θ



Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Two genes $x \in S_i$ and $y \in S_j$ are in relation θ if

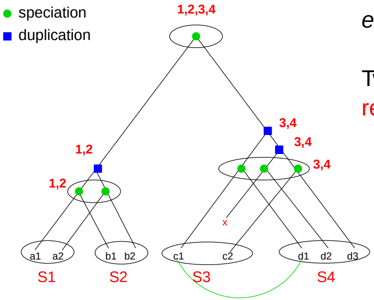
$$lca(x,y) \leq lca(x,g') \text{ for all } g' \in S_j$$

and

$$lca(x,y) \leq lca(g',y) \text{ for all } g' \in S_i.$$

If in addition $e(lca(x,y)) = speciation$ then we write: $x\theta^*y$.

Orthology and Relation θ



Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Two genes $x \in S_i$ and $y \in S_j$ are in relation θ if

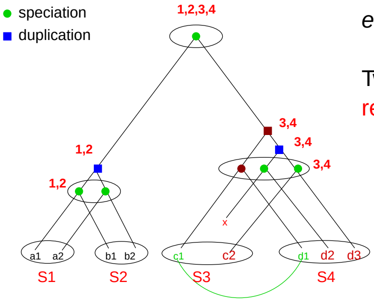
$$lca(x,y) \leq lca(x,g') \text{ for all } g' \in S_j$$

and

$$lca(x,y) \leq lca(g',y) \text{ for all } g' \in S_i.$$

If in addition $e(lca(x,y)) = speciation$ then we write: $x\theta^*y$.

Orthology and Relation θ



Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Two genes $x \in S_i$ and $y \in S_j$ are in relation θ if

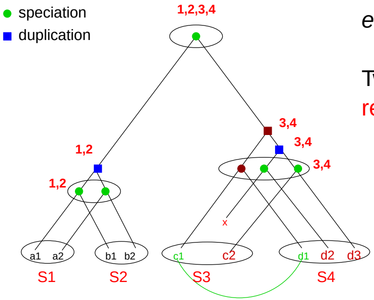
$$lca(x,y) \leq lca(x,g') \text{ for all } g' \in S_j$$

and

$$lca(x,y) \leq lca(g',y) \text{ for all } g' \in S_i.$$

If in addition $e(lca(x,y)) = speciation$ then we write: $x\theta^* y$.

Orthology and Relation θ



Two genes $x \in S_i$ and $y \in S_j$ are orthologous if $e(lca(x,y)) = speciation$.

Two genes $x \in S_i$ and $y \in S_j$ are in relation θ if

$$lca(x,y) \leq lca(x,g') \text{ for all } g' \in S_j$$

and

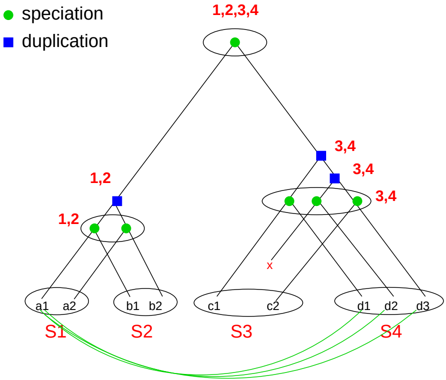
$$lca(x,y) \leq lca(g',y) \text{ for all } g' \in S_i.$$

If in addition $e(lca(x,y)) = speciation$ then we write: $x\theta^*y$.

Lemma

If x and y are orthologous in Π then $x\theta y$.
Genes x and y are orthologous in Π without losses if and only if $x\theta y$, i.e. $\theta = \theta^*$.

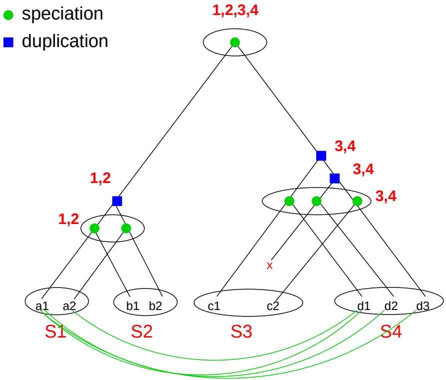
Properties of θ



Lemma

Given an arbitrary reconciliation tree Π *and two vertices* $x, x' \in S_i$ *with* $x\theta Y = \{y_1, \ldots, y_n\}$ *and* $x'\theta y_i$ *with* $y_i \in Y$*. Then it holds* $x'\theta Y$*.*

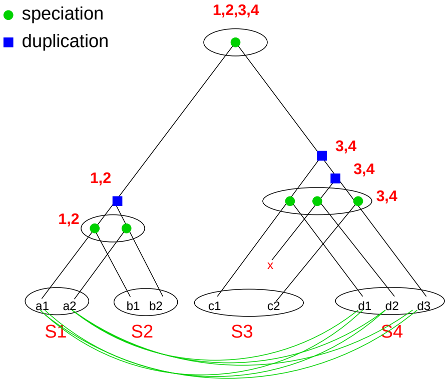
Properties of θ



Lemma

Given an arbitrary reconciliation tree Π *and two vertices* $x, x' \in S_i$ *with* $x\theta Y = \{y_1, \ldots, y_n\}$ *and* $x'\theta y_i$ *with* $y_i \in Y$*. Then it holds* $x'\theta Y$*.*

Properties of θ



Lemma

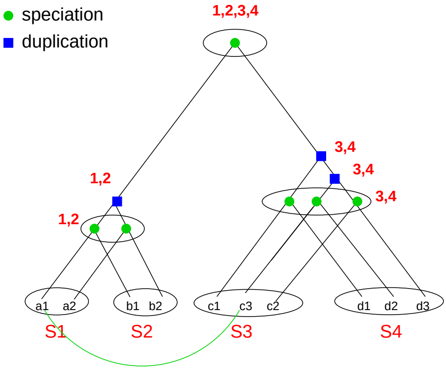
Given an arbitrary reconciliation tree Π *and two vertices* $x, x' \in S_i$ *with* $x\theta Y = \{y_1, \ldots, y_n\}$ *and* $x'\theta y_i$ *with* $y_i \in Y$. *Then it holds* $x'\theta Y$.

Properties of θ

Lemma

Given a reconciliation tree Π without losses. Then for all vertices $x \in S_i$ exists a vertex $y \in S_j$ such that $x \theta y$.

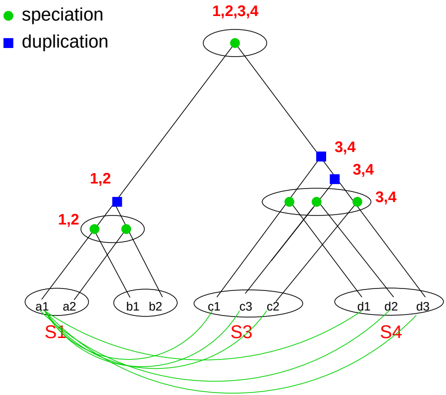
Properties of θ



Lemma

Given a reconciliation tree Π without losses. Let $x \in S_i$ and $y \in S_j$ such that $x \theta y$ and let $v = lca(x, y)$. Then it holds for all $w_1, w_2 \in L(\Pi)$ with $w_1 \leq v.1$ and $w_2 \leq v.2$ that $w_1 \theta w_2$.

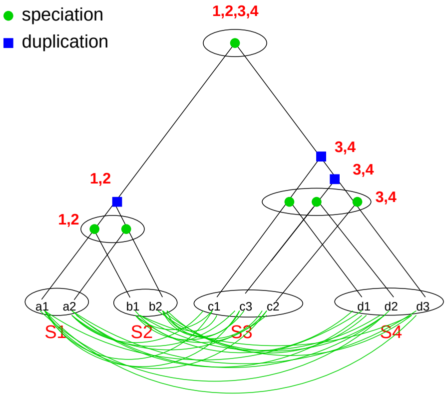
Properties of θ



Lemma

Given a reconciliation tree Π without losses. Let $x \in S_i$ and $y \in S_j$ such that $x\theta y$ and let $v = lca(x,y)$. Then it holds for all $w_1, w_2 \in L(\Pi)$ with $w_1 \leq v.1$ and $w_2 \leq v.2$ that $w_1 \theta w_2$.

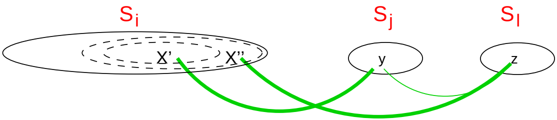
Properties of θ



Lemma

Given a reconciliation tree Π without losses. Let $x \in S_i$ and $y \in S_j$ such that $x\theta y$ and let $v = lca(x,y)$. Then it holds for all $w_1, w_2 \in L(\Pi)$ with $w_1 \leq v.1$ and $w_2 \leq v.2$ that $w_1 \theta w_2$.

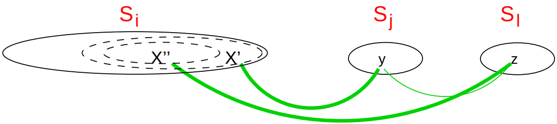
Properties of θ



Lemma

Given a reconciliation tree Π *without losses. Let* $y \in S_j$ *be an arbitrary, but fixed vertex of* $L(\Pi)$ *and and let* $z \in S_l$ *be a fixed vertex of* $L(\Pi)$ *with* $y\theta z$*. Let* $X' = \{x \in S_i \mid x\theta y\}$ *and* $X'' = \{x \in S_i \mid x\theta z\}$*. Then it holds* $X' \subseteq X''$ *or* $X'' \subseteq X'$*.*

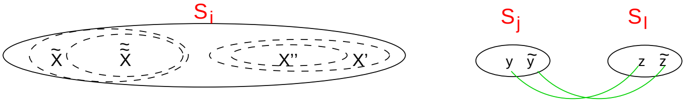
Properties of θ



Lemma

Given a reconciliation tree Π *without losses. Let* $y \in S_j$ *be an arbitrary, but fixed vertex of* $L(\Pi)$ *and and let* $z \in S_l$ *be a fixed vertex of* $L(\Pi)$ *with* $y\theta z$. *Let* $X' = \{x \in S_i \mid x\theta y\}$ *and* $X'' = \{x \in S_i \mid x\theta z\}$. *Then it holds* $X' \subseteq X''$ *or* $X'' \subseteq X'$.

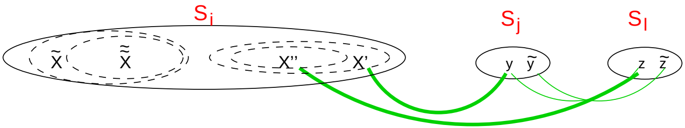
Properties of θ



Lemma

Given a reconciliation tree Π without losses. Let $y, \tilde{y} \in S_j$ be arbitrary, but fixed vertices of $L(\Pi)$ and and let $z, \tilde{z} \in S_l$ be fixed vertices of $L(\Pi)$ with $y\theta z$ and $\tilde{y}\theta\tilde{z}$.
Let $X' = \{x \in S_i \mid x\theta y\}$ and $X'' = \{x \in S_i \mid x\theta z\}$. Let $\widetilde{X} = \{x \in S_i \mid x\theta\tilde{y}\}$ and $\widetilde{\widetilde{X}} = \{x \in S_i \mid x\theta\tilde{z}\}$. If $X' \subseteq X''$ then $\widetilde{X} \subseteq \widetilde{\widetilde{X}}$ and if $X'' \subseteq X'$ then $\widetilde{\widetilde{X}} \subseteq \widetilde{X}$.

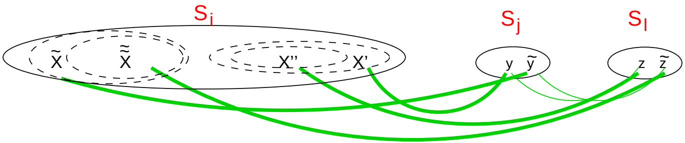
Properties of θ



Lemma

Given a reconciliation tree Π without losses. Let $y, \tilde{y} \in S_j$ be arbitrary, but fixed vertices of $L(\Pi)$ and and let $z, \tilde{z} \in S_l$ be fixed vertices of $L(\Pi)$ with $y \theta z$ and $\tilde{y} \theta \tilde{z}$.
Let $X' = \{x \in S_i \mid x \theta y\}$ and $X'' = \{x \in S_i \mid x \theta z\}$. Let $\widetilde{X} = \{x \in S_i \mid x \theta \tilde{y}\}$ and $\widetilde{\widetilde{X}} = \{x \in S_i \mid x \theta \tilde{z}\}$. If $X' \subseteq X''$ then $\widetilde{X} \subseteq \widetilde{\widetilde{X}}$ and if $X'' \subseteq X'$ then $\widetilde{\widetilde{X}} \subseteq \widetilde{X}$.

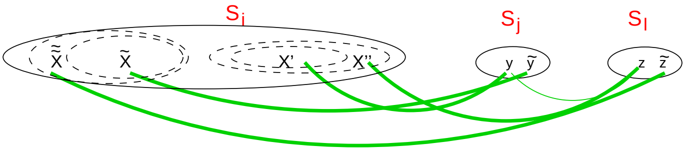
Properties of θ



Lemma

Given a reconciliation tree Π without losses. Let $y, \tilde{y} \in S_j$ be arbitrary, but fixed vertices of $L(\Pi)$ and and let $z, \tilde{z} \in S_l$ be fixed vertices of $L(\Pi)$ with $y \theta z$ and $\tilde{y} \theta \tilde{z}$.
Let $X' = \{x \in S_i \mid x \theta y\}$ and $X'' = \{x \in S_i \mid x \theta z\}$. Let $\widetilde{X} = \{x \in S_i \mid x \theta \tilde{y}\}$ and $\widetilde{\widetilde{X}} = \{x \in S_i \mid x \theta \tilde{z}\}$. If $X' \subseteq X''$ then $\widetilde{X} \subseteq \widetilde{\widetilde{X}}$ and if $X'' \subseteq X'$ then $\widetilde{\widetilde{X}} \subseteq \widetilde{X}$.

Properties of θ



Lemma

Given a reconciliation tree Π *without losses. Let* $y, \tilde{y} \in S_j$ *be arbitrary, but fixed vertices of* $L(\Pi)$ *and and let* $z, \tilde{z} \in S_l$ *be fixed vertices of* $L(\Pi)$ *with* $y \theta z$ *and* $\tilde{y} \theta \tilde{z}$.
Let $X' = \{x \in S_i \mid x \theta y\}$ *and* $X'' = \{x \in S_i \mid x \theta z\}$. *Let* $\widetilde{X} = \{x \in S_i \mid x \theta \tilde{y}\}$ *and* $\widetilde{\widetilde{X}} = \{x \in S_i \mid x \theta \tilde{z}\}$. *If* $X' \subseteq X''$ *then* $\widetilde{X} \subseteq \widetilde{\widetilde{X}}$ *and if* $X'' \subseteq X'$ *then* $\widetilde{\widetilde{X}} \subseteq \widetilde{X}$.

What we know:

$\Pi \implies \theta$ with particular properties.

Open Questions:

Given a binary relation R.
Does there exist a reconciliation tree Π such that $R = \theta$?
Does there exist a reconciliation tree Π without losses such that $R = \theta$?

Future Plans:

1. Given evolutionary distance matrix D then we have to compute the binary relation R in a good way.
2. For given R find θ such that $d(R, \theta) \rightarrow \min$
3. Use binary relation θ to construct Π.

Thank you for your attention.