Do horizontal gene transfer relations possess simple charateristics (like cographs)?

John Anders

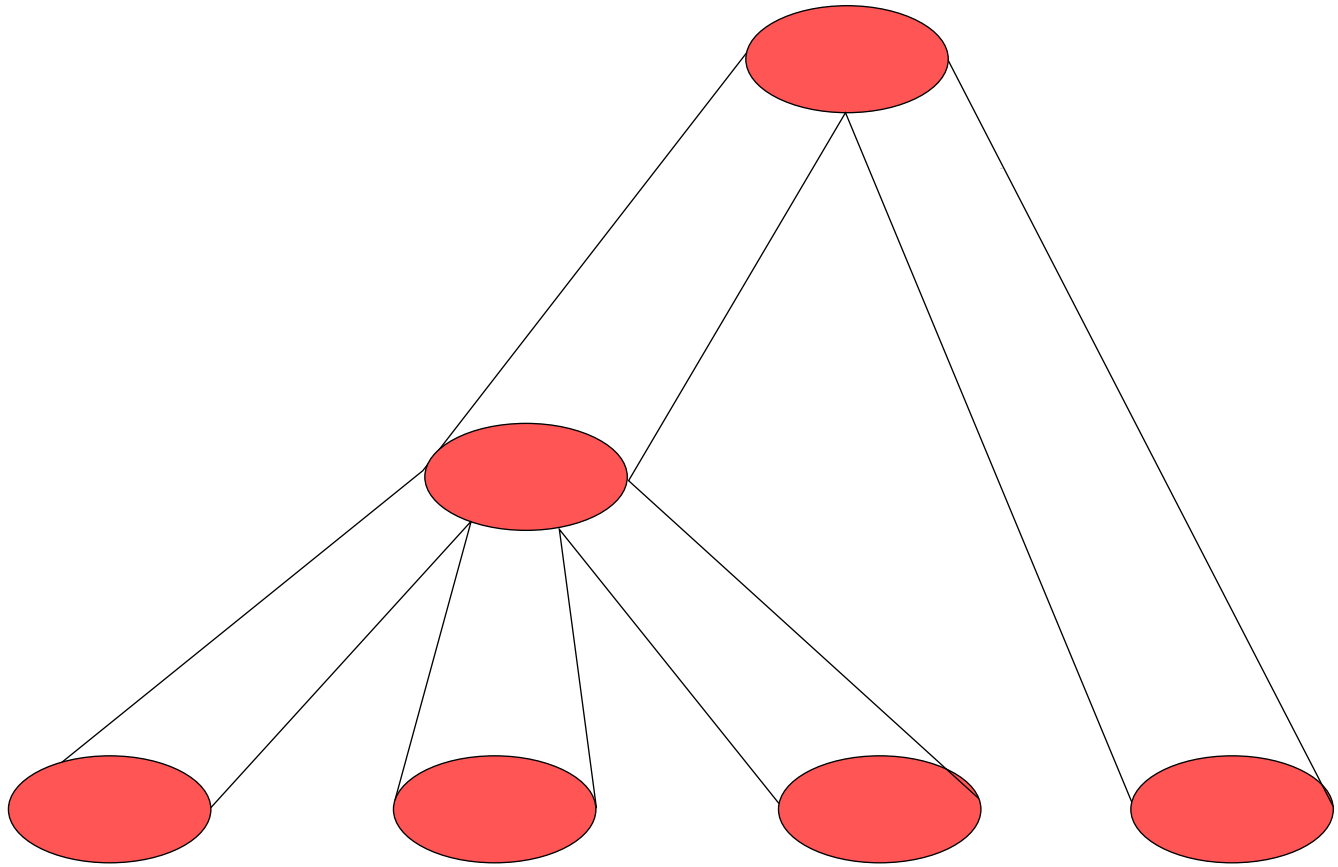
Department of Bioinformatics
University of Leipzig
johnanders@posteo.de

32nd TBI Winterseminar in Bled, 2017

why gene trees?

why gene trees?

- evolution!



Caracal
Caracal Caracal

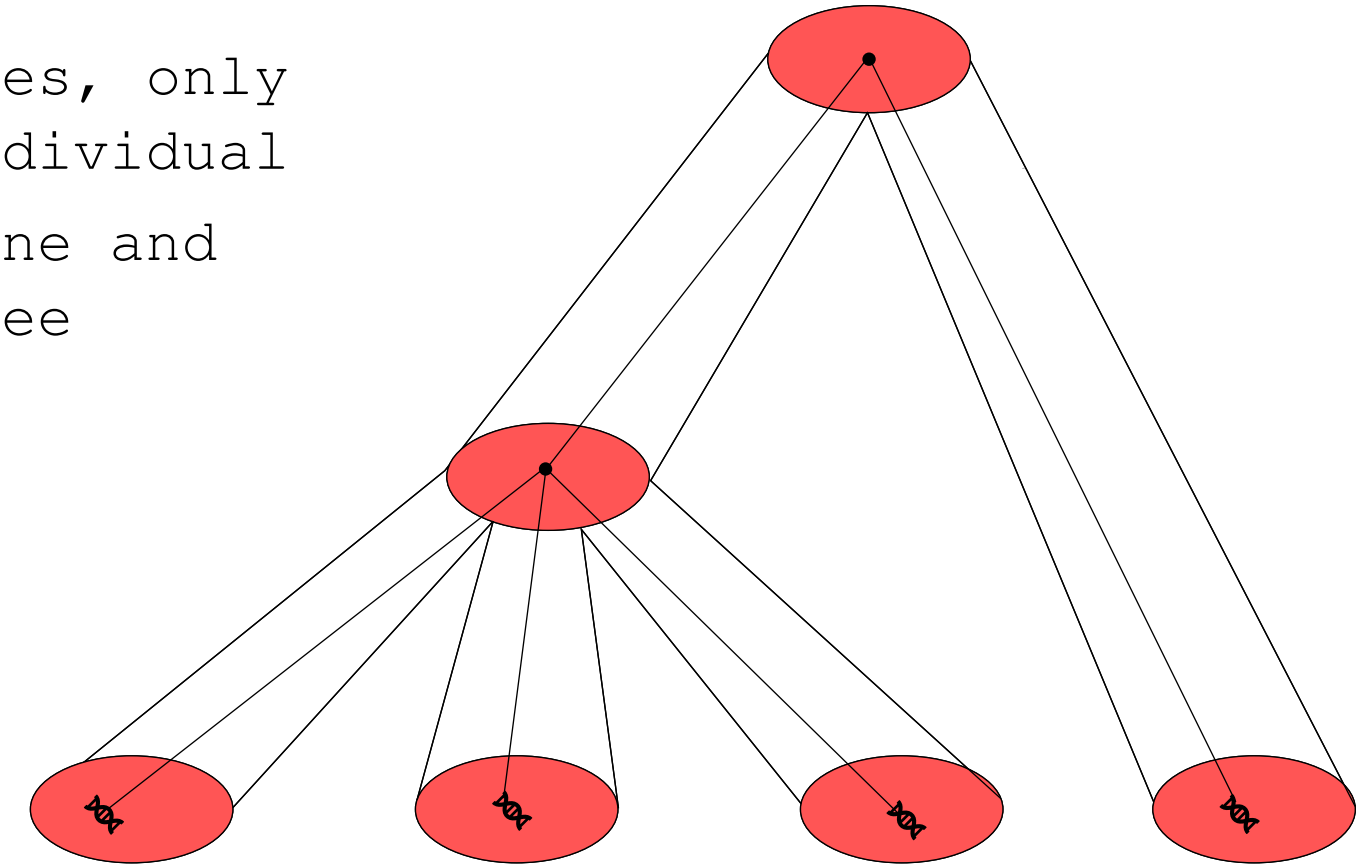
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

why gene trees?

- evolution!
- 16SRNA genes, only one per individual
- 1:1 map gene and species tree



Caracal
Caracal Caracal

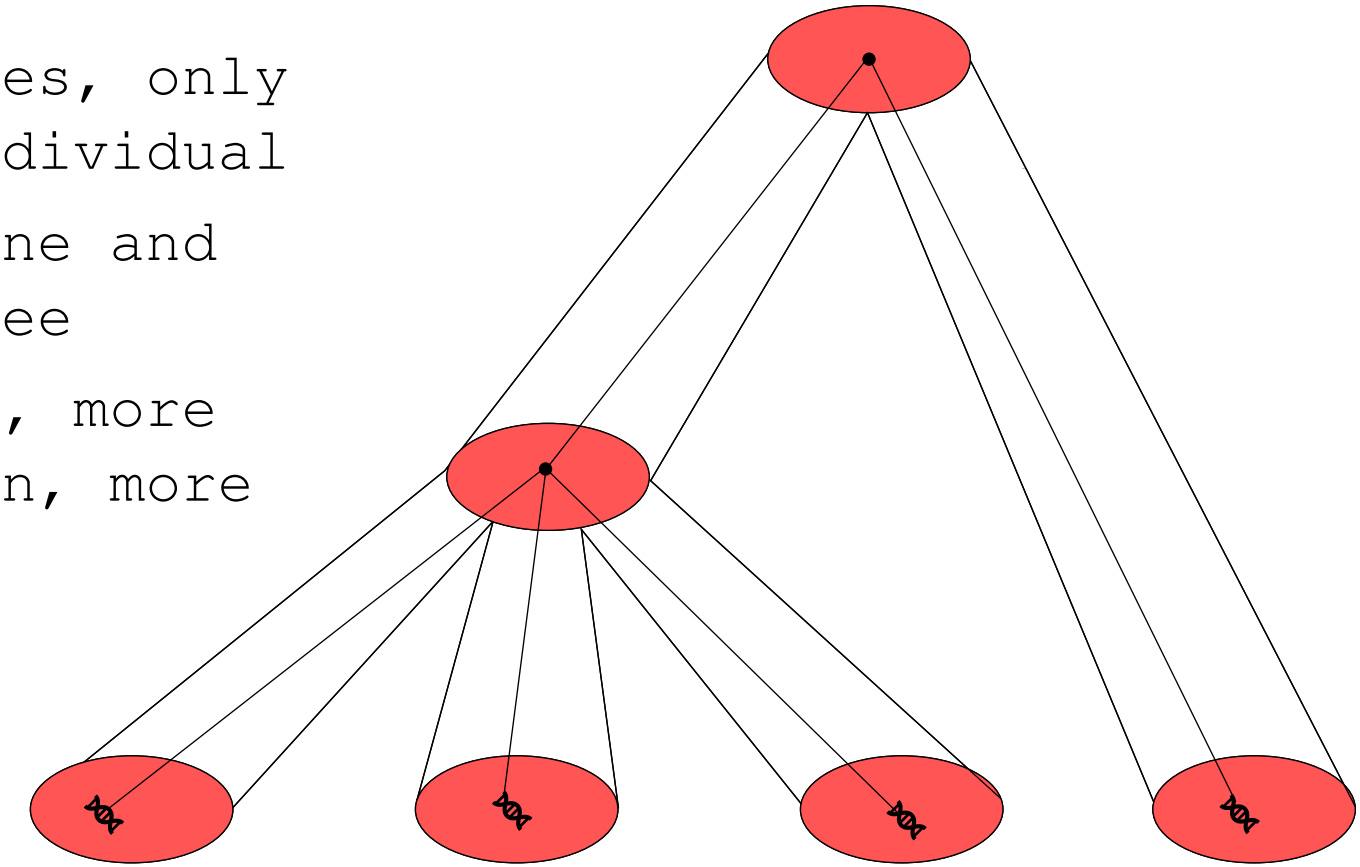
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

why gene trees?

- evolution!
- 16SRNA genes, only one per individual
- 1:1 map gene and species tree
- more genes, more information, more hazzle...



Caracal
Caracal Caracal

Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

paralogs and orthologs

- paralog ⇒ duplication in a specie
- ortholog ⇒ speciation

Caracal
Caracal Caracal

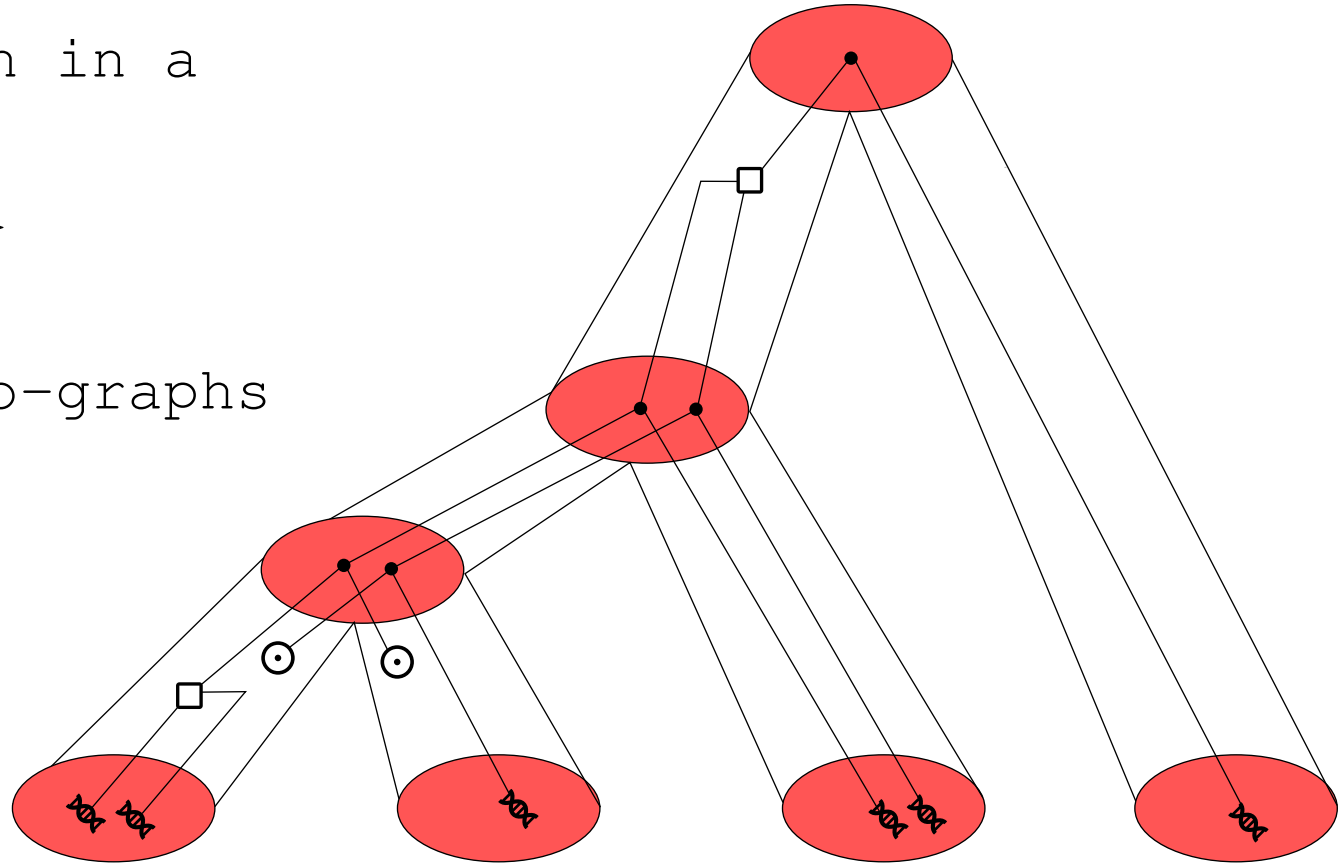
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

paralogs and orthologs

- paralog ⇒ duplication in a specie
- ortholog ⇒ speciation
- co-tree/ co-graphs



Caracal
Caracal Caracal

Clouded Leopard
Neofelis nebulosa

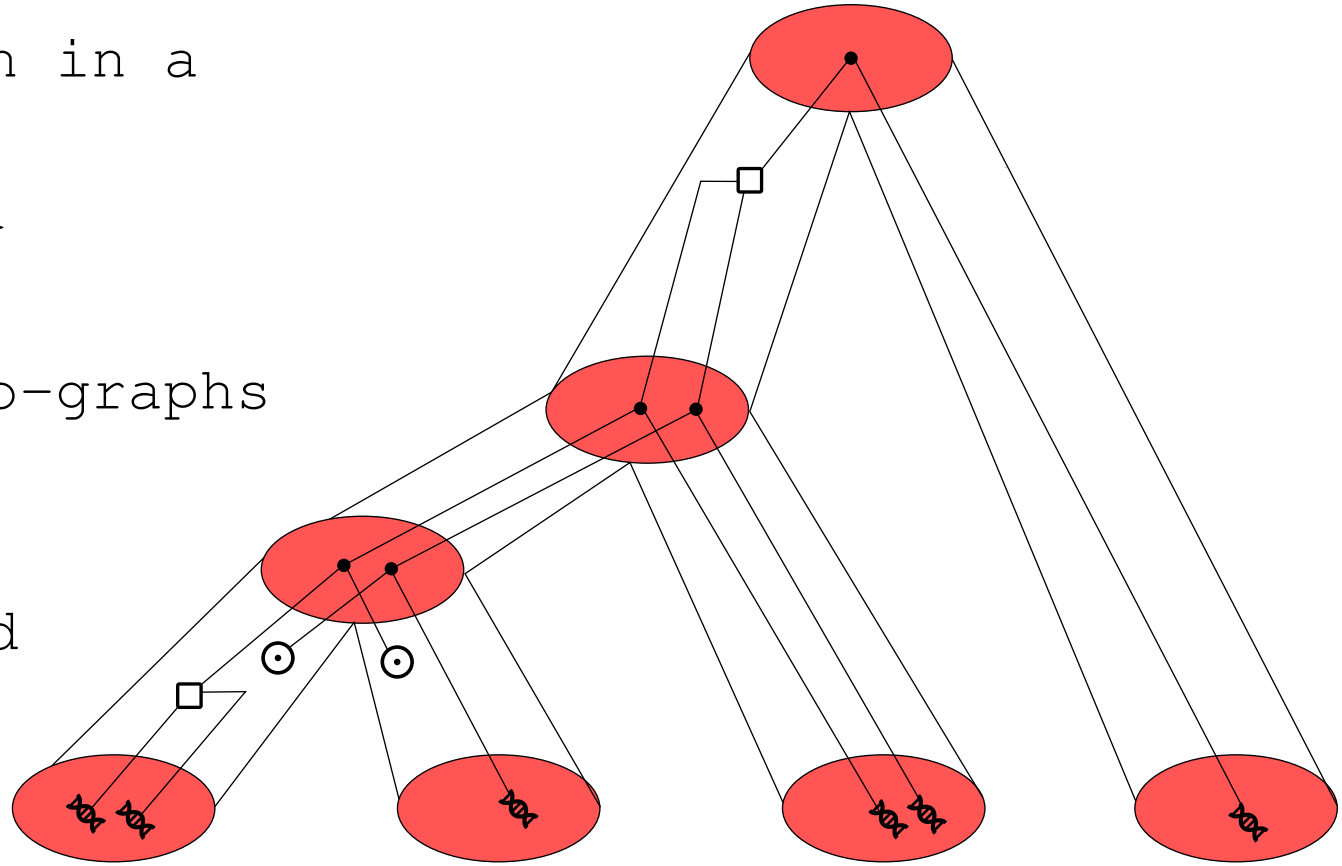
Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

paralogs and orthologs

- paralog ⇒ duplication in a specie
- ortholog ⇒ speciation
- co-tree/ co-graphs

- even more complicated



Caracal
Caracal Caracal

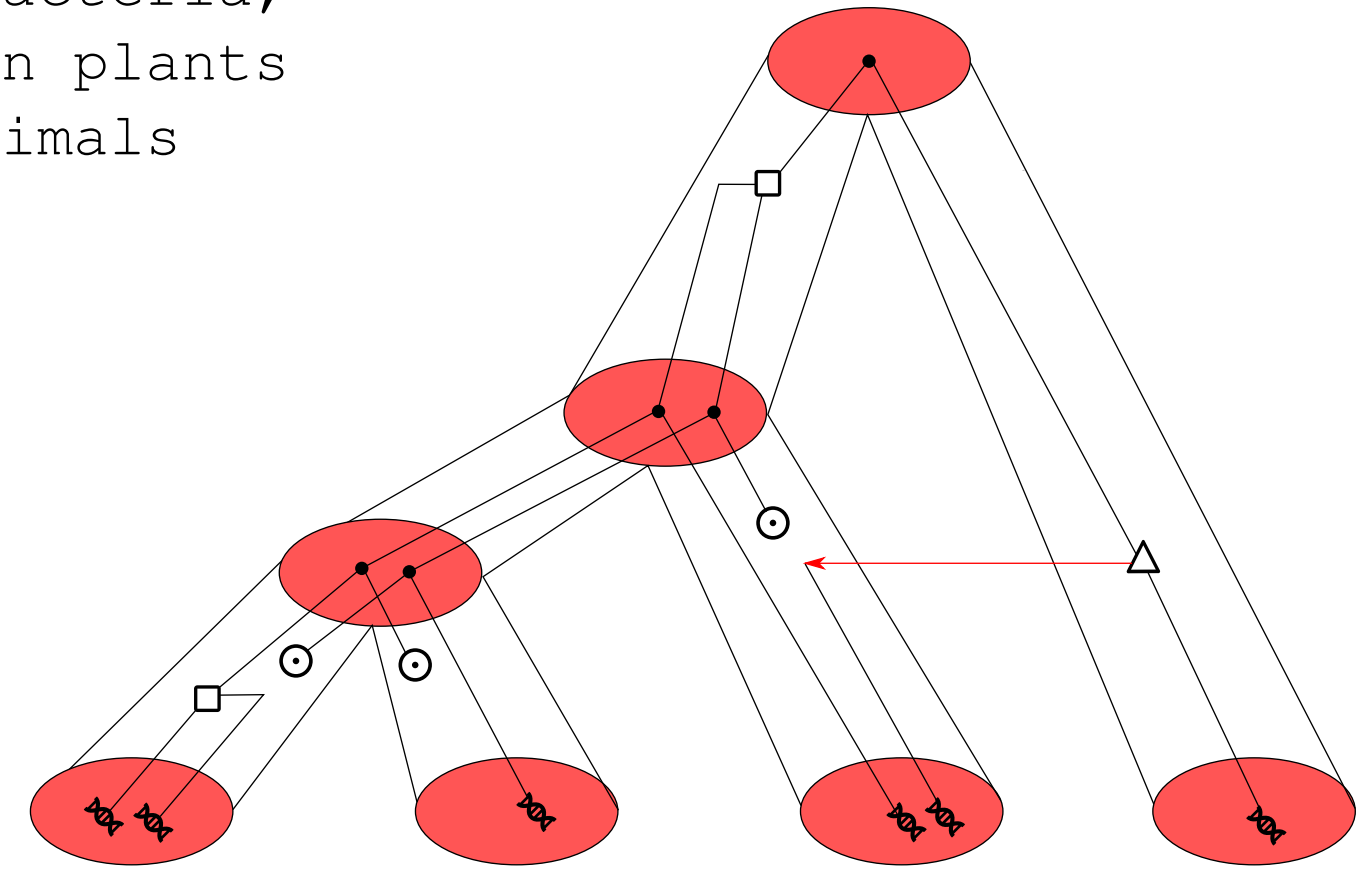
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

horizontal gene transfer (hgt)

- often in bacteria,
  but also in plants
  even in animals



Caracal
Caracal Caracal

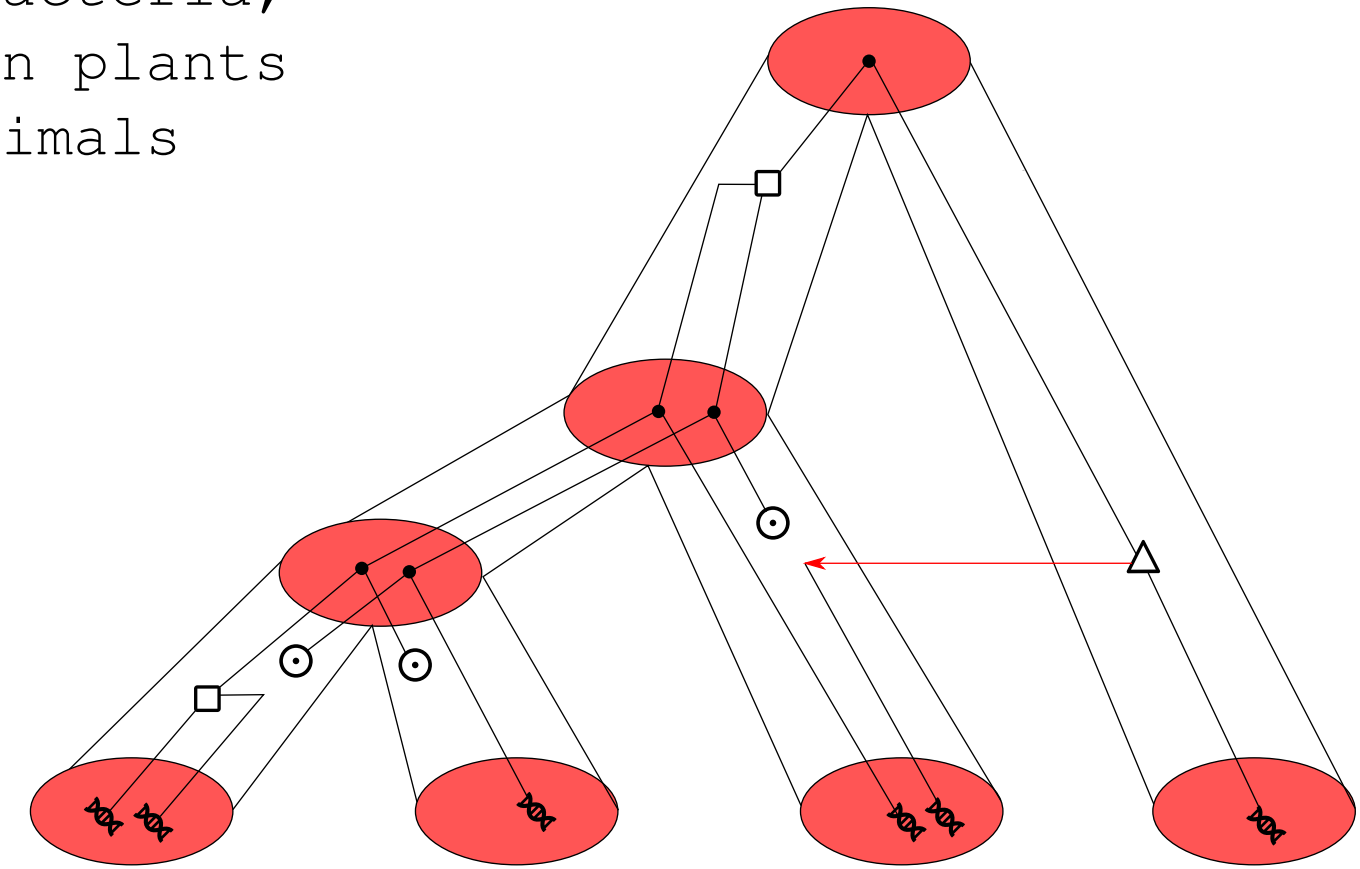
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

horizontal gene transfer (hgt)

- often in bacteria,
  but also in plants
  even in animals
- hgt has a
  direction



Caracal
Caracal Caracal

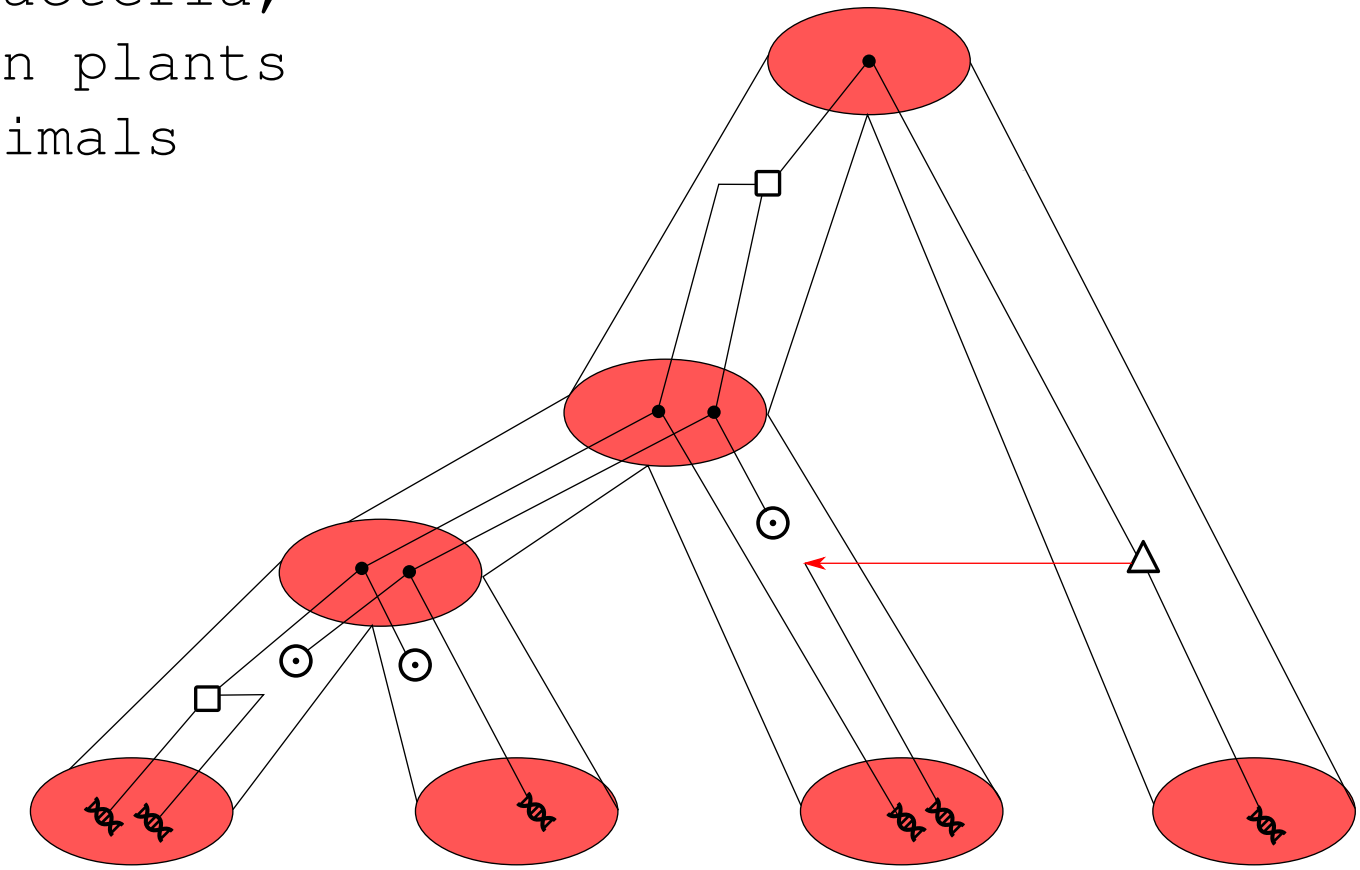
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

horizontal gene transfer (hgt)

- often in bacteria, but also in plants even in animals
- hgt has a direction



---------- observable surface ----------

Caracal
Caracal Caracal

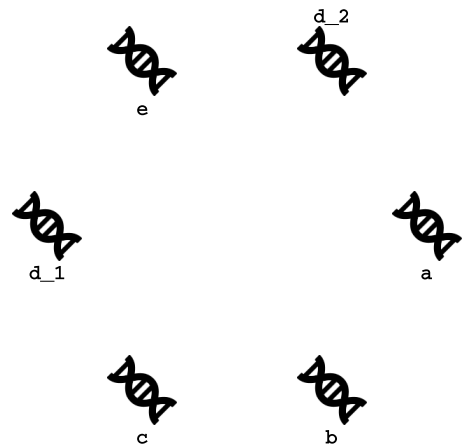
Clouded Leopard
Neofelis nebulosa

Snow Leopard
Panthera uncia

Pallas Cat
Otocolobus manul

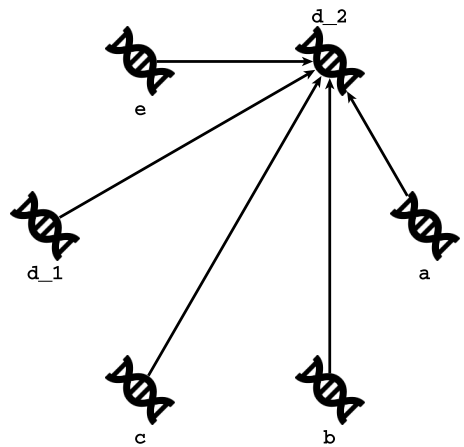
the whole problem

- compare genes (e.g. distance, characteristic)



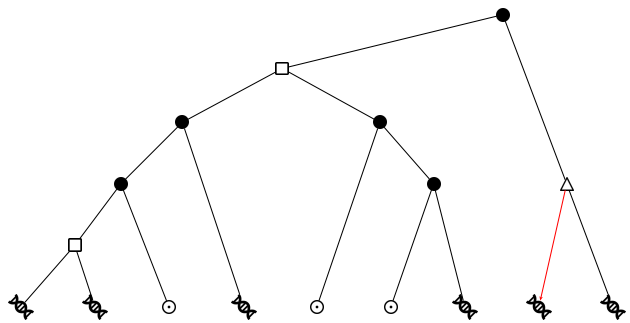
the whole problem

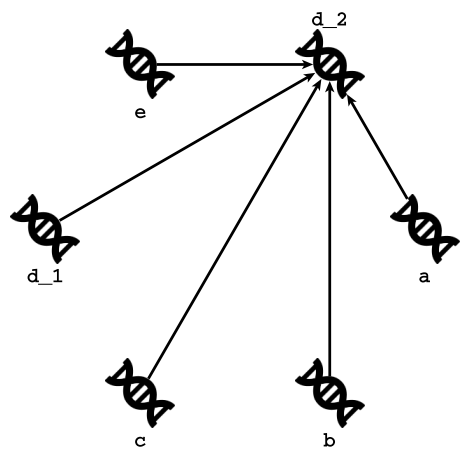
- compare genes (e.g. distance, characteristic)



the whole problem

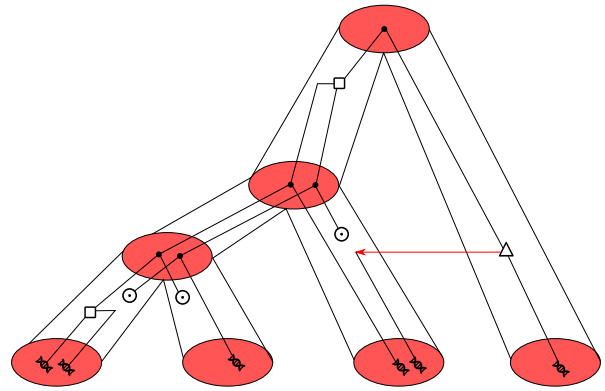
- compare genes (e.g. distance, characteristic)
- infere gene tree by relatation of genes

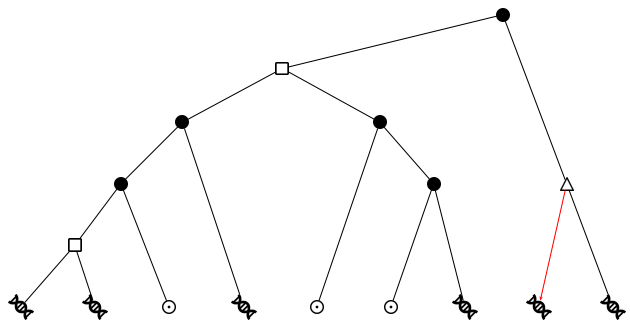


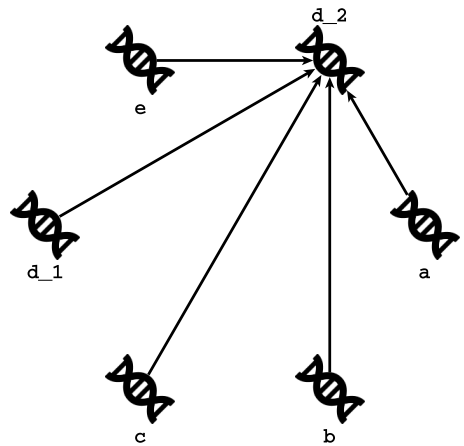


the whole problem

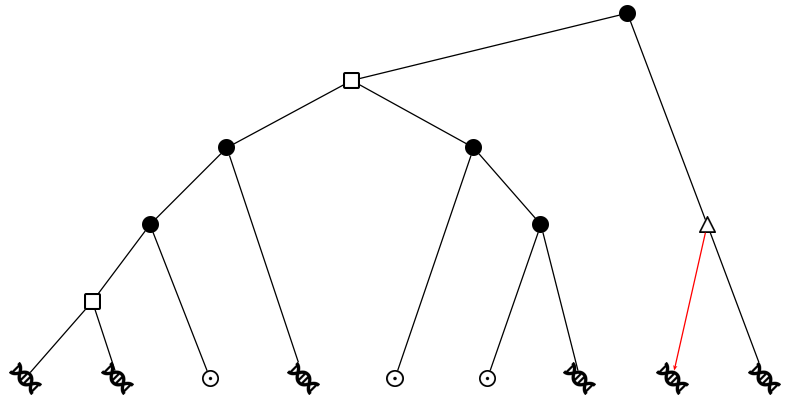
- compare genes (e.g. distance, characteristic)
- infere gene tree by relatation of genes
- find a map and a species tree in which gene tree can be embeded



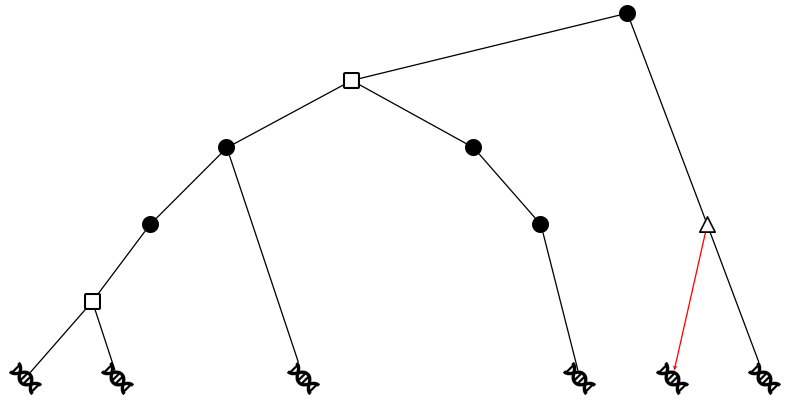




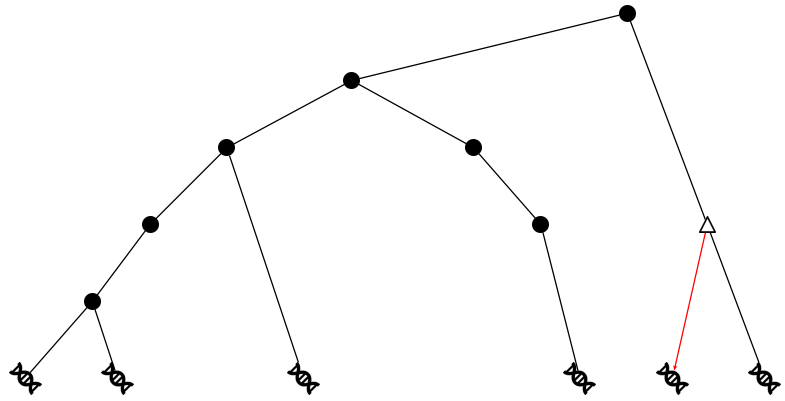
new construction



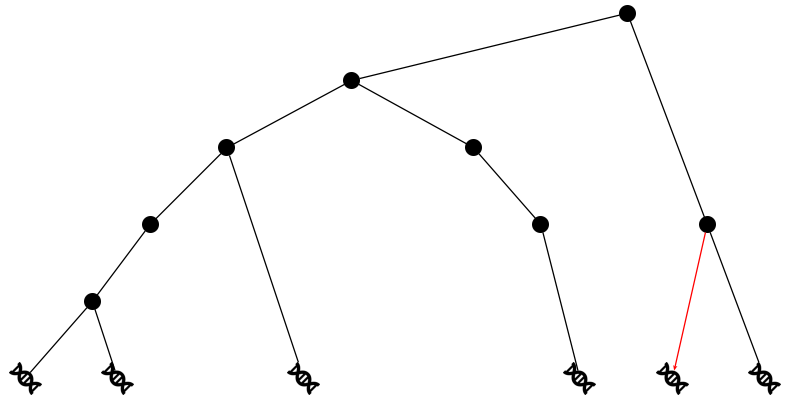
new construction



new construction



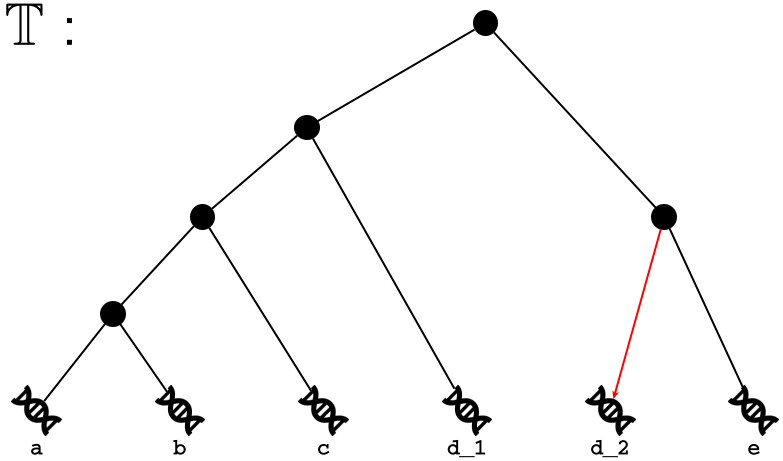
new construction



new construction

some basic notation

$\mathbb{T}$:

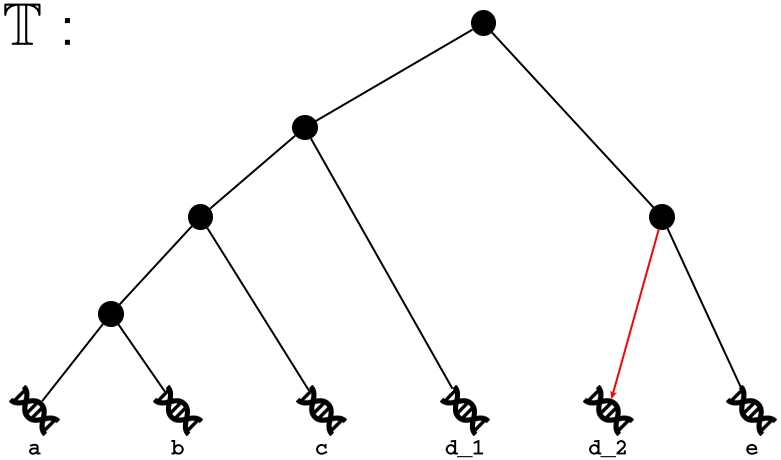


new construction

some basic notation

- $\mathbb{T} = (V_t, E_t)$
  - undirected, acylic, rooted
  - lca(a,b) a,b is in the set of leafes
  - lab(e) e is an edge
    - red $\rightarrow$ hgt
    - black $\rightarrow$ para-, ortholog

$\mathbb{T}$:



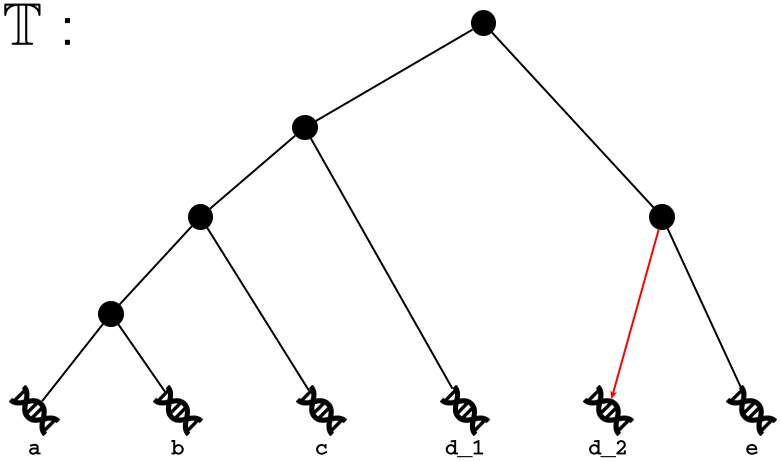
- $L(\mathbb{T})$ = Genes

new construction

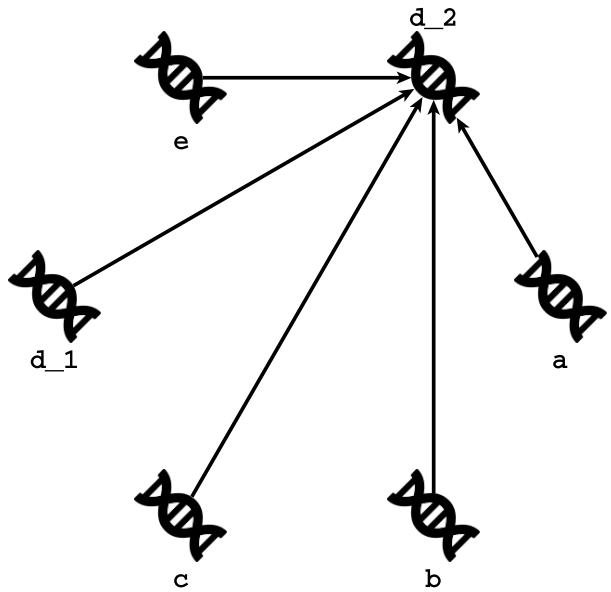
some basic notation

- $\mathbb{T} = (V_t, E_t)$
  - undirected, acylic, rooted
  - lca(a,b) a,b is in the set of leafes
  - lab(e) e is an edge
    - red $\rightarrow$ hgt
    - black $\rightarrow$ para-, ortholog
- $\mathbb{R} = (V_r, E_r)$
  - directed
- $L(\mathbb{T})$ = Genes $= V(\mathbb{R})$

$\mathbb{T}$:



$\mathbb{R}$:

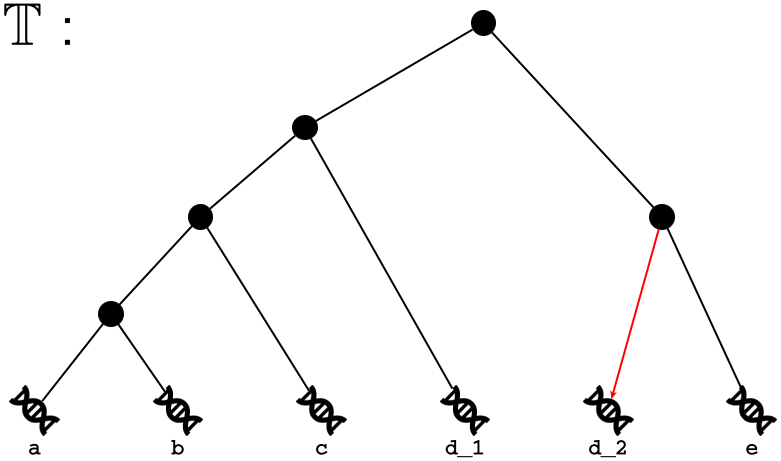


new construction

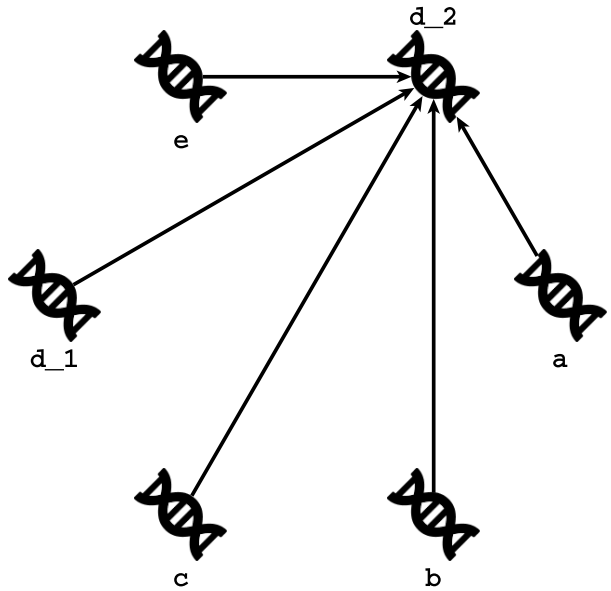
some basic notation

- $\mathbb{T} = (V_t, E_t)$
  - undirected, acylic, rooted
  - lca(a,b) a,b is in the set of leafes
  - lab(e) e is an edge
    - red $\rightarrow$ hgt
    - black $\rightarrow$ para-, ortholog
- $\mathbb{R} = (V_r, E_r)$
  - directed
- $L(\mathbb{T})$ = Genes $= V(\mathbb{R})$
- $(a, b) \in E(\mathbb{R})$
  $\Leftrightarrow$
  there is a red labelled edge in path(lca(a,b),b)

$\mathbb{T}$:



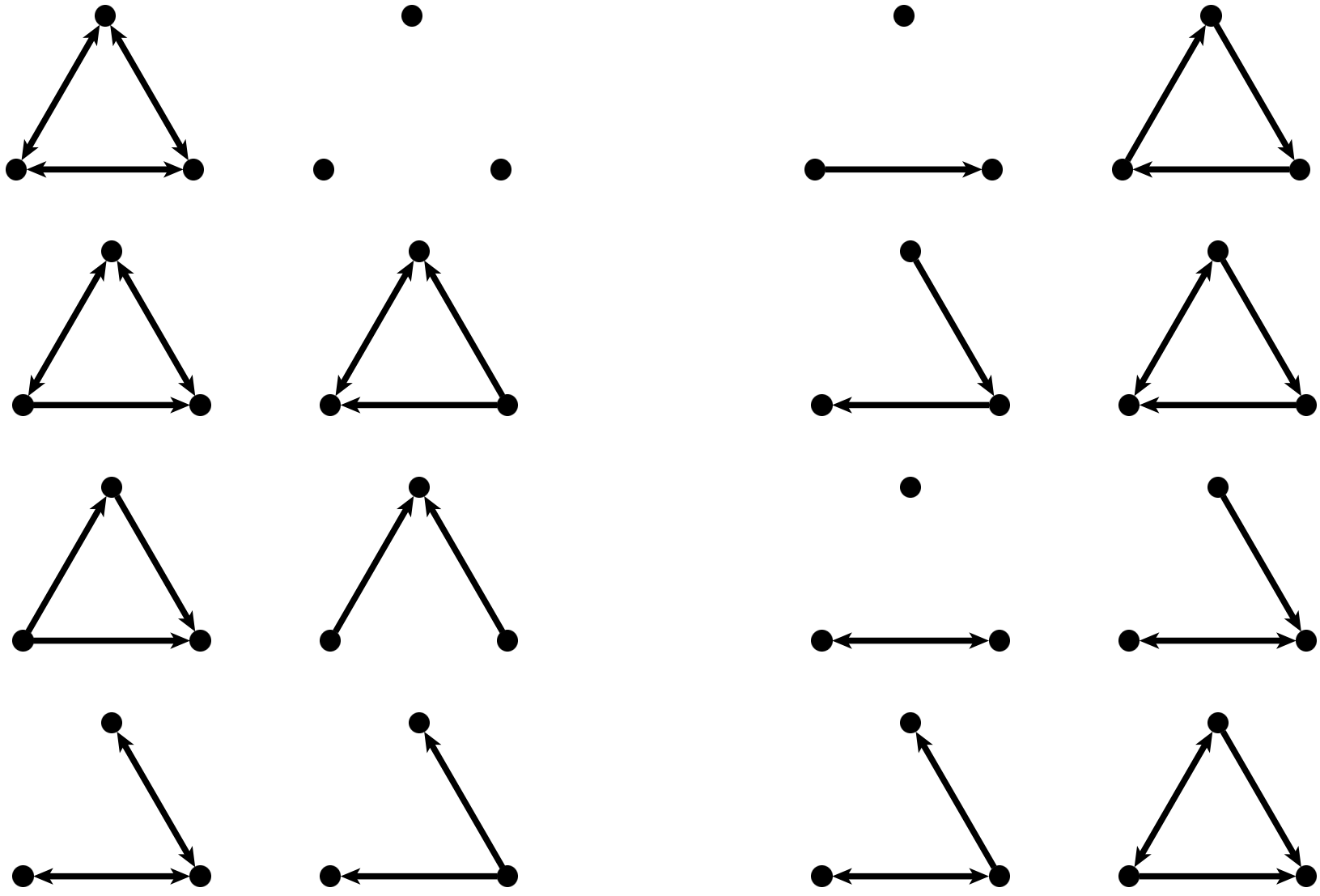
$\mathbb{R}$:



what do we want to know?

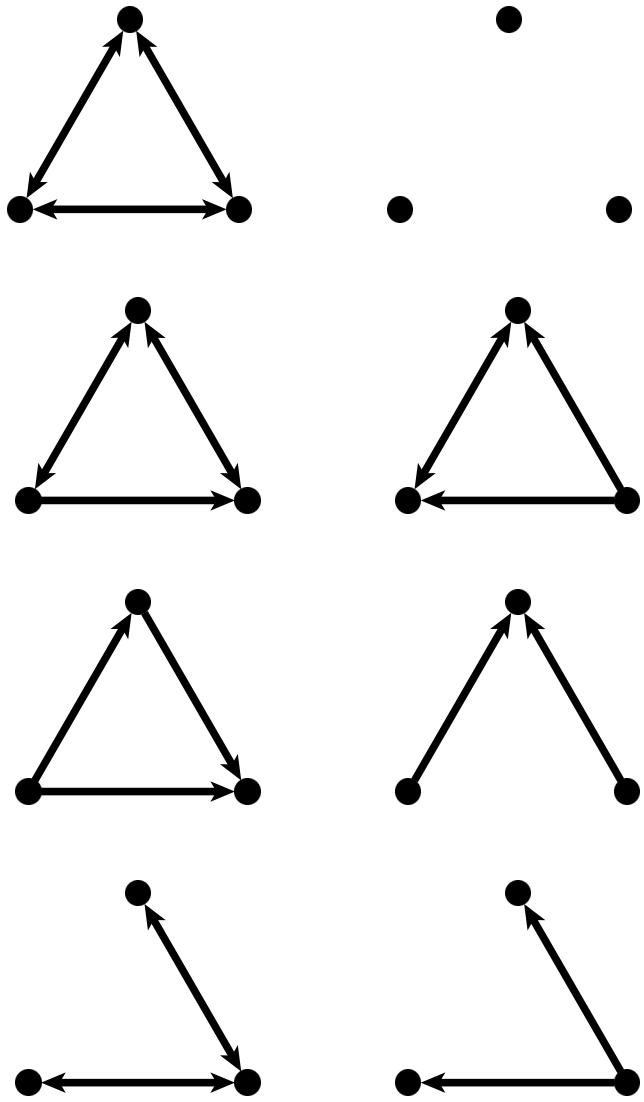
- how do we identify allowed $\mathbb{R}$?
  - there is at least one $\mathbb{T}$ that displays $\mathbb{R}$
- which information does a $\mathbb{R}$ hold about $\mathbb{T}$?
  - which $\mathbb{T}$ are displaying $\mathbb{R}$?

identify simple(n=3), allowed $\mathbb{R}$

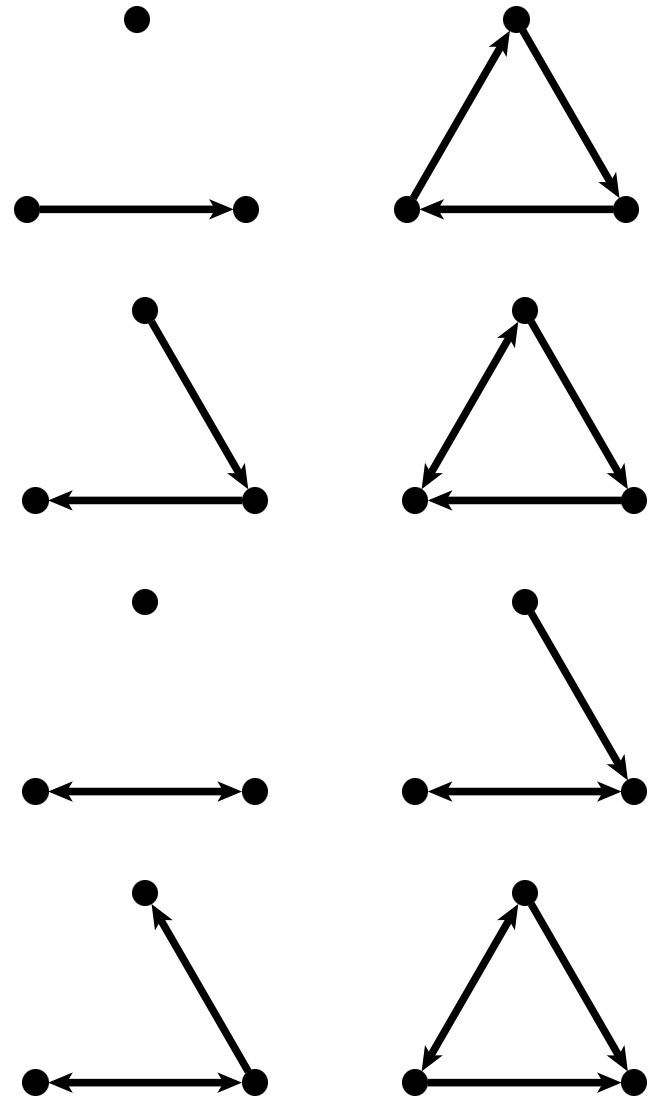


identify simple(n=3), allowed $\mathbb{R}$

allowed Triangles:

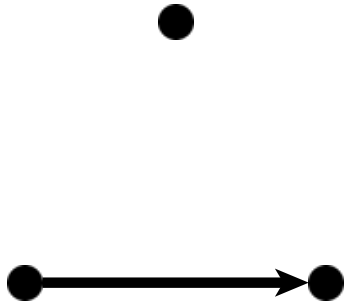


forbidden Triangles:



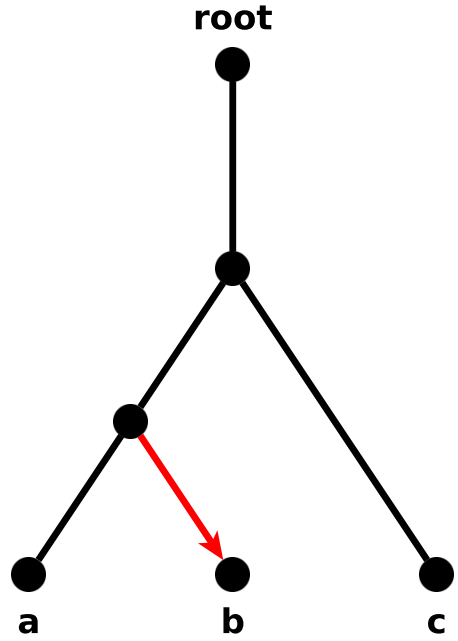
one example

$\mathbb{R}$:

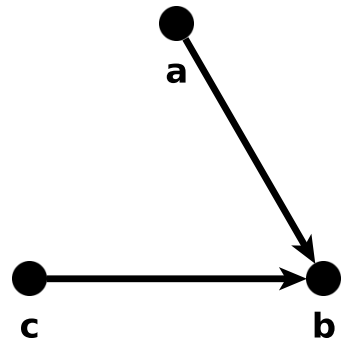


(forbidden)

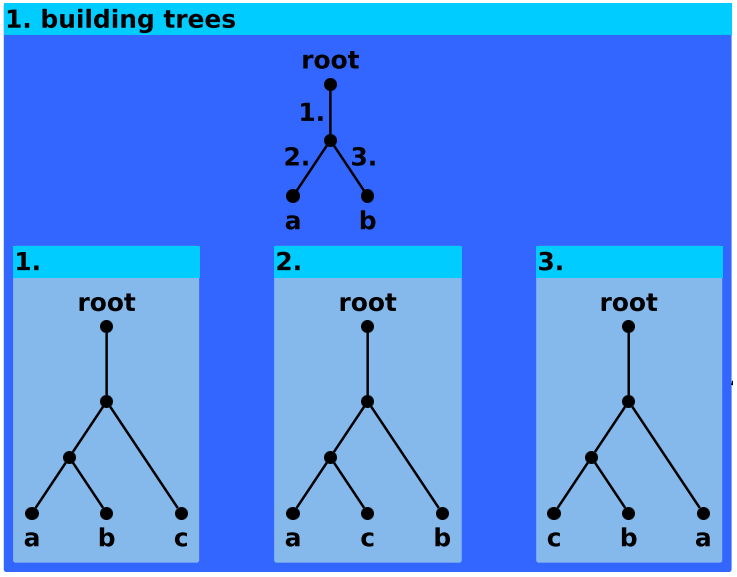
$\mathbb{T}$:



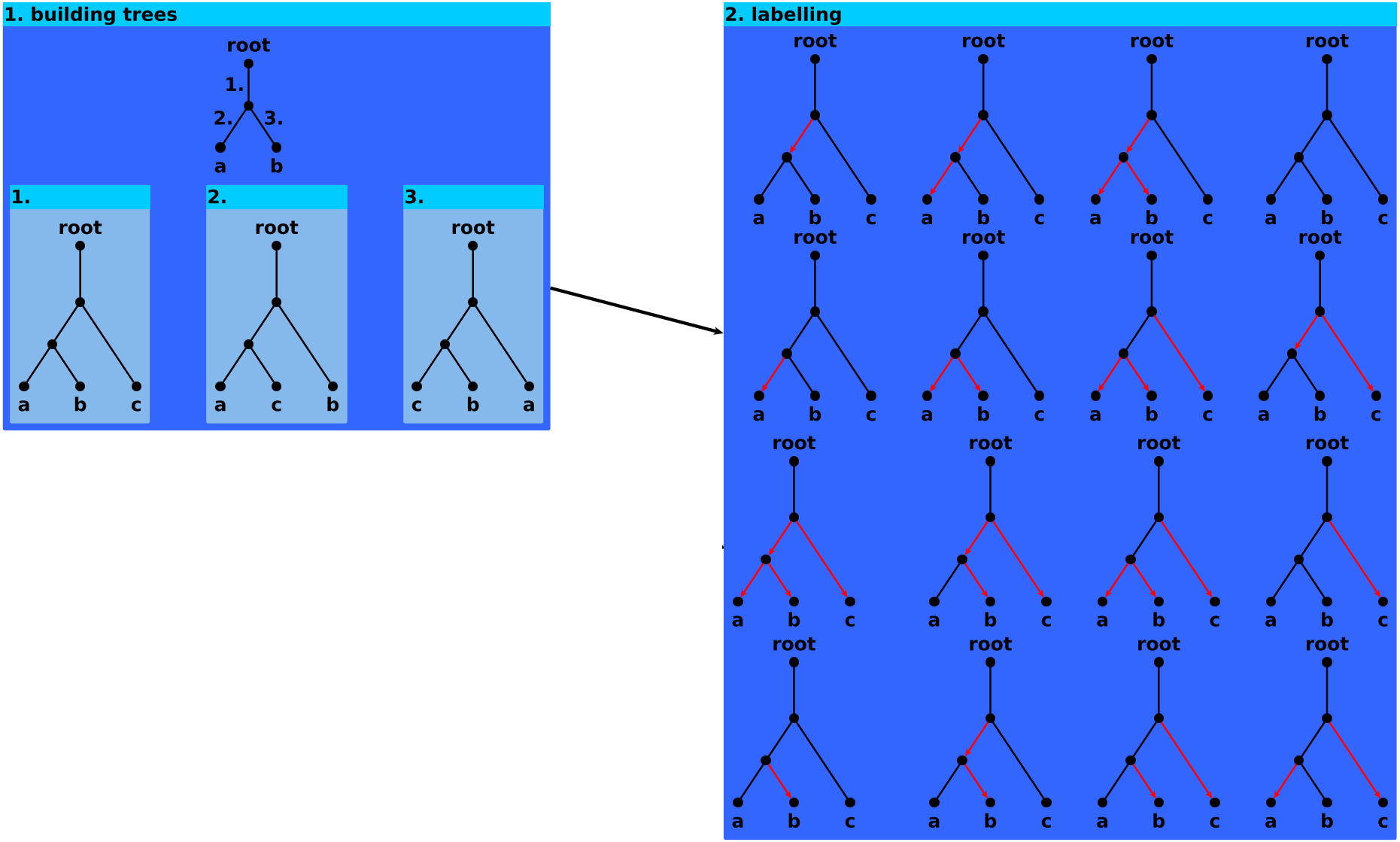
$\mathbb{R}$:



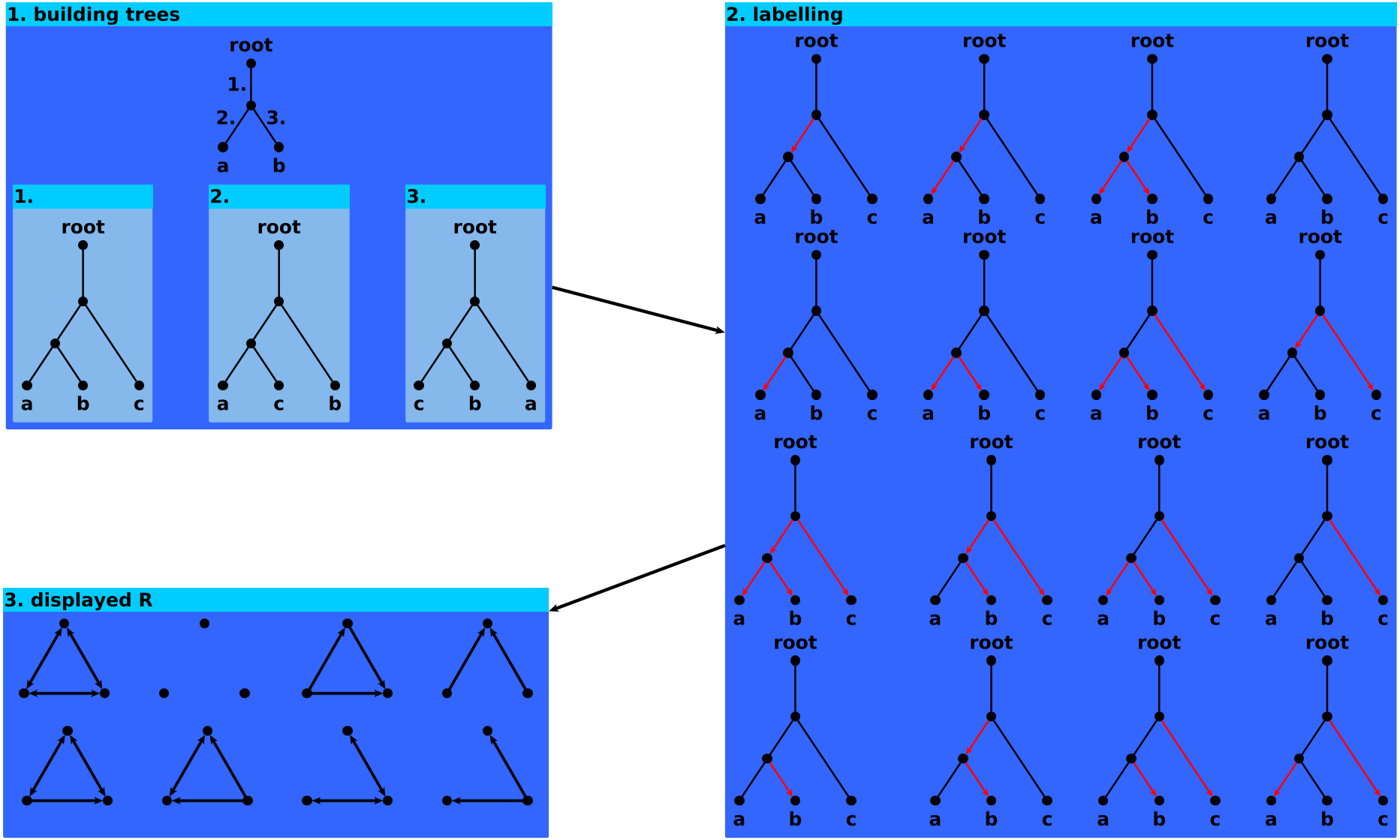
identify allowed, complex(n>3) $\mathbb{R}$



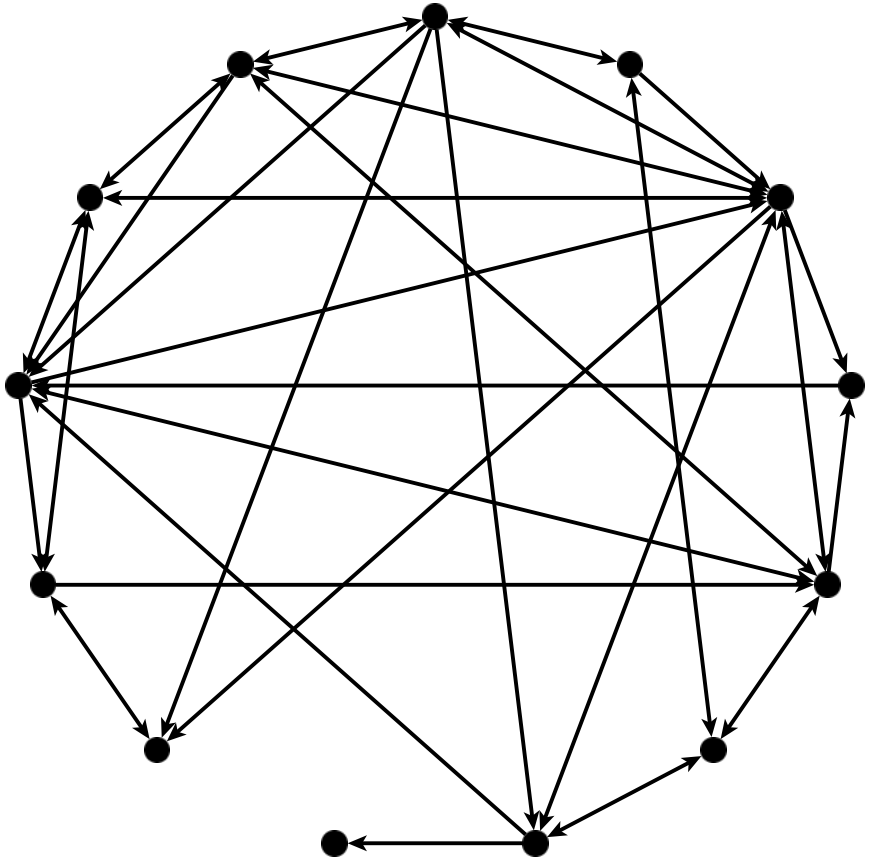
identify allowed, complex(n>3) $\mathbb{R}$

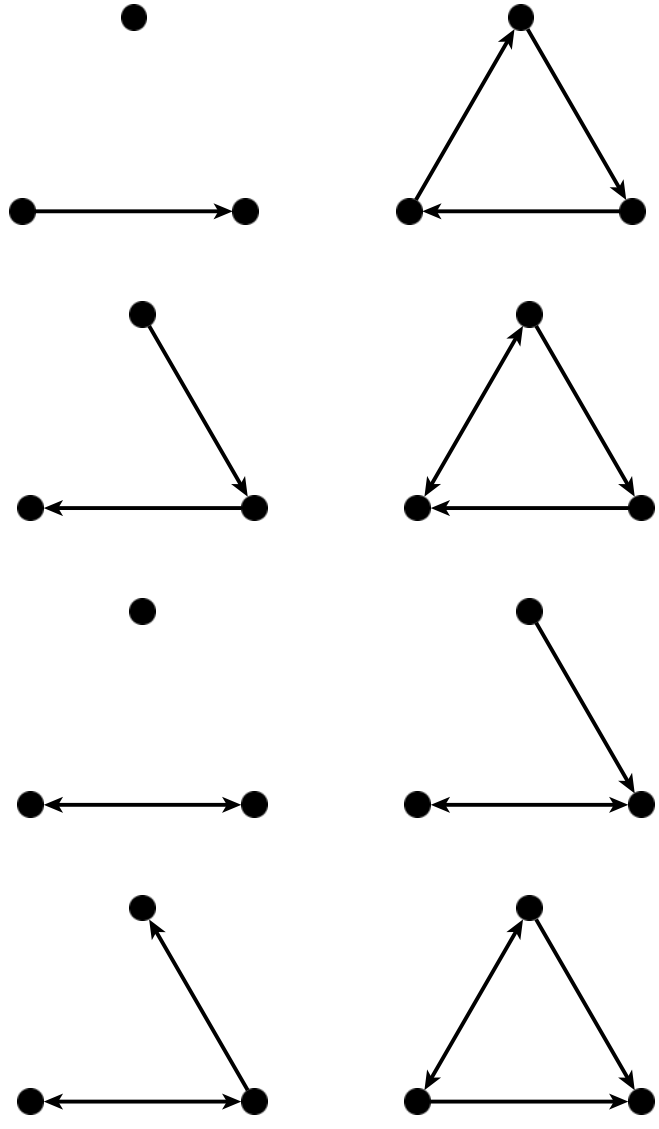


identify allowed, complex(n>3) $\mathbb{R}$

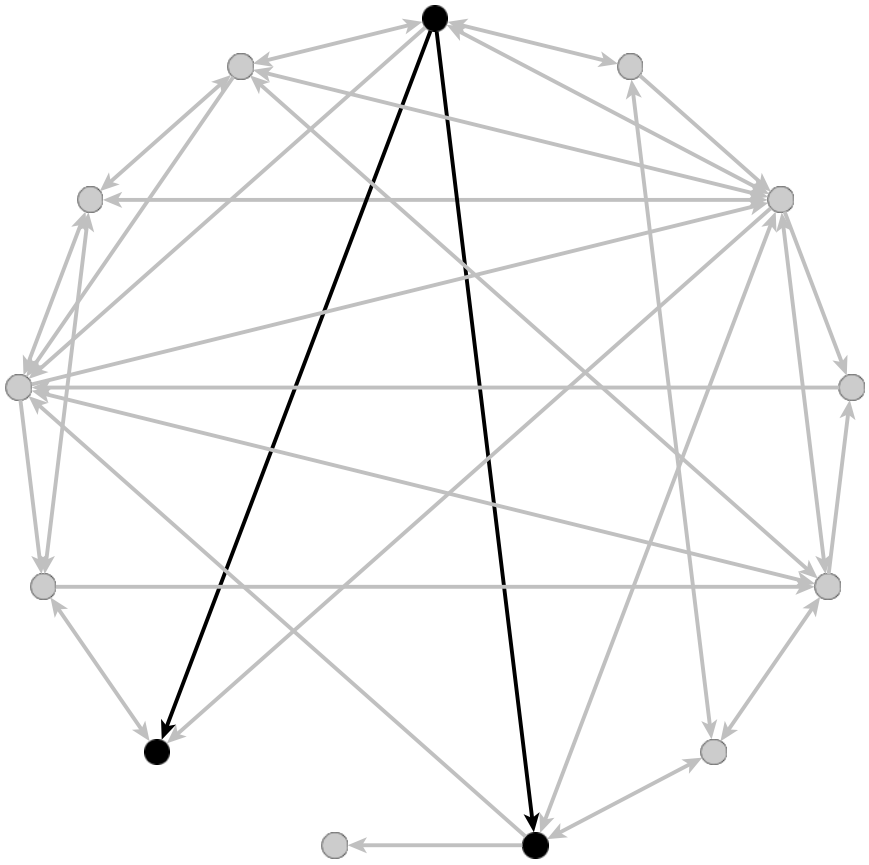


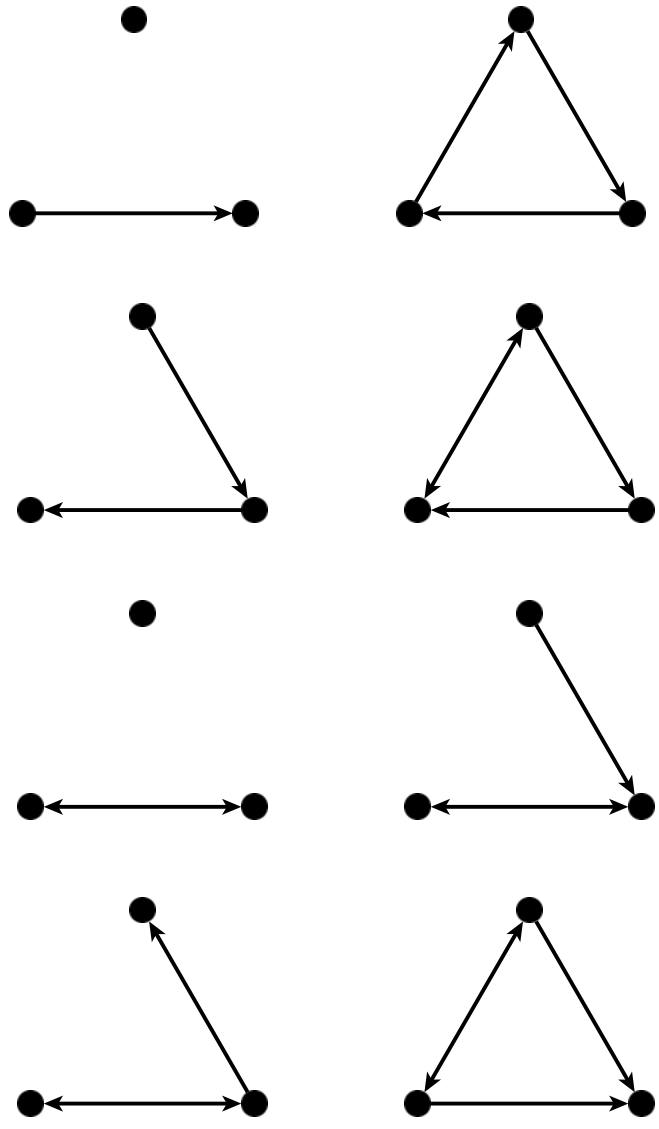
identify forbidden, complex(n>3) $\mathbb{R}$



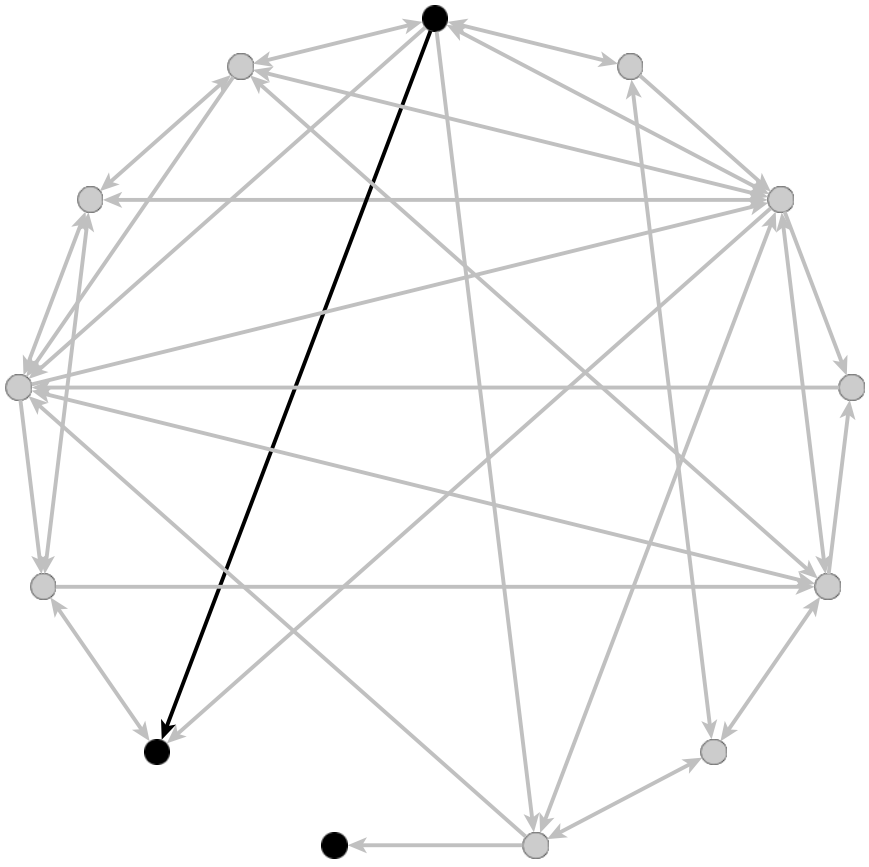


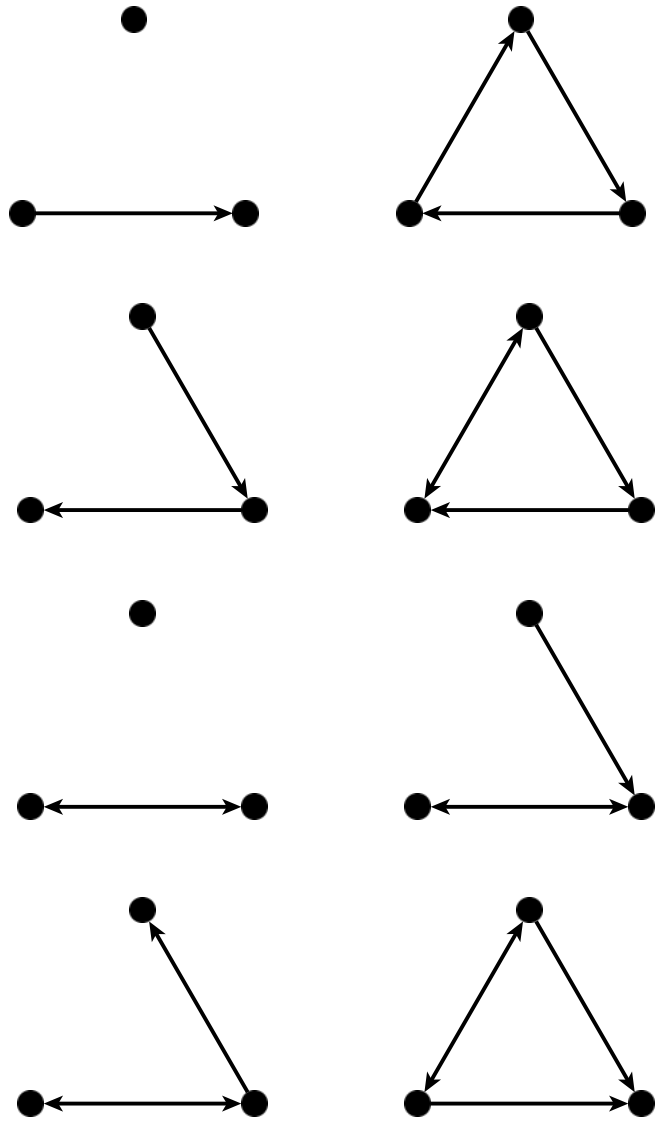
identify forbidden, complex(n>3) $\mathbb{R}$





identify forbidden, complex(n>3) $\mathbb{R}$





a conjecture

Every $\mathbb{R}$ that is not displayed by any $\mathbb{T}$ contains at least one forbidden Triplet.

end

thank you!
(special thanks to Peter, Marc and Felix)