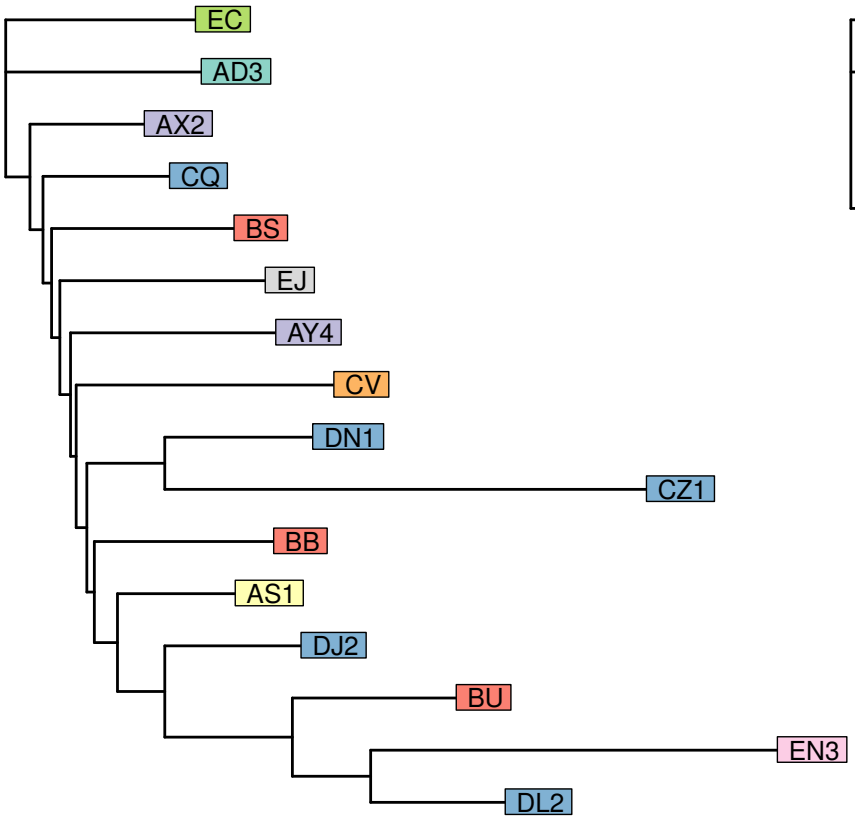
Distances between phylogenetic time trees

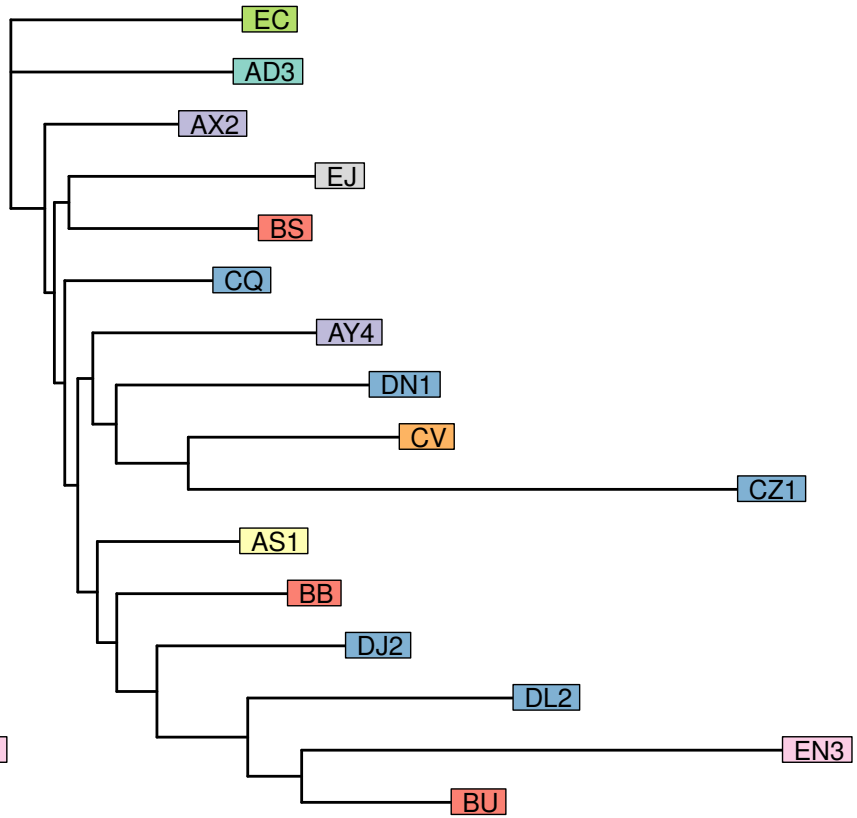
Lena Collienne

Biological Data Science Lab
Department of Computer Science
University of Otago

26/11/2021

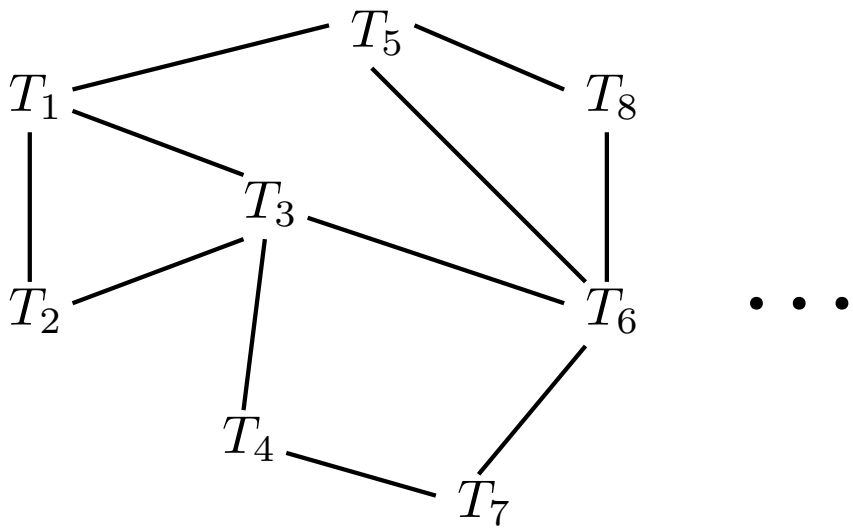
Time trees



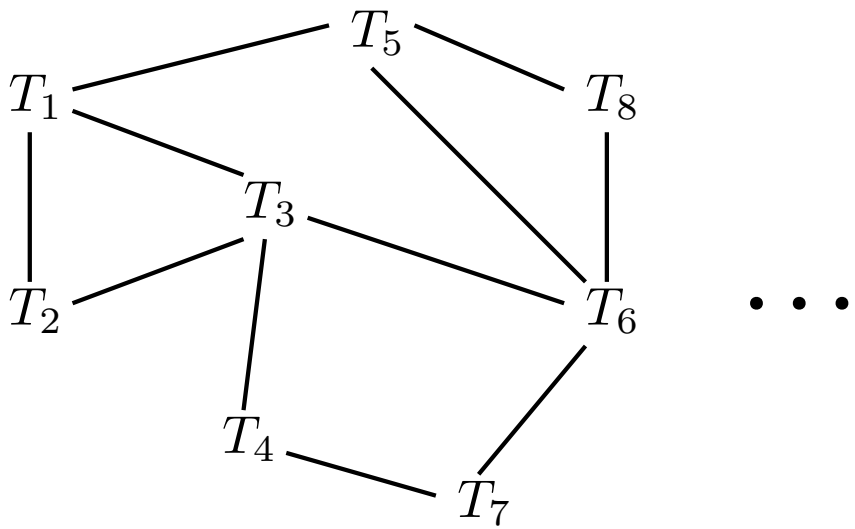


Breast
Thyroid
Lung (main tumour)
Pancreas
Vertebra C2
Kidney
Lung lesion
Eye
Brain

Discrete Tree Space

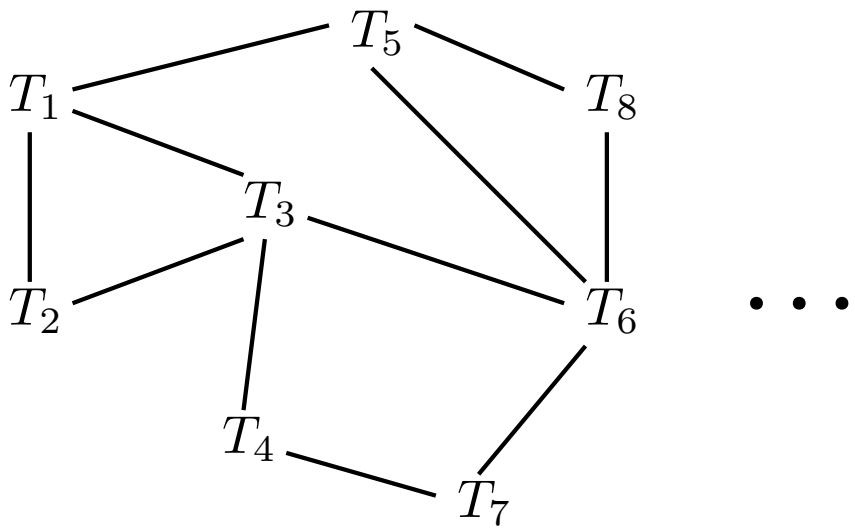


Discrete Tree Space



Popular tree re-arrangement operations: NNI, SPR, TBR

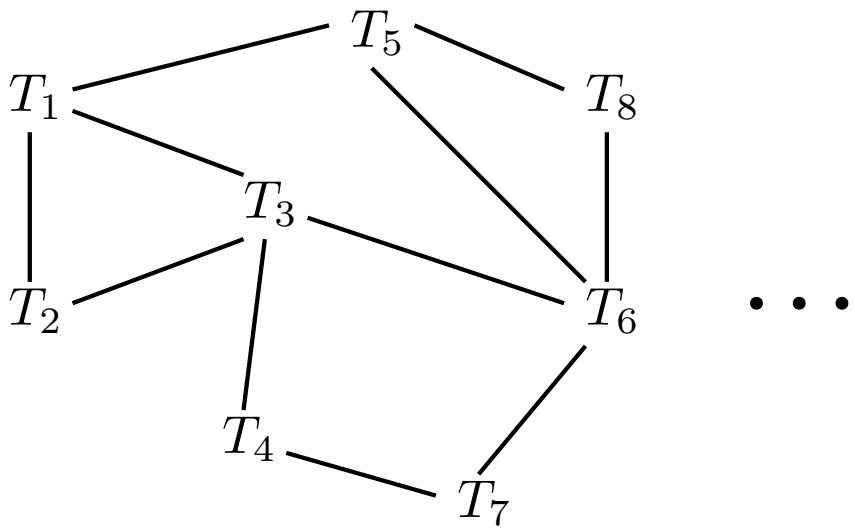
Discrete Tree Space



Popular tree re-arrangement operations: NNI, SPR, TBR

- Similarity measure

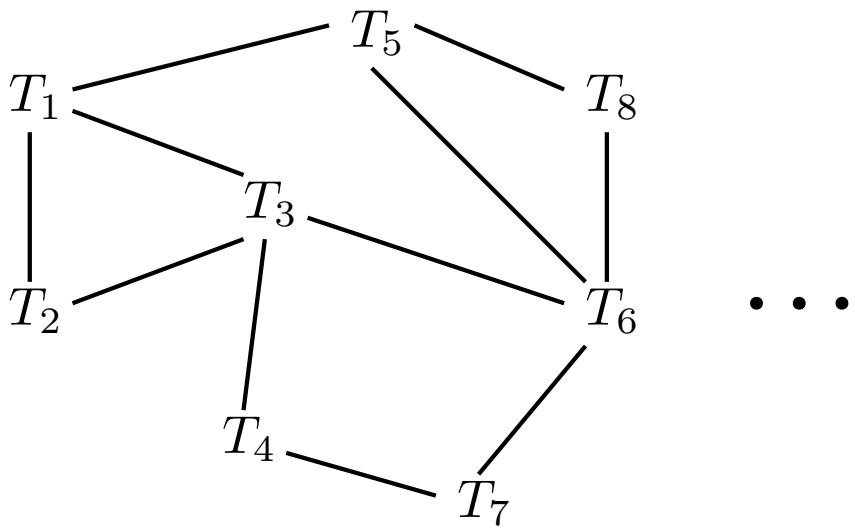
Discrete Tree Space



Popular tree re-arrangement operations: NNI, SPR, TBR

- Similarity measure
- Tree search algorithms

Discrete Tree Space



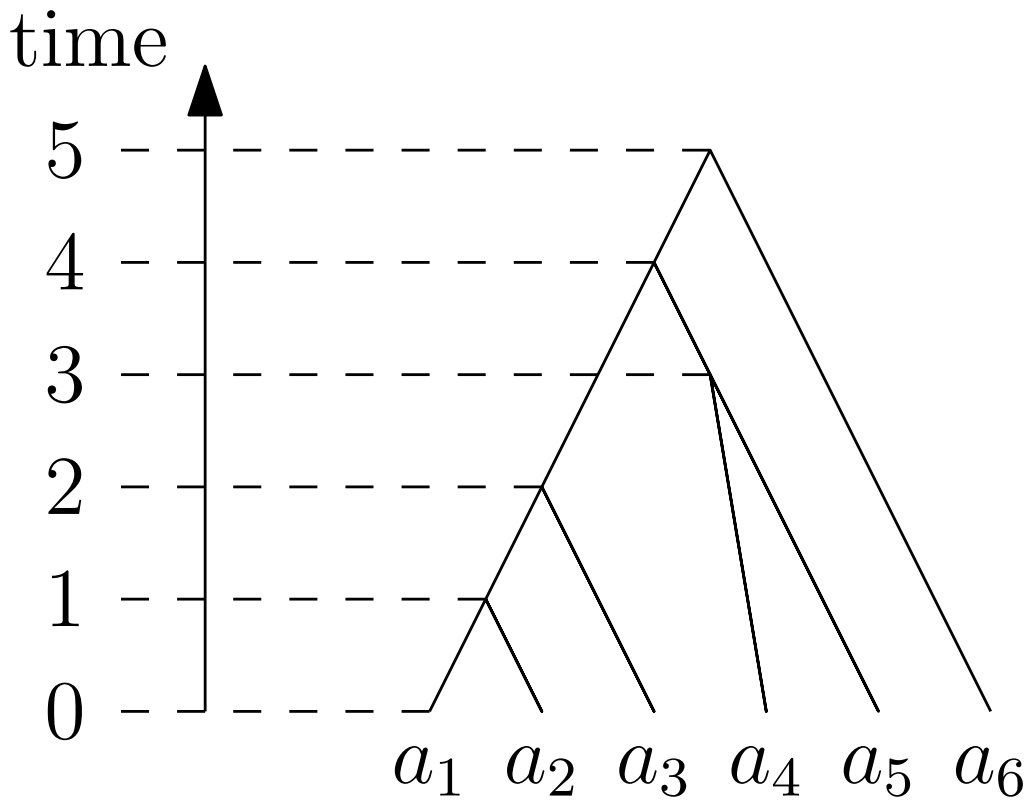
Popular tree re-arrangement operations: NNI, SPR, TBR

- Similarity measure
- Tree search algorithms

Problem: Computing distances is $\mathcal{NP}$-hard

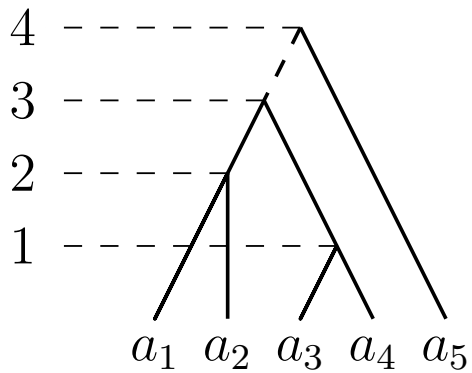
Discretising Time Trees

Ranked trees



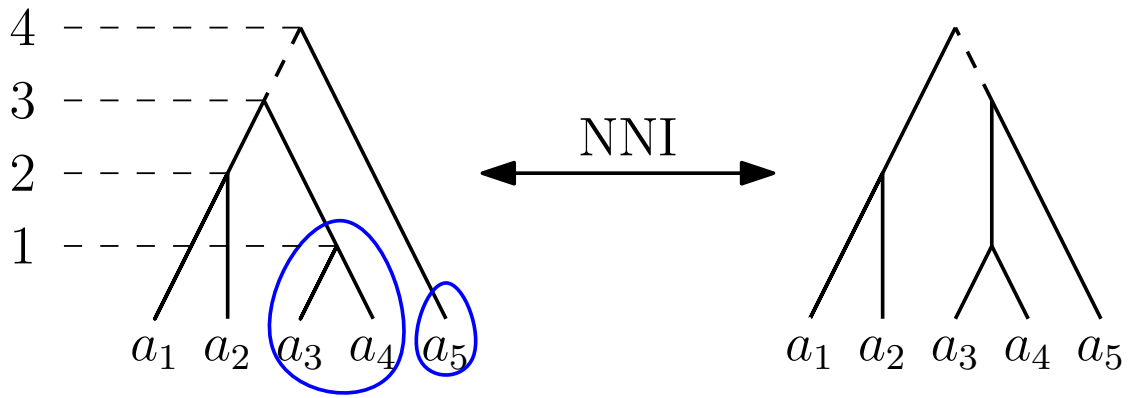
RNNI – Ranked Nearest Neighbour Interchange

NNI Move



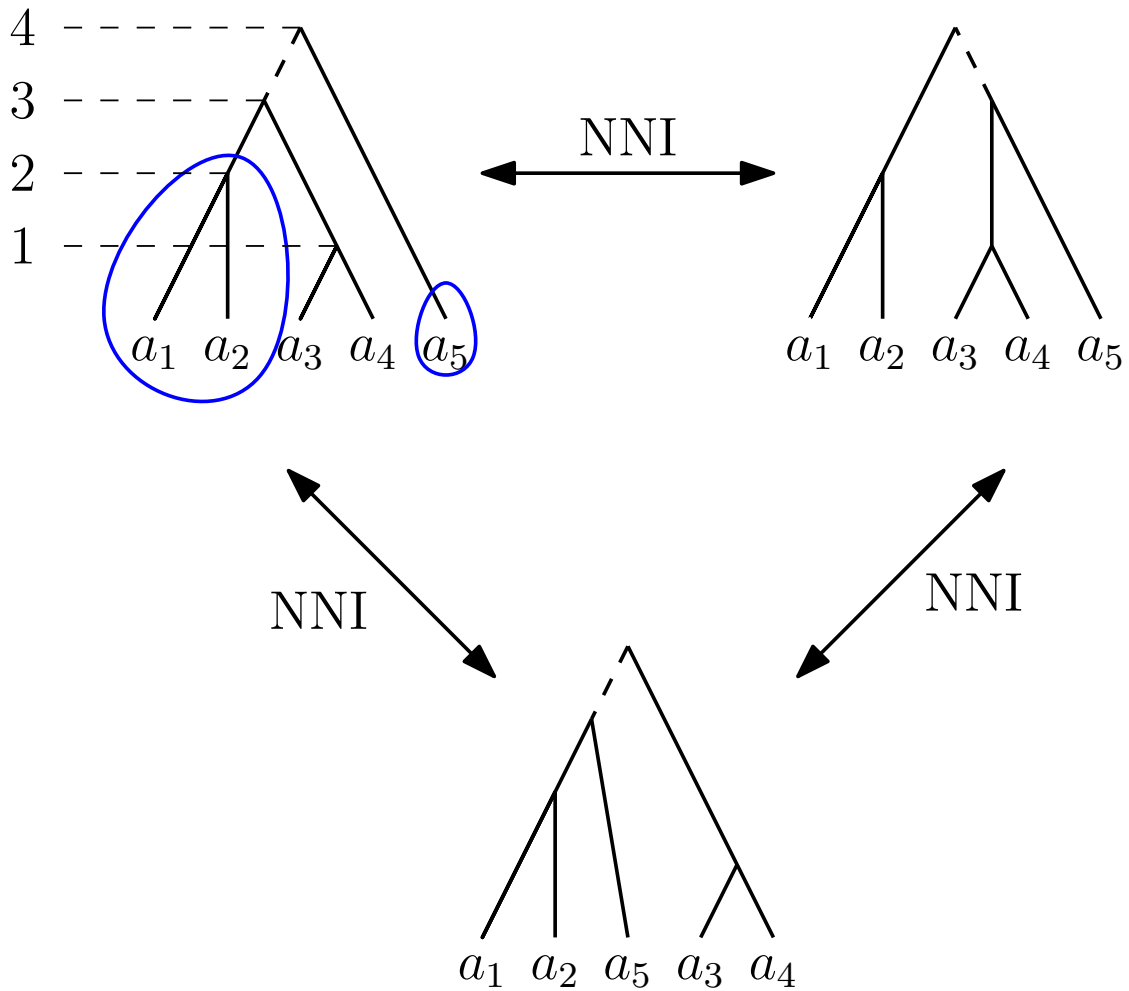
RNNI – Ranked Nearest Neighbour Interchange

NNI Move



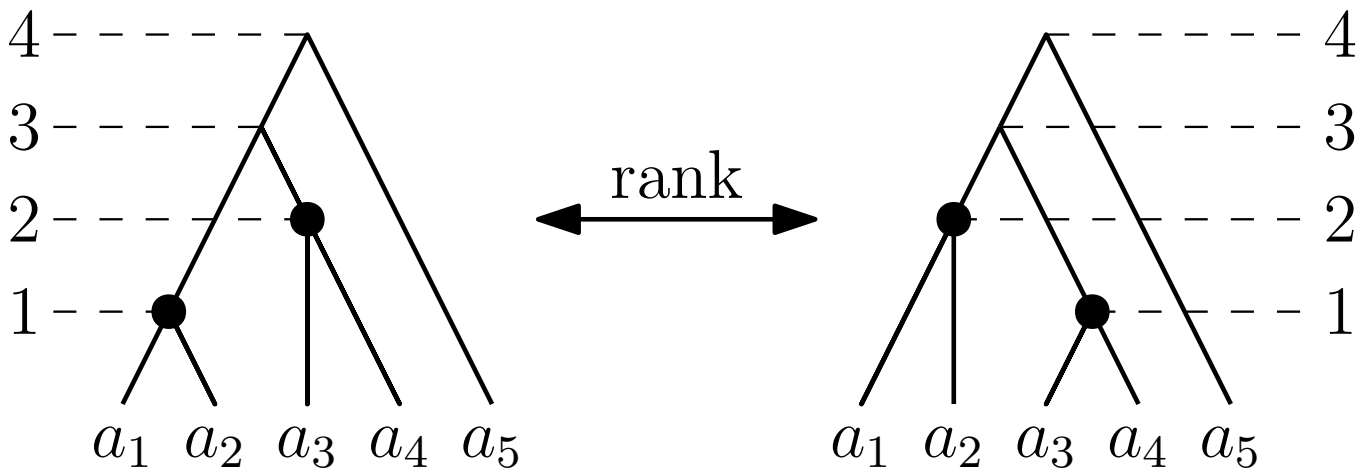
RNNI – Ranked Nearest Neighbour Interchange

NNI Move

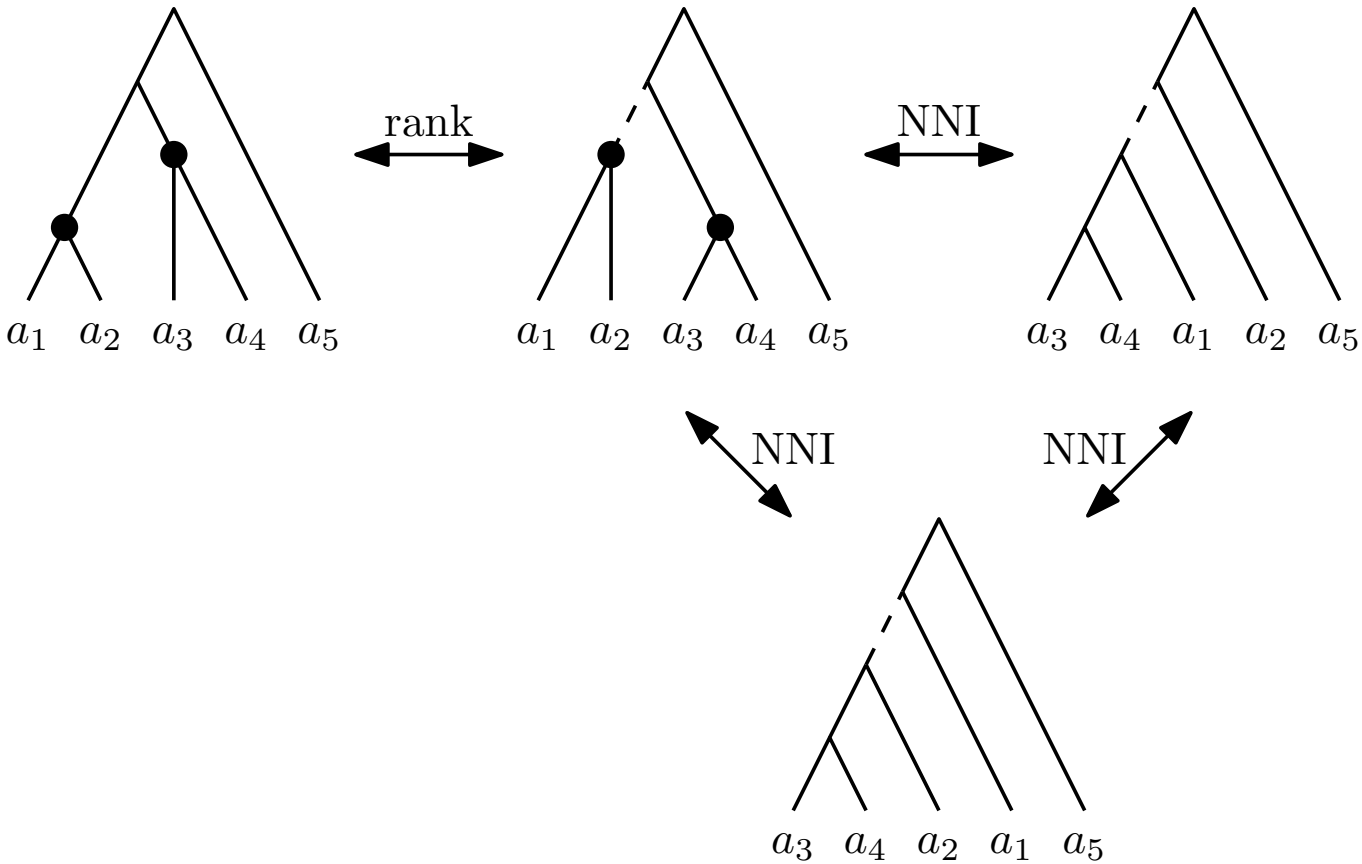


RNNI – Ranked Nearest Neighbour Interchange

Rank Move



RNNI – Ranked Nearest Neighbour Interchange

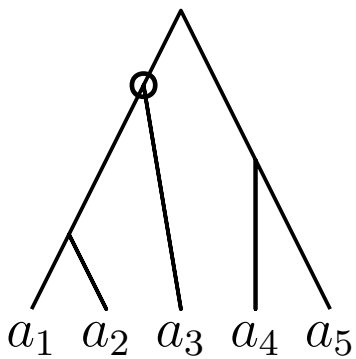


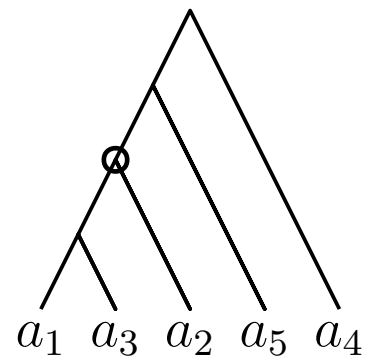
Computing shortest paths

Theorem

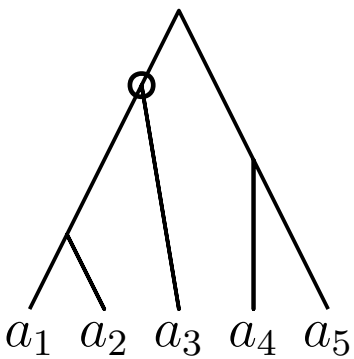
Shortest paths in RNNI *can be computed in time* $\mathcal{O}(n^2)$.

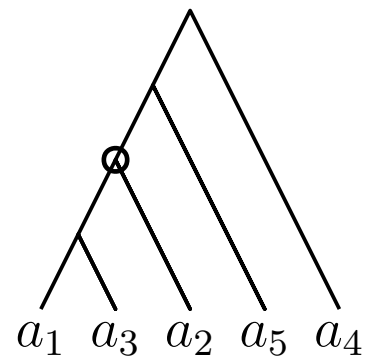
Clusters





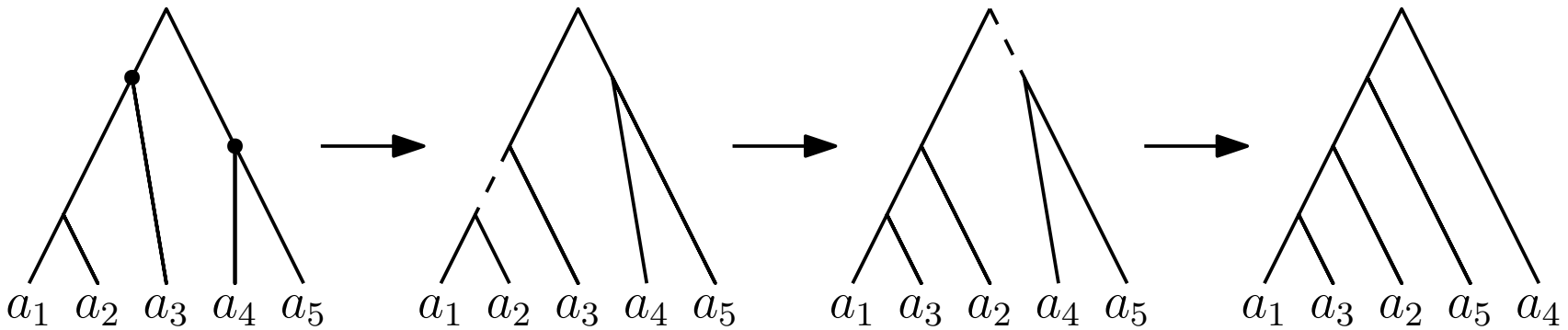
Clusters





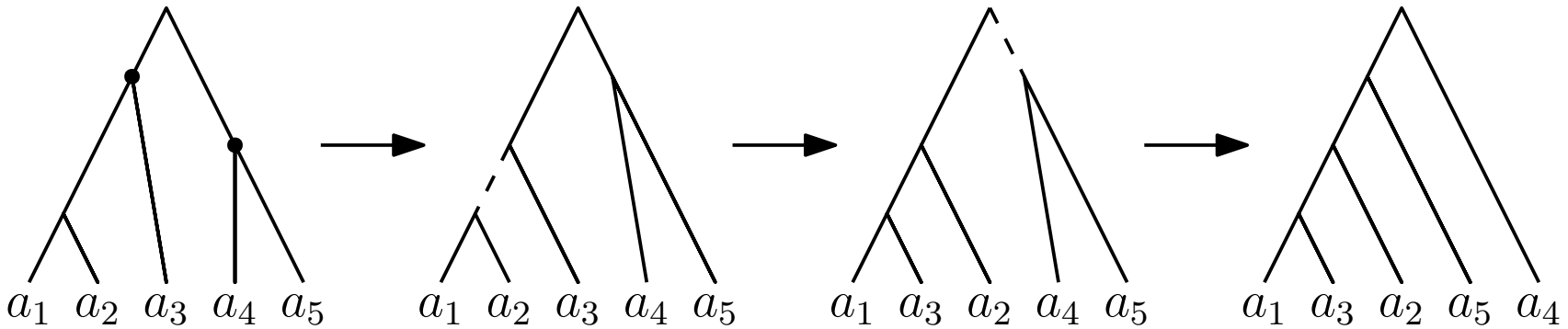
shared cluster: $\{a_1, a_2, a_3\}$

Clusters



shared cluster: $\{a_1, a_2, a_3\}$

Clusters

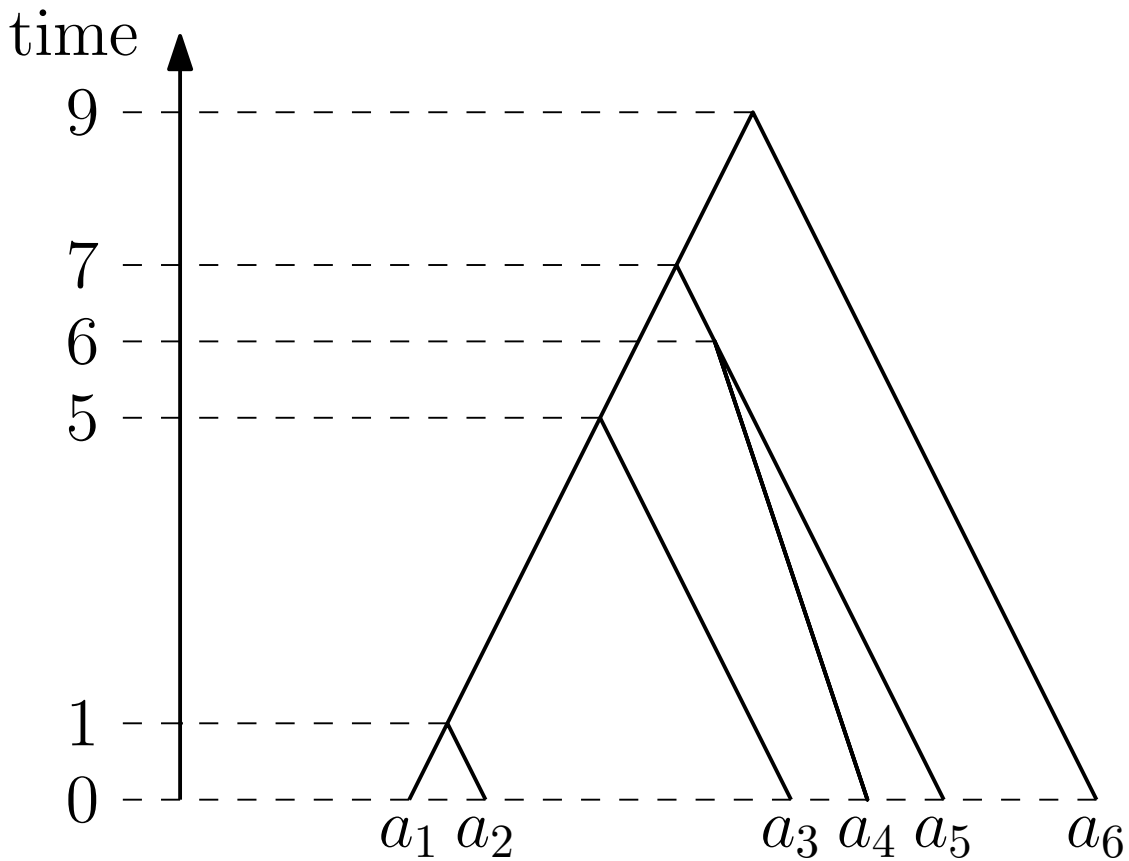


shared cluster: $\{a_1, a_2, a_3\}$

Theorem

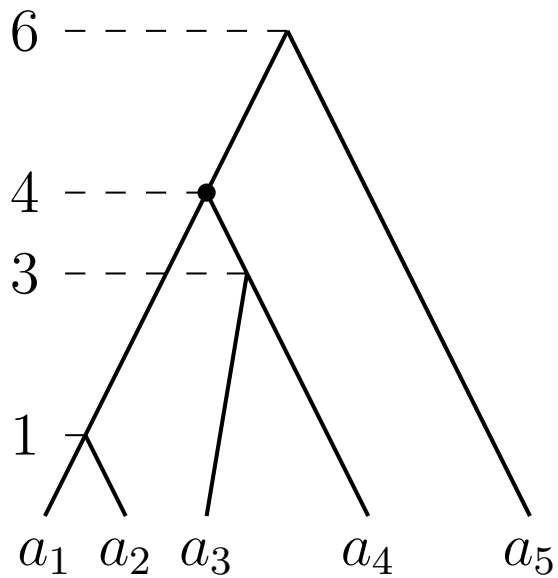
RNNI *has the cluster property, i.e. a cluster shared by two trees T and R is present in every tree on every shortest path between T and R.*

Discrete Coalescent Trees



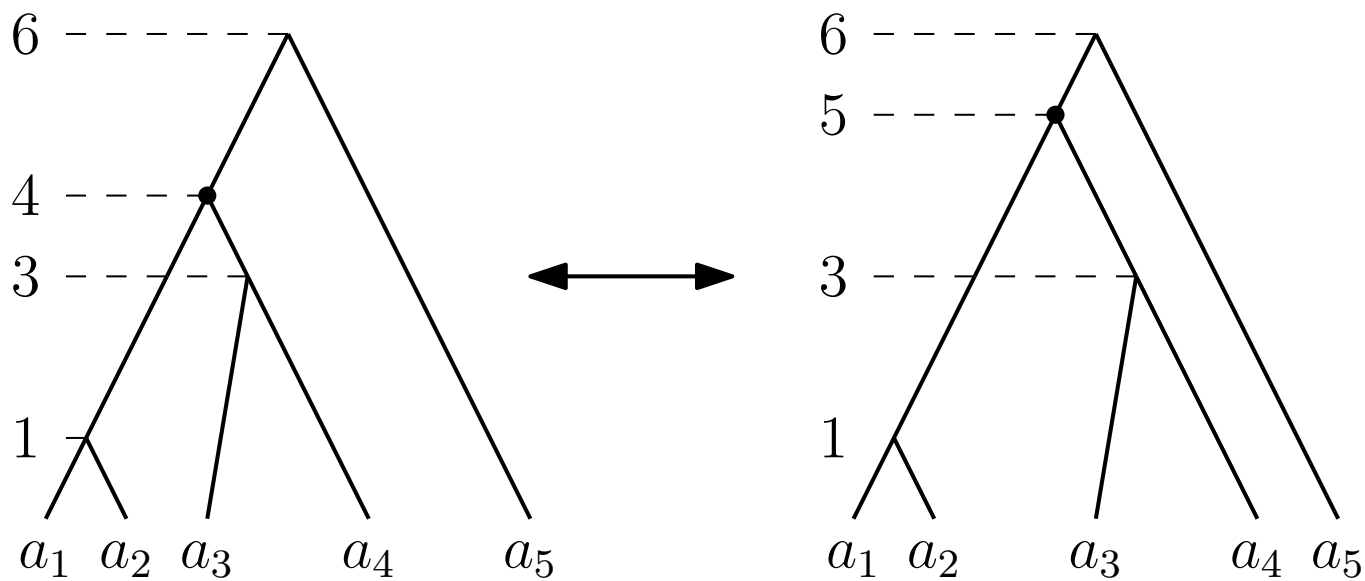
Discrete Coalescent Trees

Length moves



Discrete Coalescent Trees

Length moves



Discrete Coalescent Trees

DCT_m – The space of discrete coalescent trees

Theorem

Shortest paths in DCT_m can be computed in $\mathcal{O}(nm)$.

Discrete Coalescent Trees

DCT_m – The space of discrete coalescent trees

Theorem

Shortest paths in DCT_m can be computed in $\mathcal{O}(nm)$.

Theorem

DCT_m has the cluster property.

Thank you

- Alex Gavryushkin (University of Otago)
- David Bryant (University of Otago)
- BioDS (University of Otago/Canterbury)