

CS 466

Introduction to Bioinformatics

Lecture 15

Mohammed El-Kebir

October 29, 2018



Course Announcements

HW 3 due Oct 29 by 11:59pm

Office hour after class in SC 3216

Outline

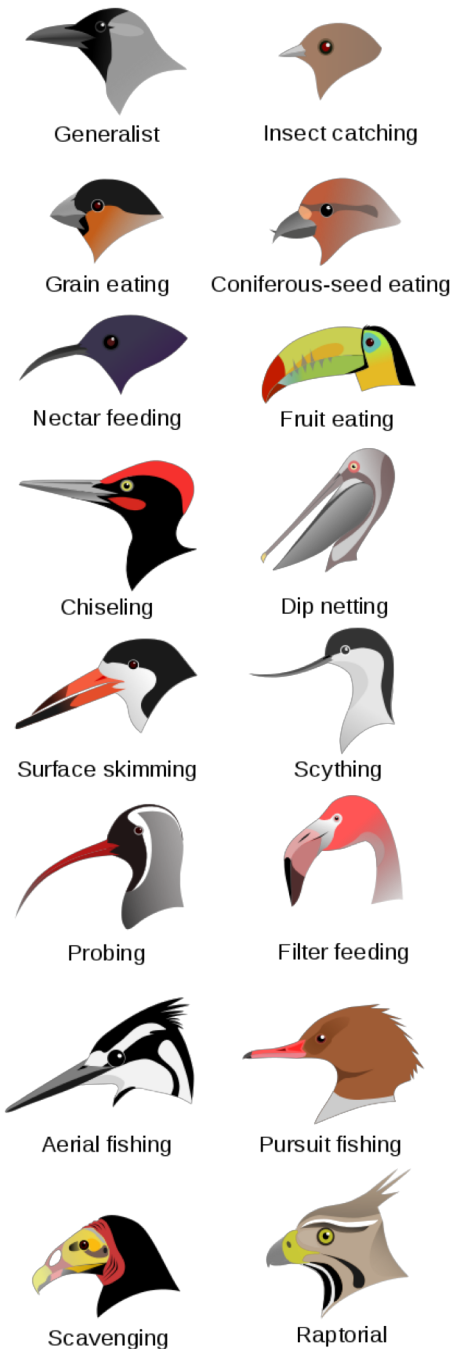
- Recap character-based phylogeny
- Application of small phylogeny maximum parsimony problem to cancer
- Compatibility

Reading:

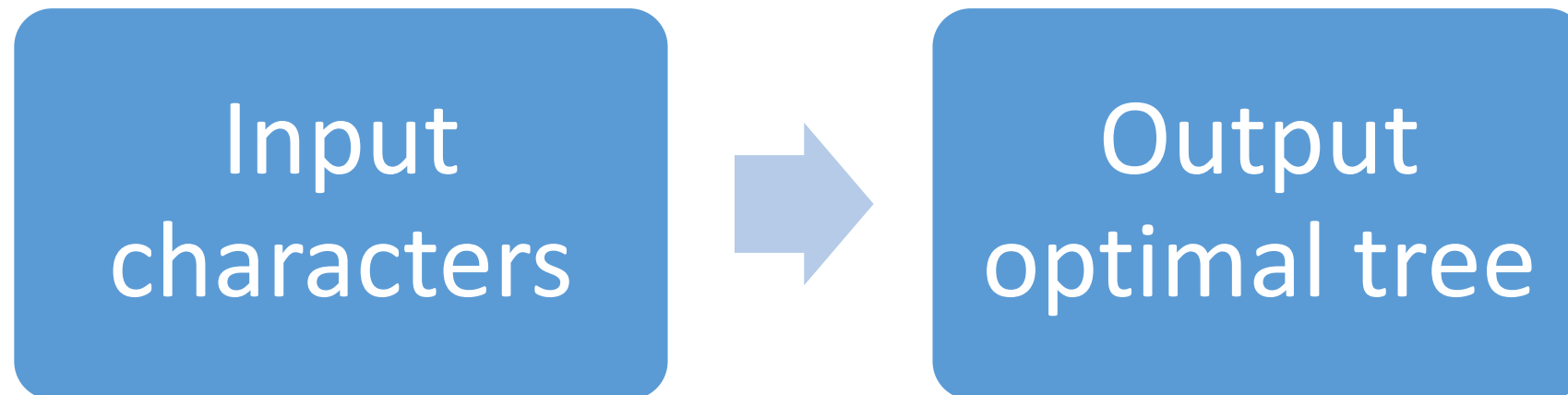
- Lecture notes

Character-Based Tree Reconstruction

- Characters may be morphological features
 - Shape of beak {generalist, insect catching, ...}
 - Number of legs {2,3,4, ..}
 - Hibernation {yes, no}
- Character may be nucleotides/amino acids
 - {A, T, C, G}
 - 20 amino acids
- Values of a character are called states
 - We assume discrete states



Character-Based Phylogeny Reconstruction

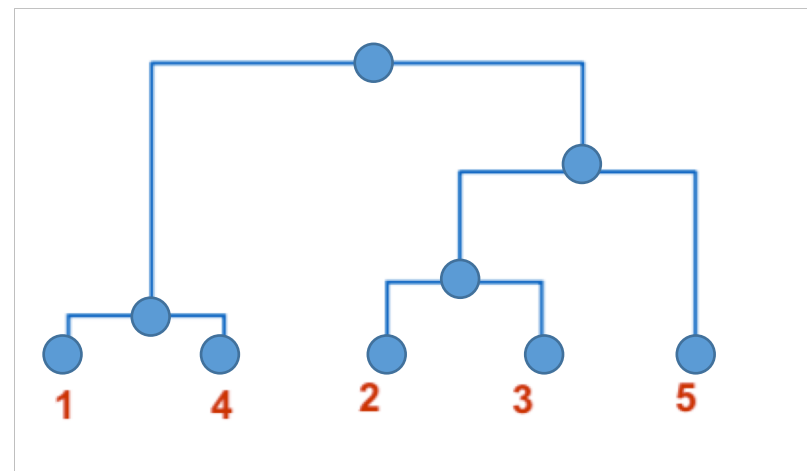


Question: What is optimal?

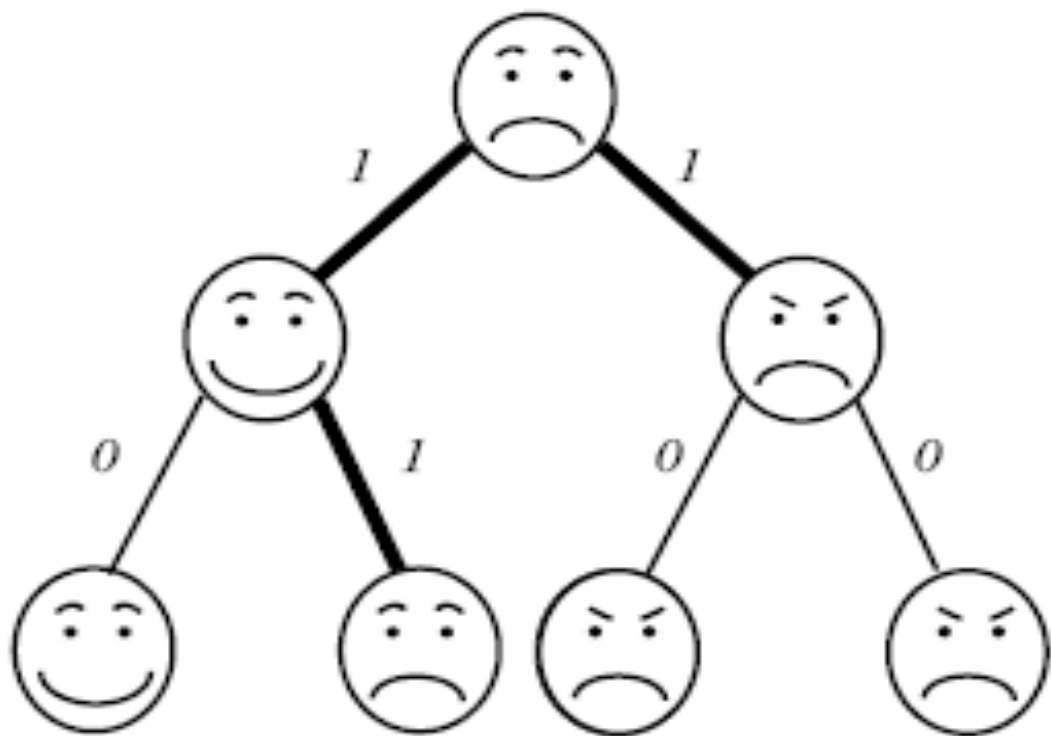
Want: Optimization criterion

Question: How to optimize this criterion?

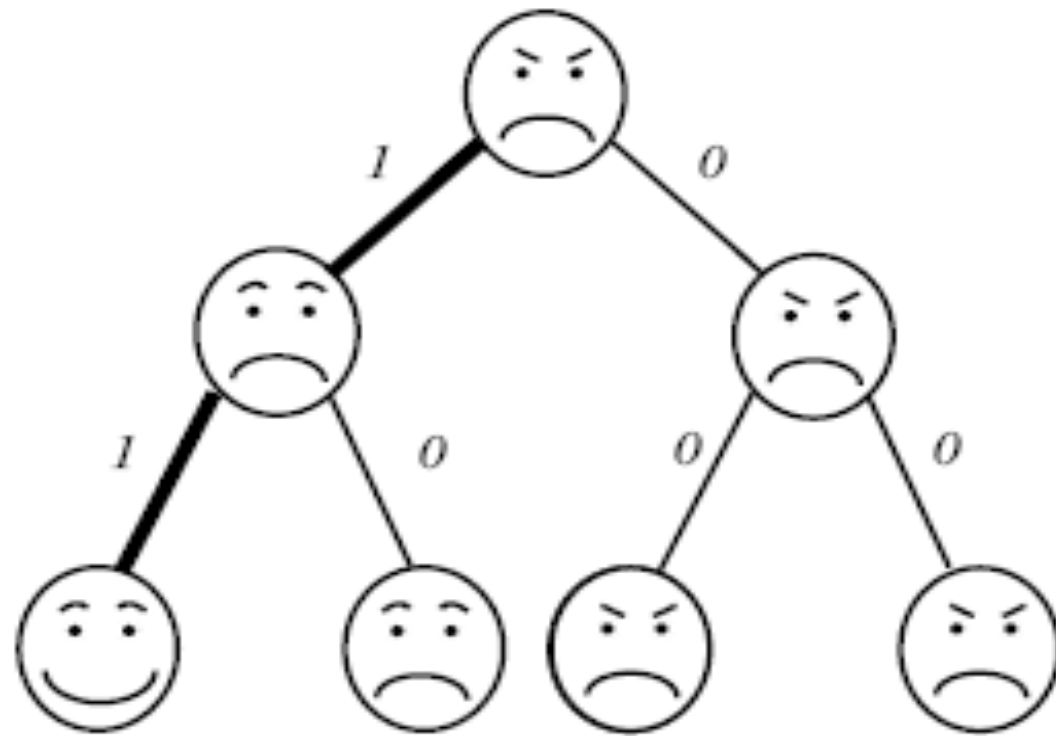
Want: Algorithm



Character-Based Phylogeny Reconstruction: Criterion



(a) *Parsimony Score=3*



(b) *Parsimony Score=2*

Parsimony: minimize number of changes on edges of tree

Again, a Small and a Large Problem

Small Maximum Parsimony Phylogeny Problem:

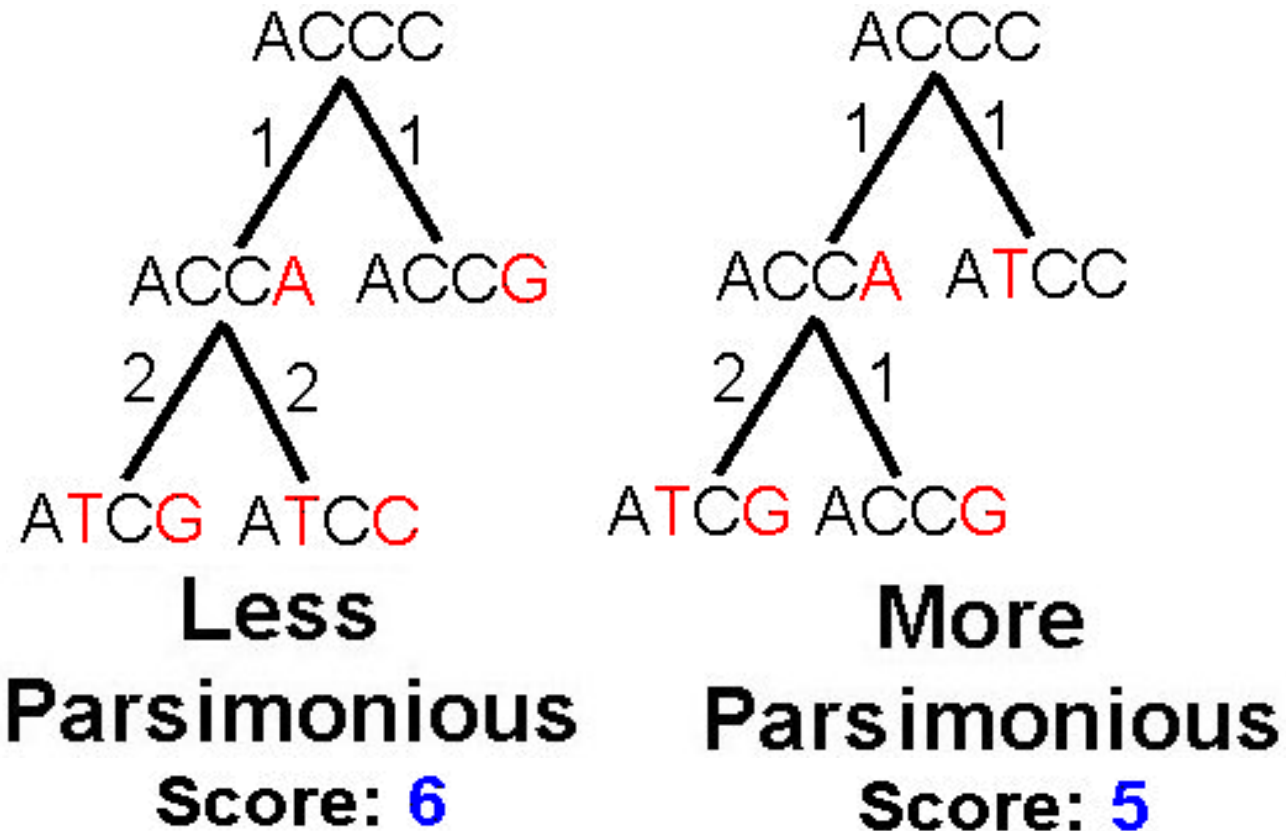
Given $m \times n$ matrix $A = [a_{i,j}]$ and tree T with m leaves, find assignment of character states to each internal vertex of T with minimum parsimony score.

Large Maximum Parsimony Phylogeny Problem:

Given $m \times n$ matrix $A = [a_{i,j}]$, find a tree T with m leaves labeled according to A and an assignment of character states to each internal vertex of T with minimum parsimony score.

Question: Are both problems easy (i.e. in P)?

Small Maximum Parsimony Phylogeny Problem



Key observations: (1) Characters can be solved independently.
(2) Optimal substructure in subtrees.

Recurrence for Small Maximum Parsimony Problem

Small Maximum Parsimony Phylogeny Problem:

Given rooted tree T whose leaves are labeled by $\sigma : L(T) \rightarrow \Sigma$, find assignment of states to each internal vertex of T with minimum parsimony score.

Let $\mu(v, s)$ be the minimum number of mutations in the subtree rooted at v when assigning state s to v .

$$c(s, t) = \begin{cases} 0, & \text{if } s = t \\ 1, & \text{if } s \neq t, \end{cases}$$

Let $\delta(v)$ be the set of children of v .

$$\mu(v, s) = \min \begin{cases} \infty, & \text{if } v \in L(T) \text{ and } s \neq \sigma(v), \\ 0, & \text{if } v \in L(T) \text{ and } s = \sigma(v), \\ \sum_{w \in \delta(v)} \min_{t \in \Sigma} \{c(s, t) + \mu(w, t)\}, & \text{if } v \notin L(T). \end{cases}$$

Filling out DP Table and Traceback

Filling out μ $\text{Fill}(T, r(T), \sigma, \Sigma)$

$\text{Fill}(T, v, \sigma, \Sigma)$ $O(m|\Sigma|^2)$

if $v \in \mathcal{L}(T)$ then

For $s \in \Sigma$

if $s = \sigma(v)$ then

$$\mu(v, s) = 0$$

else $\mu(v, s) = \infty$

else

For $w \in \delta(v)$

$\text{Fill}(T, w, \sigma, \Sigma)$

// children

$$\mu(v, s) = 0$$

For $w \in \delta(v)$

$$\mu(v, s) += \min_{t \in \Sigma} \{ c(s, t) + \mu(w, t) \}$$

Backtrace (T, v, μ)

if $v = r(T)$

$$\sigma(r(T)) = \arg \min_{s \in \Sigma} \{ \mu(r(T), s) \}$$

else let u be the parent of v and let $s = \sigma(u)$

$$\sigma(v) = \arg \min_{t \in \Sigma} \{ c(s, t) + \mu(v, t) \}$$

For $w \in \delta(v)$

Backtrace (T, w, μ)

Let $r(T)$ be the root vertex

Outline

- Recap character-based phylogeny
- Application of small phylogeny maximum parsimony problem to cancer
- Compatibility

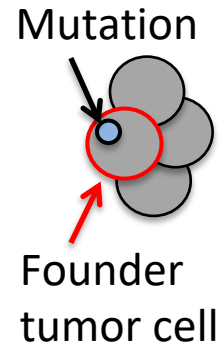
Reading:

- Lecture notes

Tumorigenesis: (i) Cell Mutation

Clonal Theory of Cancer

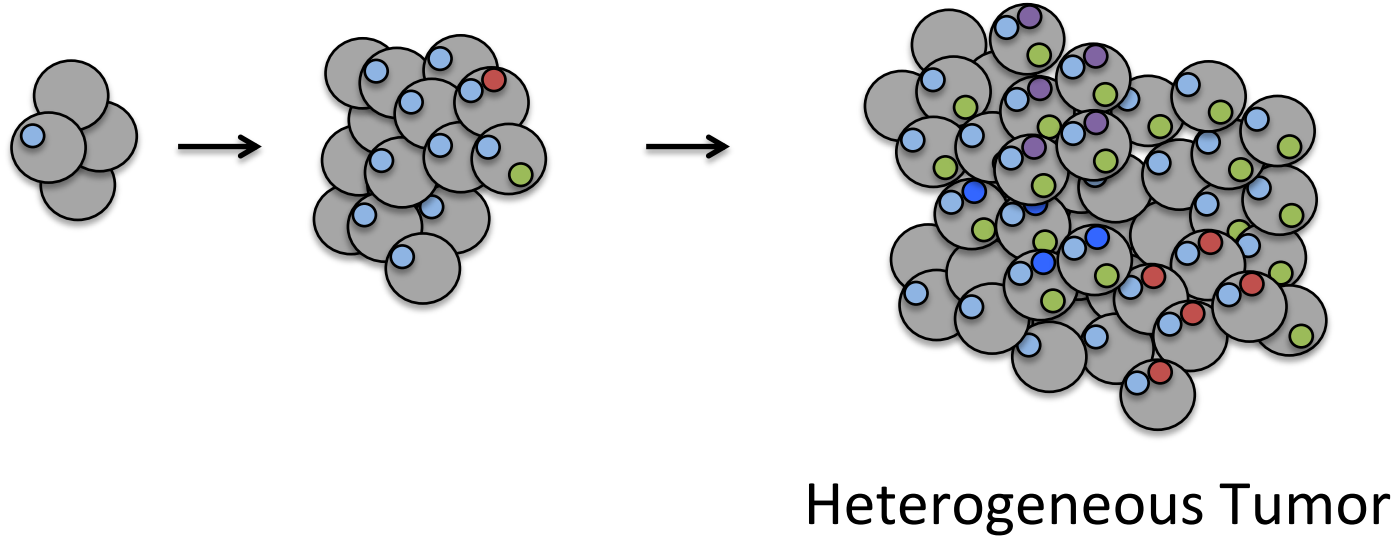
[Nowell, 1976]



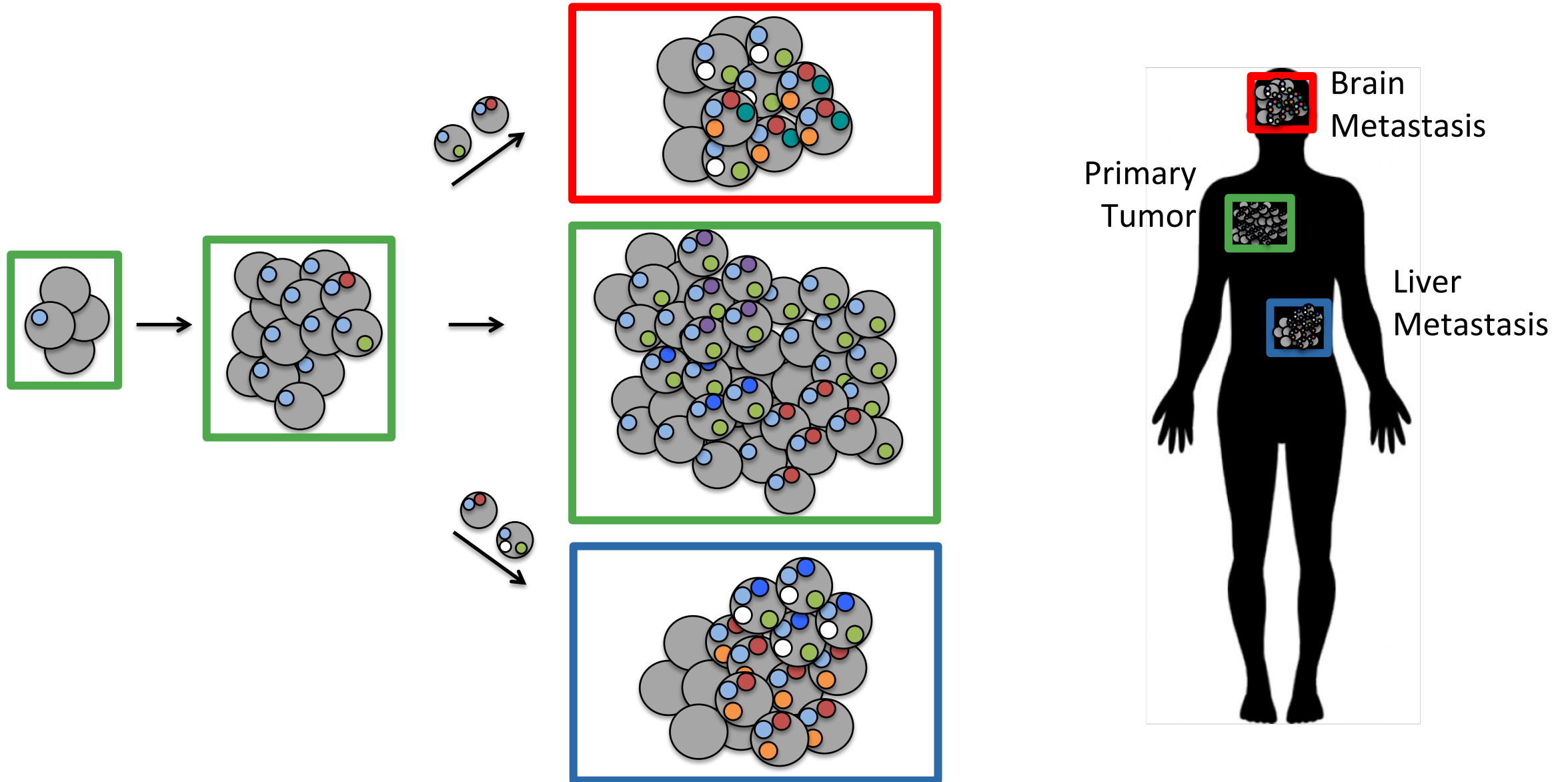
Tumorigenesis: (i) Cell Mutation, (ii) Cell Division

Clonal Theory of Cancer

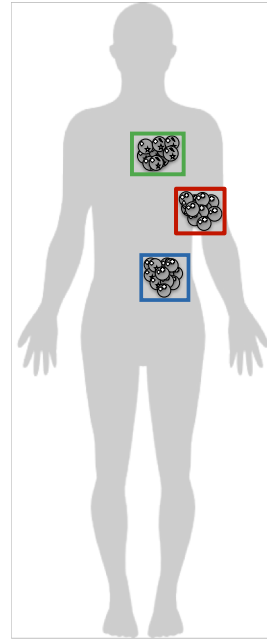
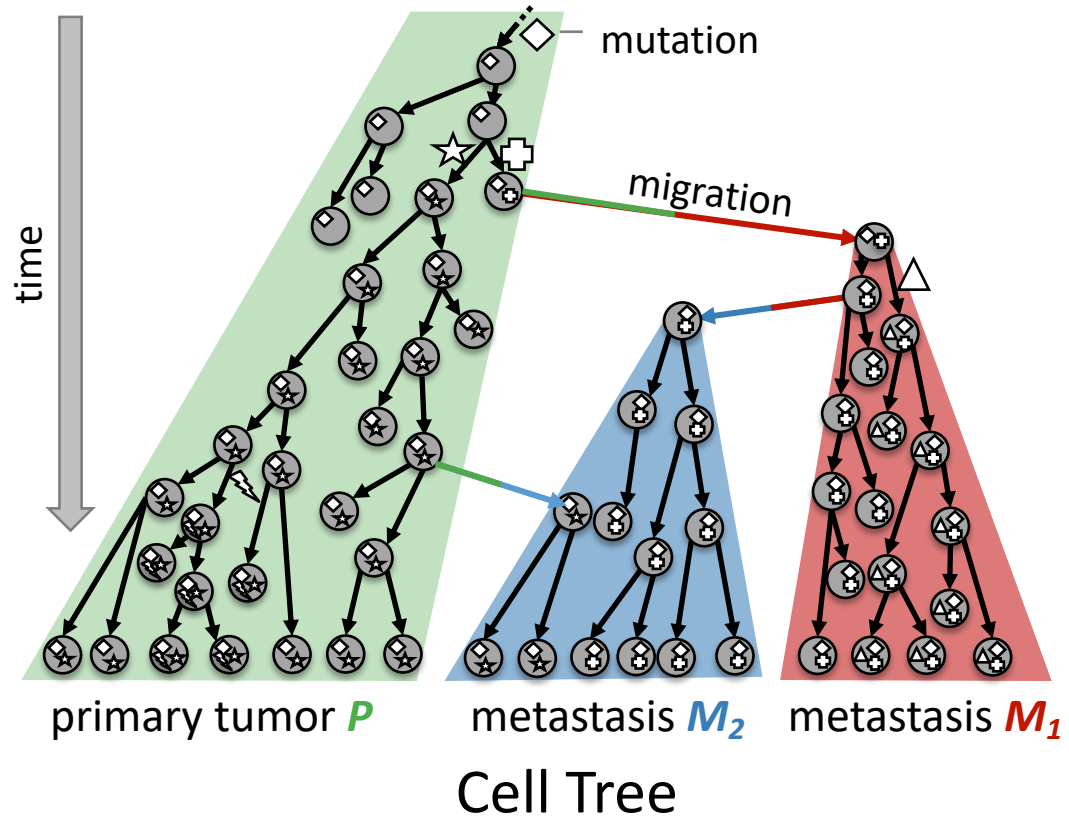
[Nowell, 1976]



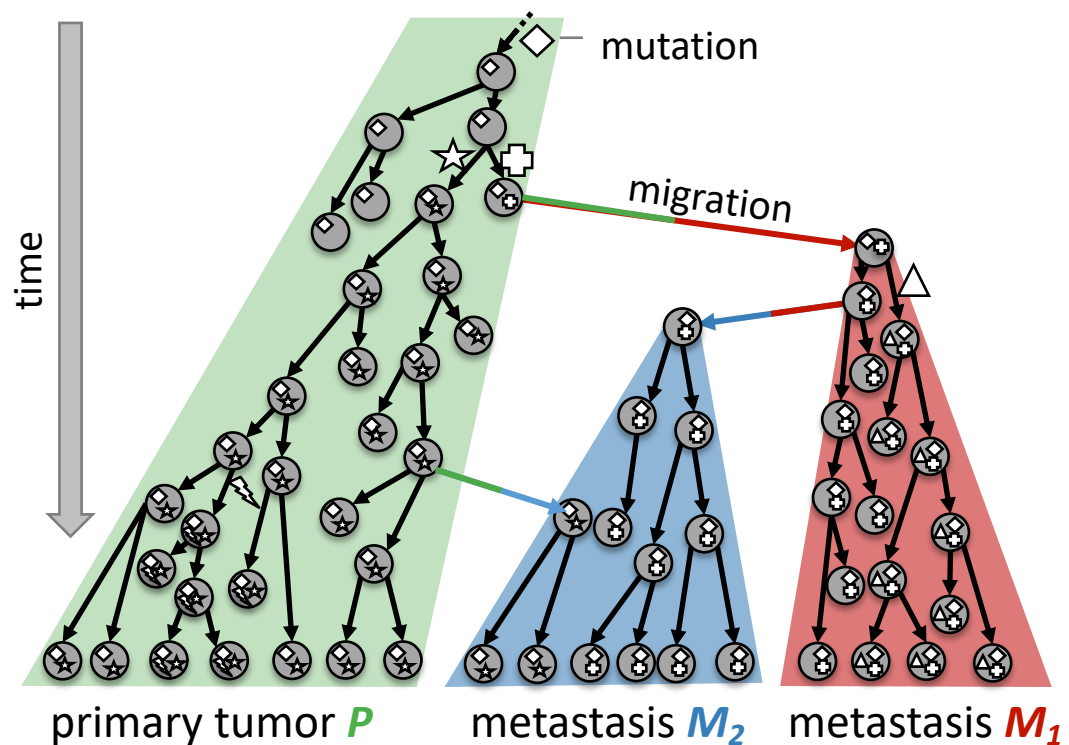
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration



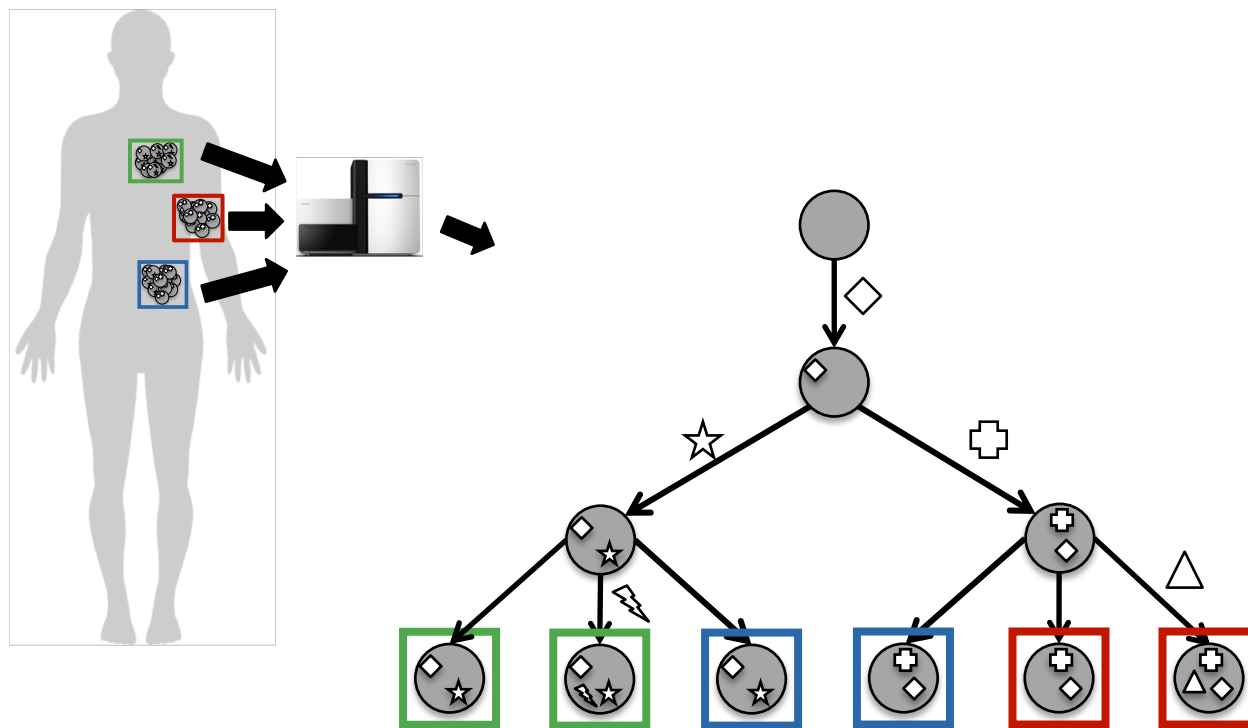
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration



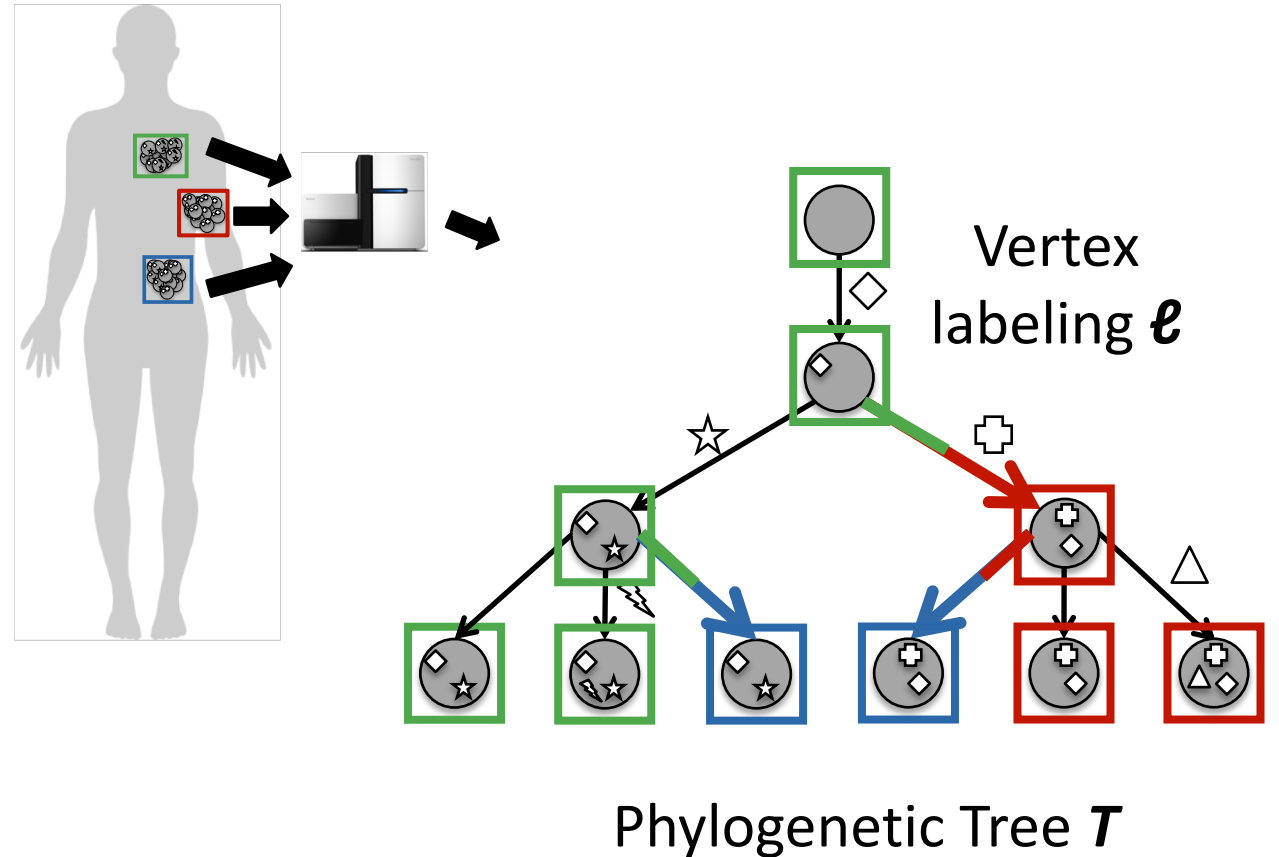
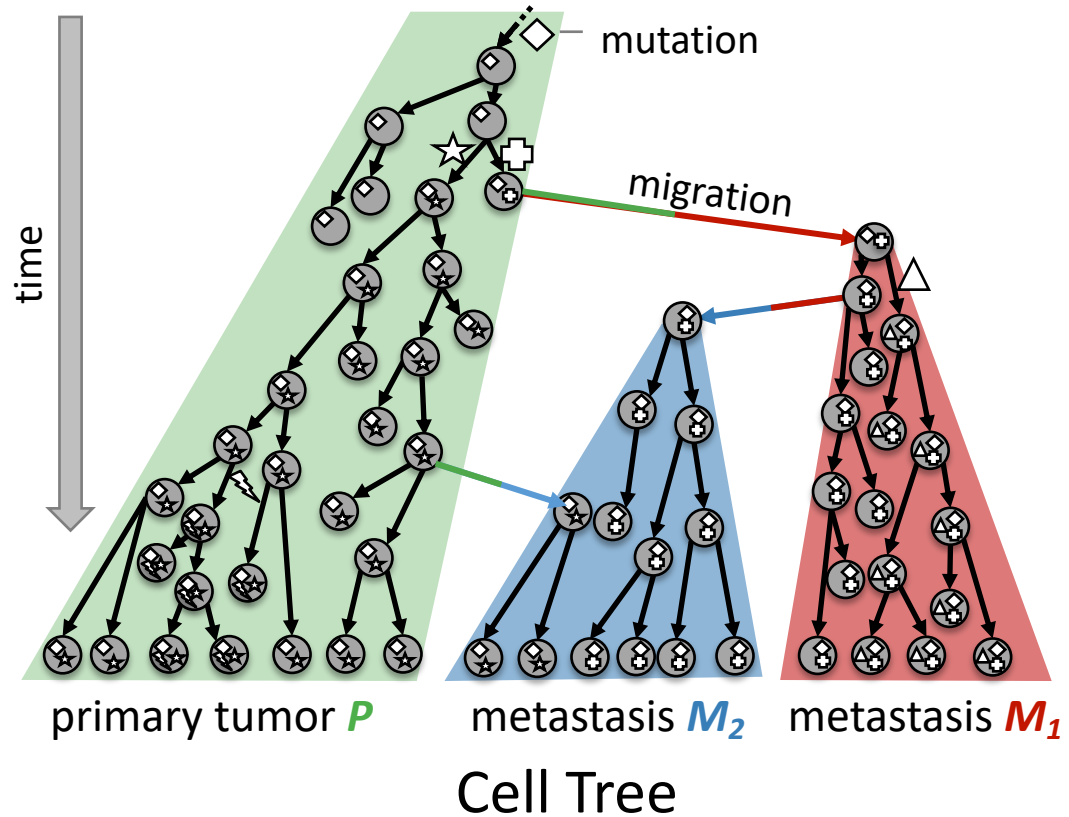
Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration



Cell Tree



Tumorigenesis: (i) Cell Division, (ii) Mutation & (iii) Migration

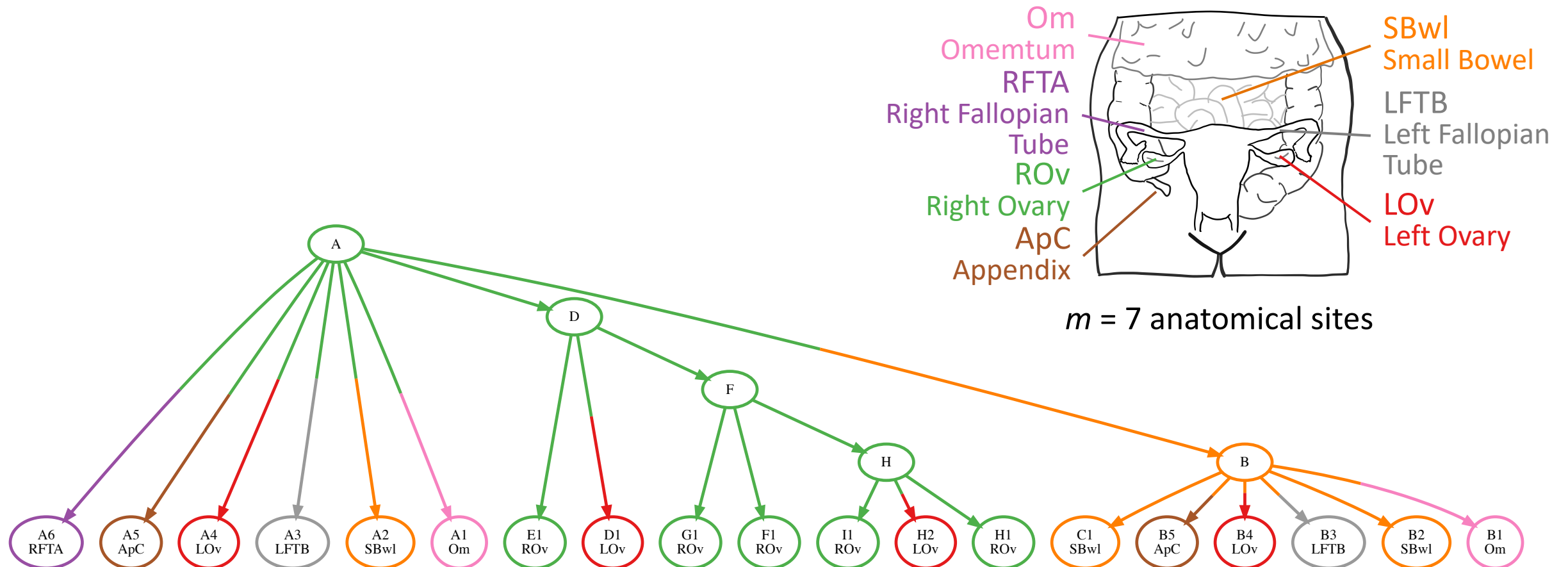


Goal: Given phylogenetic tree T , find *parsimonious* vertex labeling ℓ with fewest migrations

Minimum Migration Analysis in Ovarian Cancer

McPherson et al. (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nature Genetics*.

- Instance of the maximum parsimony small phylogeny problem [Fitch, 1971; Sankoff, 1975]

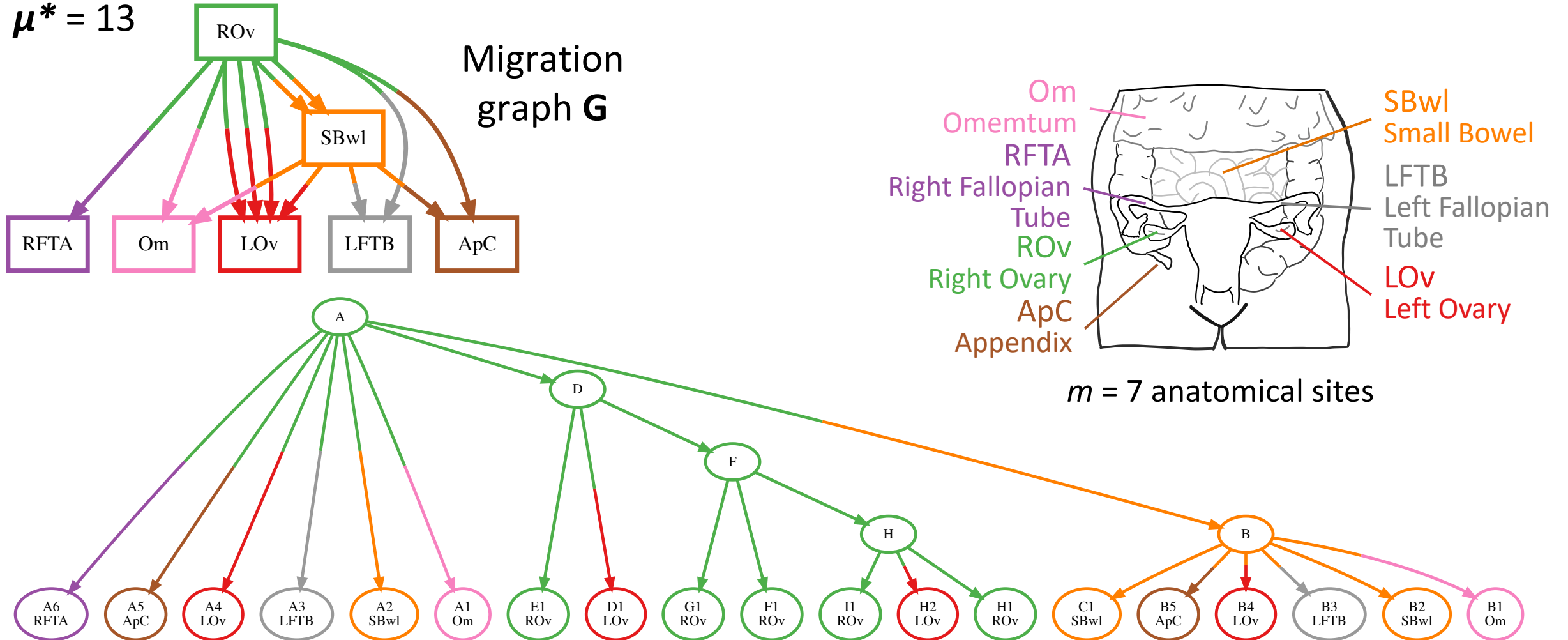


Minimum Migration Analysis in Ovarian Cancer

McPherson et al. (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nature Genetics*.

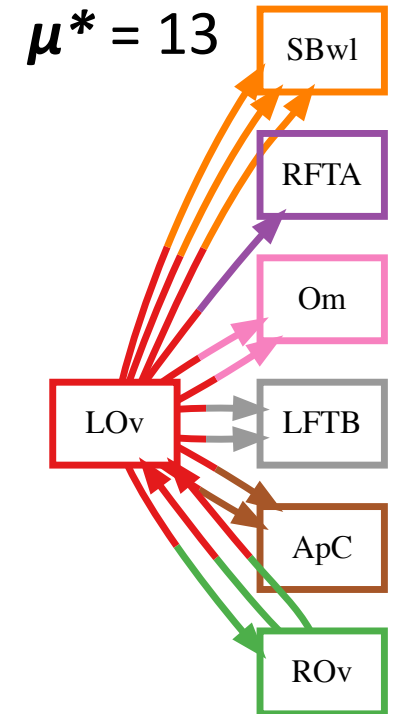
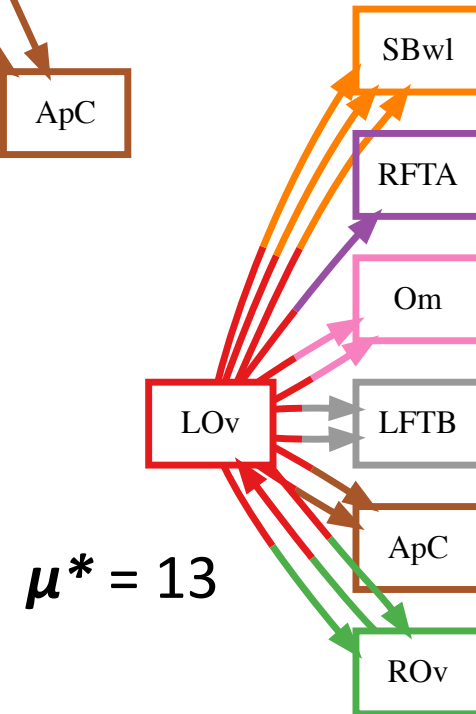
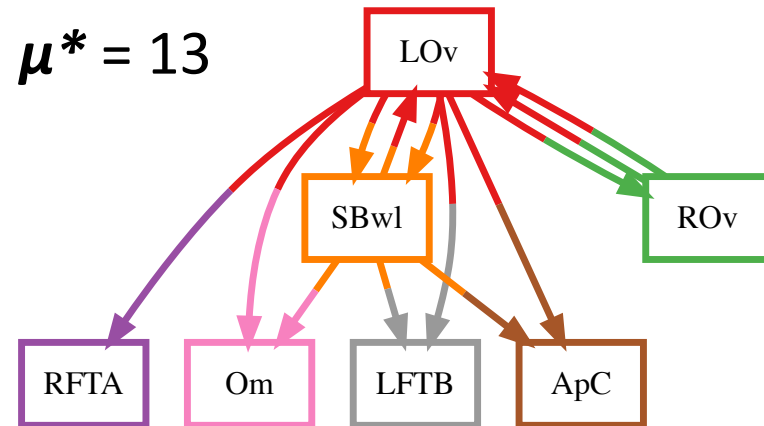
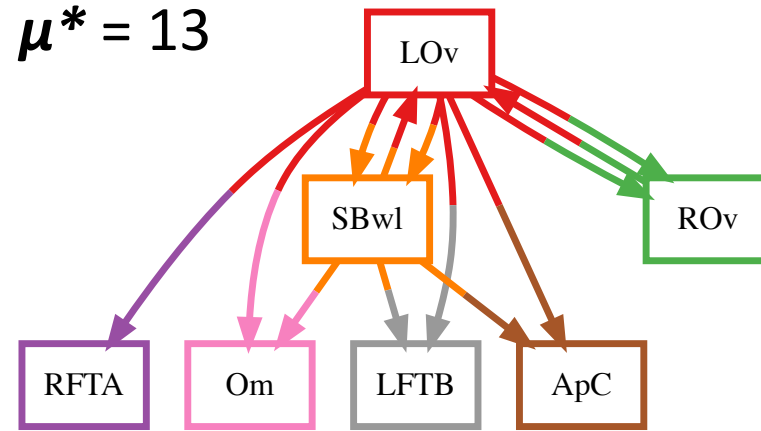
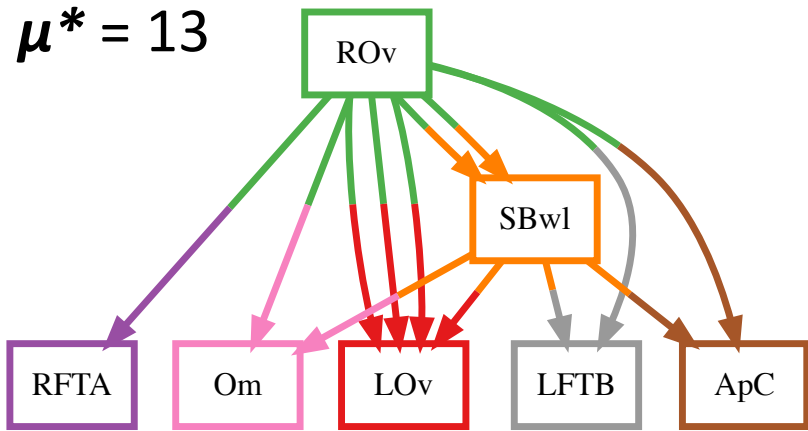
- Instance of the maximum parsimony small phylogeny problem [Fitch, 1971; Sankoff, 1975]

$\mu^* = 13$



Minimum Migration History is *Not* Unique

- Enumerate all minimum-migration vertex labelings in the backtrace step

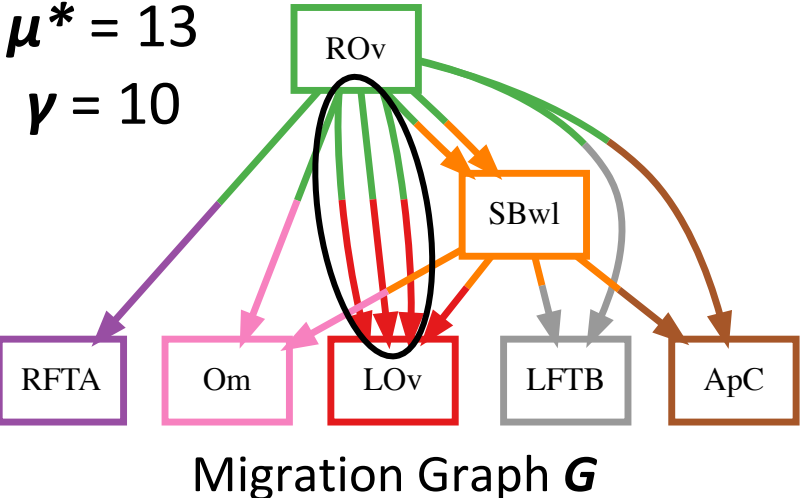


ApC	Appendix
LFTB	Left Fallopian Tube
LOv	Left Ovary
RFTA	Right Fallopian Tube
ROv	Right Ovary
SBwl	Small Bowel
Om	Omentum

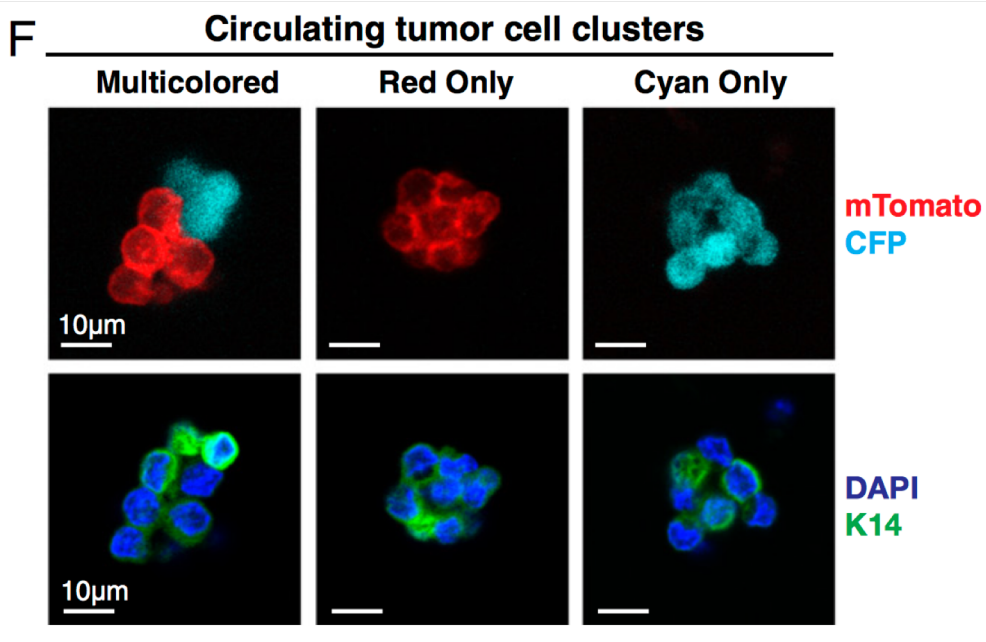
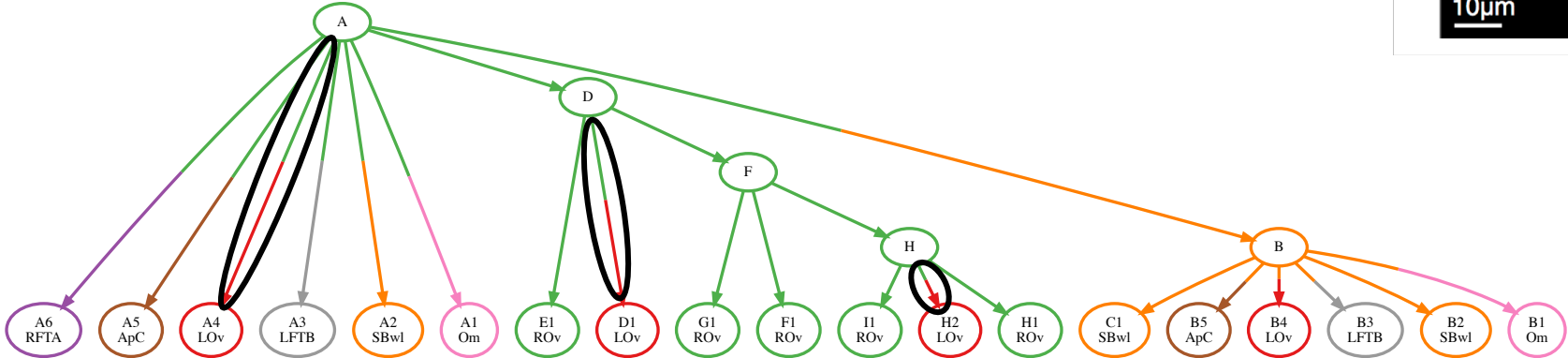
Comigrations: Simultaneous Migrations of Multiple Clones

- Multiple tumor cells migrate simultaneously through the blood stream [Cheung et al., 2016]
- Second objective: number γ of **comigrations** is the number of multi-edges in migration graph G^\dagger

† Not necessarily true in the case of directed cycles



Clone Tree T

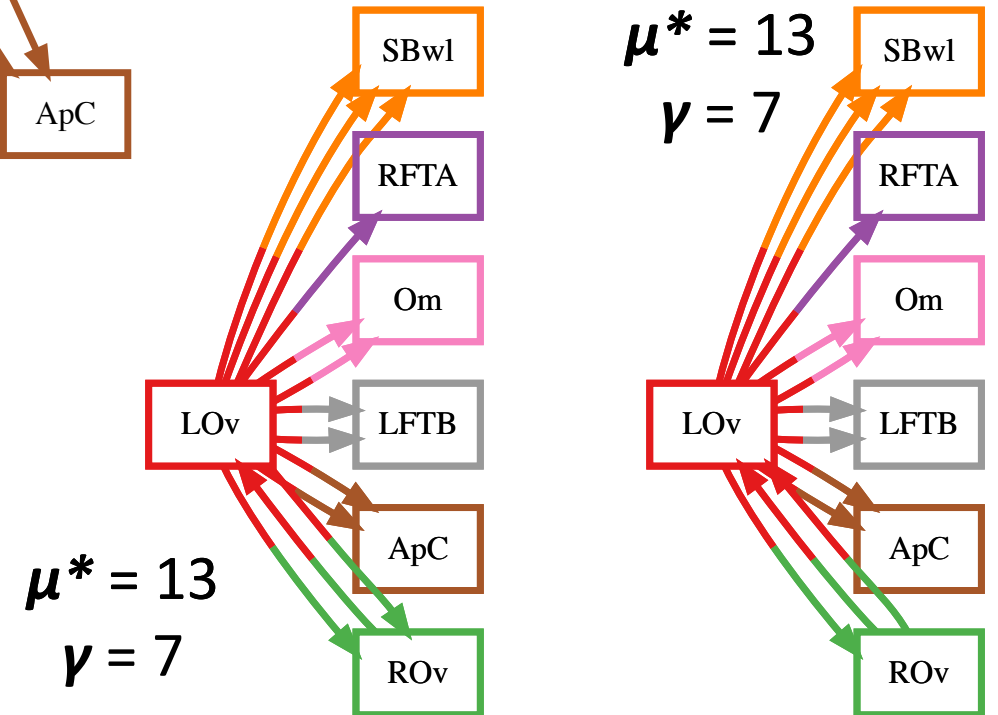
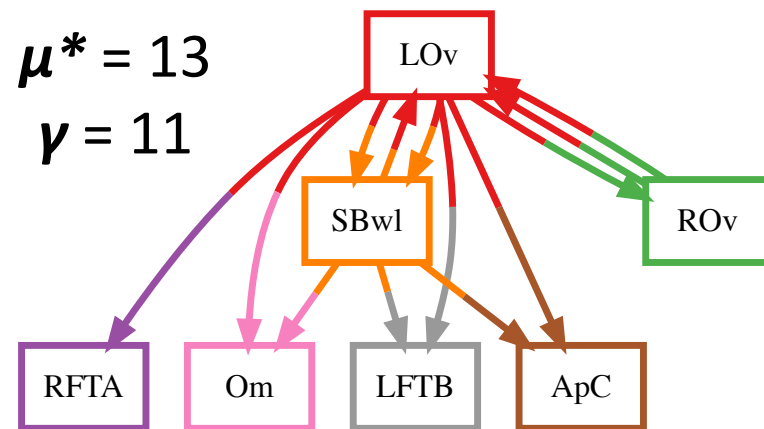
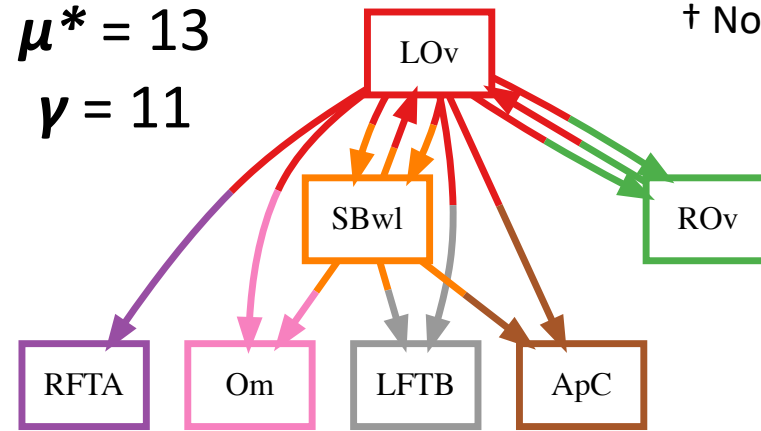
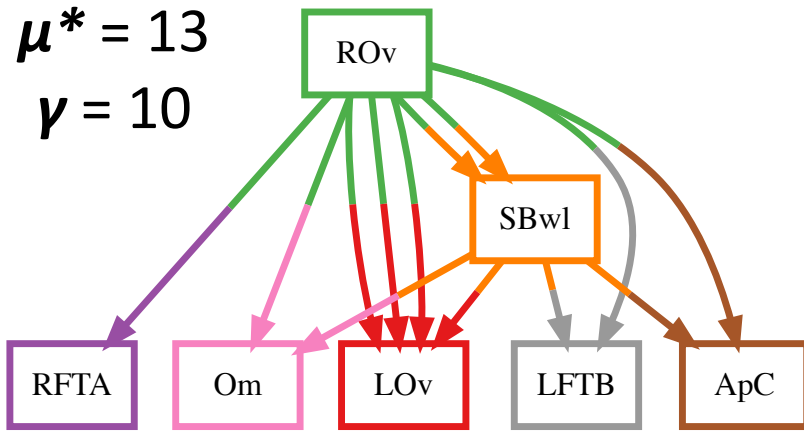


ApC	Appendix
LFTB	Left Fallopian Tube
LOv	Left Ovary
RFTA	Right Fallopian Tube
ROv	Right Ovary
SBwl	Small Bowel
Om	Omentum

Comigrations: Simultaneous Migrations of Multiple Clones

- Multiple tumor cells migrate simultaneously through the blood stream [Cheung et al., 2016]
- Second objective: number γ of **comigrations** is the number of multi-edges in migration graph G^\dagger

† Not necessarily true in the case of directed cycles

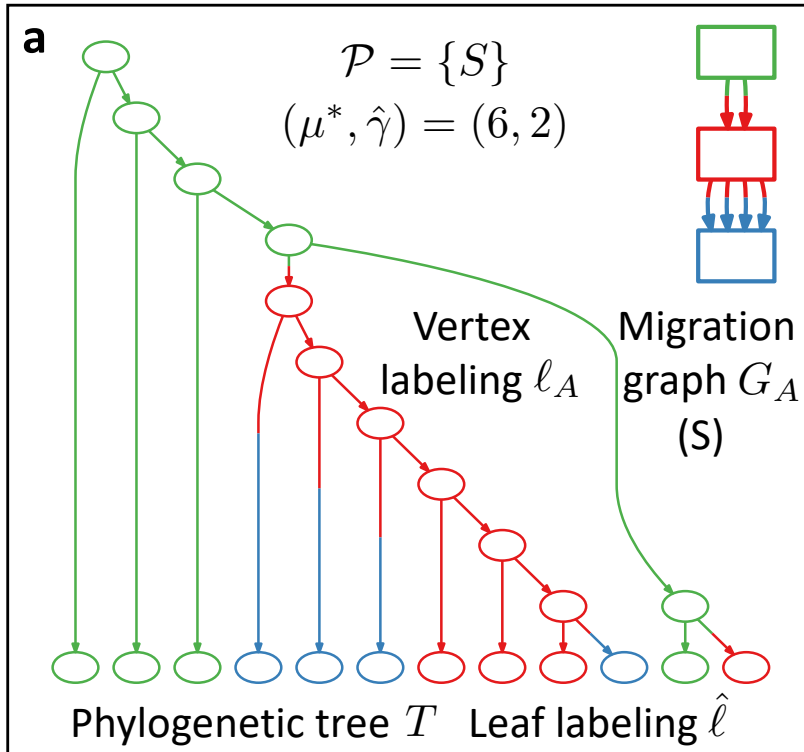


ApC	Appendix
LFTB	Left Fallopian Tube
LOv	Left Ovary
RFTA	Right Fallopian Tube
ROv	Right Ovary
SBwl	Small Bowel
Om	Omentum

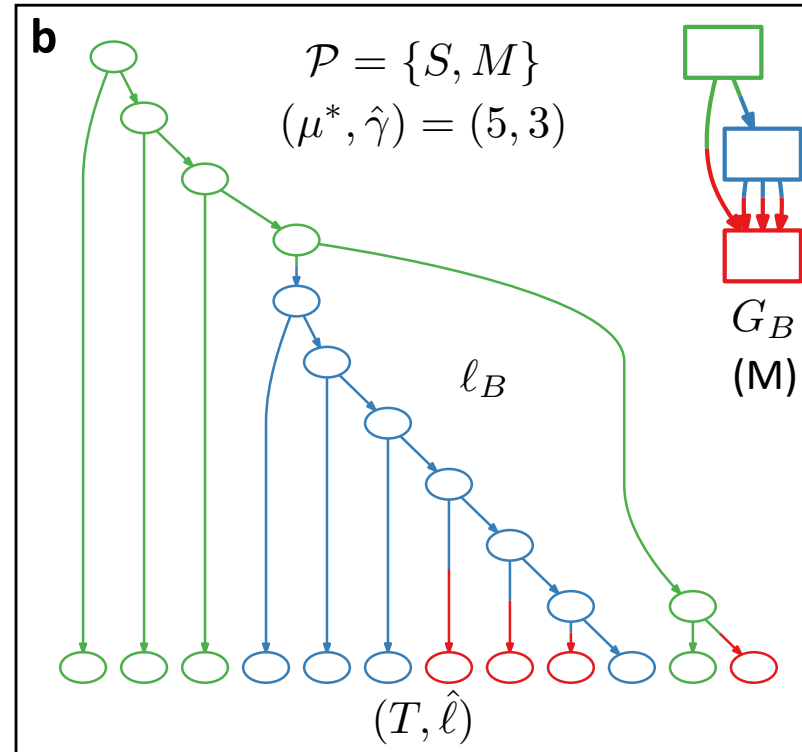
Constrained Multi-objective Optimization Problem

Parsimonious Migration History (PMH): Given a phylogenetic tree T and a set $\mathcal{P} \subseteq \{S, M, R\}$ of allowed migration patterns, find vertex labeling ℓ with minimum migration number $\mu^*(T)$ and smallest comigration number $\hat{\gamma}(T)$.

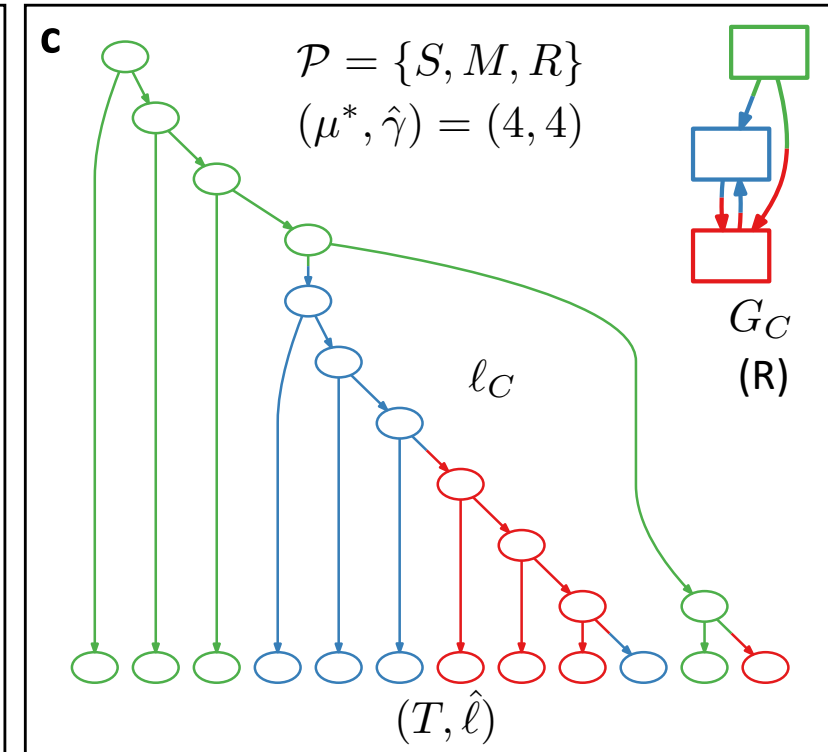
single-source seeding (S)



multi-source seeding (M)



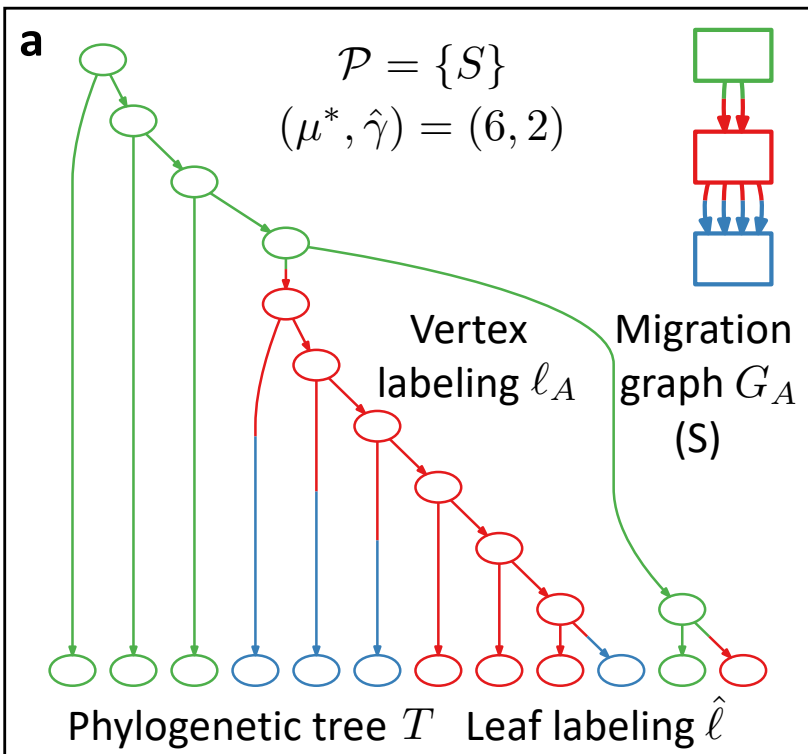
reseeding (R)



Results [El-Kebir, WABI 2018]

Parsimonious Migration History (PMH): Given a phylogenetic tree T and a set $\mathcal{P} \subseteq \{S, M, R\}$ of allowed migration patterns, find vertex labeling ℓ with minimum migration number $\mu^*(T)$ and smallest comigration number $\hat{\gamma}(T)$.

single-source seeding (S)

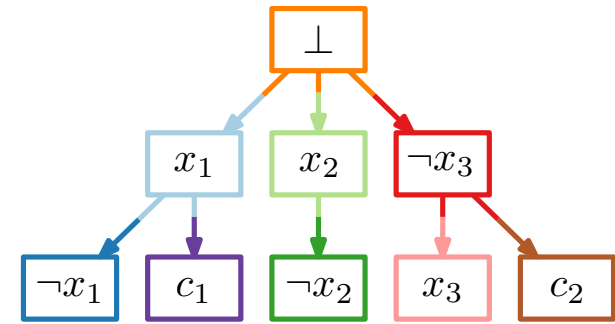


Theorem 1: PMH is NP-hard when $\mathcal{P} = \{S\}$

Theorem 2: PMH is fixed parameter tractable in the number m of locations when $\mathcal{P} = \{S\}$

PMH is NP-hard when $\mathcal{P} = \{S\}$

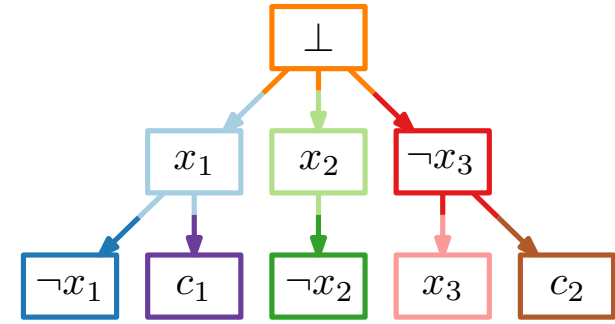
3-SAT: Given $\varphi = \bigwedge_{i=1}^k (y_{i,1} \vee y_{i,2} \vee y_{i,3})$ with variables $\{x_1, \dots, x_n\}$ and k clauses, find $\phi : [n] \rightarrow \{0,1\}$ satisfying φ



$$\Sigma = \{x_1, \dots, x_n, \neg x_1, \dots, \neg x_n, c_1, \dots, c_k, \perp\}$$

PMH is NP-hard when $\mathcal{P} = \{S\}$

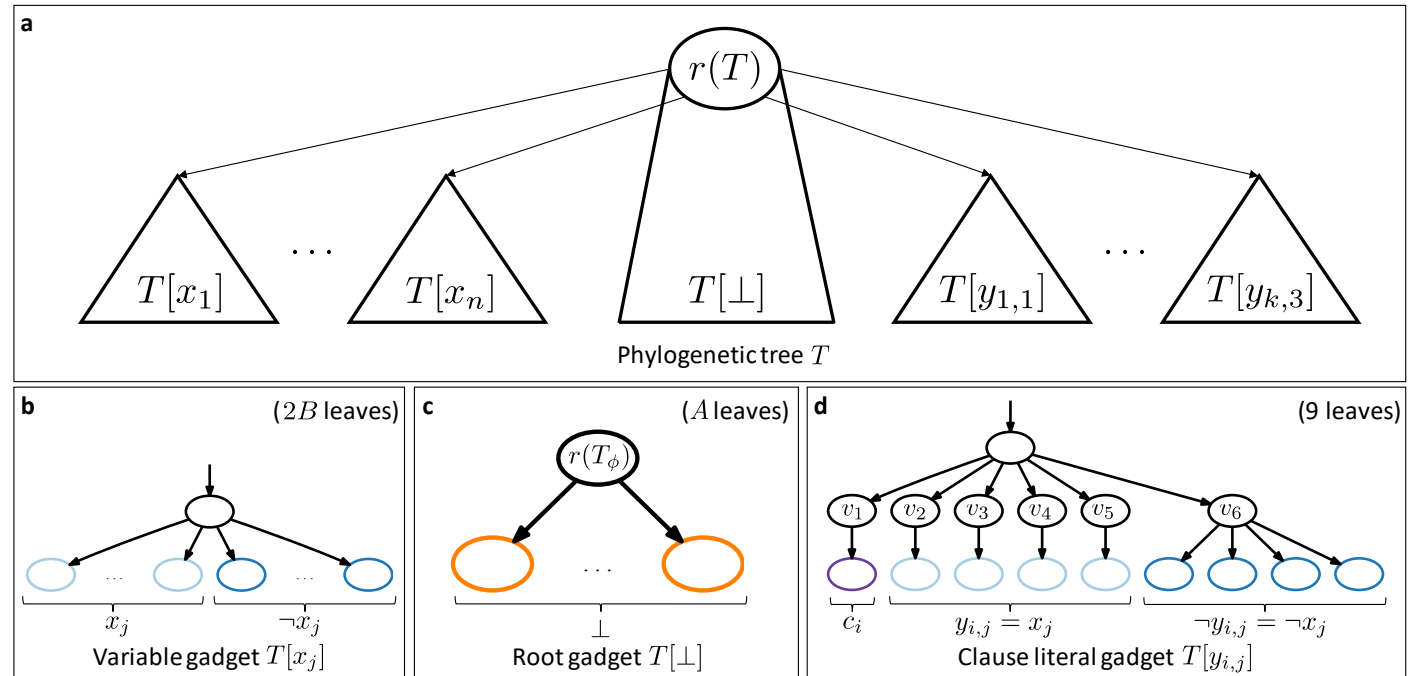
3-SAT: Given $\varphi = \bigwedge_{i=1}^k (y_{i,1} \vee y_{i,2} \vee y_{i,3})$ with variables $\{x_1, \dots, x_n\}$ and k clauses, find $\phi : [n] \rightarrow \{0,1\}$ satisfying φ



$$\Sigma = \{x_1, \dots, x_n, \neg x_1, \dots, \neg x_n, c_1, \dots, c_k, \perp\}$$

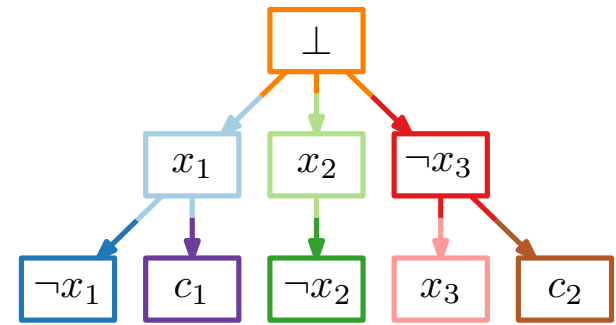
Three ideas:

1. Ensure that $(x, \neg x) \in E(G)$ or $(\neg x, x) \in E(G)$
2. Ensure that $\ell^*(r(T)) = \perp$
3. Ensure that φ is satisfiable if and only if ℓ^* encodes a satisfying truth assignment



PMH is NP-hard when $\mathcal{P} = \{S\}$

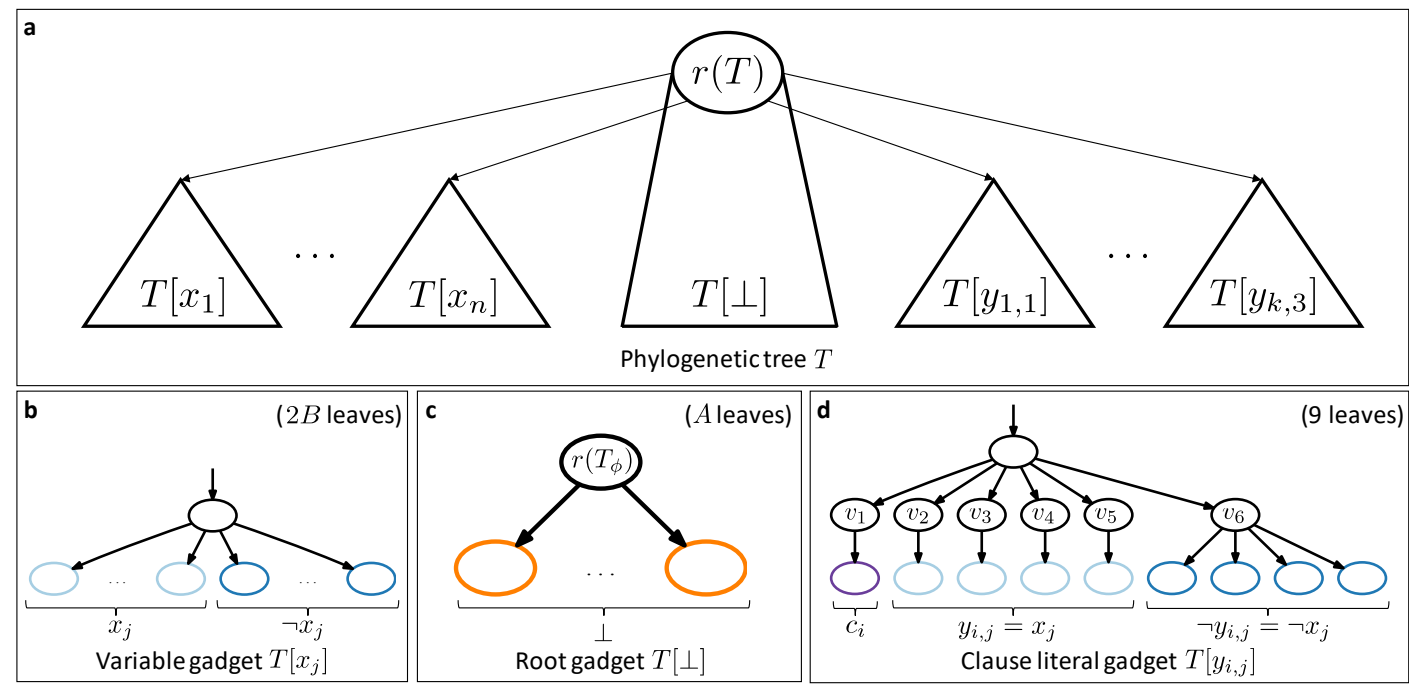
3-SAT: Given $\varphi = \bigwedge_{i=1}^k (y_{i,1} \vee y_{i,2} \vee y_{i,3})$ with variables $\{x_1, \dots, x_n\}$ and k clauses, find $\phi : [n] \rightarrow \{0,1\}$ satisfying φ



$$\Sigma = \{x_1, \dots, x_n, \neg x_1, \dots, \neg x_n, c_1, \dots, c_k, \perp\}$$

Three ideas:

1. Ensure that $(x, \neg x) \in E(G)$ or $(\neg x, x) \in E(G)$
2. Ensure that $\ell^*(r(T)) = \perp$
3. Ensure that φ is satisfiable if and only if ℓ^* encodes a satisfying truth assignment



Lemma: Let $B > 10k + 1$ and $A > 2Bn + 27k$.
 Then, φ is satisfiable if and only if $\mu^*(T) = (B + 1)n + 25k$

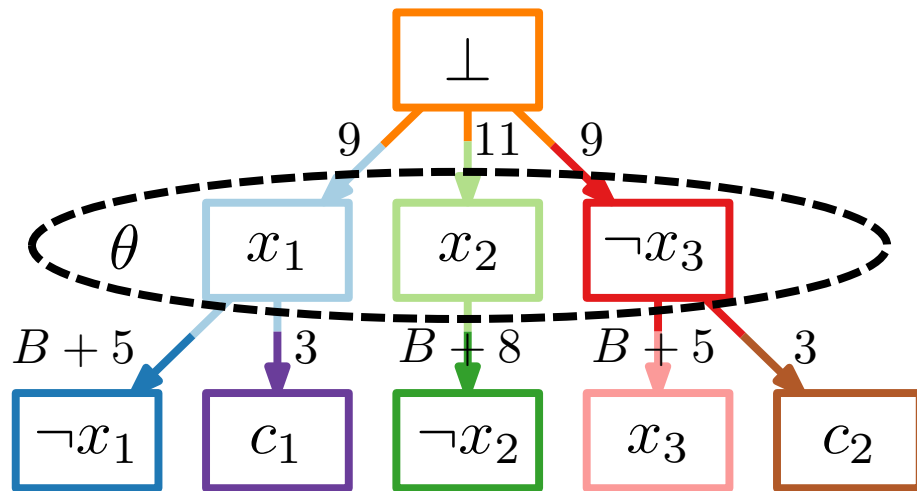
PMH is NP-hard when $\mathcal{P} = \{S\}$

$$\varphi = (x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_1, \neg x_2, \neg x_3)$$

$$k = 2, n = 3$$

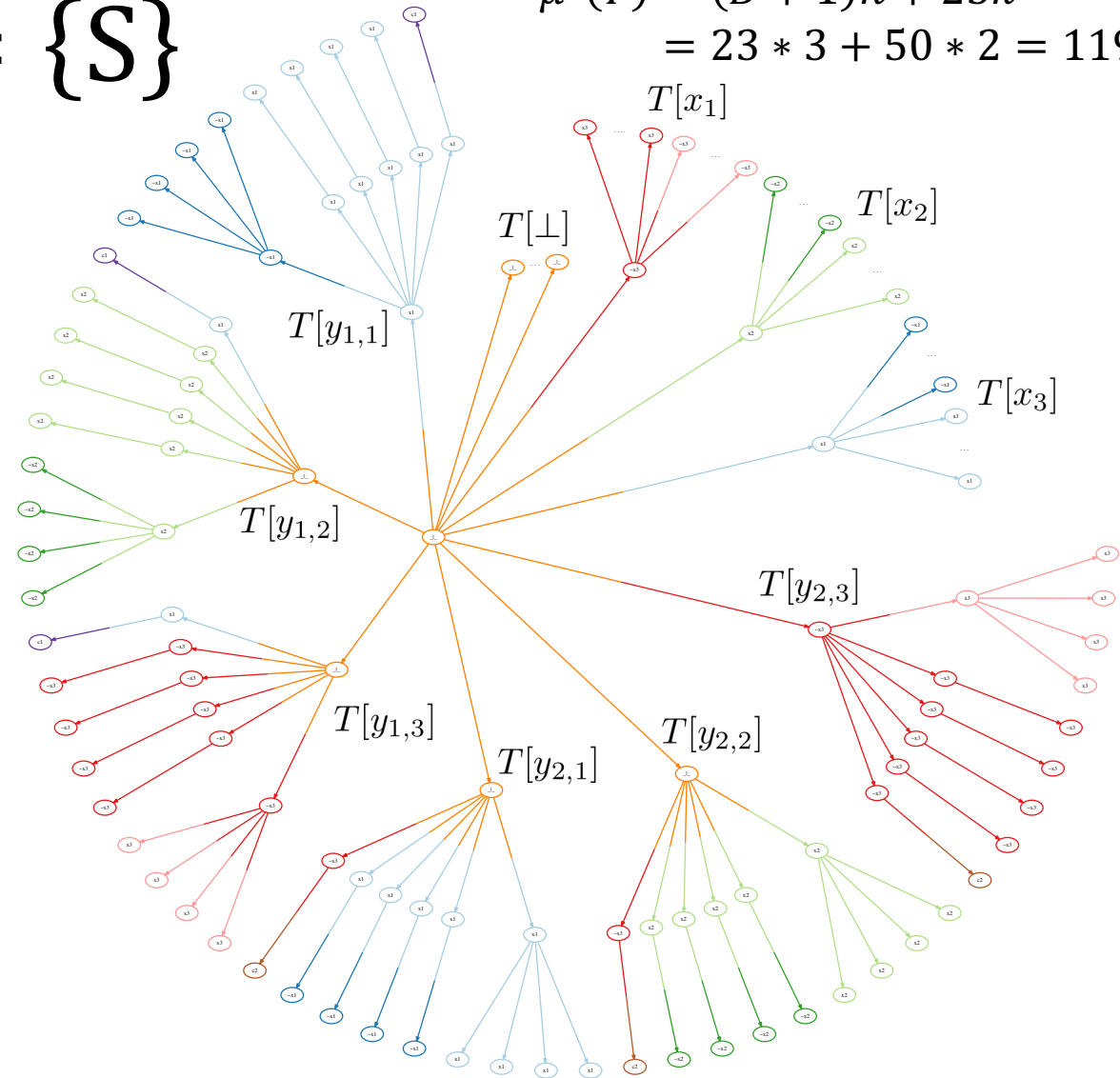
$$B = 10k + 2 = 22$$

$$A = 2Bn + 27k + 1 = 187$$



$$\Sigma = \{x_1, x_2, x_3, \neg x_1, \neg x_2, \neg x_3, c_1, c_2, \perp\}$$

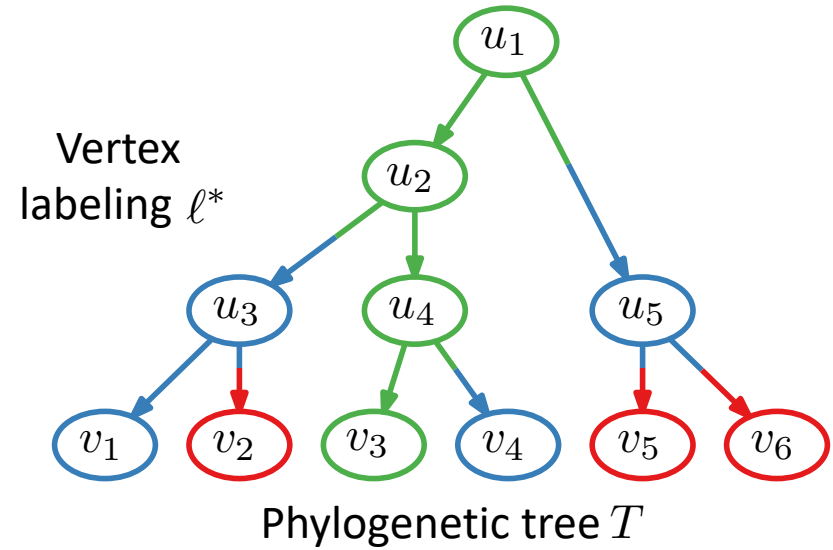
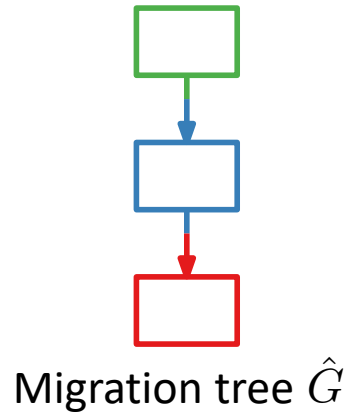
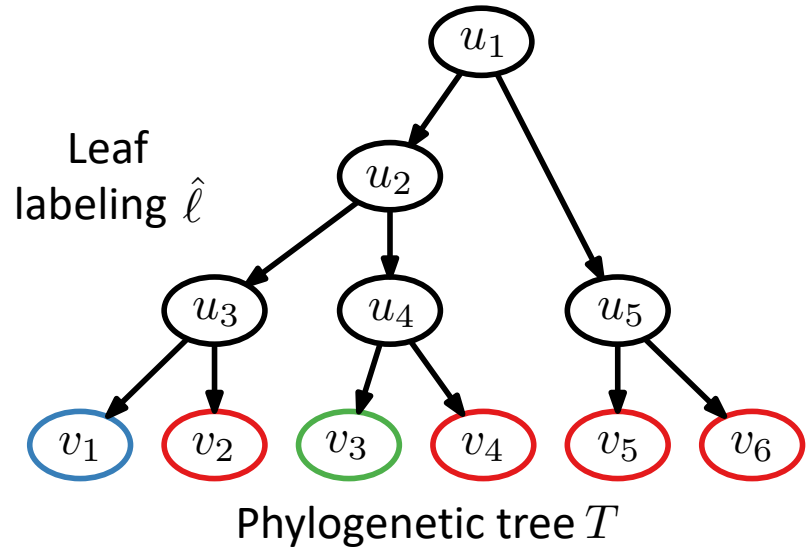
$$\begin{aligned} \mu^*(T) &= (B + 1)n + 25k \\ &= 23 * 3 + 50 * 2 = 119 \end{aligned}$$



Lemma: Let $B > 10k + 1$ and $A > 2Bn + 27k$.

Then, φ is satisfiable if and only if $\mu^*(T) = (B + 1)n + 25k$

PMH is FPT in number m of locations when $\mathcal{P} = \{S\}$



Lemma: If there exists labeling ℓ consistent with \hat{G} then

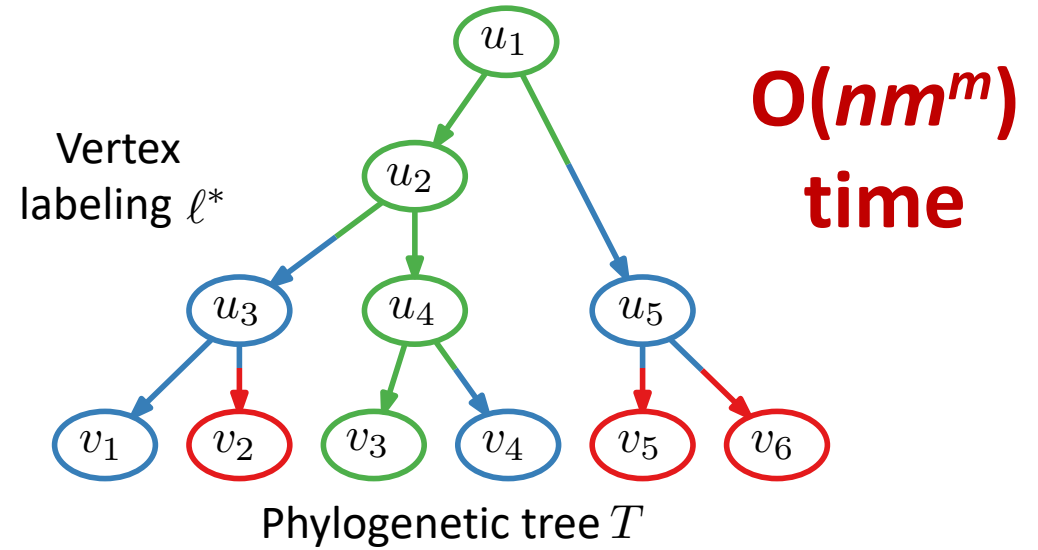
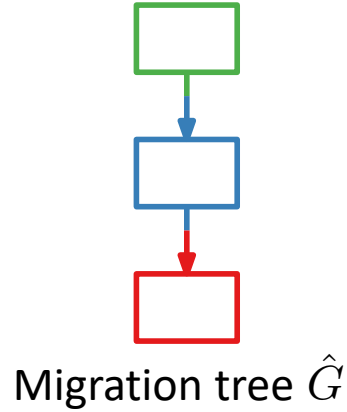
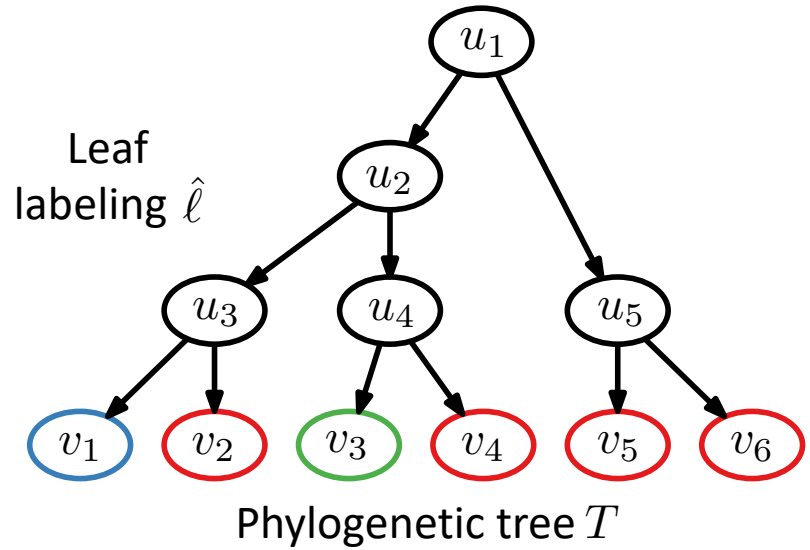
$$d_T(u, v) \geq d_{\hat{G}}(\text{lca}_{\hat{G}}(u), \hat{\ell}(v)) \quad \forall u, v \in V(T) \text{ such that } u \preceq_T v. \quad (1)$$

$$\ell^*(v) = \begin{cases} \text{LCA}_{\hat{G}}(r(T)), & \text{if } v = r(T), \\ \sigma(\ell^*(\pi(v)), \text{LCA}_{\hat{G}}(v)), & \text{if } v \neq r(T), \end{cases}$$

where $\sigma(s, t) = s$ if $s = t$ and otherwise $\sigma(s, t)$ is the unique child of s that lies on the path from s to t in \hat{G} .

Lemma: If (1) holds then ℓ^* is a minimum migration labeling consistent with \hat{G} .

PMH is FPT in number m of locations when $\mathcal{P} = \{S\}$



Lemma: If there exists labeling ℓ consistent with \hat{G} then

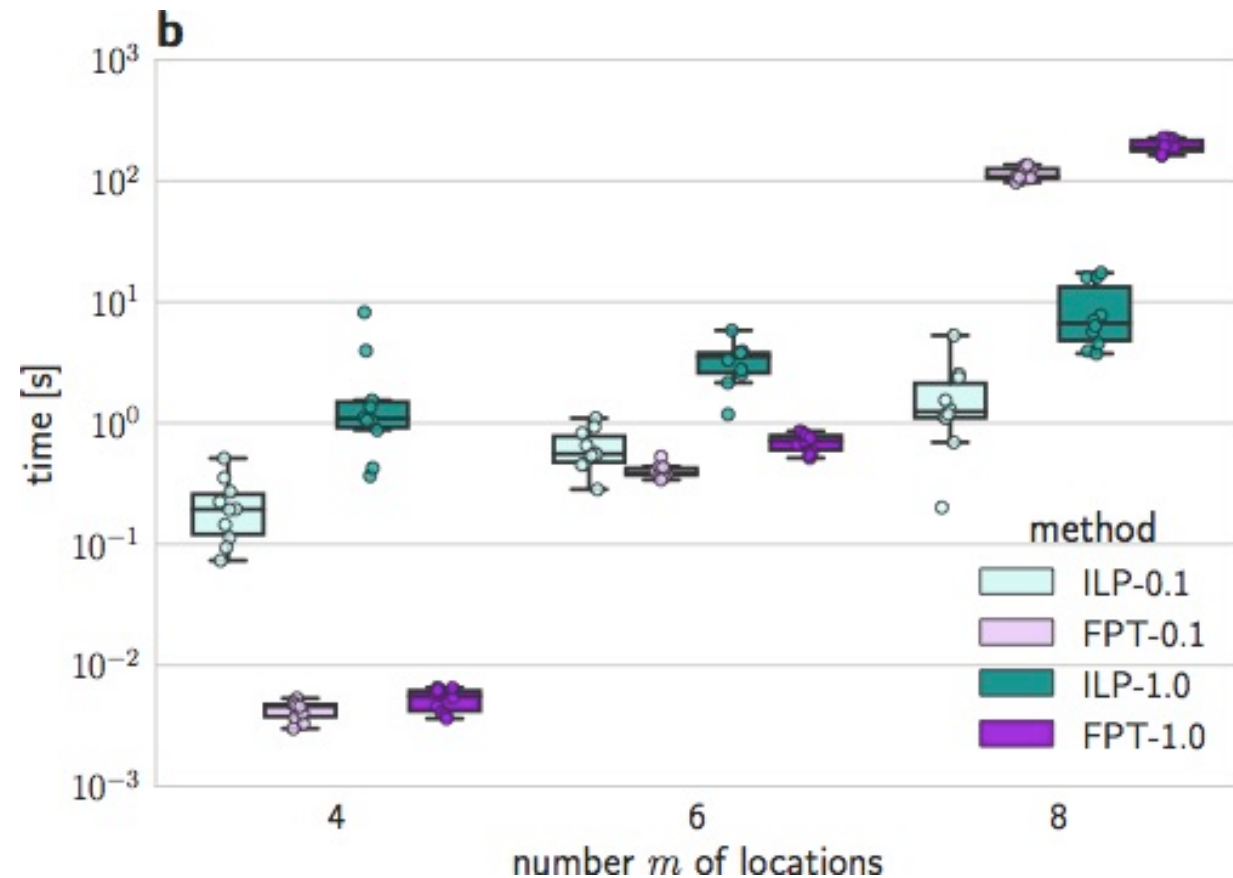
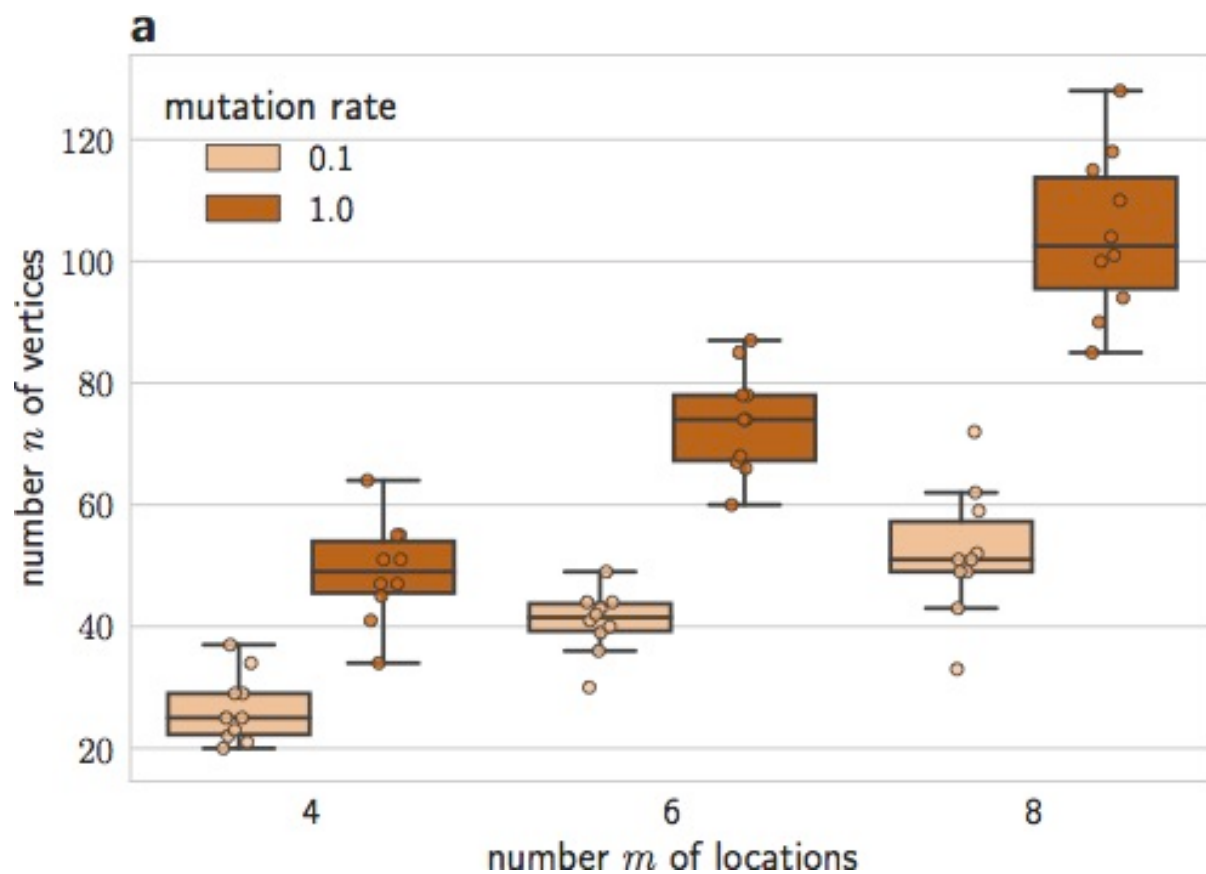
$$d_T(u, v) \geq d_{\hat{G}}(\text{lca}_{\hat{G}}(u), \hat{\ell}(v)) \quad \forall u, v \in V(T) \text{ such that } u \preceq_T v. \quad (1)$$

$$\ell^*(v) = \begin{cases} \text{LCA}_{\hat{G}}(r(T)), & \text{if } v = r(T), \\ \sigma(\ell^*(\pi(v)), \text{LCA}_{\hat{G}}(v)), & \text{if } v \neq r(T), \end{cases}$$

where $\sigma(s, t) = s$ if $s = t$ and otherwise $\sigma(s, t)$ is the unique child of s that lies on the path from s to t in \hat{G} .

Lemma: If (1) holds then ℓ^* is a minimum migration labeling consistent with \hat{G} .

Simulations



Available on: <https://github.com/elkebir-group/PMH-S>

Outline

- Recap character-based phylogeny
- Application of small phylogeny maximum parsimony problem to cancer
- **Compatibility**

Reading:

- Lecture notes

Maximum Parsimony

Small Maximum Parsimony Phylogeny Problem:

Given $m \times n$ matrix $A = [a_{i,j}]$ and tree T with m leaves, find assignment of character states to each internal vertex of T with minimum parsimony score.

Large Maximum Parsimony Phylogeny Problem:

Given $m \times n$ matrix $A = [a_{i,j}]$, find a tree T with m leaves labeled according to A and an assignment of character states to each internal vertex of T with minimum parsimony score.

Binary Characters

		Characters				
		1	2	3	4	5
Species	A	0	1	1	0	0
	B	0	0	1	1	0
	C	1	1	1	1	0
	D	1	1	0	1	1

Characters only have
two possible states

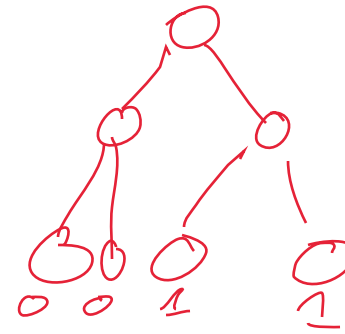
Possible Encoding:
0 : not-mutated
1 : mutated

Possible Encoding:
0 : no wings
1 : wings

Binary Characters

C
 Characters

	1	2	3	4	5
A	0	1	1	0	0
B	0	0	1	1	0
C	1	1	1	1	0
D	1	1	0	1	1



Consider a character c
 (1) Partitioning [in] taxa

leaves
 state 1

state 0

Characters only have two possible states

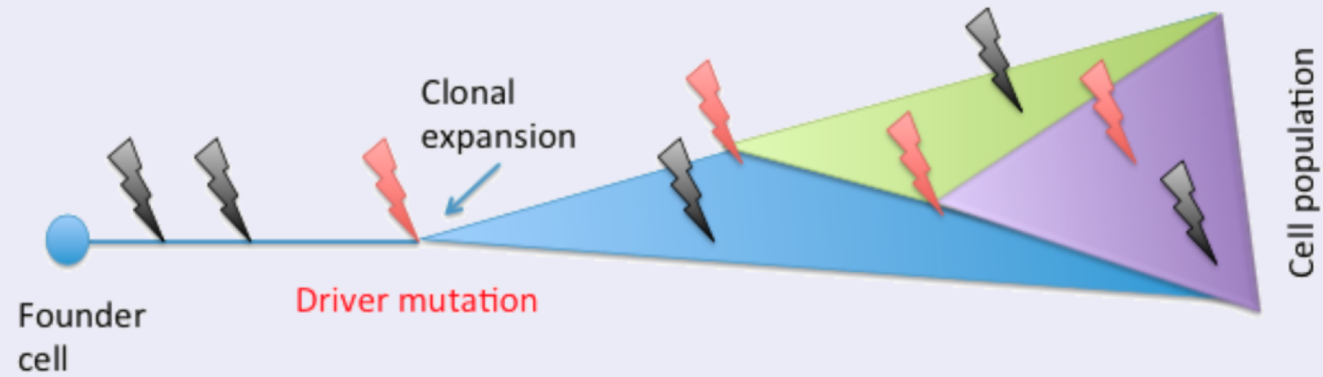
Possible Encoding:
 0 : not-mutated
 1 : mutated

Possible Encoding:
 0 : no wings
 1 : wings

Question: Given n binary characters, what is the smallest parsimony score?

Somatic Mutations and Cancer

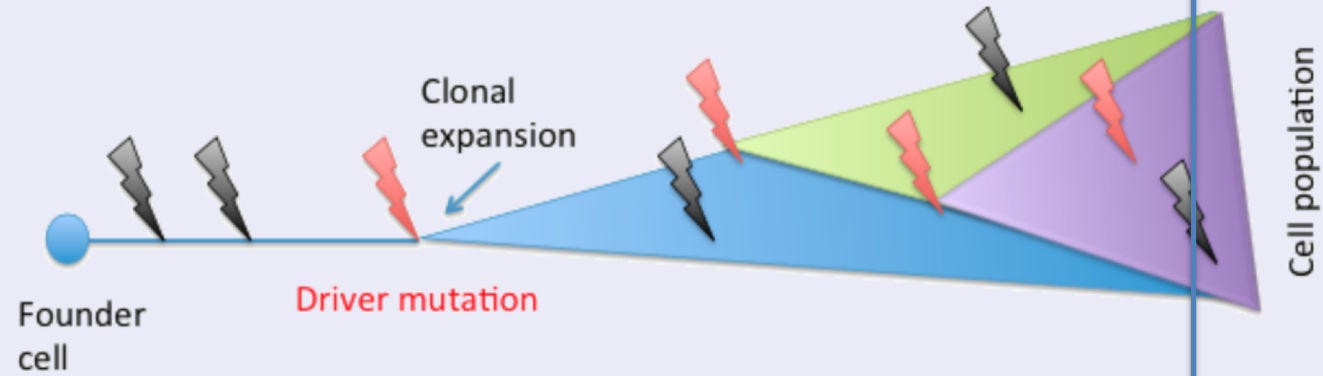
Clonal theory of cancer (Nowell, 1976)



“typical tumor”:
~10 driver mutations
100’s – 1000’s of passenger mutations

Somatic Mutations and Cancer

Clonal theory of cancer (Nowell, 1976)



“typical tumor”: ~10 driver mutations
100’s – 1000’s of passenger mutations

THE CANCER GENOME ATLAS 

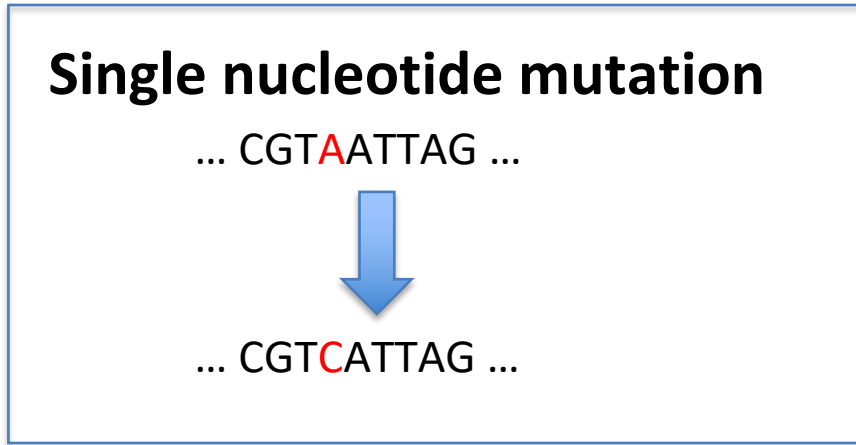


International
Cancer Genome
Consortium

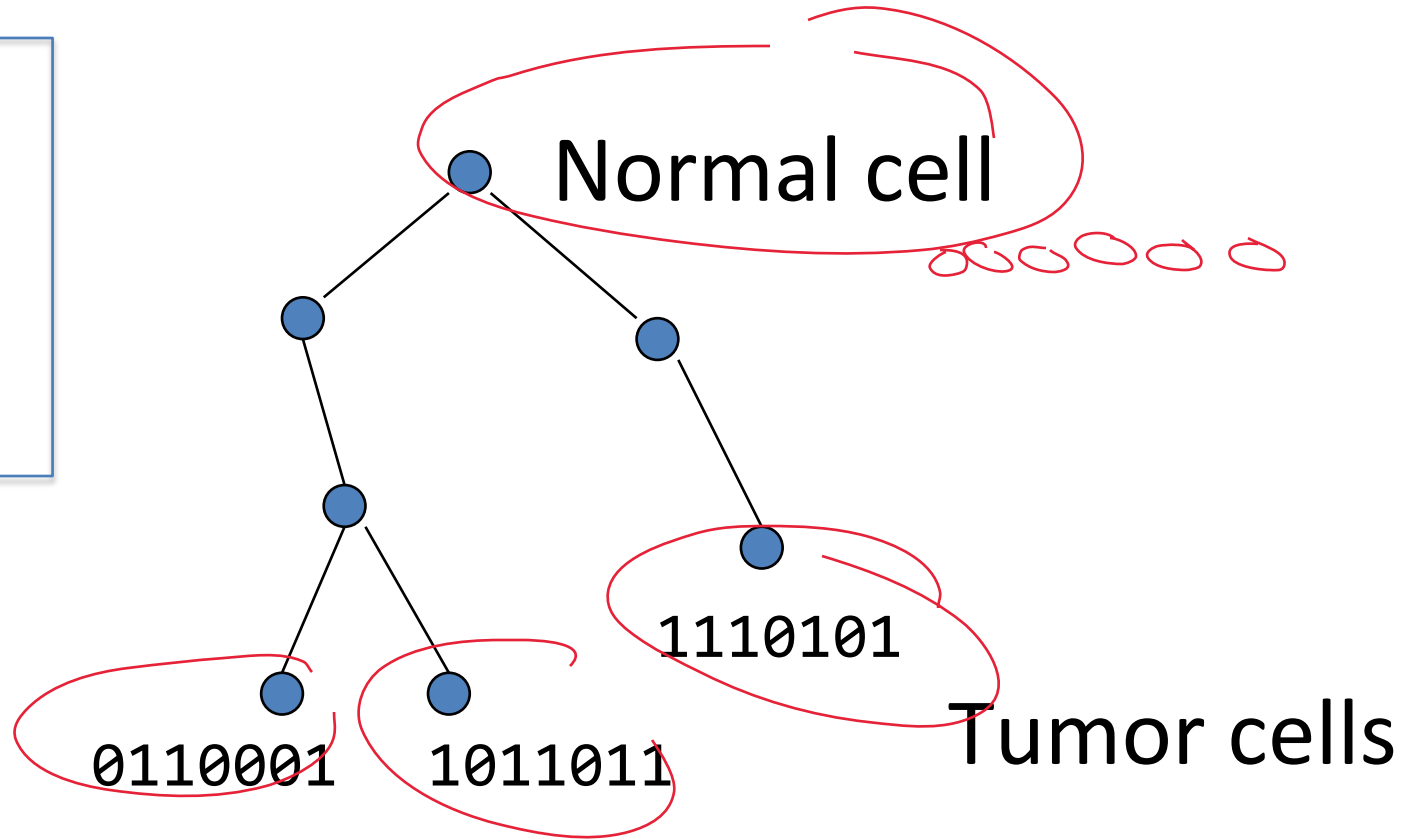


Sequence genome

Progression of Somatic Mutations

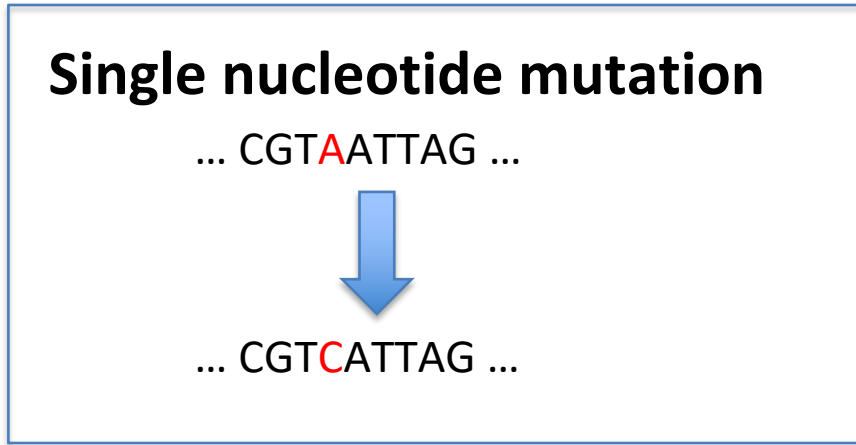


0 = normal
1 = mutated

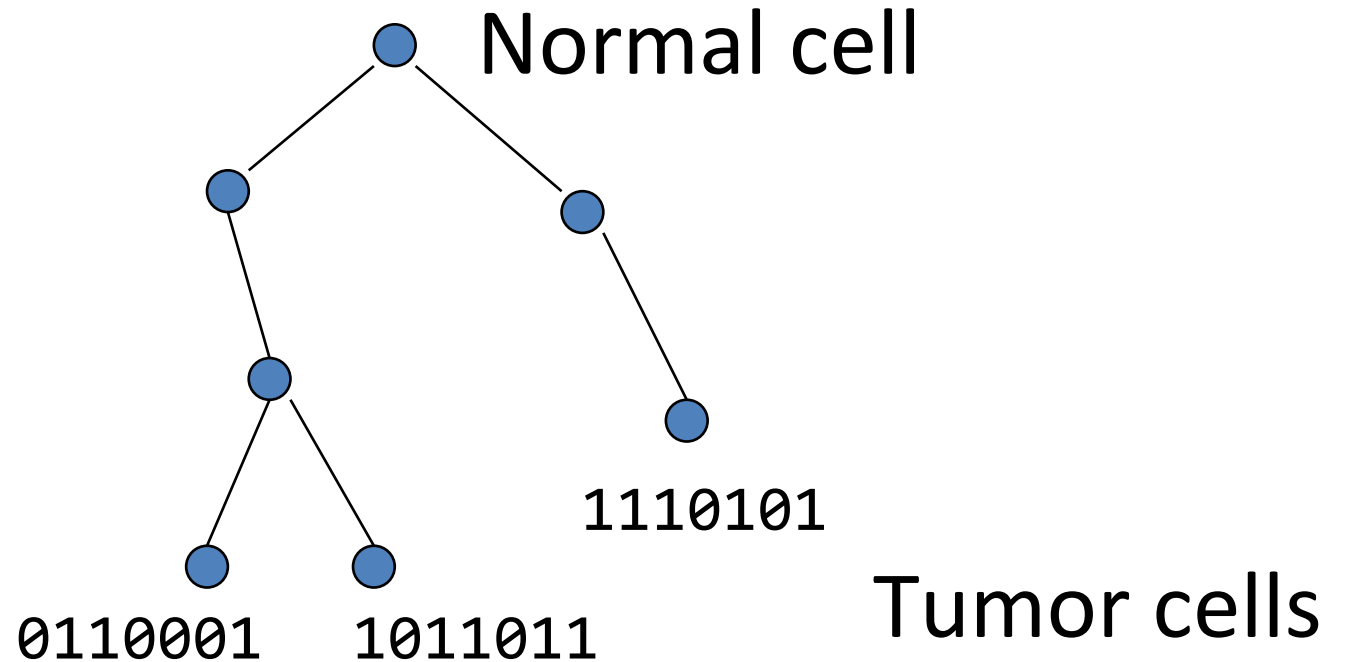


Root is the normal, founder cell and leaves are cells in tumor.

Progression of Somatic Mutations



0 = normal
1 = mutated

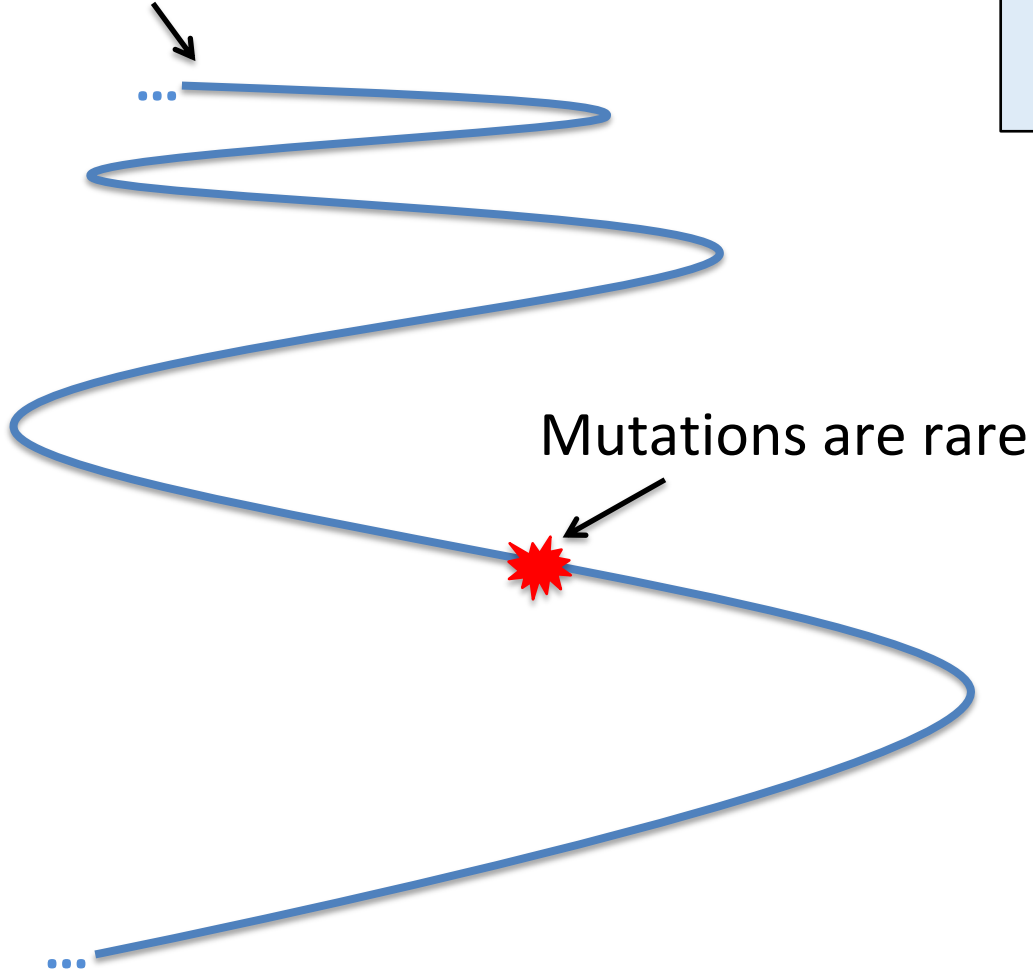


Root is the normal, founder cell and leaves are cells in tumor.

Infinite sites assumption: each locus mutates only once.

Infinite Sites Model

The genome is large



[Kimura, 1969]

Infinite sites model: multiple mutations never occur at the same position

Mutated Loci

Species (cancer cells)	Red	Blue	Green	Purple	Orange	Yellow
A	0	0	0	0	1	1
B	0	0	0	1	1	1
C	0	0	1	0	1	0
D	1	0	0	0	0	0
E	1	1	0	0	0	0

1: mutated

0: not

All sites are bi-allelic: mutated or not.

Two-state Perfect Phylogeny

Matrix $M \in \{0, 1\}^{n \times m}$ has n taxa and m characters

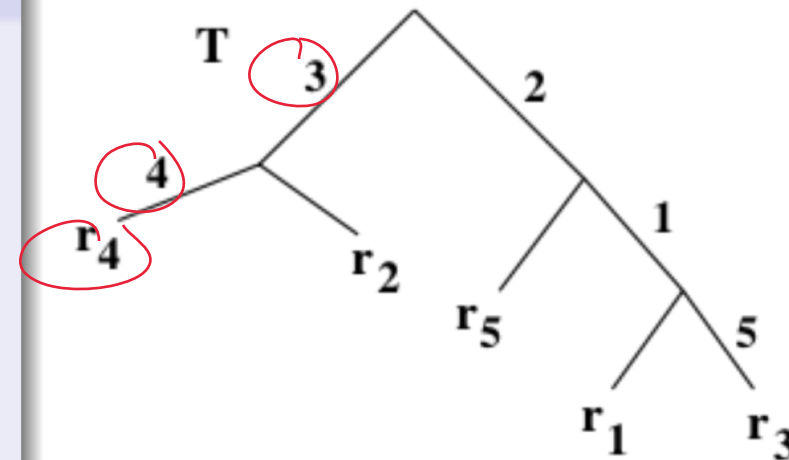
- Taxon f has state 1 for character c
 $\Leftrightarrow f$ possesses character c

	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	0	1
r_4	0	0	1	1	0
r_5	0	1	0	0	0

Definition

A perfect phylogeny for M is a rooted tree T with n leaves such that:

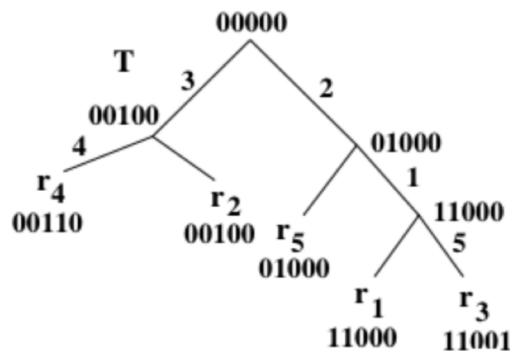
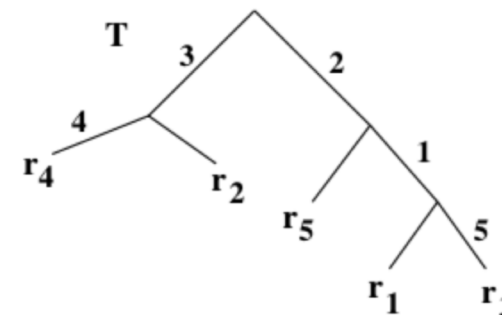
- 1 Each taxon labels only one leaf
- 2 Each character labels only one edge
- 3 Character possessed by a taxon are on unique path to root



Root node is all zero ancestor

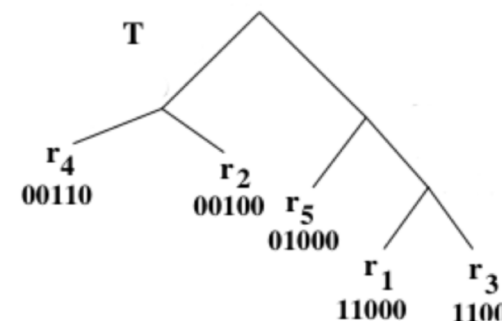
Two-state Perfect Phylogeny – Alternative Definitions

- 1 Each taxon labels exactly one leaf
- 2 Each character labels exactly one edge
- 3 Character possessed by a taxon are on unique path to root



- 1 Each taxon labels exactly one leaf
- 2 Each node is labeled by $\{0, 1\}^m$
- 3 Nodes labeled with state i for character c form a connected subtree

- 1 Each taxon labels exactly one leaf
- 2 $T_c(i)$ is smallest subtree connecting all leaves labeled with state i for character c
- 3 $T_c(0)$ and $T_c(1)$ are disjoint for all c



Two-state Perfect Phylogeny Problem

Input:

Matrix $M \in \{0, 1\}^{n \times m}$ has n **taxa** and m **characters**

- Taxon f has state 1 for character c
 $\Leftrightarrow f$ **possesses** character c

	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	0	1
r_4	0	0	1	1	0
r_5	0	1	0	0	0

Problem

Given $M \in \{0, 1\}^{n \times m}$ does M have a perfect phylogeny?

Try it yourself!

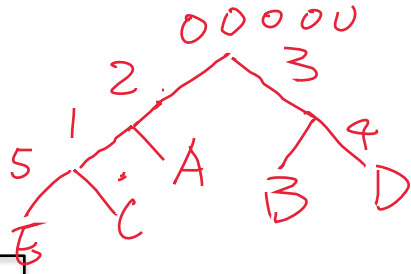
$I(c)$ is the set of taxa possessing c
 $\forall c, d \in \{n\}$:
 ① $I(c) \subseteq I(d)$, or
 ② $I(d) \subseteq I(c)$, or
 ③ $I(c) \cap I(d) = \emptyset$

Only one of these matrices can be used to build a perfect phylogeny.

- (1) As a group, **decide on an approach** to try to determine which one is which.
- (2) Try out your approach to see if you can construct the tree.
- (3) What did you learn from your attempt? (mn)

1 0
0 1
1 1

$I(A) = \{C, E\}$



$M_1 =$
Species

	Characters				
	C_1	C_2	C_3	C_4	C_5
A	0	1	0	0	0
B	0	0	1	0	0
C	1	1	0	0	0
D	0	0	1	1	0
E	1	1	0	0	1

$M_2 =$

	Characters				
	C_1	C_2	C_3	C_4	C_5
A	0	0	1	1	0
B	0	0	1	0	1
C	1	1	0	0	1
D	1	1	0	0	0
E	0	1	0	0	1

The Perfect Phylogeny Problem – Preliminaries

Problem

Given $M \in \{0, 1\}^{n \times m}$ does M have a perfect phylogeny?

Definition

$I(c)$ is the set of taxa that possess character c ; and $\sigma(f)$ is the set of characters possessed by taxon f .

	c_1	c_2	c_3	c_4	c_5		c_1 (2)	c_2 (1)	c_3 (3)	c_4 (5)	c_5 (4)
r_1	1	1	0	0	0	\Rightarrow	1	1	0	0	0
r_2	0	0	1	0	0		0	0	1	0	0
r_3	1	1	0	0	1		1	1	0	1	0
r_4	0	0	1	1	0		0	0	1	0	1
r_5	0	1	0	0	0		1	0	0	0	0

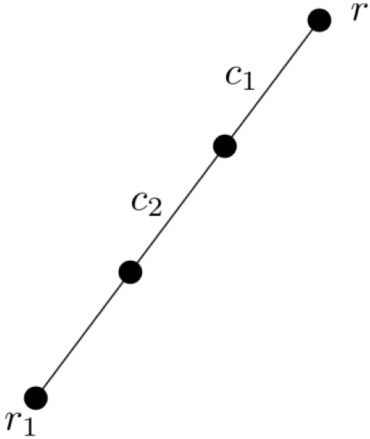
$$I(c_1) = \{r_1, r_3\}$$
$$\sigma(r_1) = \{c_1, c_2\}$$

Sort columns of M s.t. $c < d$ iff $|I(c)| \geq |I(d)|$. Break ties arbitrarily.

- Consider rows of M iteratively
 - ▶ T_i is tree of first i rows of M
- T_1 is a path graph
 - ▶ Terminal nodes r and 1
 - ▶ $|\sigma(1)| + 1$ edges labeled by $\sigma(1)$

$$c < d \text{ iff } |I(c)| \geq |I(d)|$$

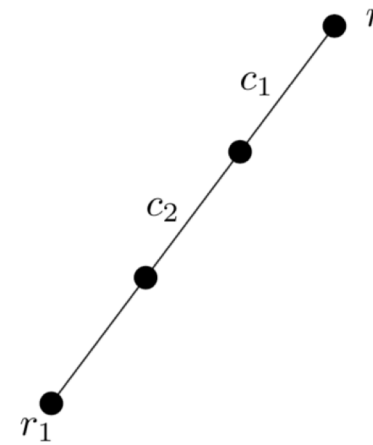
	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	1	0
r_4	0	0	1	0	1
r_5	1	0	0	0	0



$$c < d \text{ iff } |I(c)| \geq |I(d)|$$

- Consider rows of M iteratively
 - ▶ T_i is tree of first i rows of M
- T_1 is a path graph
 - ▶ Terminal nodes r and 1
 - ▶ $|\sigma(1)| + 1$ edges labeled by $\sigma(1)$
- T_{i+1} is a supertree of T_i
 - ▶ Let v be last node on walk from r matching characters $\sigma(i + 1)$
 - ★ Character d is the last match
 - ★ Unmatched characters $\tau(i + 1)$

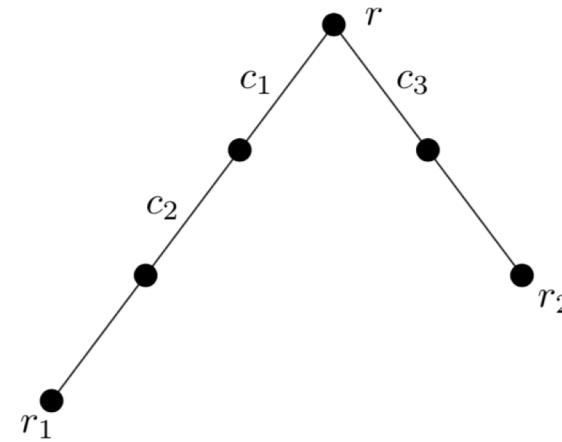
	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	1	0
r_4	0	0	1	0	1
r_5	1	0	0	0	0



- Consider rows of M iteratively
 - ▶ T_i is tree of first i rows of M
- T_1 is a path graph
 - ▶ Terminal nodes r and 1
 - ▶ $|\sigma(1)| + 1$ edges labeled by $\sigma(1)$
- T_{i+1} is a supertree of T_i
 - ▶ Let v be last node on walk from r matching characters $\sigma(i + 1)$
 - ★ Character d is the last match
 - ★ Unmatched characters $\tau(i + 1)$
 - ▶ Extend T_i with path Π
 - ★ Π has terminals v and $i + 1$
 - ★ Π has $|\tau(i + 1)| + 1$ edges labeled by $\tau(i + 1)$

$$c < d \text{ iff } |I(c)| \geq |I(d)|$$

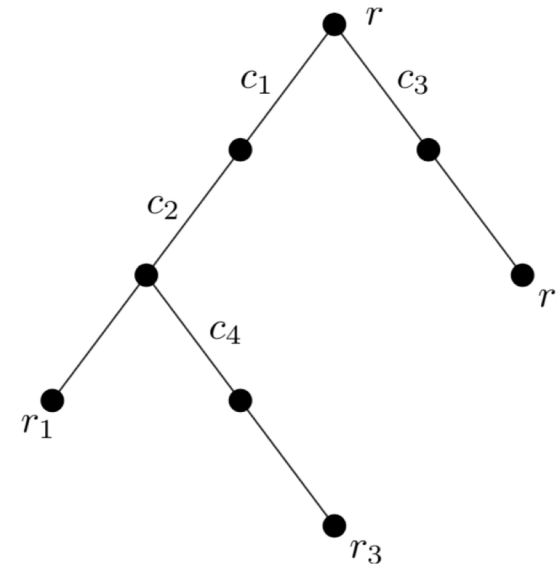
	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	1	0
r_4	0	0	1	0	1
r_5	1	0	0	0	0



$$c < d \text{ iff } |I(c)| \geq |I(d)|$$

- Consider rows of M iteratively
 - ▶ T_i is tree of first i rows of M
- T_1 is a path graph
 - ▶ Terminal nodes r and 1
 - ▶ $|\sigma(1)| + 1$ edges labeled by $\sigma(1)$
- T_{i+1} is a supertree of T_i
 - ▶ Let v be last node on walk from r matching characters $\sigma(i + 1)$
 - ★ Character d is the last match
 - ★ Unmatched characters $\tau(i + 1)$
 - ▶ Extend T_i with path Π
 - ★ Π has terminals v and $i + 1$
 - ★ Π has $|\tau(i + 1)| + 1$ edges labeled by $\tau(i + 1)$

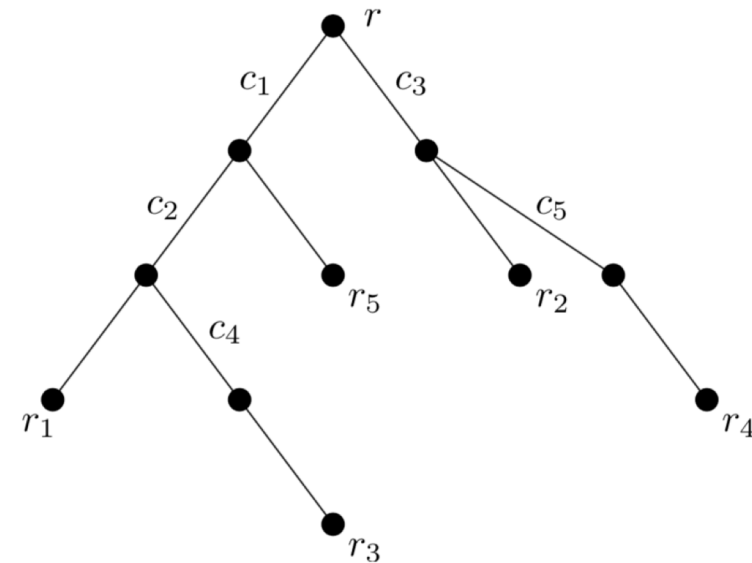
	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	1	0
r_4	0	0	1	0	1
r_5	1	0	0	0	0



- Consider rows of M iteratively
 - ▶ T_i is tree of first i rows of M
- T_1 is a path graph
 - ▶ Terminal nodes r and 1
 - ▶ $|\sigma(1)| + 1$ edges labeled by $\sigma(1)$
- T_{i+1} is a supertree of T_i
 - ▶ Let v be last node on walk from r matching characters $\sigma(i + 1)$
 - ★ Character d is the last match
 - ★ Unmatched characters $\tau(i + 1)$
 - ▶ Extend T_i with path Π
 - ★ Π has terminals v and $i + 1$
 - ★ Π has $|\tau(i + 1)| + 1$ edges labeled by $\tau(i + 1)$

$$c < d \text{ iff } |I(c)| \geq |I(d)|$$

	c_1	c_2	c_3	c_4	c_5
r_1	1	1	0	0	0
r_2	0	0	1	0	0
r_3	1	1	0	1	0
r_4	0	0	1	0	1
r_5	1	0	0	0	0



Lemma

Let $M_i \in \{0, 1\}^{i \times m}$ be a submatrix of M . If M is conflict-free then T_i is a perfect phylogeny for M_i .

Outline

- Recap character-based phylogeny
- Application of small phylogeny maximum parsimony problem to cancer
- Compatibility

Reading:

- Lecture notes

HW 3 due Oct 29 by 11:59pm