CS 598MEB
Introduction to Bioinformatics

Lecture 6

Mohammed El-Kebir

February 6, 2020

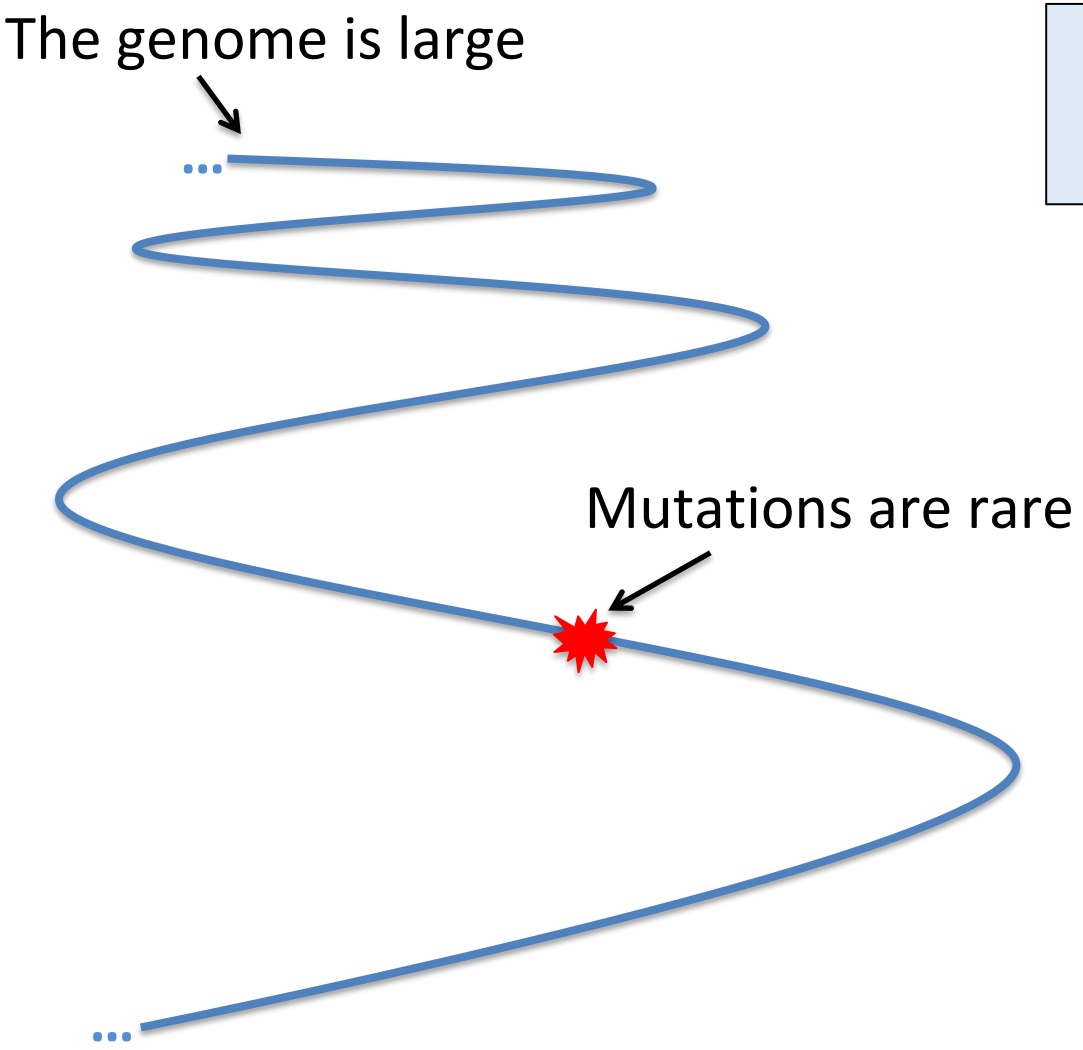
Outline

- Multi-State Perfect Phylogeny
- Tumor Phylogeny Inference from Single-cell DNA-seq

Reading:

- Lecture notes
- M. El-Kebir. SPhyR: Tumor Phylogeny Estimation from Single-Cell Sequencing Data under Loss and Error. Bioinformatics (ECCB 2018), 34(17):i671-679, 2018.

Infinite Sites Model = Two-state Perfect Phylogeny



[Kimura, 1969]

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

Mutated Loci

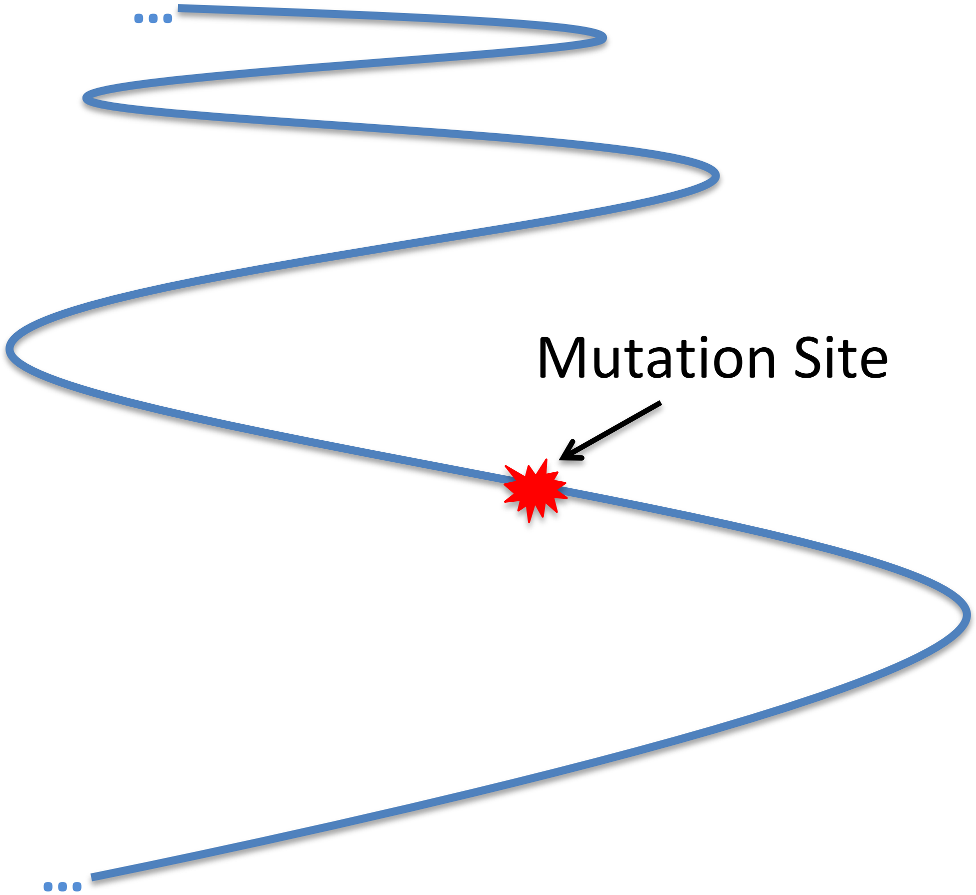
Species (cancer cells)

	✸	✸	✸	✸	✸	✸
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

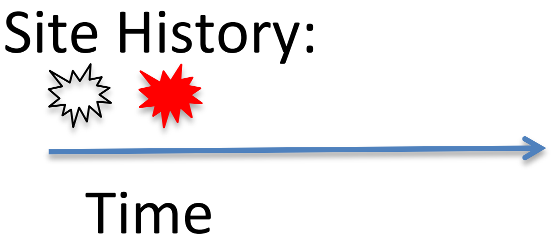
All sites are bi-allelic: mutated or not.

Infinite Alleles Model = Multi-state Perfect Phylogeny



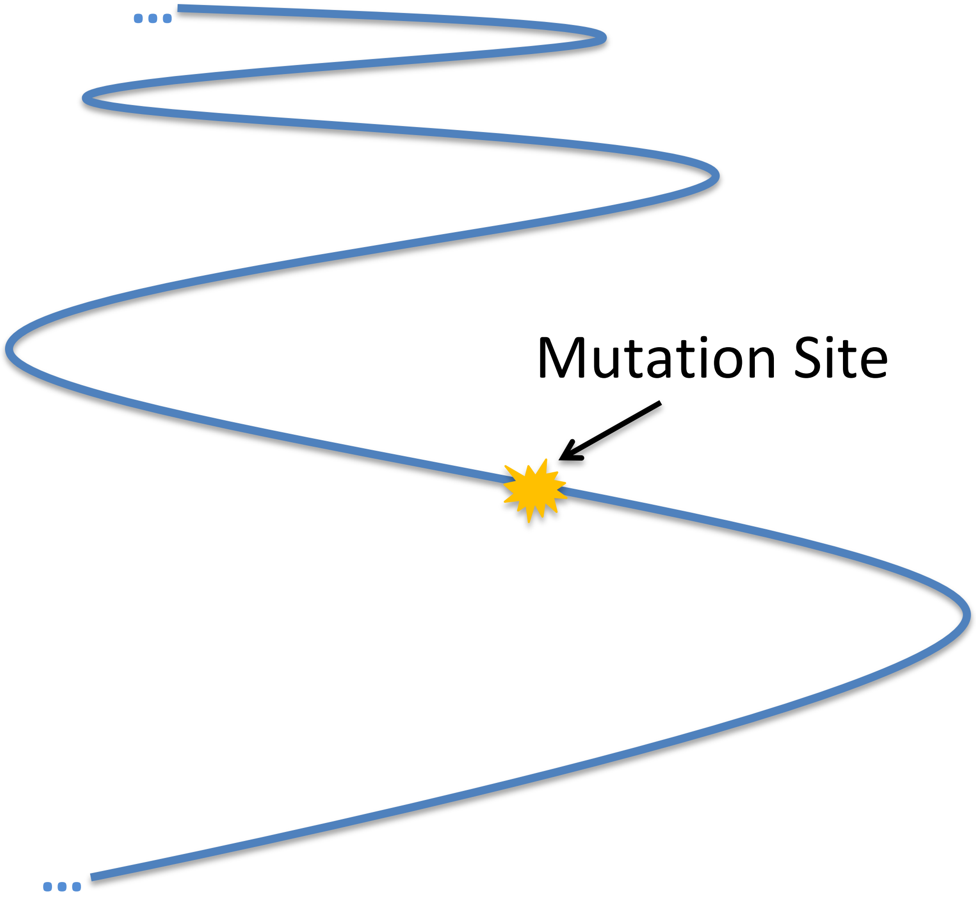
Infinite alleles model:

- For any mutation, there are an infinite number of possibilities of what mutation looks like (states).
- So, the same position can be mutated multiple times, but it never mutates to the same "allele" or state.



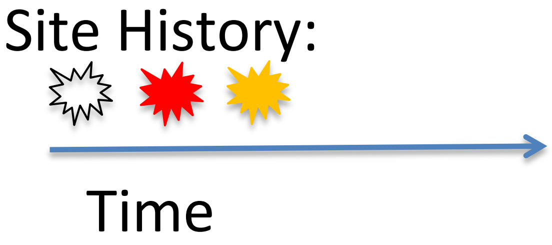
Characters have integer states

Infinite Alleles Model = Multi-state Perfect Phylogeny



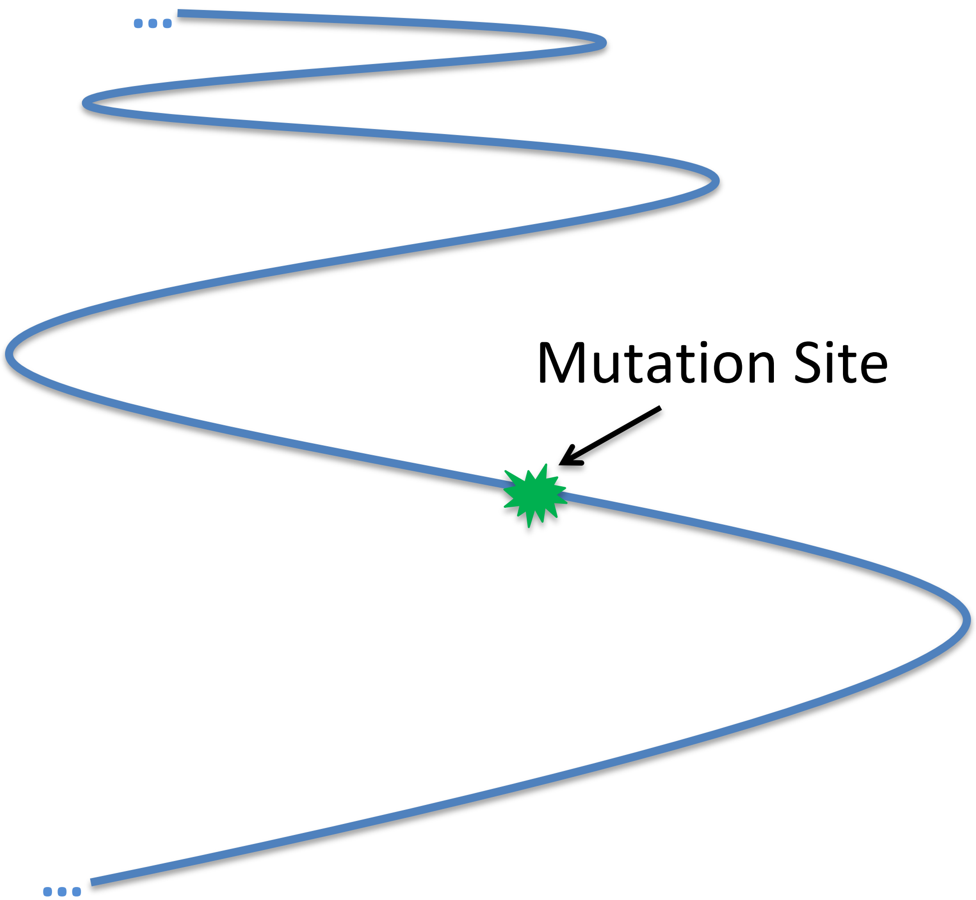
Infinite alleles model:

- For any mutation, there are an infinite number of possibilities of what mutation looks like (states).
- So, the same position can be mutated multiple times, but it never mutates to the same "allele" or state.



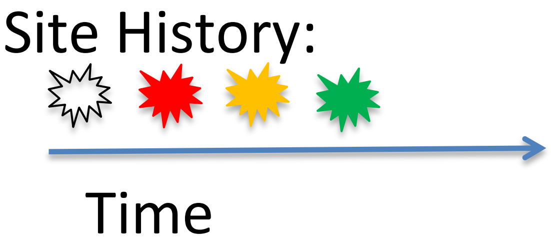
Characters have integer states

Infinite Alleles Model = Multi-state Perfect Phylogeny



Infinite alleles model:

- For any mutation, there are an infinite number of possibilities of what mutation looks like (states).
- So, the same position can be mutated multiple times, but it never mutates to the same "allele" or state.



Characters have integer states

Multi-state Perfect Phylogeny

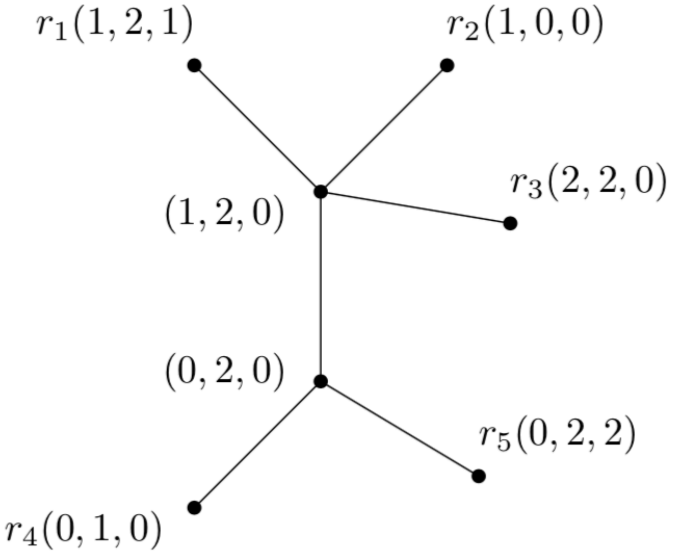
Matrix $M \in \{0, \ldots, k-1\}^{n \times m}$ has n taxa and m characters

	c_1	c_2	c_3
r_1	1	2	1
r_2	1	0	0
r_3	2	2	0
r_4	0	1	0
r_5	0	2	2

Definition

A multi-state perfect phylogeny for M is a tree T with n leaves such that:

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



Theorem (Bodlaender et al., 1992) [Bodlaender, Fellows and Warnow]

For general k, the multi-state perfect phylogeny problem is NP-complete

Cladistic vs. Qualitative Characters

Definition

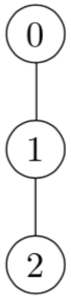
A multi-state perfect phylogeny for M is a tree T with n leaves such that:

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

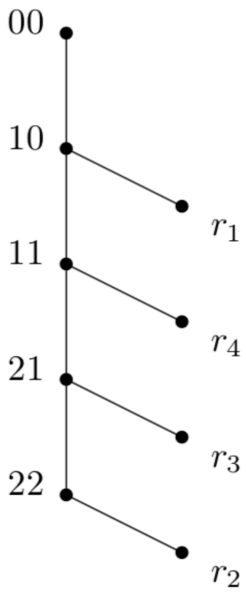
A cladistic character c has a state tree t_c on its states

A phylogeny T is consistent if the reduced tree $\sigma(T, c)$ is identical with t_c for all c

	a	b
r_1	1	0
r_2	2	2
r_3	2	1
r_4	1	1



t_a, t_b



Cladistic vs. Qualitative Characters

Definition

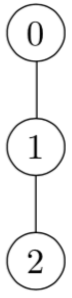
A multi-state perfect phylogeny for M is a tree T with n leaves such that:

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

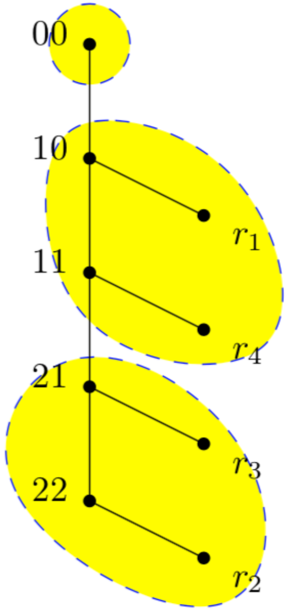
A cladistic character c has a state tree t_c on its states

A phylogeny T is consistent if the reduced tree $\sigma(T, c)$ is identical with t_c for all c

	a	b
r_1	1	0
r_2	2	2
r_3	2	1
r_4	1	1



t_a, t_b



Cladistic vs. Qualitative Characters

Definition

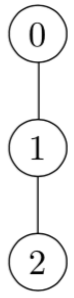
A multi-state perfect phylogeny for M is a tree T with n leaves such that:

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

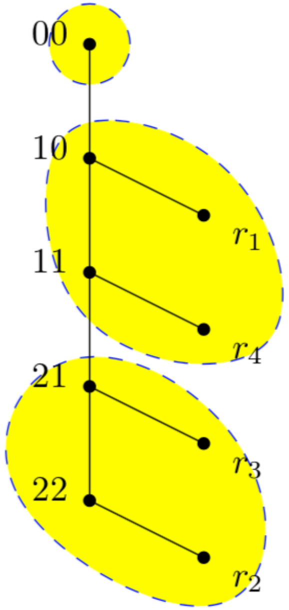
A cladistic character c has a state tree t_c on its states

A phylogeny T is consistent if the reduced tree $\sigma(T, c)$ is identical with t_c for all c

	a	b
r_1	1	0
r_2	2	2
r_3	2	1
r_4	1	1



t_a, t_b



Outline

- Multi-State Perfect Phylogeny
- Tumor Phylogeny Inference from Single-cell DNA-seq

Reading:

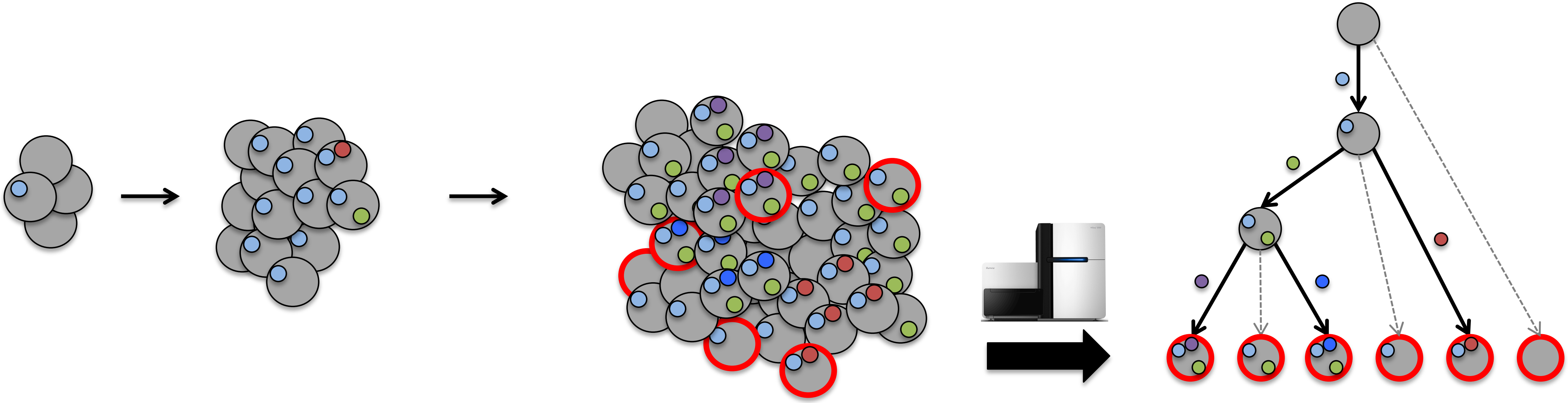
- Lecture notes
- M. El-Kebir. SPhyR: Tumor Phylogeny Estimation from Single-Cell Sequencing Data under Loss and Error. Bioinformatics (ECCB 2018), 34(17):i671-679, 2018.

SPhyR: Tumor Phylogeny Estimation from Single-Cell Sequencing Data under Loss and Error

Mohammed El-Kebir – University of Illinois at Urbana Champaign, Department of Computer Science

ECCB 2018

Single-cell Tumor Phylogeny Inference



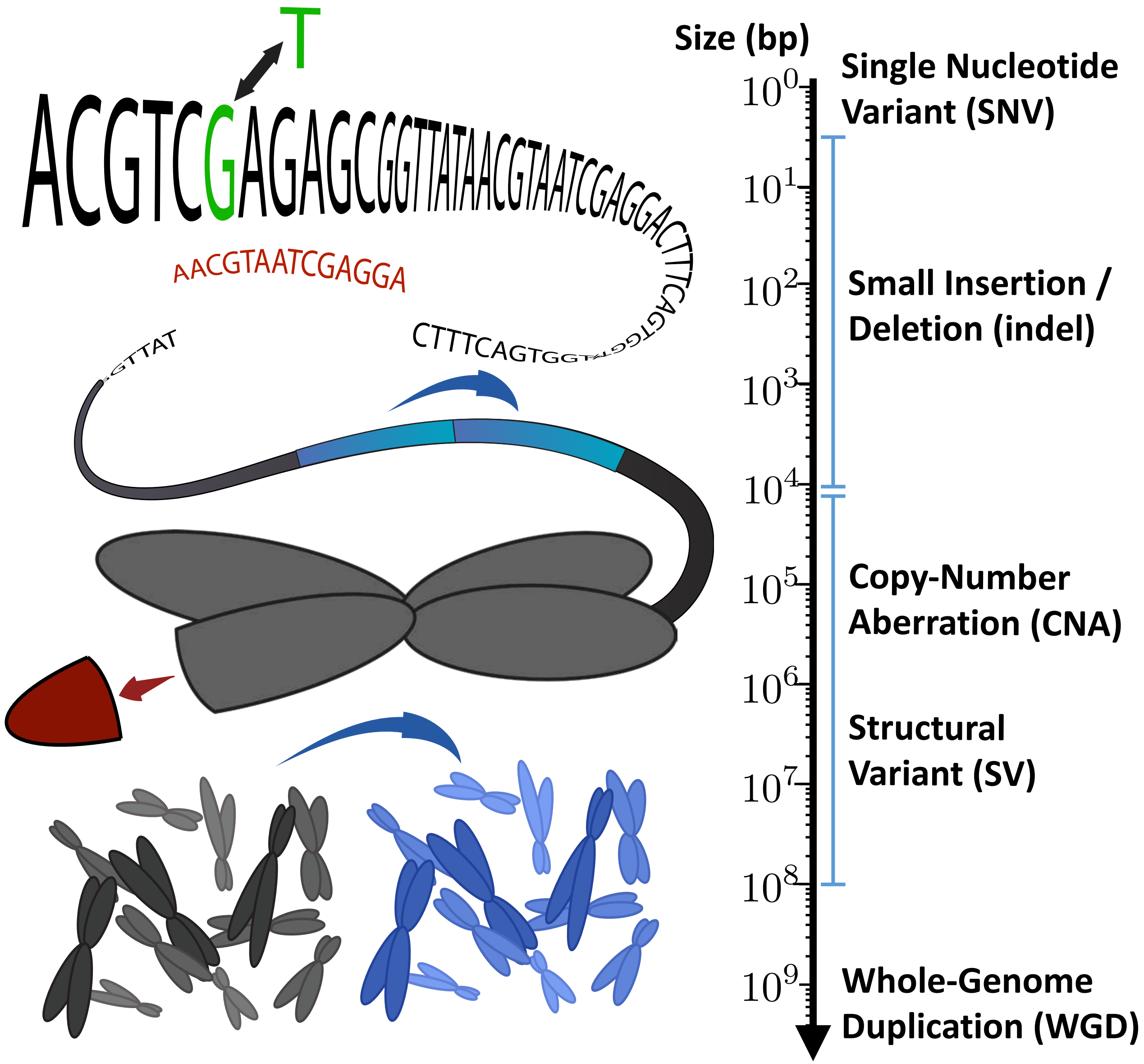
Heterogeneous Tumor

Phylogenetic Tree ***T***

Goal: Given single-cell DNA sequencing data, find phylogenetic tree ***T***

Requirement: Evolutionary model for somatic mutations

Somatic Mutations Occur at Different Genomic Scales



Size (bp)

Single Nucleotide Variant (SNV)

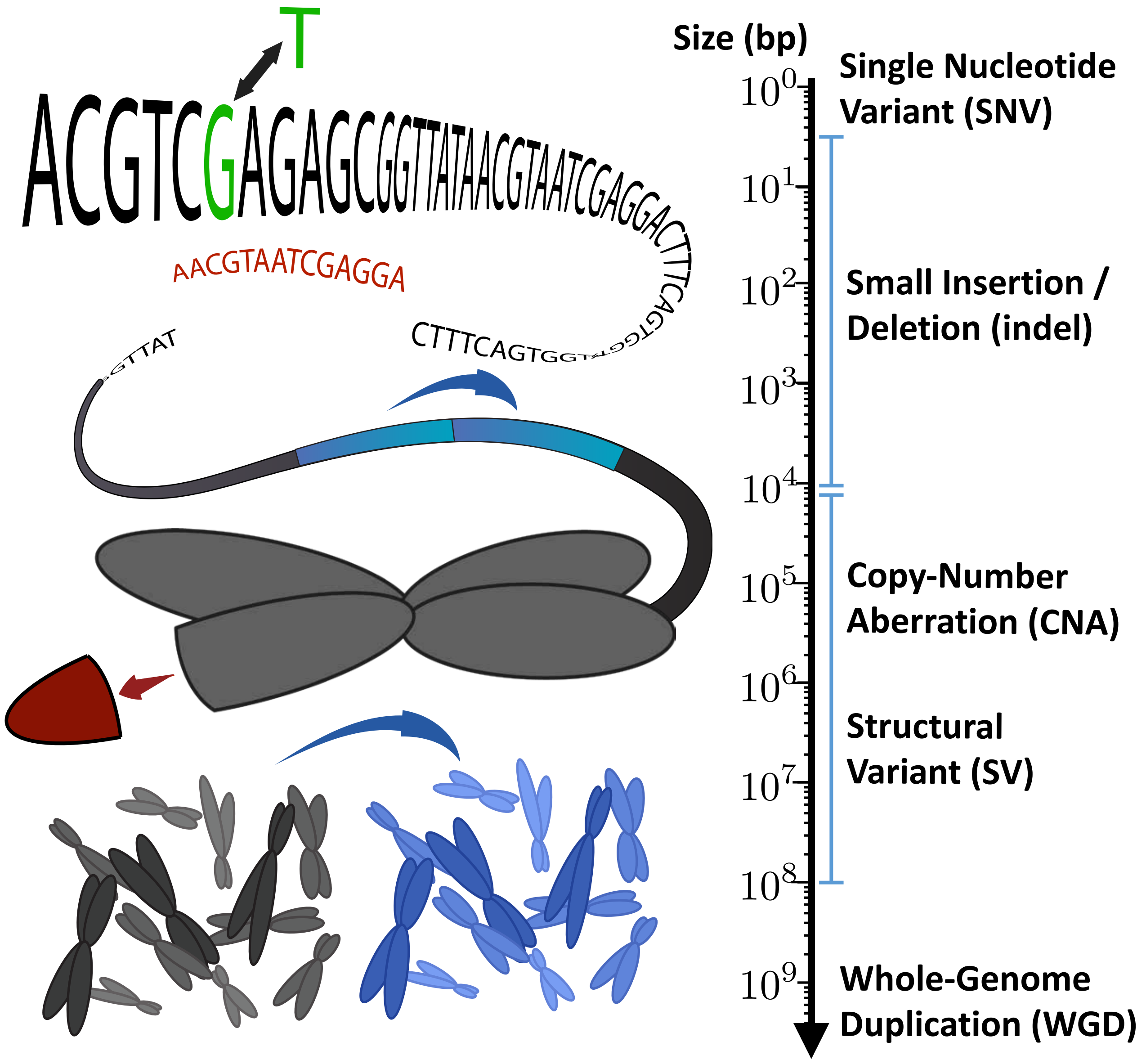
Small Insertion / Deletion (indel)

Copy-Number Aberration (CNA)

Structural Variant (SV)

Whole-Genome Duplication (WGD)

Infinite Sites Assumption is too Restrictive for SNVs



Single Nucleotide Variant (SNV)

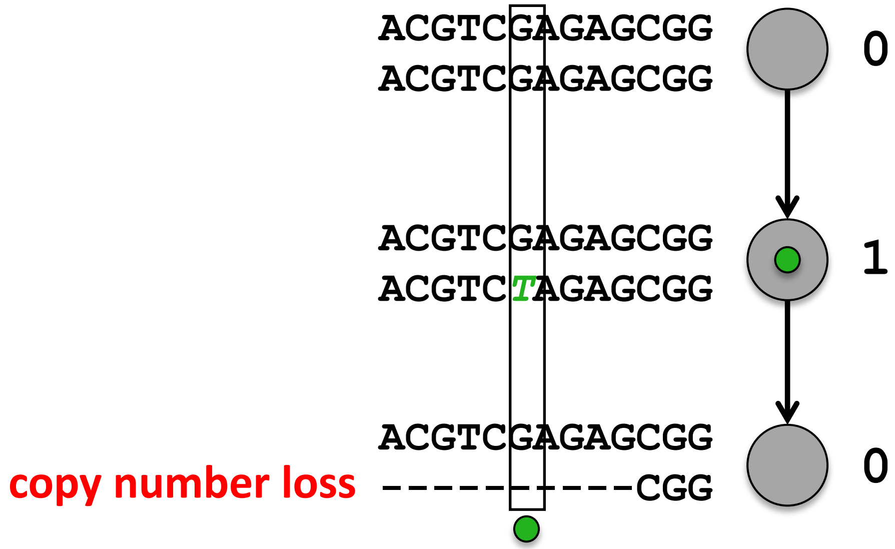
Small Insertion / Deletion (indel)

Copy-Number Aberration (CNA)

Structural Variant (SV)

Whole-Genome Duplication (WGD)

SNVs can be **lost** due to CNAs



```
ACGTCGAGAGCGG    0
ACGTCGAGAGCGG

ACGTCGAGAGCGG    1
ACGTCTAGAGCGG

ACGTCGAGAGCGG    0
----------CGG
```

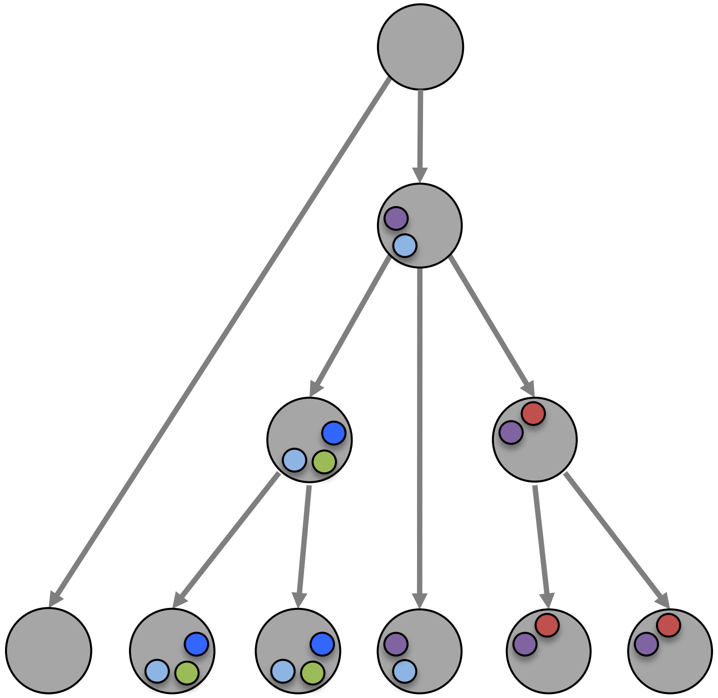
copy number loss

Infinite sites assumption:

- No parallel evolution of SNVs
- No loss of SNVs
- SCITE [Jahn et al. 2016]
- OncoNEM [Ross and Markowetz, 2016]

Outline

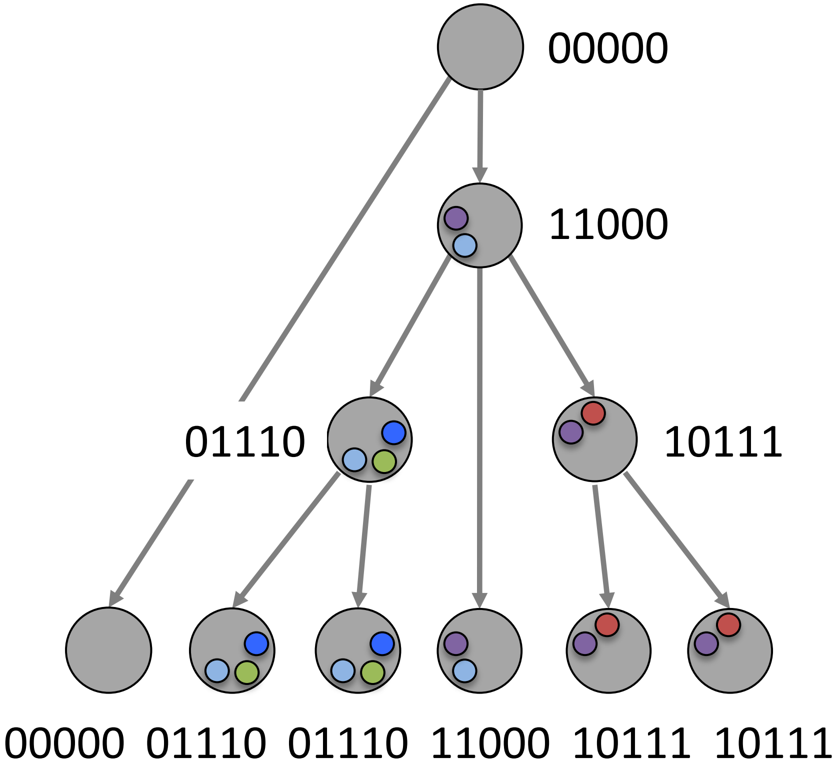
- Perfect data (error free)
    - Problem statement
    - Combinatorial characterization of solutions
    - Exact algorithm
    - Results
- Real data (with errors)
    - Problem statement
    - Heuristic algorithm
    - Results
- Conclusions



k-Dollo Phylogeny (*k*-DP) Problem

Definition 1. A *k-Dollo phylogeny* T is a rooted, node-labeled tree subject to the following conditions.

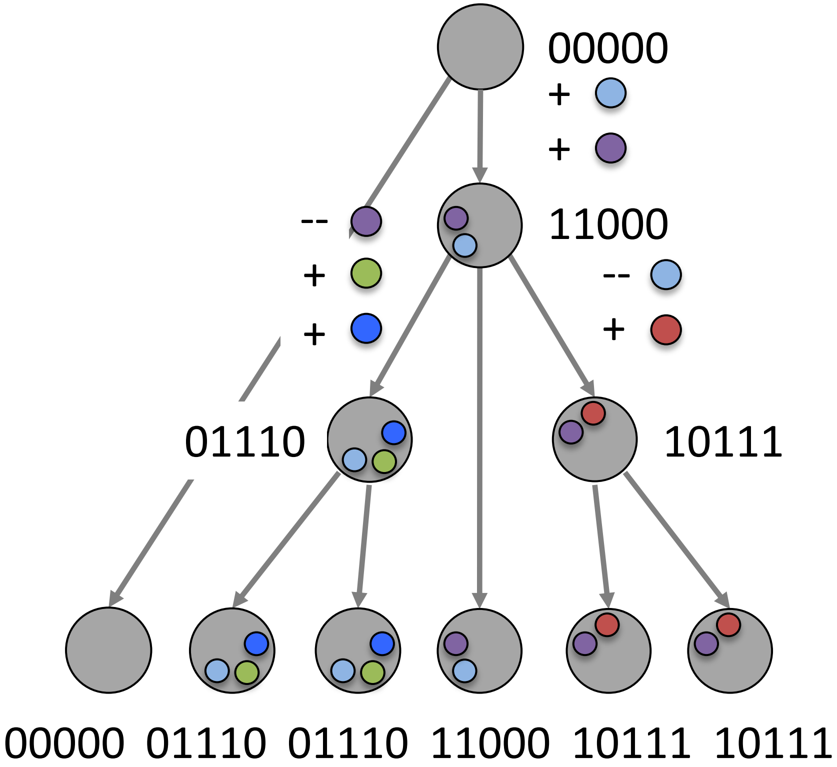
1. Each node v of T is labeled by a vector $\mathbf{b}_v \in \{0,1\}^n$.
2. The root r of T is labeled by vector $\mathbf{b}_r = [0,\ldots,0]^T$.



k-Dollo Phylogeny (*k*-DP) Problem

Definition 1. A *k-Dollo phylogeny* T is a rooted, node-labeled tree subject to the following conditions.

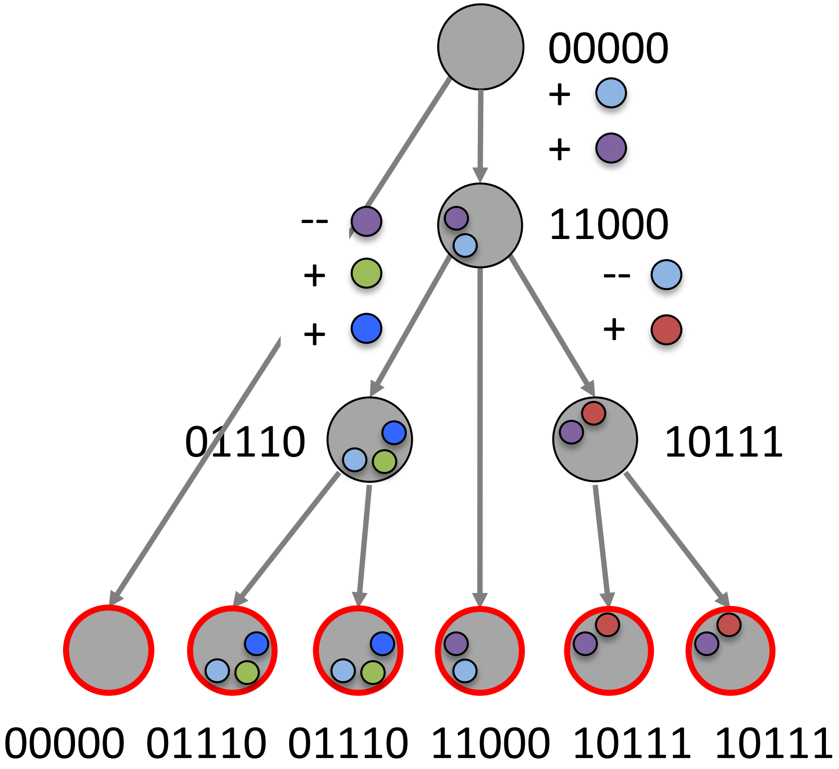
1. Each node v of T is labeled by a vector $\mathbf{b}_v \in \{0,1\}^n$.
2. The root r of T is labeled by vector $\mathbf{b}_r = [0,\ldots,0]^T$.
3. For each character $c \in [n]$, there is exactly one *gain edge* (v,w) in T such that $b_{v,c} = 0$ and $b_{w,c} = 1$.
4. For each character $c \in [n]$, there are at most k *loss edges* (v,w) in T such that $b_{v,c} = 1$ and $b_{w,c} = 0$.



k-Dollo Phylogeny (*k*-DP) Problem

Definition 1. A *k-Dollo phylogeny* T is a rooted, node-labeled tree subject to the following conditions.

1. Each node v of T is labeled by a vector $\mathbf{b}_v \in \{0,1\}^n$.
2. The root r of T is labeled by vector $\mathbf{b}_r = [0,\ldots,0]^T$.
3. For each character $c \in [n]$, there is exactly one *gain edge* (v,w) in T such that $b_{v,c} = 0$ and $b_{w,c} = 1$.
4. For each character $c \in [n]$, there are at most k *loss edges* (v,w) in T such that $b_{v,c} = 1$ and $b_{w,c} = 0$.



***k*-Dollo Phylogeny problem** (k-DP). Given a binary matrix $B \in \{0,1\}^{m \times n}$ and parameter $k \in \mathbb{N}$, determine whether there exists a k-Dollo phylogeny for B, and if so construct one.

n SNVs

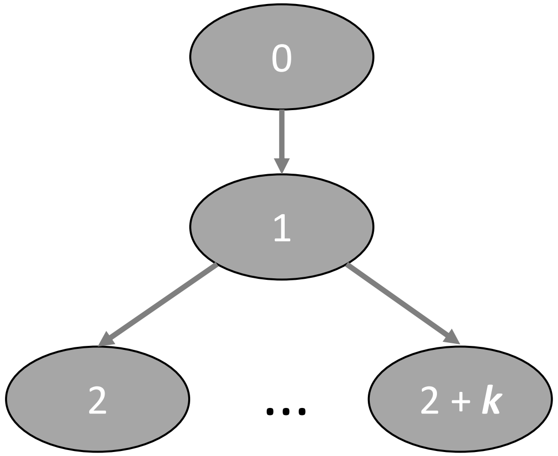
$$\boldsymbol{B} = \begin{array}{ccccc|c} ● & ● & ● & ● & ● & \\ 0 & 0 & 0 & 0 & \mathbf{0} & ◉ \\ 1 & 0 & 1 & 1 & 0 & ◉ \\ 1 & 0 & \mathbf{1} & 1 & 0 & ◉ \\ 1 & \mathbf{1} & 0 & 0 & 0 & ◉ \\ \mathbf{0} & 1 & 0 & 0 & 1 & ◉ \\ 0 & 1 & 0 & 0 & 1 & ◉ \end{array}$$

m cells

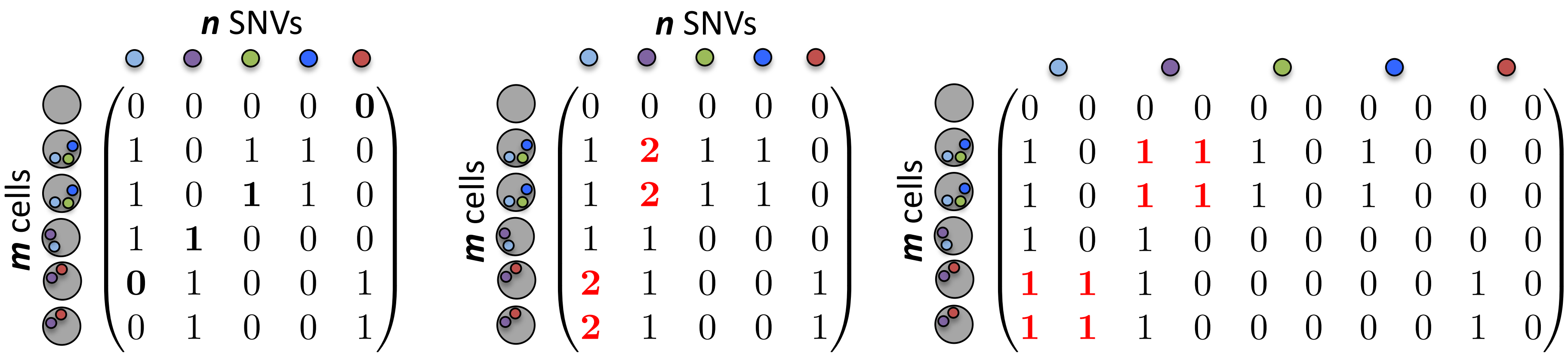
Combinatorial Characterization of *k*-DP

Theorem 3. Let $B \in \{0,1\}^{m \times n}$. The following statements are equivalent.

1. There exists a k-Dollo phylogeny T for B.
2. There exists a k-Dollo completion A of B.
3. There exists a k-completion A of B such that the binary factor matrix B' of $(A, \mathcal{S}[k])$ is a perfect phylogeny matrix.
4. There exists a k-completion A of B, and perfect phylogeny T for A whose characters are consistent with $S[k]$.



***k*-Dollo State Tree *S*[*k*]**



$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & \mathbf{0} \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & \mathbf{1} & 1 & 0 \\ 1 & \mathbf{1} & 0 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

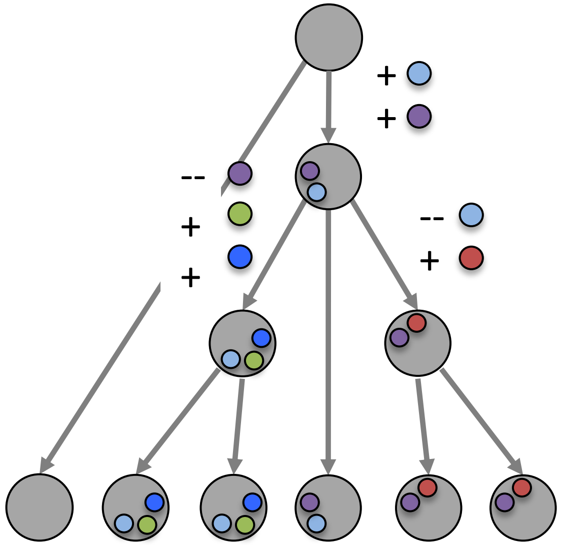
Input Matrix ***B***

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & \mathbf{2} & 1 & 1 & 0 \\ 1 & \mathbf{2} & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \mathbf{2} & 1 & 0 & 0 & 1 \\ \mathbf{2} & 1 & 0 & 0 & 1 \end{pmatrix}$$

k*-Dollo Completion *A

$$B' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \mathbf{1} & \mathbf{1} & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & \mathbf{1} & \mathbf{1} & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \mathbf{1} & \mathbf{1} & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Binary Factor Matrix ***B’***



k*-Dollo Phylogeny *T

Forbidden Submatrices in Solutions *A* to *k*-DP

$$\begin{pmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{pmatrix}$$

k **= 0**

$$\begin{pmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 1\\ 1 & 2\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 2\\ 1 & 1\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 2\\ 1 & 2\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 1\\ 2 & 1\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 1\\ 2 & 2\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 2\\ 2 & 1\end{pmatrix} \begin{pmatrix}1 & 0\\ 0 & 2\\ 2 & 2\end{pmatrix}$$

$$\begin{pmatrix}2 & 0\\ 0 & 1\\ 1 & 1\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 1\\ 1 & 2\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 2\\ 1 & 1\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 2\\ 1 & 2\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 1\\ 2 & 1\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 1\\ 2 & 2\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 2\\ 2 & 1\end{pmatrix} \begin{pmatrix}2 & 0\\ 0 & 2\\ 2 & 2\end{pmatrix}$$

$$\begin{pmatrix}1 & 1\\ 0 & 2\\ 1 & 2\end{pmatrix} \begin{pmatrix}1 & 1\\ 0 & 2\\ 2 & 2\end{pmatrix} \begin{pmatrix}2 & 1\\ 0 & 2\\ 1 & 2\end{pmatrix} \begin{pmatrix}2 & 1\\ 0 & 2\\ 2 & 2\end{pmatrix} \begin{pmatrix}2 & 0\\ 1 & 1\\ 2 & 1\end{pmatrix} \begin{pmatrix}2 & 0\\ 1 & 1\\ 2 & 2\end{pmatrix} \begin{pmatrix}2 & 0\\ 1 & 2\\ 2 & 1\end{pmatrix} \begin{pmatrix}2 & 0\\ 1 & 2\\ 2 & 2\end{pmatrix}$$

$$\begin{pmatrix}2 & 1\\ 1 & 2\\ 2 & 2\end{pmatrix}$$

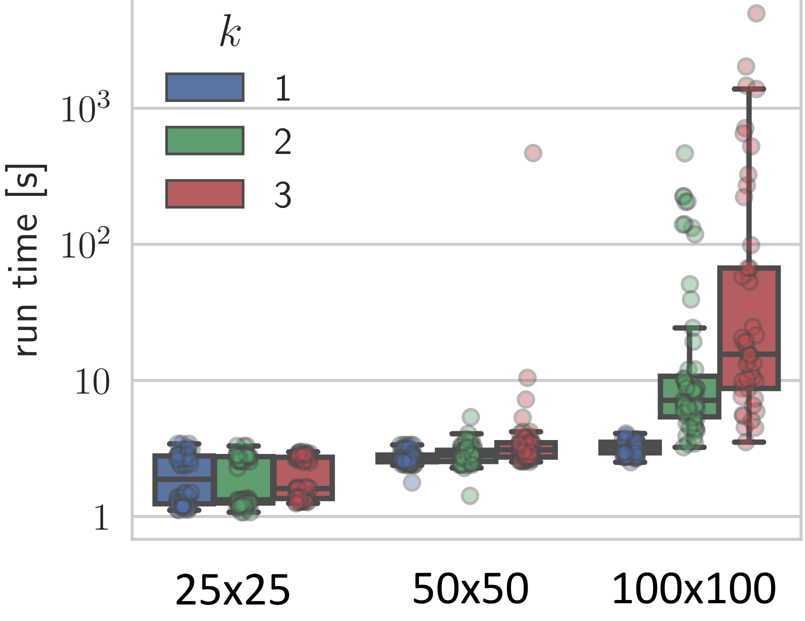
k **= 1**

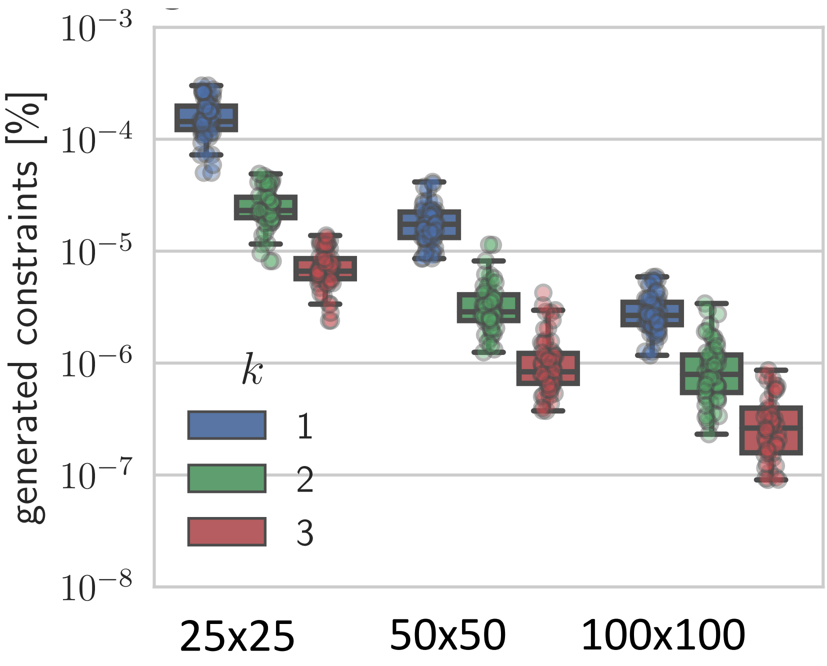
Number of forbidden submatrices is $4k^4 + 8k^3 + 8k^2 + 4k + 1$

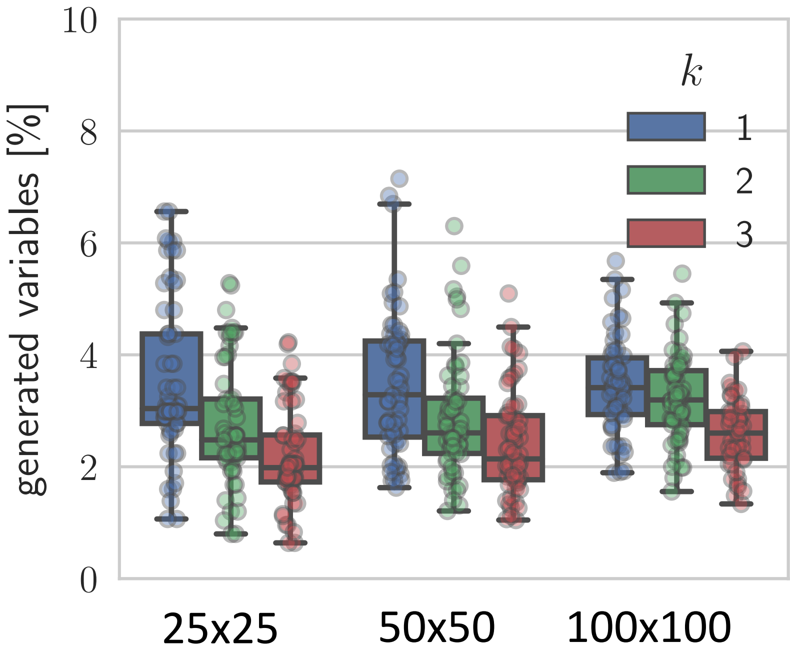
Open question: Hardness of deciding whether ***B*** admits a ***k***-Dollo completion ***A***

Results for *k*-DP

- Naive ILP does not scale and has O(*mnk*) variables and O($m^3n^2k^4$) constraints
- Column and cutting plane generation
  - Introduce variables and constraints only when needed
- Simulations with 60 instances for each *m*, *n* and *k*





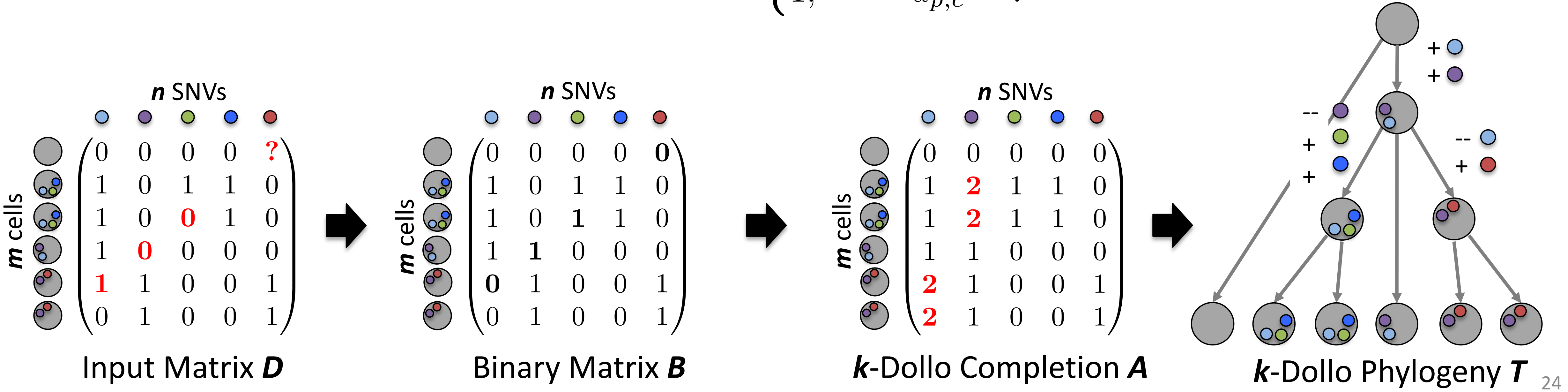


Outline

- Perfect data (error free)
    - Problem statement
    - Combinatorial characterization of solutions
    - Exact algorithm
    - Results
- Real data (with errors)
    - Problem statement
    - Heuristic algorithm
    - Results
- Conclusions

k-Dollo Phylogeny Flip and Cluster (k-DPFC) problem. Given matrix $D \in \{0,1,?\}^{m \times n}$, error rates $\alpha, \beta \in [0,1]$, integers $k, s, t \in \mathbb{N}$, find matrix $B \in \{0,1\}^{m \times n}$ and tree T such that: (1) B has at most s unique rows and at most t unique columns; (2) $\Pr(D \mid B, \alpha, \beta)$ is maximum; and (3) T is a k-Dollo phylogeny for B.

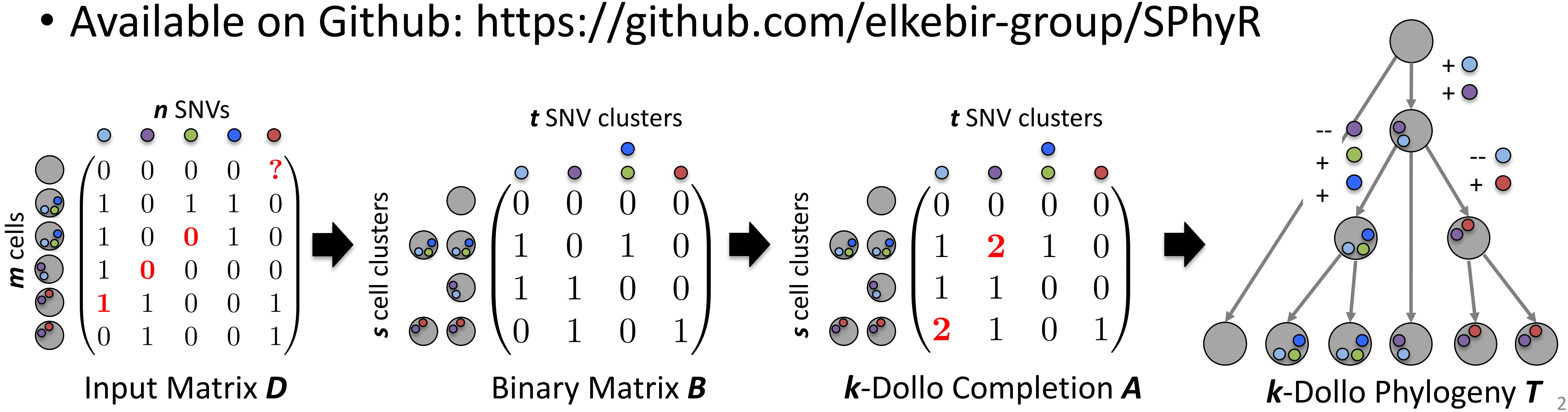
$$\Pr(D \mid B, \alpha, \beta) = \prod_{p=1}^{m} \prod_{c=1}^{n} \begin{cases} \alpha, & d_{p,c} = 1 \text{ and } b_{p,c} = 0 \\ 1-\alpha, & d_{p,c} = 1 \text{ and } b_{p,c} = 1, \\ \beta, & d_{p,c} = 0 \text{ and } b_{p,c} = 1, \\ 1-\beta, & d_{p,c} = 0 \text{ and } b_{p,c} = 0, \\ 1, & d_{p,c} = ? \end{cases}$$



$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & ? \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 0 & 0 & 0 & \mathbf{0} \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & \mathbf{1} & 1 & 0 \\ 1 & \mathbf{1} & 0 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & \mathbf{2} & 1 & 1 & 0 \\ 1 & \mathbf{2} & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \mathbf{2} & 1 & 0 & 0 & 1 \\ \mathbf{2} & 1 & 0 & 0 & 1 \end{pmatrix}$$

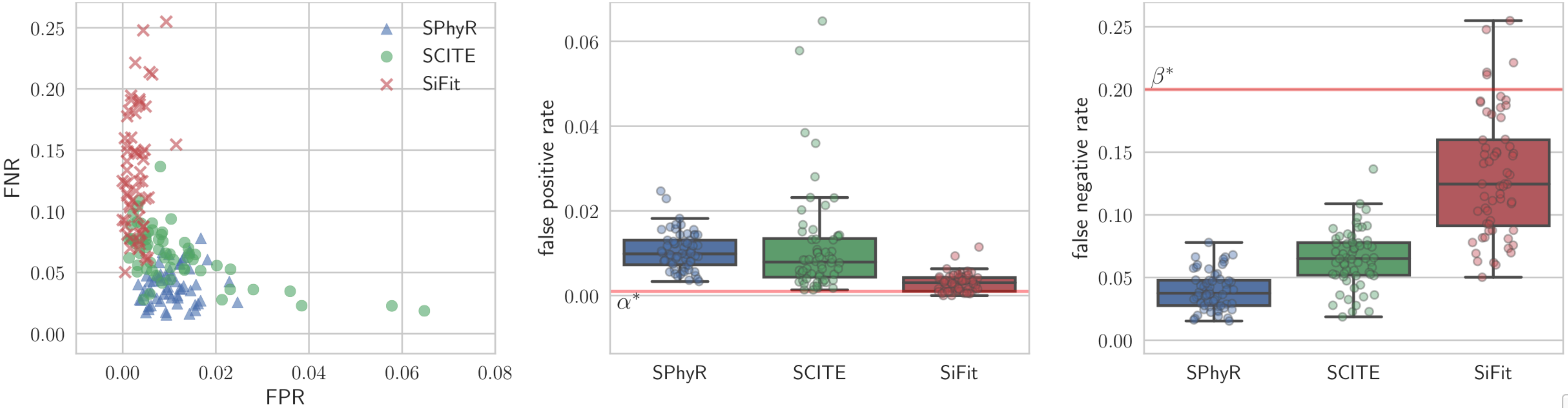
SPhyR: Single-cell Phylogeny Reconstruction

- Coordinate ascent:
  1. k-Means with random seed to obtain cell clustering ***π*** and SNV clustering ***ψ***
  2. ILP to obtain maximum likelihood ***k***-Dollo completion ***A*** given ***D***, ***π*** and ***ψ***
  3. Identify maximum likelihood ***π*** given ***A*** and ***ψ***
  4. Identify maximum likelihood ***ψ*** given ***A*** and ***π***
  5. Repeat until convergence
- Available on Github: https://github.com/elkebir-group/SPhyR

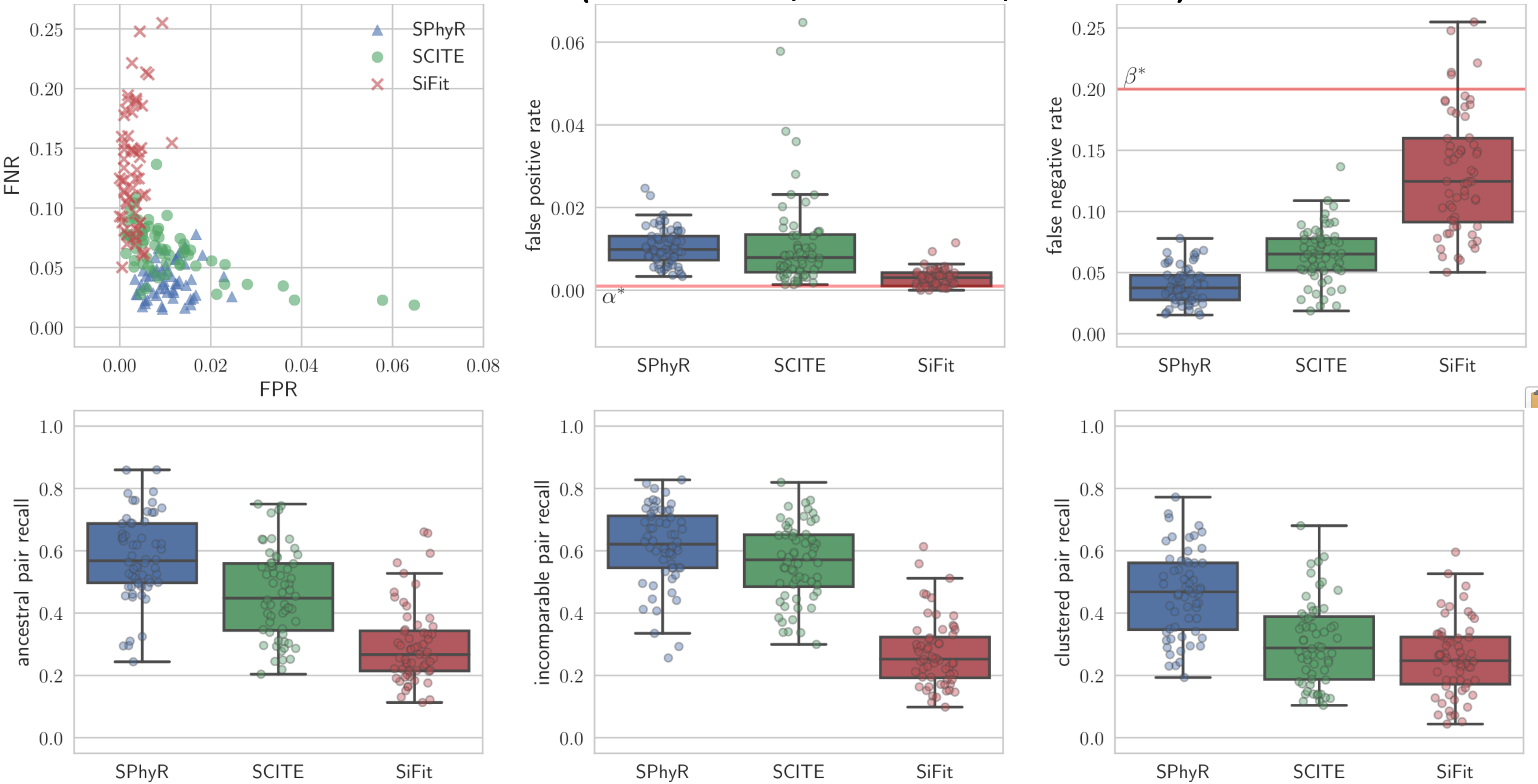


$$\begin{pmatrix} 0 & 0 & 0 & 0 & ? \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} \rightarrow$$

Simulation Results (*m* = 50, *n* = 50, *k* = 1)



Simulation Results (m = 50, n = 50, k = 1)



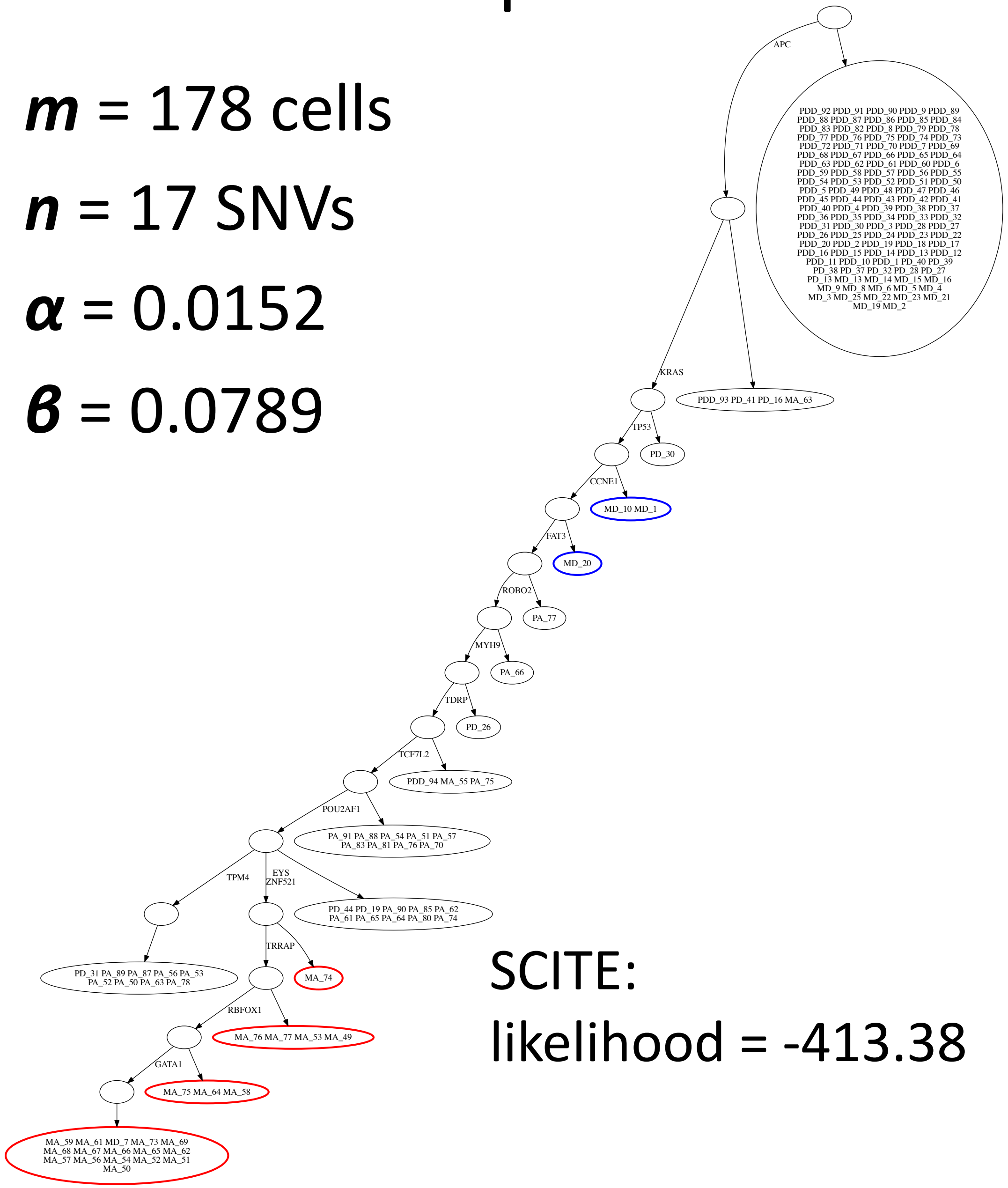
Colorectal patient CRC1 [Leung et al., 2017]

m = 178 cells

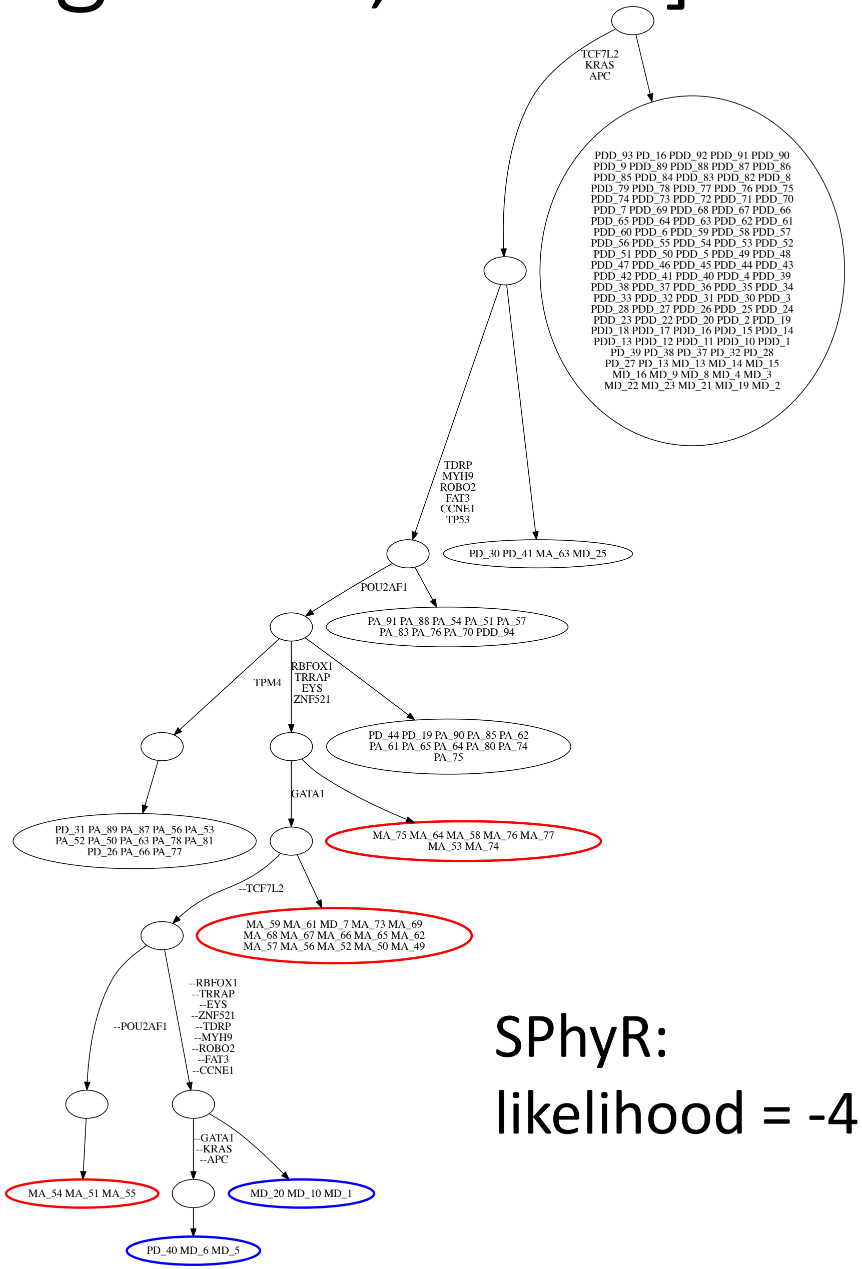
n = 17 SNVs

α = 0.0152

β = 0.0789


APC
PDD_92 PDD_91 PDD_90 PDD_9 PDD_89
PDD_88 PDD_87 PDD_86 PDD_85 PDD_84
PDD_83 PDD_82 PDD_8 PDD_79 PDD_78
PDD_77 PDD_76 PDD_75 PDD_74 PDD_73
PDD_72 PDD_71 PDD_70 PDD_7 PDD_69
PDD_68 PDD_67 PDD_66 PDD_65 PDD_64
PDD_63 PDD_62 PDD_61 PDD_60 PDD_6
PDD_59 PDD_58 PDD_57 PDD_56 PDD_55
PDD_54 PDD_53 PDD_52 PDD_51 PDD_50
PDD_5 PDD_49 PDD_48 PDD_47 PDD_46
PDD_45 PDD_44 PDD_43 PDD_42 PDD_41
PDD_40 PDD_4 PDD_39 PDD_38 PDD_37
PDD_36 PDD_35 PDD_34 PDD_33 PDD_32
PDD_31 PDD_30 PDD_3 PDD_28 PDD_27
PDD_26 PDD_25 PDD_24 PDD_23 PDD_22
PDD_20 PDD_2 PDD_19 PDD_18 PDD_17
PDD_16 PDD_15 PDD_14 PDD_13 PDD_12
PDD_11 PDD_10 PDD_1 PD_40 PD_39
PD_38 PD_37 PD_32 PD_28 PD_27
PD_13 MD_13 MD_14 MD_15 MD_16
MD_9 MD_8 MD_6 MD_5 MD_4
MD_3 MD_25 MD_22 MD_23 MD_21
MD_19 MD_2
KRAS
PDD_93 PD_41 PD_16 MA_63
TP53
PD_30
CCNE1
MD_10 MD_1
FAT3
MD_20
ROBO2
PA_77
MYH9
PA_66
TDRP
PD_26
TCF7L2
PDD_94 MA_55 PA_75
POU2AF1
PA_91 PA_88 PA_54 PA_51 PA_57
PA_83 PA_81 PA_76 PA_70
TPM4
EYS
ZNF521
PD_44 PD_19 PA_90 PA_85 PA_62
PA_61 PA_65 PA_64 PA_80 PA_74
PD_31 PA_89 PA_87 PA_56 PA_53
PA_52 PA_50 PA_63 PA_78
TRRAP
MA_74
RBFOX1
MA_76 MA_77 MA_53 MA_49
GATA1
MA_75 MA_64 MA_58
MA_59 MA_61 MD_7 MA_73 MA_69
MA_68 MA_67 MA_66 MA_65 MA_62
MA_57 MA_56 MA_54 MA_52 MA_51
MA_50

SCITE:
likelihood = -413.38


TCF7L2
KRAS
APC
PDD_93 PD_16 PDD_92 PDD_91 PDD_90
PDD_9 PDD_89 PDD_88 PDD_87 PDD_86
PDD_85 PDD_84 PDD_83 PDD_82 PDD_8
PDD_79 PDD_78 PDD_77 PDD_76 PDD_75
PDD_74 PDD_73 PDD_72 PDD_71 PDD_70
PDD_7 PDD_69 PDD_68 PDD_67 PDD_66
PDD_65 PDD_64 PDD_63 PDD_62 PDD_61
PDD_60 PDD_6 PDD_59 PDD_58 PDD_57
PDD_56 PDD_55 PDD_54 PDD_53 PDD_52
PDD_51 PDD_50 PDD_5 PDD_49 PDD_48
PDD_47 PDD_46 PDD_45 PDD_44 PDD_43
PDD_42 PDD_41 PDD_40 PDD_4 PDD_39
PDD_38 PDD_37 PDD_36 PDD_35 PDD_34
PDD_33 PDD_32 PDD_31 PDD_30 PDD_3
PDD_28 PDD_27 PDD_26 PDD_25 PDD_24
PDD_23 PDD_22 PDD_20 PDD_2 PDD_19
PDD_18 PDD_17 PDD_16 PDD_15 PDD_14
PDD_13 PDD_12 PDD_11 PDD_10 PDD_1
PD_39 PD_38 PD_37 PD_32 PD_28
PD_27 PD_13 MD_13 MD_14 MD_15
MD_16 MD_9 MD_8 MD_4 MD_3
MD_22 MD_23 MD_21 MD_19 MD_2
TDRP
MYH9
ROBO2
FAT3
CCNE1
TP53
PD_30 PD_41 MA_63 MD_25
POU2AF1
PA_91 PA_88 PA_54 PA_51 PA_57
PA_83 PA_76 PA_70 PDD_94
TPM4
RBFOX1
TRRAP
EYS
ZNF521
PD_44 PD_19 PA_90 PA_85 PA_62
PA_61 PA_65 PA_64 PA_80 PA_74
PA_75
PD_31 PA_89 PA_87 PA_56 PA_53
PA_52 PA_50 PA_63 PA_78 PA_81
PD_26 PA_66 PA_77
GATA1
MA_75 MA_64 MA_58 MA_76 MA_77
MA_53 MA_74
--TCF7L2
MA_59 MA_61 MD_7 MA_73 MA_69
MA_68 MA_67 MA_66 MA_65 MA_62
MA_57 MA_56 MA_52 MA_50 MA_49
--POU2AF1
--RBFOX1
--TRRAP
--EYS
--ZNF521
--TDRP
--MYH9
--ROBO2
--FAT3
--CCNE1
MA_54 MA_51 MA_55
--GATA1
--KRAS
--APC
MD_20 MD_10 MD_1
PD_40 MD_6 MD_5

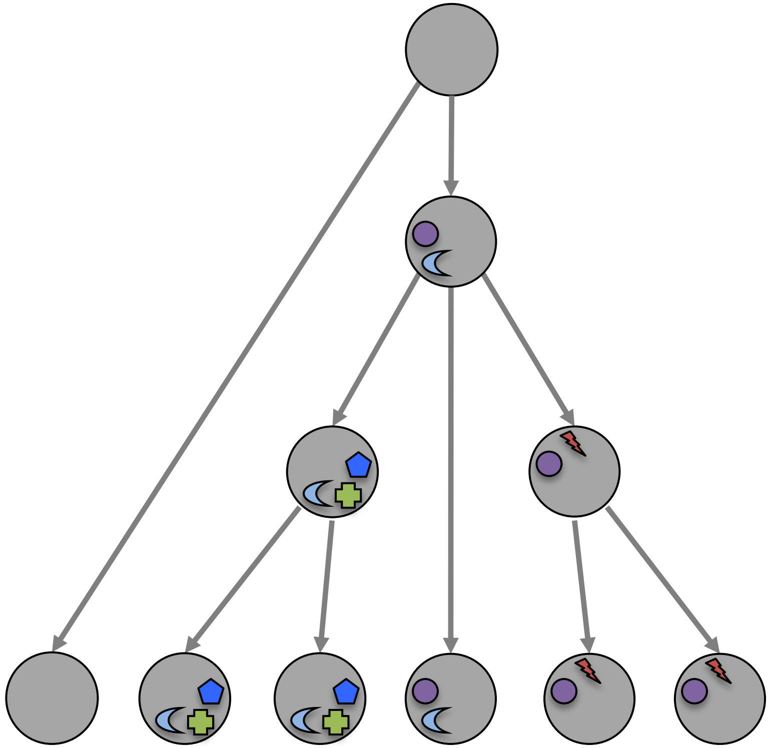
SPhyR:
likelihood = -450.70

Conclusions

- ***k***-Dollo parsimony model strikes a balance between realistic and yet sufficiently constrained
- Solutions are integer matrix completions
- SPhyR outperformed existing methods

Future work:

- Include ***α*** and ***β*** into optimization
- Model selection for ***s***, ***t*** and ***k***
- Hardness is open



Acknowledgments

- Experiments were run on NCSA Blue Waters