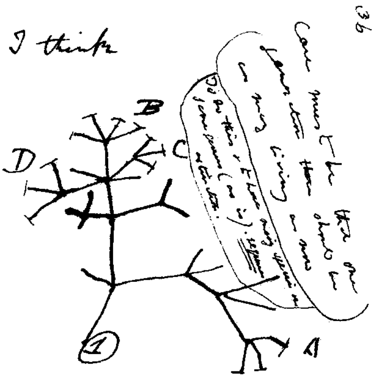
Semi-algebraic conditions for phylogenetic reconstruction

Marina Garrote-López

Seminari de Geometria Algebraica de Barcelona

Phylogenetics


36
I think
Case must be that one generation then should be as many living as now
To do this & to have many species in same genus (as is). requires extinction.
B
C
D
A
1

Charles Darwin, 1859

Phylogenetics

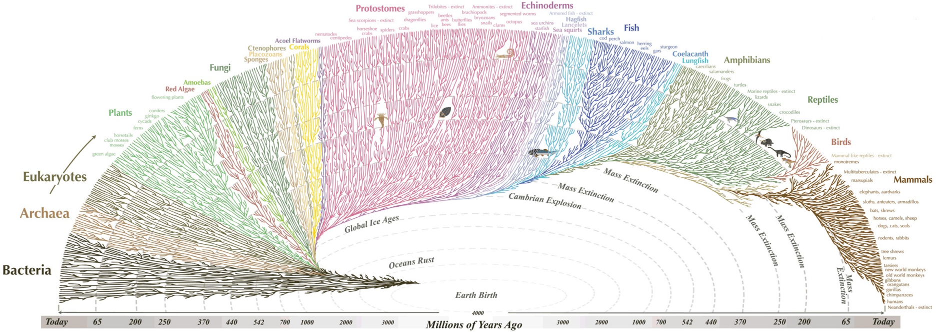


Table of contents

1 Modelling evolution

2 Phylogenetic varieties

3 Phylogenetic reconstruction methods

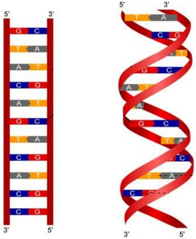
4 Stochastic phylogenetic regions

5 SAQ: Semi-algebraic quartet reconstruction method

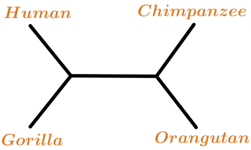
Phylogenetic reconstruction

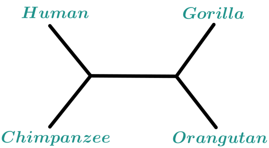
Given an alignment of DNA sequences for some species,

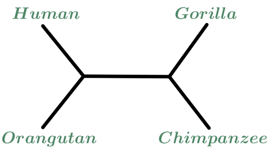
Gorilla	AACTTCGAGGCTTACCGCTG
Human	AACGTCTATGCTCACCGATG
Chimpanzee	AAGGTCGATGCTCACCGATG
Orangutan	ATTGTCGCAACTCGTCGACG



our goal is to reconstruct the topology of the phylogenetic tree that relates them.



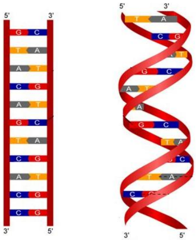




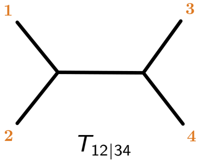
Phylogenetic reconstruction

Given an alignment of DNA sequences for some species,

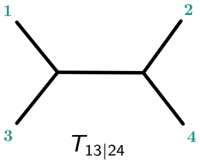
Gorilla	AACTTCGAGGCTTACCGCTG
Human	AACGTCTATGCTCACCGATG
Chimpanzee	AAGGTCGATGCTCACCGATG
Orangutan	ATTGTCGCAACTCGTCGACG



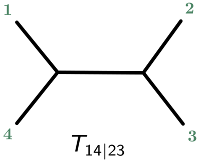
our goal is to reconstruct the topology of the phylogenetic tree that relates them.



$T_{12|34}$

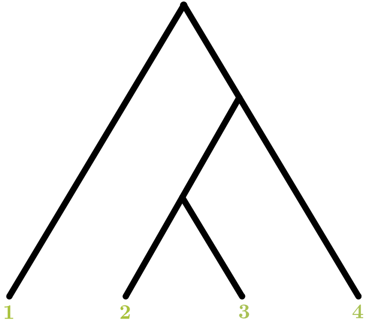


$T_{13|24}$

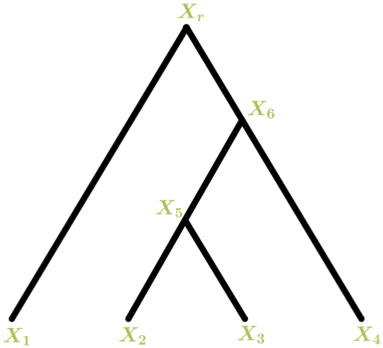


$T_{14|23}$

Modeling Evolution

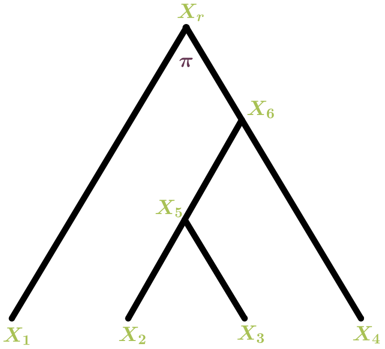


Modeling Evolution



Random variables at the nodes
$X_i \in \mathcal{K} = \{\texttt{A},\texttt{C},\texttt{G},\texttt{T}\}$

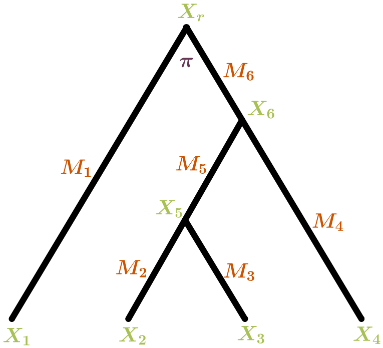
Modeling Evolution



Random variables at the nodes
$X_i \in \mathcal{K} = \{\texttt{A},\texttt{C},\texttt{G},\texttt{T}\}$

Distribution at the root
$\pi = (\pi_{\texttt{A}}, \pi_{\texttt{C}}, \pi_{\texttt{G}}, \pi_{\texttt{T}});\ \ \sum_{i \in \mathcal{K}} \pi_i = 1$

Modeling Evolution



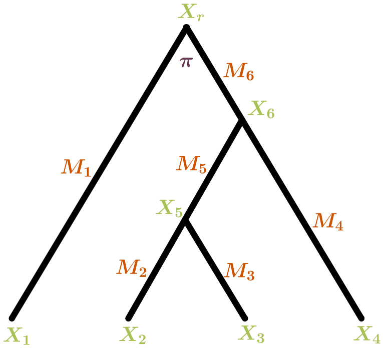
Random variables at the nodes
$X_i \in \mathcal{K} = \{\mathtt{A},\mathtt{C},\mathtt{G},\mathtt{T}\}$

Distribution at the root
$\pi = (\pi_\mathtt{A}, \pi_\mathtt{C}, \pi_\mathtt{G}, \pi_\mathtt{T});\ \sum_{i\in\mathcal{K}} \pi_i = 1$

Transition matrices at the edges

$$Me = \begin{pmatrix} P(\mathtt{A}\to\mathtt{A}|e) & \dots & P(\mathtt{A}\to\mathtt{T}|e) \\ P(\mathtt{C}\to\mathtt{A}|e) & \dots & P(\mathtt{C}\to\mathtt{T}|e) \\ P(\mathtt{G}\to\mathtt{A}|e) & \dots & P(\mathtt{T}\to\mathtt{G}|e) \\ P(\mathtt{T}\to\mathtt{A}|e) & \dots & P(\mathtt{T}\to\mathtt{T}|e) \end{pmatrix}$$

Modeling Evolution



Random variables at the nodes
$X_i \in \mathcal{K} = \{\texttt{A},\texttt{C},\texttt{G},\texttt{T}\}$

Distribution at the root
$\pi = (\pi_{\texttt{A}}, \pi_{\texttt{C}}, \pi_{\texttt{G}}, \pi_{\texttt{T}});\ \ \sum_{i \in \mathcal{K}} \pi_i = 1$

Transition matrices at the edges

$$Me = \begin{pmatrix} P(\texttt{A} \to \texttt{A}|e) & \dots & P(\texttt{A} \to \texttt{T}|e) \\ P(\texttt{C} \to \texttt{A}|e) & \dots & P(\texttt{C} \to \texttt{T}|e) \\ P(\texttt{G} \to \texttt{A}|e) & \dots & P(\texttt{T} \to \texttt{G}|e) \\ P(\texttt{T} \to \texttt{A}|e) & \dots & P(\texttt{T} \to \texttt{T}|e) \end{pmatrix}$$

A transition matrix is a square matrix with nonnegative entries and rows summing up to one.

Evolutionary models

Jukes Cantor Model

$$M_e = \begin{pmatrix} a_e & b_e & b_e & b_e \\ b_e & a_e & b_e & b_e \\ b_e & b_e & a_e & b_e \\ b_e & b_e & b_e & a_e \end{pmatrix},$$

where $3a_e + b_e = 1$.

Evolutionary models

Jukes Cantor Model

$$M_e = \begin{pmatrix} a_e & b_e & b_e & b_e \\ b_e & a_e & b_e & b_e \\ b_e & b_e & a_e & b_e \\ b_e & b_e & b_e & a_e \end{pmatrix},$$

where $3a_e + b_e = 1$.

Evolutionary models

Jukes Cantor Model

$$M_e = \begin{pmatrix} a_e & b_e & b_e & b_e \\ b_e & a_e & b_e & b_e \\ b_e & b_e & a_e & b_e \\ b_e & b_e & b_e & a_e \end{pmatrix},$$

where $3a_e + b_e = 1$.

Strand Symmetric Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ e_e & f_e & g_e & h_e \\ h_e & g_e & f_e & e_e \\ d_e & c_e & b_e & a_e \end{pmatrix},$$

where rows sum up to 1.

Evolutionary models

Jukes Cantor Model

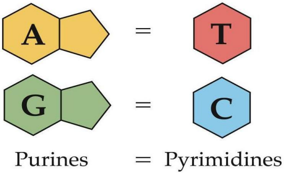
$$M_e = \begin{pmatrix} a_e & b_e & b_e & b_e \\ b_e & a_e & b_e & b_e \\ b_e & b_e & a_e & b_e \\ b_e & b_e & b_e & a_e \end{pmatrix},$$

where $3a_e + b_e = 1$.

Strand Symmetric Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ e_e & f_e & g_e & h_e \\ h_e & g_e & f_e & e_e \\ d_e & c_e & b_e & a_e \end{pmatrix},$$

where rows sum up to 1.



Evolutionary models

Jukes Cantor Model

$$M_e = \begin{pmatrix} a_e & b_e & b_e & b_e \\ b_e & a_e & b_e & b_e \\ b_e & b_e & a_e & b_e \\ b_e & b_e & b_e & a_e \end{pmatrix},$$

where $3a_e + b_e = 1$.

Strand Symmetric Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ e_e & f_e & g_e & h_e \\ h_e & g_e & f_e & e_e \\ d_e & c_e & b_e & a_e \end{pmatrix},$$

where rows sum up to 1.

Kimura Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ b_e & a_e & d_e & c_e \\ c_e & d_e & a_e & b_e \\ d_e & c_e & b_e & a_e \end{pmatrix},$$

where $a_e + b_e + c_e + d_e = 1$.

Evolutionary models

Jukes Cantor Model

$$M_e = \begin{pmatrix} a_e & b_e & b_e & b_e \\ b_e & a_e & b_e & b_e \\ b_e & b_e & a_e & b_e \\ b_e & b_e & b_e & a_e \end{pmatrix},$$

where $3a_e + b_e = 1$.

Strand Symmetric Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ e_e & f_e & g_e & h_e \\ h_e & g_e & f_e & e_e \\ d_e & c_e & b_e & a_e \end{pmatrix},$$

where rows sum up to 1.

Kimura Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ b_e & a_e & d_e & c_e \\ c_e & d_e & a_e & b_e \\ d_e & c_e & b_e & a_e \end{pmatrix},$$

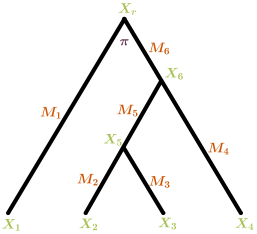
where $a_e + b_e + c_e + d_e = 1$.

General Markov Model

$$M_e = \begin{pmatrix} a_e & b_e & c_e & d_e \\ e_e & f_e & g_e & h_e \\ j_e & k_e & l_e & m_e \\ n_e & o_e & p_e & q_e \end{pmatrix},$$

where rows sum up to 1.

Joint distribution

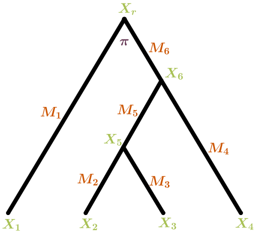


Definition

The **joint distribution** $p_{s_1,\dots,s_n}$ at the leaves of a rooted phylogenetic tree T, which is the probability that the random variables $X_1,\dots,X_n$ of the leaves take the states $s_1,\dots,s_n$

$$p_{s_1\dots s_n} = Prob(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n).$$

Joint distribution



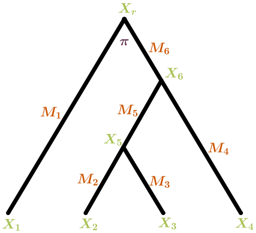
Definition

The **joint distribution** $p_{s_1,\dots,s_n}$ at the leaves of a rooted phylogenetic tree T, which is the probability that the random variables $X_1,\dots,X_n$ of the leaves take the states $s_1,\dots,s_n$

$$p_{s_1\dots s_n} = Prob(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n).$$

$$p_{x_1\dots x_n} = \sum_{x_v, v\in Int(T)} \pi_{x_r} \prod_{e\in E(T)} M_e(x_{pa(e)}, x_{ch(e)}),$$

Joint distribution



Definition

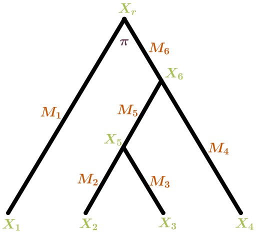
The **joint distribution** $p_{s_1,\dots,s_n}$ at the leaves of a rooted phylogenetic tree T, which is the probability that the random variables $X_1,\dots,X_n$ of the leaves take the states $s_1,\dots,s_n$

$$p_{s_1\dots s_n} = Prob(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n).$$

$$p_{\mathtt{A},\mathtt{T},\mathtt{C},\mathtt{C}} = \sum_{x_r,x_5,x_6\in\{\mathtt{A},\mathtt{C},\mathtt{G},\mathtt{T}\}} \pi_{x_r}\cdot M_1(x_r,\mathtt{A})\cdot M_6(x_r,x_6)\cdot M_5(x_6,x_5)\cdot$$

$$\cdot M_2(x_5,\mathtt{T})\cdot M_3(x_5,\mathtt{C})\cdot M_4(x_6,\mathtt{C})$$

Joint distribution



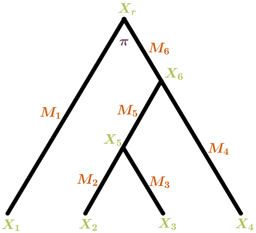
Definition

We denote by $p_{s_1,\dots,s_n}$ the **joint distribution** at the leaves of a rooted phylogenetic tree T, which is the probability that the random variables $X_1,\dots,X_n$ of the leaves take the states $s_1,\dots,s_n$

$$p_{s_1\dots s_n} = Prob(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n).$$

- The entries of the joint distribution at the leaves $p^T = \left(p^T_{s_1\dots s_n}\right)_{s_1,\dots,s_n}$ can be expressed as a **polynomial** in terms of the parameters of the model.

Joint distribution



Definition

We denote by $p_{s_1,\dots,s_n}$ the **joint distribution** at the leaves of a rooted phylogenetic tree T, which is the probability that the random variables $X_1,\dots,X_n$ of the leaves take the states $s_1,\dots,s_n$

$$p_{s_1\dots s_n} = Prob(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n).$$

- The entries of the joint distribution at the leaves $p^T = \left(p^T_{s_1\dots s_n}\right)_{s_1,\dots,s_n}$ can be expressed as a **polynomial** in terms of the parameters of the model.
- We can estimate p^T easily (by the relative frequencies in an alignment) but **NOT** the parameters.

Phylogenetic variety

Definition

For fixed tree T and model $\mathcal{M}$, fixed the position of the root r we use φ_T to denote the **parametrization map**,

$$\begin{array}{rcl} \varphi_T : \mathbb{R}^d & \longrightarrow & \mathbb{R}^{4^n} \\ (\pi, \{M_e\}_{e \in E(T)}) & \mapsto & P = (p_{x_1,x_1,\ldots,x_1}, p_{x_1,x_1,\ldots,x_2}, \ldots, p_{x_n,x_n,\ldots,x_n}) \end{array}$$

Phylogenetic variety

Definition

For fixed tree T and model $\mathcal{M}$, fixed the position of the root r we use φ_T to denote the **parametrization map**,

$$\begin{array}{rcl} \varphi_T : \mathbb{R}^d & \longrightarrow & \mathbb{R}^{4^n} \\ (\pi, \{M_e\}_{e\in E(T)}) & \mapsto & P = (p_{x_1,x_1,\ldots,x_1}, p_{x_1,x_1,\ldots,x_2}, \ldots, p_{x_n,x_n,\ldots,x_n}) \end{array}$$

The **phylogenetic algebraic variety** associated to a tree T and a model $\mathcal{M}$ is

$$\mathcal{V}_T = \overline{\operatorname{Im}\varphi_T}.$$

$I_T = I(\mathcal{V}_T)$ is the **phylogenetic ideal** of T and $\mathcal{M}$.

Phylogenetic variety

Definition

For fixed tree T and model $\mathcal{M}$, fixed the position of the root r we use φ_T to denote the **parametrization map**,

$$\begin{array}{rcl} \varphi_T : \mathbb{R}^d & \longrightarrow & \mathbb{R}^{4^n} \\ (\pi, \{M_e\}_{e \in E(T)}) & \mapsto & P = (p_{x_1,x_1,\dots,x_1}, p_{x_1,x_1,\dots,x_2}, \dots, p_{x_n,x_n,\dots,x_n}) \end{array}$$

The **phylogenetic algebraic variety** associated to a tree T and a model $\mathcal{M}$ is

$$\mathcal{V}_T = \overline{\operatorname{Im} \varphi_T}.$$

$I_T = I(\mathcal{V}_T)$ is the **phylogenetic ideal** of T and $\mathcal{M}$.

Polynomials $f \in I_T$ are called **phylogenetic invariants** of T.

Phylogenetic variety

Definition

For fixed tree T and model $\mathcal{M}$, fixed the position of the root r we use φ_T to denote the **parametrization map**,

$$\begin{array}{rcl} \varphi_T : \mathbb{R}^d & \longrightarrow & \mathbb{R}^{4^n} \\ (\pi, \{M_e\}_{e\in E(T)}) & \mapsto & P = (p_{x_1,x_1,\dots,x_1}, p_{x_1,x_1,\dots,x_2}, \dots, p_{x_n,x_n,\dots,x_n}) \end{array}$$

The **phylogenetic algebraic variety** associated to a tree T and a model $\mathcal{M}$ is

$$\mathcal{V}_T = \overline{\operatorname{Im}\varphi_T}.$$

$I_T = I(\mathcal{V}_T)$ is the **phylogenetic ideal** of T and $\mathcal{M}$.

Polynomials $f \in I_T$ are called **phylogenetic invariants** of T.

Polynomials $f \in I_T$ and $f \notin I_{T'}$, with $T \neq T'$ are the **topology invariants** of T.

Using algebraic varieties in phylogenetics

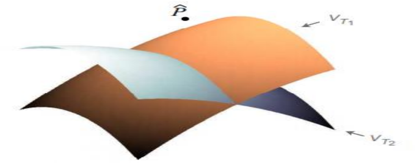
- An alignment produces a point $\hat{p} = (p_{\mathtt{AA\ldots A}}, p_{\mathtt{AA\ldots C}}, \ldots, p_{\mathtt{TT\ldots T}})$ in $\mathbb{R}^{4^n}$.

Using algebraic varieties in phylogenetics

- An alignment produces a point $\hat{p} = (p_{\mathtt{AA\ldots A}}, p_{\mathtt{AA\ldots C}}, \ldots, p_{\mathtt{TT\ldots T}})$ in $\mathbb{R}^{4^n}$.
- $\hat{p}$ should be *close* to $\mathcal{V}_{T_0}$ (if the tree T_0 and model $\mathcal{M}$ fit the data).

Using algebraic varieties in phylogenetics

- An alignment produces a point $\hat{p} = (p_{\texttt{AA}\ldots\texttt{A}}, p_{\texttt{AA}\ldots\texttt{C}}, \ldots, p_{\texttt{TT}\ldots\texttt{T}})$ in $\mathbb{R}^{4^n}$.
- $\hat{p}$ should be *close* to $\mathcal{V}_{T_0}$ (if the tree T_0 and model $\mathcal{M}$ fit the data).
- **Tree topology reconstruction using algebraic geometry.** For each possible topology T, evaluate elements of $I(\mathcal{V}_T)$ at $\hat{p}$: the polynomials of $I(\mathcal{V}_{T_0})$ should be ≈ 0 when evaluated at $\hat{p}$.



Problem: computation of invariants

Computational algebra softwares fail to compute the ideal for ≥ 4 species!

Problem: computation of invariants

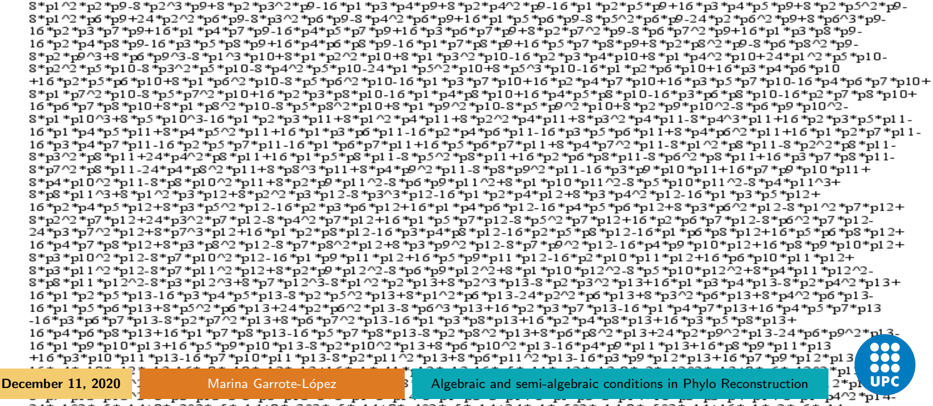
Computational algebra softwares fail to compute the ideal for ≥ 4 species!

For example, Kimura 3-parameter with 4 species is a toric variety with 8002 generators like,

Problem: computation of invariants

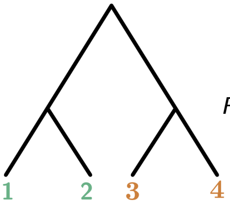
Computational algebra softwares fail to compute the ideal for ≥ 4 species!

For example, Kimura 3-parameter with 4 species is a toric variety with 8002 generators like,



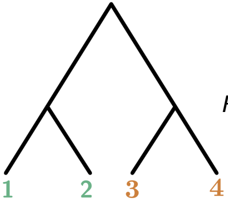
```
8*p1^2*p2*p9-8*p2^3*p9+8*p2*p3^2*p9-16*p1*p3*p4*p9+8*p2*p4^2*p9-16*p1*p2*p5*p9+16*p3*p4*p5*p9+8*p2*p5^2*p9-
8*p1^2*p6*p9+24*p2^2*p6*p9-8*p3^2*p6*p9-8*p4^2*p6*p9+16*p1*p5*p6*p9-8*p5^2*p6*p9-24*p2*p6^2*p9+8*p6^3*p9-
16*p2*p3*p7*p9+16*p1*p4*p7*p9-16*p4*p5*p7*p9+16*p3*p6*p7*p9+8*p2*p7^2*p9-8*p6*p7^2*p9+16*p1*p3*p8*p9-
16*p2*p4*p8*p9-16*p3*p5*p8*p9+16*p4*p6*p8*p9-16*p1*p7*p8*p9+16*p5*p7*p8*p9+8*p2*p8^2*p9-8*p6*p8^2*p9-
8*p2*p9^3+8*p6*p9^3-8*p1^3*p10+8*p1*p2^2*p10+8*p1*p3^2*p10-16*p2*p3*p4*p10+8*p1*p4^2*p10+24*p1^2*p5*p10-
8*p2^2*p5*p10-8*p3^2*p5*p10-8*p4^2*p5*p10-24*p1*p5^2*p10+8*p5^3*p10-16*p1*p2*p6*p10+16*p3*p4*p6*p10
+16*p2*p5*p6*p10+8*p1*p6^2*p10-8*p5*p6^2*p10-16*p1*p3*p7*p10+16*p2*p4*p7*p10+16*p3*p5*p7*p10-16*p4*p6*p7*p10+
8*p1*p7^2*p10-8*p5*p7^2*p10+16*p2*p3*p8*p10-16*p1*p4*p8*p10+16*p4*p5*p8*p10-16*p3*p6*p8*p10-16*p2*p7*p8*p10+
16*p6*p7*p8*p10+8*p1*p8^2*p10-8*p5*p8^2*p10+8*p1*p9^2*p10-8*p5*p9^2*p10+8*p2*p9*p10^2-8*p6*p9*p10^2-
8*p1*p10^3+8*p5*p10^3-16*p1*p2*p3*p11+8*p1^2*p4*p11+8*p2^2*p4*p11+8*p3^2*p4*p11-8*p4^3*p11+16*p2*p3*p5*p11-
16*p1*p4*p5*p11+8*p4*p5^2*p11+16*p1*p3*p6*p11-16*p2*p4*p6*p11-16*p3*p5*p6*p11+8*p4*p6^2*p11+16*p1*p2*p7*p11-
16*p3*p4*p7*p11-16*p2*p5*p7*p11-16*p1*p6*p7*p11+16*p5*p6*p7*p11+8*p4*p7^2*p11-8*p1^2*p8*p11-8*p2^2*p8*p11-
8*p3^2*p8*p11+24*p4^2*p8*p11+16*p1*p5*p8*p11-8*p5^2*p8*p11+16*p2*p6*p8*p11-8*p6^2*p8*p11+16*p3*p7*p8*p11-
8*p7^2*p8*p11-24*p4*p8^2*p11+8*p8^3*p11+8*p4*p9^2*p11-8*p8*p9^2*p11-16*p3*p9*p10*p11+16*p7*p9*p10*p11+
8*p4*p10^2*p11-8*p8*p10^2*p11+8*p2*p9*p11^2-8*p6*p9*p11^2+8*p1*p10*p11^2-8*p5*p10*p11^2-8*p4*p11^3+
8*p8*p11^3+8*p1^2*p3*p12+8*p2^2*p3*p12-8*p3^3*p12-16*p1*p2*p4*p12+8*p3*p4^2*p12-16*p1*p3*p5*p12+
16*p2*p4*p5*p12+8*p3*p5^2*p12-16*p2*p3*p6*p12+16*p1*p4*p6*p12-16*p4*p5*p6*p12+8*p3*p6^2*p12-8*p1^2*p7*p12+
8*p2^2*p7*p12+24*p3^2*p7*p12-8*p4^2*p7*p12+16*p1*p5*p7*p12-8*p5^2*p7*p12+16*p2*p6*p7*p12-8*p6^2*p7*p12-
24*p3*p7^2*p12+8*p7^3*p12+16*p1*p2*p8*p12-16*p3*p4*p8*p12-16*p2*p5*p8*p12-16*p1*p6*p8*p12+16*p5*p6*p8*p12+
16*p4*p7*p8*p12+8*p3*p8^2*p12-8*p7*p8^2*p12+8*p3*p9^2*p12-8*p7*p9^2*p12-16*p4*p9*p10*p12+16*p8*p9*p10*p12+
8*p3*p10^2*p12-8*p7*p10^2*p12-16*p1*p9*p11*p12+16*p5*p9*p11*p12-16*p2*p10*p11*p12+16*p6*p10*p11*p12+
8*p3*p11^2*p12-8*p7*p11^2*p12+8*p2*p9*p12^2-8*p6*p9*p12^2+8*p1*p10*p12^2-8*p5*p10*p12^2+8*p4*p11*p12^2-
8*p8*p11*p12^2-8*p3*p12^3+8*p7*p12^3-8*p1^2*p2*p13+8*p2^3*p13-8*p2*p3^2*p13+16*p1*p3*p4*p13-8*p2*p4^2*p13+
16*p1*p2*p5*p13-16*p3*p4*p5*p13-8*p2*p5^2*p13+8*p1^2*p6*p13-24*p2^2*p6*p13+8*p3^2*p6*p13+8*p4^2*p6*p13-
16*p1*p5*p6*p13+8*p5^2*p6*p13+24*p2*p6^2*p13-8*p6^3*p13+16*p2*p3*p7*p13-16*p1*p4*p7*p13+16*p4*p5*p7*p13
-16*p3*p6*p7*p13-8*p2*p7^2*p13+8*p6*p7^2*p13-16*p1*p3*p8*p13+16*p2*p4*p8*p13+16*p3*p5*p8*p13+
16*p4*p6*p8*p13+16*p1*p7*p8*p13-16*p5*p7*p8*p13-8*p2*p8^2*p13+8*p6*p8^2*p13+24*p2*p9^2*p13-24*p6*p9^2*p13-
16*p1*p9*p10*p13+16*p5*p9*p10*p13-8*p2*p10^2*p13+8*p6*p10^2*p13-16*p4*p9*p11*p13+16*p8*p9*p11*p13
+16*p3*p10*p11*p13-16*p7*p10*p11*p13-8*p2*p11^2*p13+8*p6*p11^2*p13-16*p3*p9*p12*p13+16*p7*p9*p12*p13
```

Flattening



$$Flatt_{12|34}(P) = \begin{array}{c} \\ \mathtt{AA} \\ AC \\ AG \\ \vdots \\ TT \end{array} \begin{array}{ccccc} \mathtt{AA} & \mathtt{AC} & \mathtt{AG} & \dots & \mathtt{TT} \\ \hline p_{\mathtt{AAAA}} & p_{\mathtt{AAAC}} & p_{\mathtt{AAAG}} & \dots & p_{\mathtt{AATT}} \\ p_{\mathtt{ACAA}} & p_{\mathtt{ACAC}} & p_{\mathtt{ACAG}} & \dots & p_{\mathtt{ACTT}} \\ p_{\mathtt{AGAA}} & p_{\mathtt{AGAC}} & p_{\mathtt{AGAG}} & \dots & p_{\mathtt{AGTT}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{\mathtt{TTAA}} & p_{\mathtt{TTAC}} & p_{\mathtt{TTAG}} & \dots & p_{\mathtt{TTTT}} \end{array}$$

Flattening



$$Flatt_{12|34}(P) = \begin{array}{c|ccccc} & \mathtt{AA} & \mathtt{AC} & \mathtt{AG} & \dots & \mathtt{TT} \\ \hline \mathtt{AA} & p_{\mathtt{AAAA}} & p_{\mathtt{AAAC}} & p_{\mathtt{AAAG}} & \cdots & p_{\mathtt{AATT}} \\ AC & p_{\mathtt{ACAA}} & p_{\mathtt{ACAC}} & p_{\mathtt{ACAG}} & \cdots & p_{\mathtt{ACTT}} \\ AG & p_{\mathtt{AGAA}} & p_{\mathtt{AGAC}} & p_{\mathtt{AGAG}} & \cdots & p_{\mathtt{AGTT}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ TT & p_{\mathtt{TTAA}} & p_{\mathtt{TTAC}} & p_{\mathtt{TTAG}} & \cdots & p_{\mathtt{TTTT}} \end{array}$$

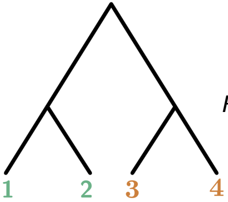
Theorem [Allman – Rhodes]

Let $P = \varphi_T(\pi, \{M_e\}_{e\in E(T)})$ where $T = T_{12|34}$. Then

$$\operatorname{rank}(Flatt_{12|34}(P)) \leq 4.$$

$Flatt_{13|24}(P)$ and $Flatt_{14|23}(P)$ have rank 16 for generic P.

Flattening



$$Flatt_{12|34}(P) = \begin{array}{c|ccccc} & \texttt{AA} & \texttt{AC} & \texttt{AG} & \dots & \texttt{TT} \\ \hline \texttt{AA} & p_{\texttt{AAAA}} & p_{\texttt{AAAC}} & p_{\texttt{AAAG}} & \dots & p_{\texttt{AATT}} \\ AC & p_{\texttt{ACAA}} & p_{\texttt{ACAC}} & p_{\texttt{ACAG}} & \dots & p_{\texttt{ACTT}} \\ AG & p_{\texttt{AGAA}} & p_{\texttt{AGAC}} & p_{\texttt{AGAG}} & \dots & p_{\texttt{AGTT}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ TT & p_{\texttt{TTAA}} & p_{\texttt{TTAC}} & p_{\texttt{TTAG}} & \dots & p_{\texttt{TTTT}} \end{array}$$

Theorem [Allman – Rhodes]

Let $P = \varphi_T(\pi, \{M_e\}_{e \in E(T)})$ where $T = T_{12|34}$. Then

$$\text{rank}(Flatt_{12|34}(P)) \leq 4.$$

$Flatt_{13|24}(P)$ and $Flatt_{14|23}(P)$ have rank 16 for generic P.

Therefore 5×5 minors of $Flatt_{12|34}(P)$ are **topology invariants**.

Algebraic phylogenetic reconstruction methods

The distance of an $m \times n$ matrix M to the set

$$\mathcal{R}_k = \{m \times n \text{ matrices of rank } \leq k\}$$

can be computed easily by,

Eckart-Young Theorem

$$d_k(M) = d_F(M, \mathcal{R}_k) = \sqrt{\sum_{i \geq k+1} \sigma_i^2},$$

where σ_i are the singular values of M.

Algebraic phylogenetic reconstruction methods

The distance of an $m \times n$ matrix M to the set

$$\mathcal{R}_k = \{m \times n \text{ matrices of rank } \leq k\}$$

can be computed easily by,

Eckart-Young Theorem

$$d_k(M) = d_F(M, \mathcal{R}_k) = \sqrt{\sum_{i \geq k+1} \sigma_i^2},$$

where σ_i are the singular values of M.

Phylogenetic reconstruction methods

Compute $d_4(Flatt_{A|B}(P))$ for the tree possible bipartitions. The lower the score is, the more it is likely that the bipartition comes from an edge of T.

Stochastic phylogenetic regions

Definition

The **stochastic phylogenetic regions** is defined as

$$\mathcal{V}_T^+ = \{P \in \mathcal{V}_T \mid P = \varphi_T(s) \text{ and } s \in S \subset [0,1]^d\},$$

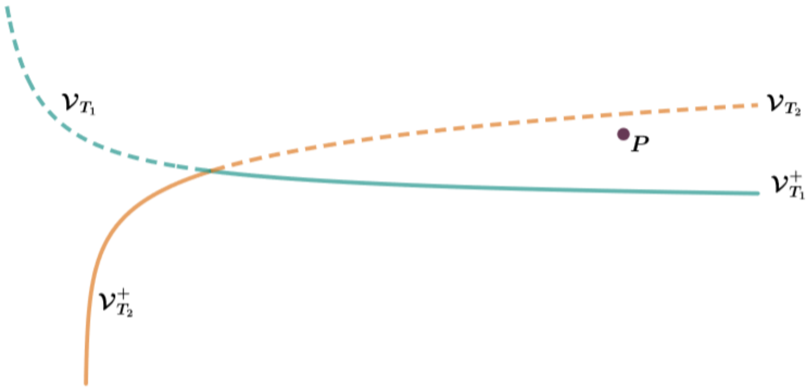
is the subset of $\mathcal{V}_T$ that contains distributions arising from stochastic parameters.

Stochastic Parameters

A vector π is stochastic iff its entries are non-negative and $\sum \pi_i = 1$.
A matrix is stochastic iff its entries are non-negative and

$$\sum_j M_e(i,j) = 1, \forall i, e$$

Could the stochastic varieties be useful for phylogenetic reconstruction?



Computing the distance to a Phylogenetic variety

Let $P = (p_1, \dots, p_{4^n}) \in \triangle^{4^n-1}$ be a distribution. We want to compute the distance of P to $\mathcal{V}_T^+$,

$$d(P, \mathcal{V}_T^+) = \min_{Q \in \mathcal{V}_T^+} d(P, Q)$$

Computing the distance to a Phylogenetic variety

Let $P = (p_1, \dots, p_{4^n}) \in \triangle^{4^n - 1}$ be a distribution. We want to compute the distance of P to $\mathcal{V}_T^+$,

$$d(P, \mathcal{V}_T^+) = \min_{Q \in \mathcal{V}_T^+} d(P, Q)$$

Since $Q \in \mathcal{V}_T^+$, we can write $Q = \varphi_T(x)$ with stochastic parameters $x \in \mathbb{R}^d$. Denote by $\Omega \subset \mathbb{R}^d$ the domain of stochastic parameters. Let

$$f_P(x) := d(P, \varphi_T(x)) = \sum_i^{4^n} (p_i - \varphi_i(x))^2.$$

Computing the distance to a Phylogenetic variety

Let $P = (p_1, \dots, p_{4^n}) \in \triangle^{4^n-1}$ be a distribution. We want to compute the distance of P to $\mathcal{V}_T^+$,

$$d(P, \mathcal{V}_T^+) = \min_{Q \in \mathcal{V}_T^+} d(P, Q)$$

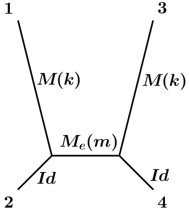
Since $Q \in \mathcal{V}_T^+$, we can write $Q = \varphi_T(x)$ with stochastic parameters $x \in \mathbb{R}^d$. Denote by $\Omega \subset \mathbb{R}^d$ the domain of stochastic parameters. Let

$$f_P(x) := d(P, \varphi_T(x)) = \sum_i^{4^n} (p_i - \varphi_i(x))^2.$$

If $P^+ = \varphi_T(x^*) \in \mathcal{V}_T^+$ is such that $d(P, P^+) = d(P, \mathcal{V}_T^+)$ then

- $(P - P^+) \perp T_P \mathcal{V}_T$, i.e. x^* is a critical point of $f_P(x)$
- x^* is not a critical point of $f_P(x)$ but $P^+ \in \partial\Omega$

Long branch attraction for JC model

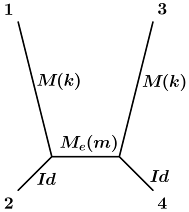


Let $P = \varphi_{12|34}\,(M, Id, M, Id, M_e)$.

Proposition [Casanellas – Fernández-Sánchez – G-L]

If M_e has negative off-diagonal entries and M is stochastic then $P^+ = \varphi_{12|34}(\tilde{M}, Id, \tilde{M}, Id, Id)$ is a local minimum of the distance function $d(P, \mathcal{V}_T^+)$.

Long branch attraction for JC model



Let $P = \varphi_{12|34}\,(M, Id, M, Id, M_e)$.

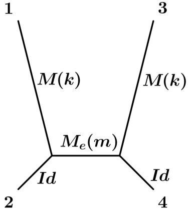
Proposition [Casanellas – Fernández-Sánchez – G-L]

If M_e has negative off-diagonal entries and M is stochastic then $P^+ = \varphi_{12|34}(\tilde{M}, Id, \tilde{M}, Id, Id)$ is a local minimum of the distance function $d(P, \mathcal{V}_T^+)$.

Conjecture: Global minumum

$$d(P, \mathcal{V}_T^+) = d\big(P, P^+\big).$$

Long branch attraction for JC model



Let $P = \varphi_{12|34}\,(M, Id, M, Id, M_e)$.

Proposition [Casanellas – Fernández-Sánchez – G-L]

If M_e has negative off-diagonal entries and M is stochastic then $P^+ = \varphi_{12|34}(\tilde{M}, Id, \tilde{M}, Id, Id)$ is a local minimum of the distance function $d(P, \mathcal{V}_T^+)$.

Conjecture: Global minumum

$$d(P, \mathcal{V}_T^+) = d\big(P, P^+\big).$$

Theorem [Casanellas – Fernández-Sánchez – G-L]

Let $P_0 = \varphi_{12|34}\,(M, Id, M, Id, M_e)$ such that $d(P_0, \mathcal{V}_T^+) = d\big(P_0, P^+\big)$ then, for any P close enough to P_0 we have

$$d(P, \mathcal{V}_T^+) \geq d(P, \mathcal{V}_{T_2}^+).$$

Computing the distance to a Phylogenetic variety

Lemma [Draisma – Horobet – Ottaviani – Sturmfels – Thomas]

For general $P \in \mathbb{C}^{4^n}$ the number of critical points of f_P on the manifold $\mathcal{V} \setminus \mathcal{V}_{sing}$ is finite and is called the **Euclidean Distance degree** of $\mathcal{V}$.

Computing the distance to a Phylogenetic variety

Lemma [Draisma – Horobet – Ottaviani – Sturmfels – Thomas]

For general $P \in \mathbb{C}^{4^n}$ the number of critical points of f_P on the manifold $\mathcal{V} \setminus \mathcal{V}_{sing}$ is finite and is called the **Euclidean Distance degree** of $\mathcal{V}$.

Computations difficulties

ED degree for the Jukes Cantor model on 4-leaf trees is 290.

- $>$ 2.5 months with *Macaluay2*.
- $\approx$ 2.5 hours with *Magma*.

Numerical Algebraic Geometry Only PHCpack founds the 290 solutions.

The computations were performed on a machine with 10 *Dual Core Intel(R) Xeon(R) Silver* 64 *Processor* 4114 *(*2.20 *GHz,* 13.75*M Cache) equipped with* 256 *GB RAM running Ubuntu* 18.04.2*.*

Computing the distance to a Phylogenetic variety

Lemma [Draisma – Horobet – Ottaviani – Sturmfels – Thomas]

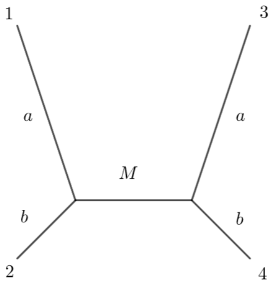
For general $P \in \mathbb{C}^{4^n}$ the number of critical points of f_P on the manifold $\mathcal{V} \setminus \mathcal{V}_{sing}$ is finite and is called the **Euclidean Distance degree** of $\mathcal{V}$.

Algorithm

1. Compute the Euclidean distance degree d for the variety $\mathcal{V}_T$.
2. Compute the d critical points x such that $\nabla f(x) = 0$ and $x \in \Omega$.
3. Compute the critical points $\nabla f = 0$ at the boundaries $\partial\Omega$.
4. Choose point with the lowest value when evaluated at f.

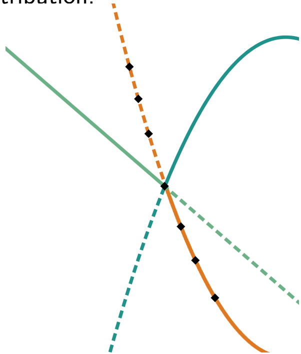
Simulations

- We took trees with branch lengths a and b at the exterior edges. M is a JC matrix with eigenvalue $m \in [0.94, 1.06]$.
- For each set of parameters we considered 100 data points, each corresponding to 10000 independent samples from the corresponding multinomial distribution.

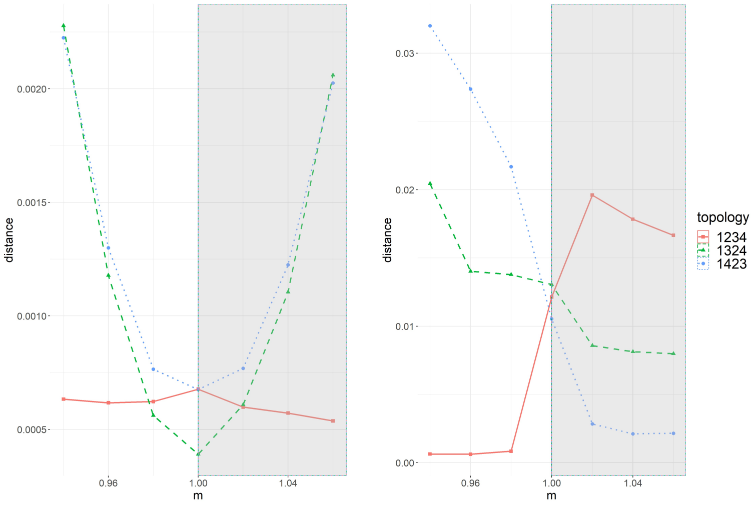


Simulations

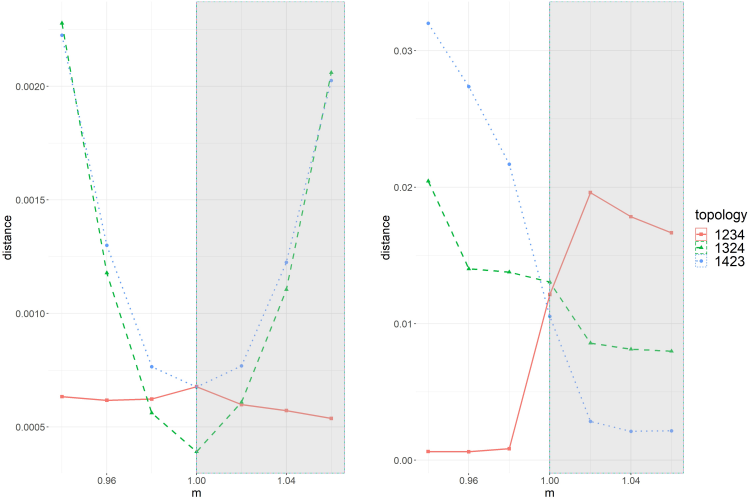
- We took trees with branch lengths a and b at the exterior edges. M is a JC matrix with eigenvalue $m \in [0.94, 1.06]$.
- For each set of parameters we considered 100 data points, each corresponding to 10000 independent samples from the corresponding multinomial distribution.



Simulations a = 0.5 & b = 0.5



Simulations a = 0.75 & b = 0.1



Stochastic conditions for the General Markov Model

Theorem [Allman – Rhodes – Taylor]

Let $P = \varphi_{\mathcal{T}}(\pi, \{M_e\}_{e\in E(\mathcal{T})})$ be a 4-tensor that arises from nonsingular real parameters for $GM(\kappa)$ model on $T_{12|34}$. If the marginalizations $P_{+\dots}$ and $P_{\dots+}$ arise from stochastic parameters and, moreover, the $\kappa^2 \times \kappa^2$ matrix

$$Flatt_{13|24}\left(P *_2 (adj(P^T_{+..+})P^T_{.+.+}) *_3 (adj(P_{.+.+})P_{.++.})\right)$$

is positive semidefinite, then P arises from stochastic parameters.

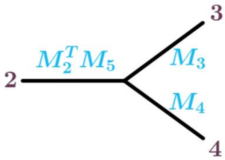
Stochastic conditions for the General Markov Model

Theorem [Allman – Rhodes – Taylor]

Let $P = \varphi_{\mathcal{T}}(\pi, \{M_e\}_{e \in E(\mathcal{T})})$ be a 4-tensor that arises from nonsingular real parameters for $GM(\kappa)$ model on $T_{12|34}$. If the marginalizations $P_{+\dots}$ and $P_{\dots+}$ arise from stochastic parameters and, moreover, the $\kappa^2 \times \kappa^2$ matrix

$$Flatt_{13|24}\left(P *_2 (adj(P^T_{+..+})P^T_{.+.+}) *_3 (adj(P_{.+.+})P_{.++.})\right)$$

is positive semidefinite, then P arises from stochastic parameters.



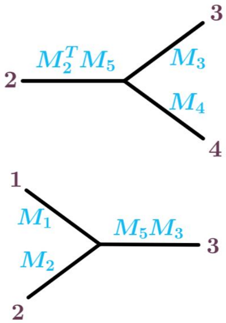
Stochastic conditions for the General Markov Model

Theorem [Allman – Rhodes – Taylor]

Let $P = \varphi_{\mathcal{T}}(\pi, \{M_e\}_{e \in E(\mathcal{T})})$ be a 4-tensor that arises from nonsingular real parameters for $GM(\kappa)$ model on $T_{12|34}$. If the marginalizations $P_{+\dots}$ and $P_{\dots+}$ arise from stochastic parameters and, moreover, the $\kappa^2 \times \kappa^2$ matrix

$$Flatt_{13|24}\left(P *_2 \left(adj(P^T_{+..+})P^T_{.+.+}\right) *_3 \left(adj(P_{.+.+})P_{.++.}\right)\right)$$

is positive semidefinite, then P arises from stochastic parameters.



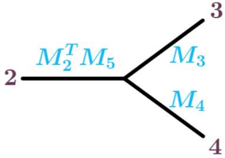
Stochastic conditions for the General Markov Model

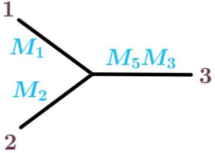
Theorem [Allman – Rhodes – Taylor]

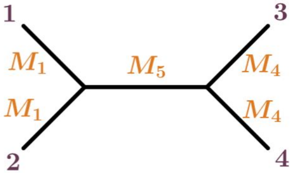
Let $P = \varphi_{\mathcal{T}}(\pi, \{M_e\}_{e \in E(\mathcal{T})})$ be a 4-tensor that arises from nonsingular real parameters for $GM(\kappa)$ model on $T_{12|34}$. If the marginalizations $P_{+\dots}$ and $P_{\dots+}$ arise from stochastic parameters and, moreover, the $\kappa^2 \times \kappa^2$ matrix

$$Flatt_{13|24}\left(P *_2 (adj(P^T_{+..+})P^T_{.+.+}) *_3 (adj(P_{.+.+})P_{.++.})\right)$$

is positive semidefinite, then P arises from stochastic parameters.







Stochastic conditions for the General Markov Model

Theorem (Casanellas, Fernández-Sánchez, G-L)

Let $P = \varphi_{\mathcal{T}}(\pi, \{M_e\}_{e \in E(\mathcal{T})})$ be a 4-tensor for $GM(\kappa)$ model on $T_{12|34}$. Let $\tilde{P}$ be constructed as in the previous theorem. Then,

$$Flat_{13|24}(\tilde{P}) = Flat_{14|23}(\tilde{P}),$$

and

$$Flatt_{12|34}(\tilde{P}) \neq Flatt_{13|24}(\tilde{P}).$$

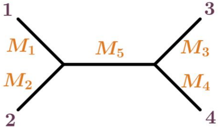
In particular

$$\begin{aligned} & det(P_{+..+})det(P_{.+.+})Flatt_{13|24}\big(P *_2 (adj(P^T_{+..+})P^T_{.+.+})) *_3 (adj(P_{.+.+})P_{.++.}))) \\ =& det(P_{+..+})det(P_{.+.+})Flatt_{14|23}\big(P *_2 (adj(P^T_{+..+})P^T_{.+.+})) *_3 (adj(P_{.+.+})P_{.++.})) \end{aligned}$$

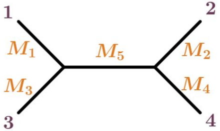
gives rise to 256 topology invariants of degree 17.

T Leaf-transformations

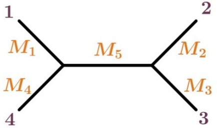
$T_{12|34}$



$T_{13|24}$

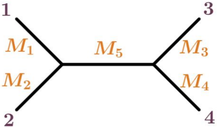


$T_{14|23}$



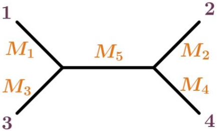
T Leaf-transformations

$T_{12|34}$



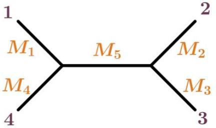
$\rightarrow \alpha_i^{12}(P_{12})$

$T_{13|24}$



$\rightarrow \alpha_i^{13}(P_{13})$

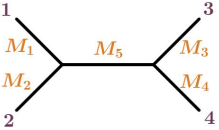
$T_{14|23}$



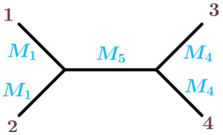
$\rightarrow \alpha_i^{14}(P_{14})$

T Leaf-transformations

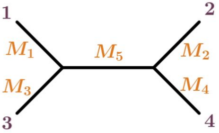
$T_{12|34}$



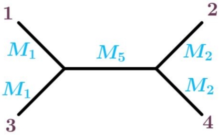
$\rightarrow \alpha_i^{12}(P_{12}) \rightarrow$



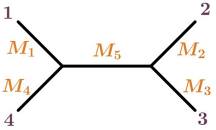
$T_{13|24}$



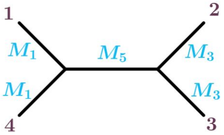
$\rightarrow \alpha_i^{13}(P_{13}) \rightarrow$



$T_{14|23}$

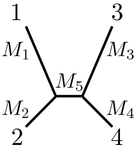


$\rightarrow \alpha_i^{14}(P_{14}) \rightarrow$

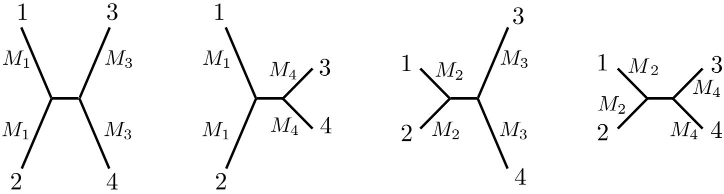


12|34 **Leaf-transformations**

Original tree T

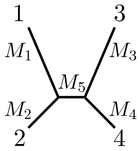


Resulting trees associated with the 12|34 leaf-transformations on the (theoretical) distribution from T

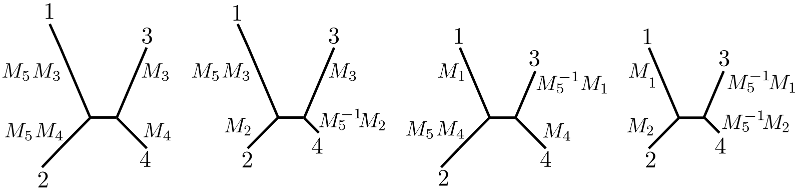


13|24 **Leaf-transformations**

Original tree T

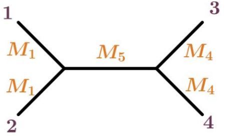


Resulting trees associated with some 13|24 leaf-transformations on the (theoretical) distribution from T

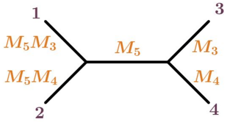


Leaf-transformations on distributions of $T = 12|34$

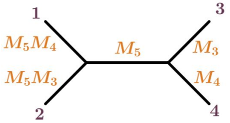
$\alpha_i^{12}(P)$



$\alpha_i^{13}(P)$

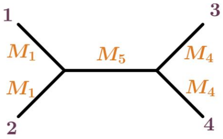


$\alpha_i^{14}(P)$



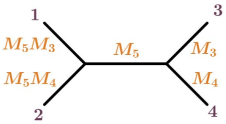
Leaf-transformations on distributions of $T = 12|34$

$\alpha_i^{12}(P)$



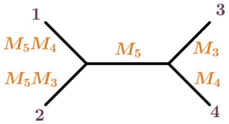
$$\Rightarrow \begin{cases} Flatt_{12|34}(\alpha_i^{12}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \checkmark \\ Flatt_{13|24}(\alpha_i^{12}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \text{✗} \\ Flatt_{14|23}(\alpha_i^{12}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \text{✗} \end{cases}$$

$\alpha_i^{13}(P)$



$$\Rightarrow \begin{cases} Flatt_{13|24}(\alpha_i^{13}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \text{✗} \\ Flatt_{12|34}(\alpha_i^{13}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \checkmark \\ Flatt_{14|32}(\alpha_i^{13}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \text{✗} \end{cases}$$

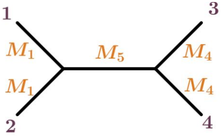
$\alpha_i^{14}(P)$



$$\Rightarrow \begin{cases} Flatt_{14|23}(\alpha_i^{14}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \text{✗} \\ Flatt_{12|43}(\alpha_i^{14}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \checkmark \\ Flatt_{13|42}(\alpha_i^{14}(P)) & \rightarrow \quad \text{rank} \leq 4 \ \text{✗} \end{cases}$$

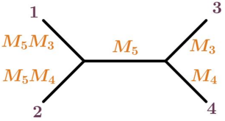
Leaf-transformations on distributions of $T = 12|34$

$\alpha_i^{12}(P)$



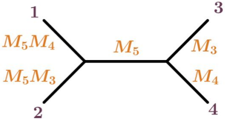
$$\Rightarrow \begin{cases} Flatt_{12|34}(\alpha_i^{12}(P)) & \rightarrow \quad \text{PSD ✗} \\ Flatt_{13|24}(\alpha_i^{12}(P)) & \rightarrow \quad \text{PSD ✓} \\ Flatt_{14|23}(\alpha_i^{12}(P)) & \rightarrow \quad \text{PSD ✓} \end{cases}$$

$\alpha_i^{13}(P)$



$$\Rightarrow \begin{cases} Flatt_{13|24}(\alpha_i^{13}(P)) & \rightarrow \quad \text{PSD ✗} \\ Flatt_{12|34}(\alpha_i^{13}(P)) & \rightarrow \quad \text{PSD ✗} \\ Flatt_{14|32}(\alpha_i^{13}(P)) & \rightarrow \quad \text{PSD ✗} \end{cases}$$

$\alpha_i^{14}(P)$



$$\Rightarrow \begin{cases} Flatt_{14|23}(\alpha_i^{14}(P)) & \rightarrow \quad \text{PSD ✗} \\ Flatt_{12|43}(\alpha_i^{14}(P)) & \rightarrow \quad \text{PSD ✗} \\ Flatt_{13|42}(\alpha_i^{14}(P)) & \rightarrow \quad \text{PSD ✗} \end{cases}$$

SAQ: semi-algebraic quartet reconstruction method

Theorem [Casanellas – Fernández-Sánchez – G-L]

The rank of the *psd* approximation of a real matrix M is less than or equal to $rank(M)$.

SAQ: semi-algebraic quartet reconstruction method

Theorem [Casanellas – Fernández-Sánchez – G-L]

The rank of the *psd* approximation of a real matrix M is less than or equal to $rank(M)$.

Lemma [Casanellas – Fernández-Sánchez – G-L]

Let P be the theoretical distribution from a 3-parameter Kimura process on the quartet tree $T = 12|34$. Then, the rank of the *psd* approximation of the flattening matrix $Flat_{T'}(\alpha^{T'}(P))$ is grater than 4 for $T' \neq T$.

SAQ: semi-algebraic quartet reconstruction method

SAQ method

Let P be a data point obtained from an alignment, then the score for $T = 12|34$ is:

$$s^{i}_{12|34} := \frac{\min\left\{\delta_4\left(psd\left(Flatt_{13|24}\left(\alpha_i^{12}(P)\right)\right)\right), \delta_4\left(psd\left(Flatt_{14|23}\left(\alpha_i^{12}(P)\right)\right)\right)\right\}}{\delta_4\left(psd\left(Flatt_{12|34}\left(\alpha_i^{12}(P)\right)\right)\right)}$$

SAQ: semi-algebraic quartet reconstruction method

SAQ method

Let P be a data point obtained from an alignment, then the score for $T = 12|34$ is:

$$s^i_{12|34} := \frac{\min\left\{\delta_4\left(psd\left(Flatt_{13|24}\left(\alpha_i^{12}(P)\right)\right)\right), \delta_4\left(psd\left(Flatt_{14|23}\left(\alpha_i^{12}(P)\right)\right)\right)\right\}}{\delta_4\left(psd\left(Flatt_{12|34}\left(\alpha_i^{12}(P)\right)\right)\right)}$$

$$\text{and } s_{12|34} := \text{mean}_i\{s^i_{12|34}\}$$

SAQ: semi-algebraic quartet reconstruction method

SAQ method

Let P be a data point obtained from an alignment, then the score for $T = 12|34$ is:

$$s^i_{12|34} := \frac{\min\left\{\delta_4\left(psd\left(Flatt_{13|24}\left(\alpha_i^{12}(P)\right)\right)\right), \delta_4\left(psd\left(Flatt_{14|23}\left(\alpha_i^{12}(P)\right)\right)\right)\right\}}{\delta_4\left(psd\left(Flatt_{12|34}\left(\alpha_i^{12}(P)\right)\right)\right)}$$

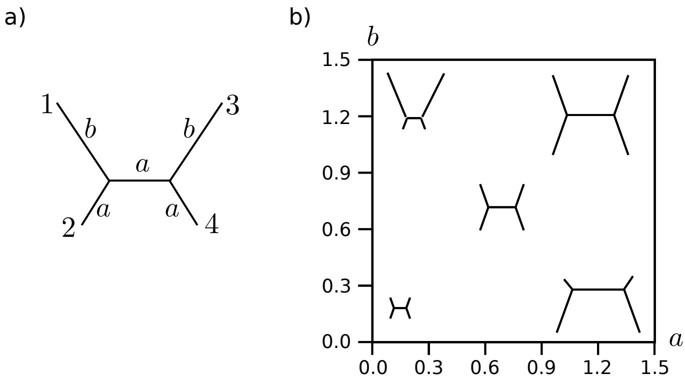
$$\text{and } s_{12|34} := \text{mean}_i\{s^i_{12|34}\}$$

$$\mathtt{SAQ}(P) := \frac{1}{s_{12|34}(P) + s_{13|24}(P) + s_{14|23}(P)}\left(s_{12|34}(P), s_{13|24}(P), s_{14|23}(P)\right).$$

If $Q \in \mathbb{R}^{256}$ is a distribution that tends to P generated on the tree $12|34$ with generic stochastic parameters, then

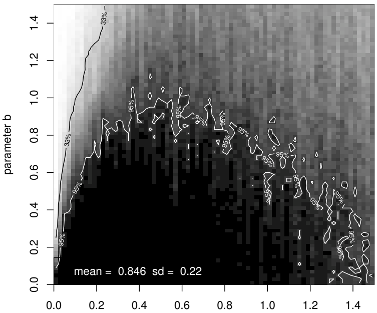
$$\lim_{Q \to P} \mathtt{SAQ}(Q) = \mathtt{SAQ}(P) = (1, 0, 0).$$

Simulations: Tree Space

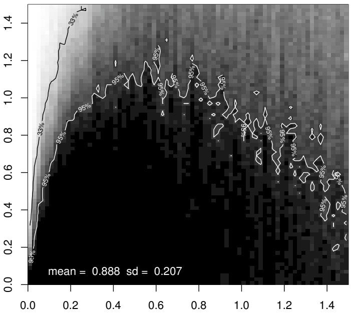


Simulations: Tree Space

GM; length 500 bp



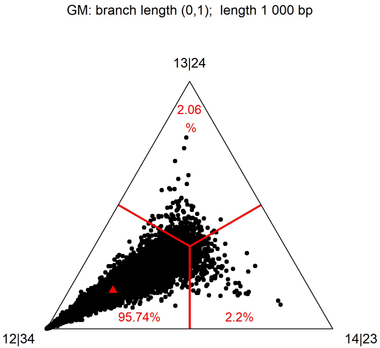
GM; length 1 000 bp



base pairs	SAQ	Erik+2	NJ	ML
500	84.6	72.4	72.5	72.1
1 000	88.8	80.3	79.7	73.6

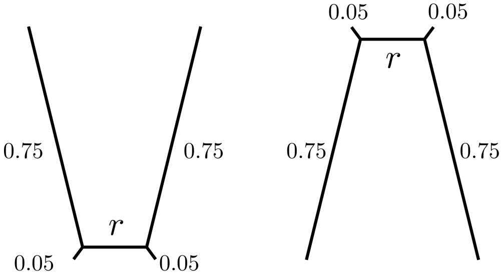
Simulations: Random branch lengths

A total of 10 000 alignments are considered, obtained from 4-taxa trees with random branch lengths uniformly distributed in the interval (0,1), and generated according to the General Markov substitution model.

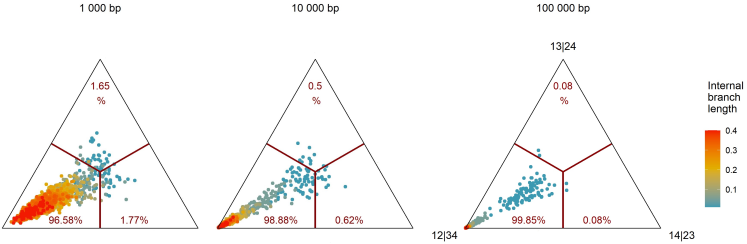


Simulations: Mixture models

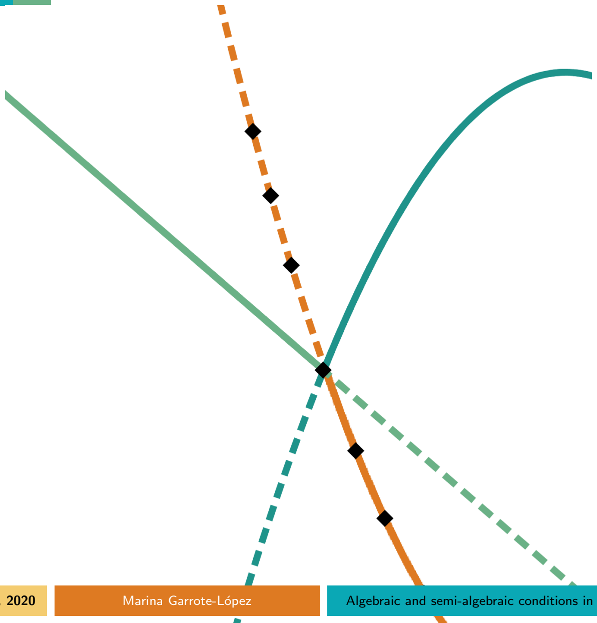
Mixture data



Simulations: Mixture models



internal branch length	0.01	0.05	0.1	0.2	0.3
SAQ	37	83	96	100	100
Erik+2 (2)	12	35	60	86	96
MP	0	2	19	76	99
ML(GTR+2 Γ)	0	4	14	77	95



Thanks for your attention!