RADICAL PHYLOGENETIC INVERSION

Mike Hendy

Institute of Fundamental Sciences
Massey University
Palmerston North
New Zealand

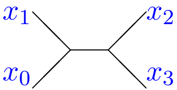
November 2010

Acknowledgements

- David Penny, Massey University
- Mike Steel, Canterbury University
- Peter Waddell, University of South Carolina
- Andreas Dress, Universität Bielefeld

Local to Global

- $X = \{x_0, x_1, \cdots, x_n\}$ is a set of $n+1$ taxa, and T is an X–tree (the leaves represent the taxa).
  Example $X = \{x_0, x_1, x_2, x_3\}$:



Local to Global

- $X = \{x_0, x_1, \cdots, x_n\}$ is a set of $n+1$ taxa, and T is an X–tree (the leaves represent the taxa).
  Example $X = \{x_0, x_1, x_2, x_3\}$:

- Nucleotide sequences evolve from an ancestral sequence, under a model of nucleotide substitution governed by **local** stochastic matrices M_e on each edge of T.

Local to Global

- $X = \{x_0, x_1, \cdots, x_n\}$ is a set of $n+1$ taxa, and T is an X–tree (the leaves represent the taxa).
  Example $X = \{x_0, x_1, x_2, x_3\}$:

- Nucleotide sequences evolve from an ancestral sequence, under a model of nucleotide substitution governed by **local** stochastic matrices M_e on each edge of T.
- We (usually) can only observe are the **global** data, the aligned homologous sequences for the taxa $x_i \in X$.

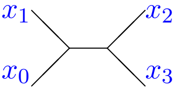
ATGGTAT ACGGTGT

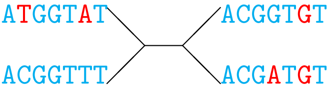
?

ACGGTTT ACGATGT

Local to Global

- $X = \{x_0, x_1, \cdots, x_n\}$ is a set of $n+1$ taxa, and T is an X–tree (the leaves represent the taxa).
  Example $X = \{x_0, x_1, x_2, x_3\}$:





- Nucleotide sequences evolve from an ancestral sequence, under a model of nucleotide substitution governed by **local** stochastic matrices M_e on each edge of T.
- We (usually) can only observe are the **global** data, the aligned homologous sequences for the taxa $x_i \in X$.

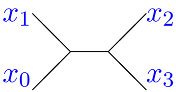
ATGGTAT ACGGTGT

?

ACGGTTT ACGATGT

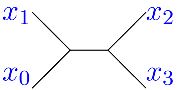
- The **inverse** problem of phylogenetics, deduce the local structure of T, and (if possible) the stochastic matrices M_e, from the global data (aligned sequences).

Global to Local - Pathset Group

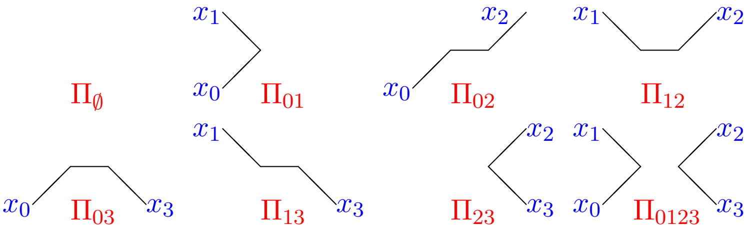


- The path Π_{ij} is the set of edges of T connecting the leaves x_i,x_j. The **pathset group** of T is the group generated by the paths Π_{01}, Π_{02},$\cdots$ Π_{0n} under disjoint union.

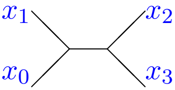
Global to Local - Pathset Group

- 

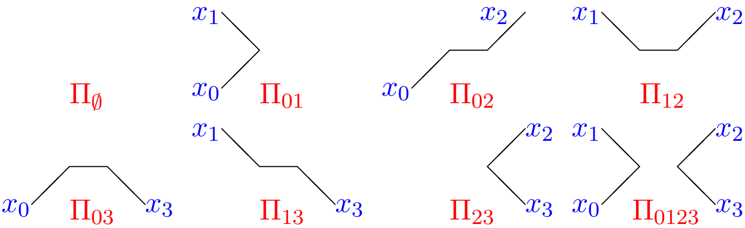
  The path Π_{ij} is the set of edges of T connecting the leaves x_i,x_j. The **pathset group** of T is the group generated by the paths Π_{01}, Π_{02},$\cdots$ Π_{0n} under disjoint union.
- The 8 pathsets of the pathset group of T_{23}:



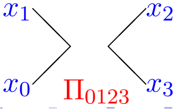
Global to Local - Pathset Group

- 

The path Π_{ij} is the set of edges of T connecting the leaves x_i,x_j. The **pathset group** of T is the group generated by the paths Π_{01}, Π_{02},$\cdots$ Π_{0n} under disjoint union.
- The 8 pathsets of the pathset group of T_{23}:

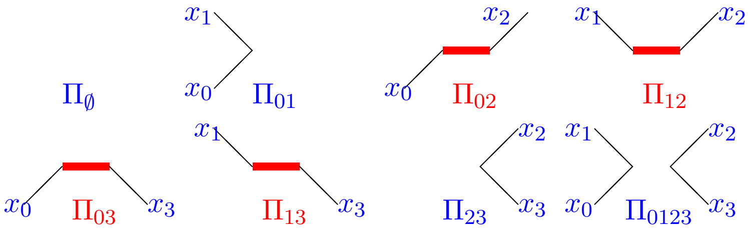


- The pathset Π_{0123} is the disjoint union of Π_{01}, Π_{02} and Π_{03}.



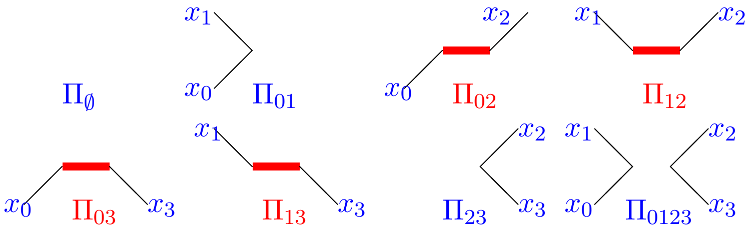
Global to Local - Pathsets

- The edge e_{23} belongs to the four pathsets Π_B where $|B \cap \{2,3\}|$ is an odd number.



Global to Local - Pathsets

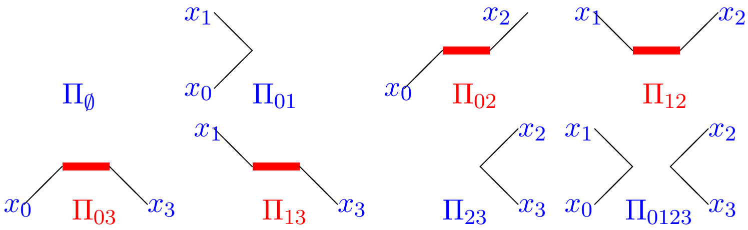
- The edge e_{23} belongs to the four pathsets Π_B where $|B \cap \{2,3\}|$ is an odd number.



- In general edge e_A in Π_B iff $|A \cap B|$ is odd.

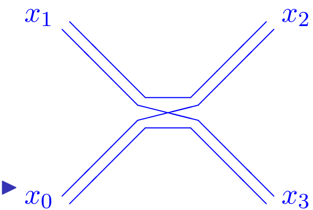
Global to Local - Pathsets

- The edge e_{23} belongs to the four pathsets Π_B where $|B \cap \{2,3\}|$ is an odd number.

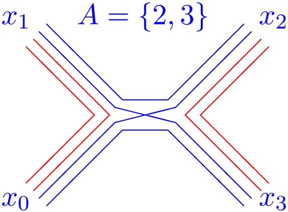


- In general edge e_A in Π_B iff $|A \cap B|$ is odd.
- An X–tree has $2^{|X|-1} = 2^n$ pathsets.

Global to Local - Pathsets

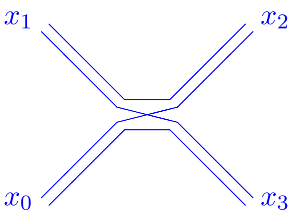


$\Pi_{02} + \Pi_{12} + \Pi_{03} + \Pi_{13}$

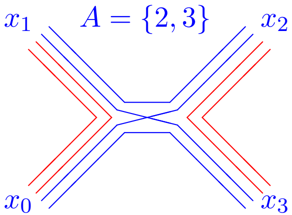


$\Pi_{01} + \Pi_{12} + \Pi_{03} + \Pi_{13}$
$-\Pi_{\emptyset} - \Pi_{01} - \Pi_{23} - \Pi_{0123}$

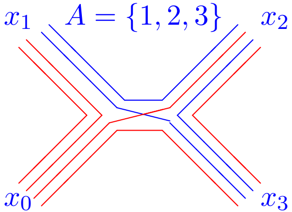
Global to Local - Pathsets



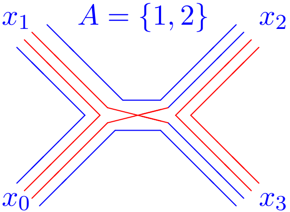
$\Pi_{02} + \Pi_{12} + \Pi_{03} + \Pi_{13}$



$\Pi_{01} + \Pi_{12} + \Pi_{03} + \Pi_{13}$
$-\Pi_{\emptyset} - \Pi_{01} - \Pi_{23} - \Pi_{0123}$



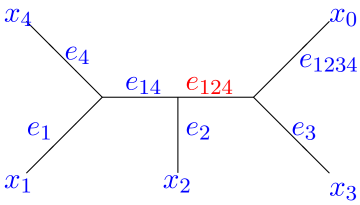
$\Pi_{01} + \Pi_{02} + \Pi_{03} + \Pi_{0123}$
$-\Pi_{\emptyset} - \Pi_{12} - \Pi_{13} - \Pi_{23}$



$\Pi_{01} + \Pi_{02} + \Pi_{13} + \Pi_{23}$
$-\Pi_{\emptyset} - \Pi_{12} - \Pi_{03} - \Pi_{0123}$

Pathsets

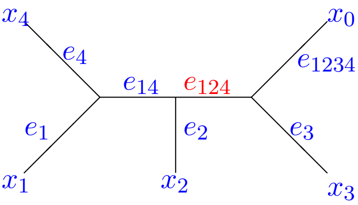
- Each edge $e_A \in E(T)$ belongs to the $2^{|X|-2}$ pathsets Π_B with $|A \cap B|$ odd.



The edge e_{124} belongs to the 8 pathsets Π_{01}, Π_{02}, Π_{13}, Π_{23}, Π_{04}, Π_{0124}, Π_{34} and Π_{1234}.

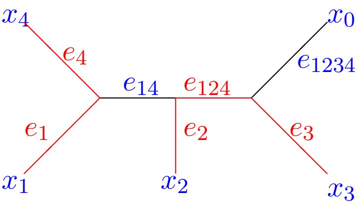
Pathsets

- Each edge $e_A \in E(T)$ belongs to the $2^{|X|-2}$ pathsets Π_B with $|A \cap B|$ odd.



The edge e_{124} belongs to the 8 pathsets Π_{01}, Π_{02}, Π_{13}, Π_{23}, Π_{04}, Π_{0124}, Π_{34} and Π_{1234}.

- Each pathset Π_B comprises all edges $e_A \in E(T)$ with $|A \cap B|$ odd.



The pathset Π_{1234} contains the edges e_1, e_2, e_3, e_4 and e_{124}.

Global to Local - Pathweights

- Suppose $w\colon E(T) \to \mathbb{R}$ is an edge-weighting function. Then for any even-ordered subset $A \subseteq X$, the weight of the pathset Π_A is
$$w(\Pi_A) = \sum_{\substack{e_B \in E(T):\\ |A \cap B| \equiv 1 \pmod 2}} w(e_B).$$

Global to Local - Pathweights

- Suppose $w\colon E(T) \to \mathbb{R}$ is an edge-weighting function. Then for any even-ordered subset $A \subseteq X$, the weight of the pathset Π_A is

$$w(\Pi_A) = \sum_{\substack{e_B \in E(T):\\ |A\cap B| \equiv 1 (\text{mod } 2)}} w(e_B).$$

- Given the set of **global** pathset weights, we can find the **local** edgeweights by

$$2^{|X|-2} w(e_B) = \sum_{\substack{A \subseteq X:\\ |A| \equiv 0 (\text{mod } 2)\\ |A\cap B| \equiv 1 (\text{mod } 2)}} w(\Pi_A) - \sum_{\substack{A \subseteq X:\\ |A| \equiv 0 (\text{mod } 2)\\ |A\cap B| \equiv 0 (\text{mod } 2)}} w(\Pi_A)$$

Global to Local - Pathweights

- Suppose $w\colon E(T) \to \mathbb{R}$ is an edge-weighting function. Then for any even-ordered subset $A \subseteq X$, the weight of the pathset Π_A is
$$w(\Pi_A) = \sum_{\substack{e_B \in E(T):\\ |A\cap B| \equiv 1 \pmod 2}} w(e_B).$$
- Given the set of **global** pathset weights, we can find the **local** edgeweights by
$$2^{|X|-2} w(e_B) = \sum_{\substack{A \subseteq X:\\ |A| \equiv 0 \pmod 2\\ |A\cap B| \equiv 1 \pmod 2}} w(\Pi_A) \;-\; \sum_{\substack{A \subseteq X:\\ |A| \equiv 0 \pmod 2\\ |A\cap B| \equiv 0 \pmod 2}} w(\Pi_A)$$
- Example: $w(e_B) = -\ln(\det(M_{e_B}))$, the log-det weight of the transition matrices. If we can measure the log-det of a pathset, we can determine the log-det of each edge of T.

Global to Local - Pathweights

- The transition matrice multiply along paths, their determinants multiply across the edges of a path set. If we can determine

$$\det(M_{\Pi_B}) = \prod_{e_A \in E(T):\ |A \cap B| \equiv 1 \pmod 2} \det(M_{e_A}),$$

from the sequences at the leaves, then we can calculate the determinants of each each stochastic matrix

Global to Local - Pathweights

- The transition matrice multiply along paths, their determinants multiply across the edges of a path set. If we can determine

$$\det(M_{\Pi_B}) = \prod_{e_A \in E(T):\ |A\cap B| \equiv 1 (\text{mod } 2)} \det(M_{e_A}),$$

  from the sequences at the leaves, then we can calculate the determinants of each each stochastic matrix

- $$(\det(M_{e_A}))^{2^{|X|-2}} = \prod_{\substack{A\subseteq X:\\ |A|\equiv 0 (\text{mod } 2)\\ |A\cap B|\equiv 1 (\text{mod } 2)}} \det(M_{\Pi_B}) \prod_{\substack{A\subseteq X:\\ |A|\equiv 0 (\text{mod } 2)\\ |A\cap B|\equiv 0 (\text{mod } 2)}} \det(M_{\Pi_B})^{-1}.$$

Global to Local - Pathweights

- The transition matrice multiply along paths, their determinants multiply across the edges of a path set. If we can determine

$$\det(M_{\Pi_B}) = \prod_{e_A \in E(T):\ |A\cap B| \equiv 1 (\text{mod } 2)} \det(M_{e_A}),$$

from the sequences at the leaves, then we can calculate the determinants of each each stochastic matrix

- $$(\det(M_{e_A}))^{2^{|X|-2}} = \prod_{\substack{A\subseteq X: \\ |A|\equiv 0 (\text{mod } 2) \\ |A\cap B|\equiv 1 (\text{mod } 2)}} \det(M_{\Pi_B}) \prod_{\substack{A\subseteq X: \\ |A|\equiv 0 (\text{mod } 2) \\ |A\cap B|\equiv 0 (\text{mod } 2)}} \det(M_{\Pi_B})^{-1}.$$

- The product above has value 1 iff there is no edge with edge-split A. This can determine the toplogy of T.

Neyman's (Cavender, Farris) Model

- Given a set $X = \{x_0, x_1, \cdots, x_n\}$ of $n + 1$ taxa on a phylogenetic tree T.

Neyman's (Cavender, Farris) Model

- Given a set $X = \{x_0, x_1, \cdots, x_n\}$ of $n+1$ taxa on a phylogenetic tree T.
- Each edge e of T is assigned the probability p_e that the character states at the endpoints of e differ. (p_e is independent of the states at its endpoints (symmetric)).

Neyman's (Cavender, Farris) Model

- Given a set $X = \{x_0, x_1, \cdots, x_n\}$ of $n+1$ taxa on a phylogenetic tree T.
- Each edge e of T is assigned the probability p_e that the character states at the endpoints of e differ. (p_e is independent of the states at its endpoints (symmetric)).
- The stochastic matrix for edge e is therefore

$$P_e = \begin{bmatrix} 1-p_e & p_e \\ p_e & 1-p_e \end{bmatrix}.$$

Neyman's (Cavender, Farris) Model

- Given a set $X = \{x_0, x_1, \cdots, x_n\}$ of $n+1$ taxa on a phylogenetic tree T.
- Each edge e of T is assigned the probability p_e that the character states at the endpoints of e differ. (p_e is independent of the states at its endpoints (symmetric)).
- The stochastic matrix for edge e is therefore

$$P_e = \begin{bmatrix} 1-p_e & p_e \\ p_e & 1-p_e \end{bmatrix}.$$

- At a site the (notional) root x_0 is assigned a character $\chi(x_0) \in \{\mathtt{R}, \mathtt{Y}\}$.

Neyman's (Cavender, Farris) Model

- Given a set $X = \{x_0, x_1, \cdots, x_n\}$ of $n+1$ taxa on a phylogenetic tree T.
- Each edge e of T is assigned the probability p_e that the character states at the endpoints of e differ. (p_e is independent of the states at its endpoints (symmetric)).
- The stochastic matrix for edge e is therefore

$$P_e = \begin{bmatrix} 1-p_e & p_e \\ p_e & 1-p_e \end{bmatrix}.$$

- At a site the (notional) root x_0 is assigned a character $\chi(x_0) \in \{\mathtt{R}, \mathtt{Y}\}$.
- Each taxon has a homologous 2–state (R and Y) sequence generated recursively by the parameters p_e.

Neyman's (Cavender, Farris) Model

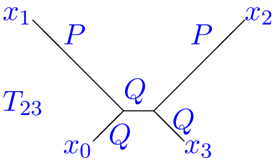
- Given a set $X = \{x_0, x_1, \cdots, x_n\}$ of $n+1$ taxa on a phylogenetic tree T.
- Each edge e of T is assigned the probability p_e that the character states at the endpoints of e differ. (p_e is independent of the states at its endpoints (symmetric)).
- The stochastic matrix for edge e is therefore

$$P_e = \begin{bmatrix} 1-p_e & p_e \\ p_e & 1-p_e \end{bmatrix}.$$

- At a site the (notional) root x_0 is assigned a character $\chi(x_0) \in \{\mathtt{R}, \mathtt{Y}\}$.
- Each taxon has a homologous 2–state (R and Y) sequence generated recursively by the parameters p_e.
- The probability of each possible **site pattern** (assignment of states at the leaves of T) is a polynomial function of the p_e terms.

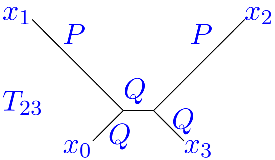
Example - Felsenstein

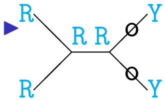
If at a site, x_0 is assigned the state R, $p_{\texttt{RRYY}}$, the probability of site pattern RRYY, is obtained by summing the probabilities of four subcases of assigning states at the internal vertices, given the substitution probabilities P and Q on the edges of T.



Example - Felsenstein

If at a site, x_0 is assigned the state R, $p_{\texttt{RRYY}}$, the probability of site pattern RRYY, is obtained by summing the probabilities of four subcases of assigning states at the internal vertices, given the substitution probabilities P and Q on the edges of T.

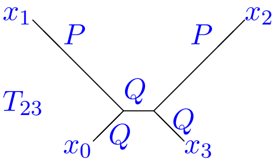


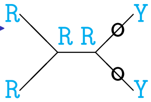


$$(1-P)(1-Q)^2PQ$$

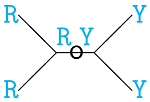
Example - Felsenstein

If at a site, x_0 is assigned the state R, $p_{\texttt{RRYY}}$, the probability of site pattern RRYY, is obtained by summing the probabilities of four subcases of assigning states at the internal vertices, given the substitution probabilities P and Q on the edges of T.



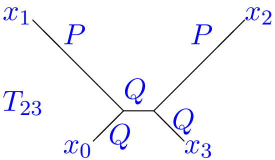
- 

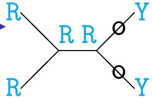
$(1-P)(1-Q)^2PQ$

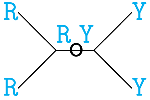
- $(1-P)^2(1-Q)^2Q$


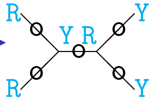
Example - Felsenstein

If at a site, x_0 is assigned the state R, $p_{\mathtt{RRYY}}$, the probability of site pattern RRYY, is obtained by summing the probabilities of four subcases of assigning states at the internal vertices, given the substitution probabilities P and Q on the edges of T.



- 

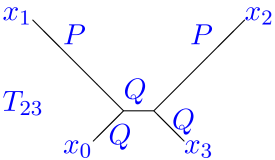
$(1-P)(1-Q)^2PQ$
- $(1-P)^2(1-Q)^2Q$


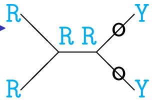
- 

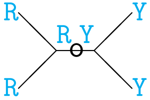
P^2Q^3

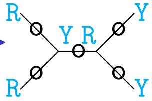
Example - Felsenstein

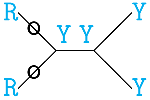
If at a site, x_0 is assigned the state R, p_{RRYY}, the probability of site pattern RRYY, is obtained by summing the probabilities of four subcases of assigning states at the internal vertices, given the substitution probabilities P and Q on the edges of T.



- 

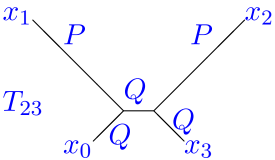
$(1-P)(1-Q)^2PQ$
- $(1-P)^2(1-Q)^2Q$


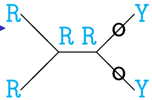
- 

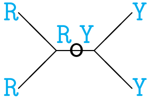
P^2Q^3
- $(1-P)(1-Q)^2PQ$


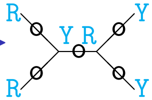
Example - Felsenstein

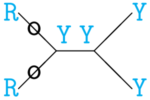
If at a site, x_0 is assigned the state R, $p_{\texttt{RRYY}}$, the probability of site pattern RRYY, is obtained by summing the probabilities of four subcases of assigning states at the internal vertices, given the substitution probabilities P and Q on the edges of T.



- 

$(1-P)(1-Q)^2PQ$
- $(1-P)^2(1-Q)^2Q$


- 

P^2Q^3
- $(1-P)(1-Q)^2PQ$


- Summing: $p_{\texttt{RRYY}} = Q(1-2Q+Q^2-P^2+2P^2Q)$.

Inversion

- Given the character $\chi(x_0) = \mathtt{R}$ at a site, the probability of site pattern $\mathtt{RRYY}$ ($\chi(x_1) = \mathtt{R}$, $\chi(x_2) = \mathtt{Y}$, $\chi(x_3) = \mathtt{Y}$) is

$$p_{\mathtt{RRYY}} = Q(1 - 2Q + Q^2 - P^2 + 2P^2Q).$$

Inversion

- Given the character $\chi(x_0) = \texttt{R}$ at a site, the probability of site pattern $\texttt{RRYY}$ ($\chi(x_1) = \texttt{R}$, $\chi(x_2) = \texttt{Y}$, $\chi(x_3) = \texttt{Y}$) is

$$p_{\texttt{RRYY}} = Q(1 - 2Q + Q^2 - P^2 + 2P^2Q).$$

- Similar formulae can be derived for $p_{\texttt{RRRR}}, \cdots, p_{\texttt{RYYY}}$.

Inversion

- Given the character $\chi(x_0) = \mathtt{R}$ at a site, the probability of site pattern $\mathtt{RRYY}$ ($\chi(x_1) = \mathtt{R}$, $\chi(x_2) = \mathtt{Y}$, $\chi(x_3) = \mathtt{Y}$) is

$$p_{\mathtt{RRYY}} = Q(1 - 2Q + Q^2 - P^2 + 2P^2Q).$$

- Similar formulae can be derived for $p_{\mathtt{RRRR}}, \cdots, p_{\mathtt{RYYY}}$.
- Inverting, P and Q can be expressed as rational functions (with radicals) of the pattern probabilities:

$$P = \frac{1}{2}\left(1 - \sqrt[4]{\frac{\mu_{12}\mu_{0123}}{\mu_{03}}}\right), \qquad Q = \frac{1}{2}\left(1 - \sqrt[4]{\frac{\mu_{03}\mu_{0123}}{\mu_{12}}}\right), \tag{1}$$

where

$$\mu_{12} = 1 - 2(p_{\mathtt{RYRR}} + p_{\mathtt{RRYR}} + p_{\mathtt{RYRY}} + p_{\mathtt{RRYY}}),$$

$$\mu_{0123} = 1 - 2(p_{\mathtt{RYRR}} + p_{\mathtt{RRYR}} + p_{\mathtt{RRRY}} + p_{\mathtt{RYYY}}),$$

$$\mu_{03} = 1 - 2(p_{\mathtt{RRRY}} + p_{\mathtt{RYRY}} + p_{\mathtt{RRYY}} + p_{\mathtt{RYYY}}).$$

Eigenvalues

- The stochastic matrices $M_{e_1} = M_{e_2} = \begin{bmatrix} 1-P & P \\ P & 1-P \end{bmatrix}$, $M_{e_3} = M_{e_{23}} = M_{e_{123}} = \begin{bmatrix} 1-Q & Q \\ Q & 1-Q \end{bmatrix}$ describe the substitution probabilities on the edges of T_{23}.

Eigenvalues

- The stochastic matrices $M_{e_1} = M_{e_2} = \begin{bmatrix} 1-P & P \\ P & 1-P \end{bmatrix}$, $M_{e_3} = M_{e_{23}} = M_{e_{123}} = \begin{bmatrix} 1-Q & Q \\ Q & 1-Q \end{bmatrix}$ describe the substitution probabilities on the edges of T_{23}.
- These matrices have a common diagonalising matrix, $H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, with $H_1^{-1} M_{e_1} H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1-2P \end{bmatrix}$.

Eigenvalues

- The stochastic matrices $M_{e_1} = M_{e_2} = \begin{bmatrix} 1-P & P \\ P & 1-P \end{bmatrix}$, $M_{e_3} = M_{e_{23}} = M_{e_{123}} = \begin{bmatrix} 1-Q & Q \\ Q & 1-Q \end{bmatrix}$ describe the substitution probabilities on the edges of T_{23}.
- These matrices have a common diagonalising matrix, $H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, with $H_1^{-1} M_{e_1} H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1-2P \end{bmatrix}$.
- Hence the M_{e_1} has eigenvalues $1, \lambda_1 = 1 - 2P$, and M_{e_3} has eigenvalues $1, \lambda_3 = 1 - 2Q$.

Eigenvalues

- The stochastic matrices $M_{e_1} = M_{e_2} = \begin{bmatrix} 1-P & P \\ P & 1-P \end{bmatrix}$, $M_{e_3} = M_{e_{23}} = M_{e_{123}} = \begin{bmatrix} 1-Q & Q \\ Q & 1-Q \end{bmatrix}$ describe the substitution probabilities on the edges of T_{23}.
- These matrices have a common diagonalising matrix, $H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, with $H_1^{-1} M_{e_1} H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1-2P \end{bmatrix}$.
- Hence the M_{e_1} has eigenvalues $1, \lambda_1 = 1-2P$, and M_{e_3} has eigenvalues $1, \lambda_3 = 1-2Q$.
- The transition matrices multiply along paths, hence so do their eigenvalues.

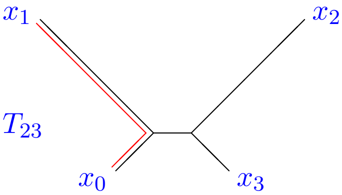
Eigenvalues

- The stochastic matrices $M_{e_1} = M_{e_2} = \begin{bmatrix} 1-P & P \\ P & 1-P \end{bmatrix}$, $M_{e_3} = M_{e_{23}} = M_{e_{123}} = \begin{bmatrix} 1-Q & Q \\ Q & 1-Q \end{bmatrix}$ describe the substitution probabilities on the edges of T_{23}.
- These matrices have a common diagonalising matrix, $H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, with $H_1^{-1} M_{e_1} H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1-2P \end{bmatrix}$.
- Hence the M_{e_1} has eigenvalues $1, \lambda_1 = 1-2P$, and M_{e_3} has eigenvalues $1, \lambda_3 = 1-2Q$.
- The transition matrices multiply along paths, hence so do their eigenvalues.
- Hence the eigenvalue across path Π_{ij} connecting x_i to x_j is

$$\mu_{ij} = \prod_{e \in \Pi_{ij}} \lambda_e$$

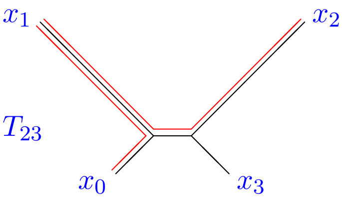
.

Pathsets



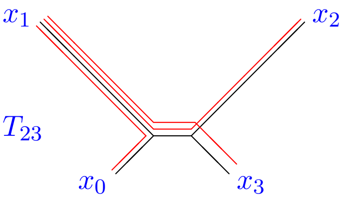
- The path Π_{01}

Pathsets



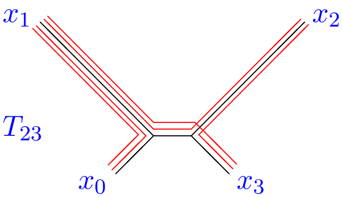
- The path Π_{01}
- The path Π_{12}

Pathsets



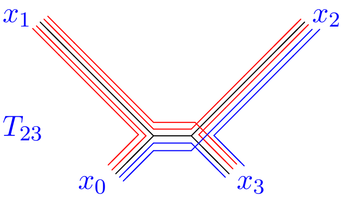
- The path Π_{01}
- The path Π_{12}
- The path Π_{13}

Pathsets



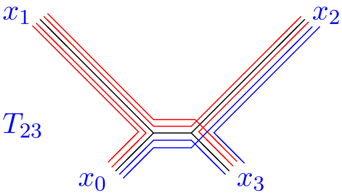
- The path Π_{01}
- The path Π_{12}
- The path Π_{13}
- The pathset $\Pi_{0123} = \Pi_{01} \cup \Pi_{23}$

Pathsets



- The path Π_{01}
- The path Π_{12}
- The path Π_{13}
- The pathset $\Pi_{0123} = \Pi_{01} \cup \Pi_{23}$
- Subtract paths Π_{02}, Π_{03}, Π_{23}.

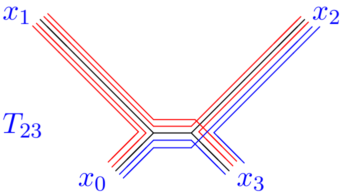
Pathsets



- The path Π_{01}
- The path Π_{12}
- The path Π_{13}
- The pathset $\Pi_{0123} = \Pi_{01} \cup \Pi_{23}$
- Subtract paths Π_{02}, Π_{03}, Π_{23}.

- Counting the occurrences of each edge in Π_{01}, Π_{12}, Π_{13} and Π_{0123}, minus the occurrences in Π_{02}, Π_{03} and Π_{23}, we see edge e_1 is counted 4 times, and each other edge cancels.

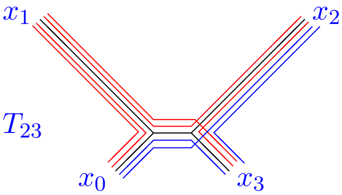
Pathsets



- The path Π_{01}
- The path Π_{12}
- The path Π_{13}
- The pathset $\Pi_{0123} = \Pi_{01} \cup \Pi_{23}$
- Subtract paths Π_{02}, Π_{03}, Π_{23}.

- Counting the occurrences of each edge in Π_{01}, Π_{12}, Π_{13} and Π_{0123}, minus the occurrences in Π_{02}, Π_{03} and Π_{23}, we see edge e_1 is counted 4 times, and each other edge cancels.
- Hence $\frac{\mu_{01}\mu_{12}\mu_{13}\mu_{0123}}{\mu_{02}\mu_{03}\mu_{23}} = \lambda_1^4$

Pathsets



- The path Π_{01}
- The path Π_{12}
- The path Π_{13}
- The pathset $\Pi_{0123} = \Pi_{01} \cup \Pi_{23}$
- Subtract paths Π_{02}, Π_{03}, Π_{23}.

- Counting the occurrences of each edge in Π_{01}, Π_{12}, Π_{13} and Π_{0123}, minus the occurrences in Π_{02}, Π_{03} and Π_{23}, we see edge e_1 is counted 4 times, and each other edge cancels.
- Hence $\frac{\mu_{01}\mu_{12}\mu_{13}\mu_{0123}}{\mu_{02}\mu_{03}\mu_{23}} = \lambda_1^4$
- $p_{e_1} = \frac{1}{2}(1 - \lambda_1)$ is a rational function (with radicals) of the eigenvalues μ_B, and thus of the site pattern probabilities $p_{\mathtt{RRRR}}, \cdots, p_{\mathtt{RYYY}}$.

- For 2–state sequences evolving under Neyman's model on any X–tree T, ($X = \{x_0, x_1, \cdots, x_n\}$), the local ($\lambda_A$) and global ($\mu_B$) eigenvalues are related by the products:

$$\mu_B = \prod_{A \subseteq X^* :\ |A \cap B| \equiv 1 (\mathrm{mod}\ 2)} \lambda_A,$$

$$(\lambda_A)^{2^{n-1}} = \prod_{\substack{B \subseteq X : \\ |B| \equiv 0 (\mathrm{mod}\ 2) \\ |A \cap B| \equiv 1 (\mathrm{mod}\ 2)}} \mu_B \prod_{\substack{B \subseteq X : \\ |B| \equiv 0 (\mathrm{mod}\ 2) \\ |A \cap B| \equiv 0 (\mathrm{mod}\ 2)}} \mu_B^{-1}.$$

- For 2–state sequences evolving under Neyman's model on any X–tree T, ($X = \{x_0, x_1, \cdots, x_n\}$), the local ($\lambda_A$) and global ($\mu_B$) eigenvalues are related by the products:

$$\mu_B = \prod_{A \subseteq X^* :\ |A \cap B| \equiv 1 (\text{mod } 2)} \lambda_A,$$

$$(\lambda_A)^{2^{n-1}} = \prod_{\substack{B \subseteq X : \\ |B| \equiv 0 (\text{mod } 2) \\ |A \cap B| \equiv 1 (\text{mod } 2)}} \mu_B \prod_{\substack{B \subseteq X : \\ |B| \equiv 0 (\text{mod } 2) \\ |A \cap B| \equiv 0 (\text{mod } 2)}} \mu_B^{-1}.$$

- The local eigenvalues (λ_A) are derived from the edge change probabilities p_{e_A}:

$$\lambda_A = \begin{cases} 1 - 2p_{e_A} & \text{if } e_A \text{ an edge of } T \\ 1 & \text{else} \end{cases}$$

- For 2–state sequences evolving under Neyman's model on any X–tree T, ($X = \{x_0, x_1, \cdots, x_n\}$), the local ($\lambda_A$) and global ($\mu_B$) eigenvalues are related by the products:

$$\mu_B = \prod_{A \subseteq X^* :\ |A \cap B| \equiv 1 (\text{mod } 2)} \lambda_A,$$

$$(\lambda_A)^{2^{n-1}} = \prod_{\substack{B \subseteq X: \\ |B| \equiv 0 (\text{mod } 2) \\ |A \cap B| \equiv 1 (\text{mod } 2)}} \mu_B \prod_{\substack{B \subseteq X: \\ |B| \equiv 0 (\text{mod } 2) \\ |A \cap B| \equiv 0 (\text{mod } 2)}} \mu_B^{-1}.$$

- The local eigenvalues (λ_A) are derived from the edge change probabilities p_{e_A}:

$$\lambda_A = \begin{cases} 1 - 2p_{e_A} & \text{if } e_A \text{ an edge of } T \\ 1 & \text{else} \end{cases}$$

- The global eigenvalues (μ_B) are derived from the site pattern probabilities s_C:

$$\mu_B = 1 - 2 \Big(\sum_{\substack{C \subseteq X^* : \\ |B \cap C| \equiv 1 (\text{mod } 2)}} s_C \Big).$$

- Given this invertible relationship between the local (λ_A) and global (μ_B) eigenvalues we can extend to a relationship between the probabilities, as

$$p_{e_A} = (1 - \lambda_A)/2$$

for each edge e_A of T –

- Given this invertible relationship between the local (λ_A) and global (μ_B) eigenvalues we can extend to a relationship between the probabilities, as

$$p_{e_A} = (1 - \lambda_A)/2$$

for each edge e_A of T –
- – and

$$s_C = 2^{-n} \left(\sum_{\substack{B \subseteq X^* : \\ |A \cap B| \equiv 0 \pmod 2}} \mu_B - \sum_{\substack{B \subseteq X^* : \\ |A \cap B| \equiv 1 \pmod 2}} \mu_B \right)$$

for each site pattern C.

- For models of 4–state sequences evolving on an X–tree T, similar relationships hold between the local and global eigenvalues if the 4×4 stochastic matrices M_{e_A} have a common diagonalising matrix.

- For models of 4–state sequences evolving on an X–tree T, similar relationships hold between the local and global eigenvalues if the 4×4 stochastic matrices M_{e_A} have a common diagonalising matrix.
- The stochastic matrices for the generalised Kimura 3–substitution types (gK3ST) model are diagonalised by the Hadamard matrix

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

For the gK3ST and submodels (Jukes Cantor, K2ST) the site pattern and edge probabilities are similarly related with polynomial and rational (with radicals) functions.

- For models of 4–state sequences evolving on an X–tree T, similar relationships hold between the local and global eigenvalues if the 4×4 stochastic matrices M_{e_A} have a common diagonalising matrix.
- The stochastic matrices for the generalised Kimura 3–substitution types (gK3ST) model are diagonalised by the Hadamard matrix

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

  For the gK3ST and submodels (Jukes Cantor, K2ST) the site pattern and edge probabilities are similarly related with polynomial and rational (with radicals) functions.
- Models with stochastic matrices with more than 3 independent parameters cannot have a common diagonalising matrix.

THANKS!