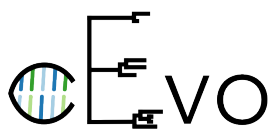
# Computational Biology

Lecturers:
Tanja Stadler, Carsten Magnus & Tim Vaughan

Teaching Assistants:
Jūlija Pečerska, Jérémie Sciré,
Sarah Nadeau & Marc Manceau

Computational Evolution
Department of Biosystems Science and Engineering

HS 2019



# Why compute the likelihood?



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);
- Why look for the ML tree?

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);
- Why look for the ML tree?
  - Take higher order sequence correlations into account (unlike phenetic approaches);
  - Take hidden mutations into account (unlike cladistic approaches);

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);
- Why look for the ML tree?
  - Take higher order sequence correlations into account (unlike phenetic approaches);
  - Take hidden mutations into account (unlike cladistic approaches);
- How to get the ML tree?

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);
- Why look for the ML tree?
  - Take higher order sequence correlations into account (unlike phenetic approaches);
  - Take hidden mutations into account (unlike cladistic approaches);
- How to get the ML tree?
  - iterate over proposed parameter and tree space;
  - calculate the (log-)likelihood of each tree;
  - select the tree with the max (log-)likelihood;

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);
- Why look for the ML tree?
  - Take higher order sequence correlations into account (unlike phenetic approaches);
  - Take hidden mutations into account (unlike cladistic approaches);
- How to get the ML tree?
  - iterate over proposed parameter and tree space;
  - calculate the (log-)likelihood of each tree;
  - select the tree with the max (log-)likelihood;
- How to compute the (log-)likelihood?

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Why compute the likelihood?



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

- Maximum likelihood tree search;
- Any other likelihood-based tree search (e.g. Bayesian);
- Why look for the ML tree?
  - Take higher order sequence correlations into account (unlike phenetic approaches);
  - Take hidden mutations into account (unlike cladistic approaches);
- How to get the ML tree?
  - iterate over proposed parameter and tree space;
  - calculate the (log-)likelihood of each tree;
  - select the tree with the max (log-)likelihood;
- How to compute the (log-)likelihood?
  - Iterate through all possible combinations of nucleotides on the tree nodes ($O(4^{(2N-1)})$);
  - Felsenstein's tree pruning algorithm.

# Full tree log-likelihood computation



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

**Data:** Sequence alignment $A$, tree $\tau$
**Result:** $\mathrm{logL}$
$\mathrm{logL} \leftarrow 0$;
**for** $i \leftarrow 1$ **to** $N$ **do**
  $L \leftarrow 0$;
  **for** $x$ *in* $\{T, C, A, G\}$ **do**
    $L \leftarrow L + \pi_x L_i(\mathrm{root}, x)$;
  **end**
  $\mathrm{logL} \leftarrow \mathrm{logL} + \log(L)$;
**end**
return($\mathrm{logL}$);

# Tree node/nucleotide likelihood computation



**Data:** Node $n$, sequence alignment $A$, tree $\tau$, transition probability matrices $P_{TN93}(t)$ for each branch length $t$ in $\tau$
**Result:** $L_i(n)$
**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** $N$ **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  **for** $i \leftarrow 1$ **to** $N$ **do**
    **for** $c$ *in* $\{T, C, A, G\}$ **do**
      $L1 \leftarrow 0$;
      $L2 \leftarrow 0$;
      **for** $y$ *in* $\{T, C, A, G\}$ **do**
        $L1 \leftarrow L1 + P(|n, child1|)[c, y] \cdot L_i(child1, y)$;
        $L2 \leftarrow L2 + P(|n, child2|)[c, y] \cdot L_i(child2, y)$;
      **end**
      $L_i(node, c) = L1 \cdot L2$;
    **end**
  **end**
**end**

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Tree node/nucleotide likelihood computation



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

For each tree node k, for each site (1 to N), we need to compute:

| | T | C | A | G |
|---|---|---|---|---|
| node **k** | | | | |
| **1** | ? | ? | ? | ? |
| **2** | ? | ? | ? | ? |
| ... | | | | |
| **N** | ? | ? | ? | ? |

# Tree node/nucleotide likelihood computation



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

For each tree node k, for each site (1 to N), we need to compute:

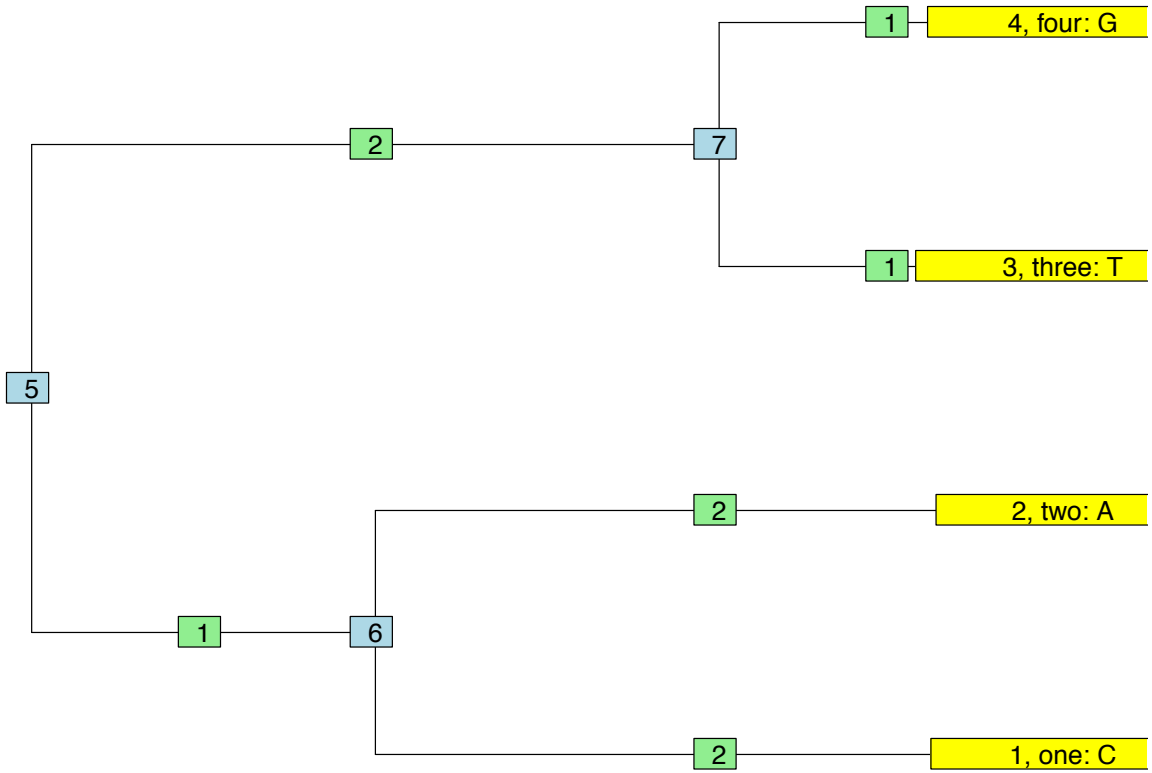
| | T | C | A | G |
|---|---|---|---|---|
| node **k** | | | | |
| **1** | ? | ? | ? | ? |
| **2** | ? | ? | ? | ? |
| ... | | | | |
| **N** | ? | ? | ? | ? |

For each node in the tree we need to know the likelihoods for both child nodes.

# Felsenstein's pruning algorithm: Example



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution




Alignment:

```
sequences=list(one="C", two="A", three="T", four="G")
```

# Felsenstein's pruning algorithm: Tips



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $\mathrm{i} \leftarrow 1$ **to** N **do**
    $\mathrm{L_i(n,)} \leftarrow [0,0,0,0]$;
    $\mathrm{L_i(n, A[n,i])} \leftarrow 1$;
  **end**
**else**
  ...;
**end**

# Felsenstein's pruning algorithm: Tips



4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** N **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  ...;
**end**

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | | | | |
| node **2** | | | | |
| **1** | | | | |
| node **3** | | | | |
| **1** | | | | |
| node **4** | | | | |
| **1** | | | | |

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Tips



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** N **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  ...;
**end**

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 0 | 0 | 0 |
| node **2** | | | | |
| **1** | | | | |
| node **3** | | | | |
| **1** | | | | |
| node **4** | | | | |
| **1** | | | | |

# Felsenstein's pruning algorithm: Tips



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** N **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  ...;
**end**

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | **1** | 0 | 0 |
| node **2** | | | | |
| **1** | | | | |
| node **3** | | | | |
| **1** | | | | |
| node **4** | | | | |
| **1** | | | | |

# Felsenstein's pruning algorithm: Tips



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** N **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  ...;
**end**

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | **1** | 0 |
| node **3** | | | | |
| **1** | | | | |
| node **4** | | | | |
| **1** | | | | |

# Felsenstein's pruning algorithm: Tips



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** N **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  ...;
**end**

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **3** | | | | |
| **1** | **1** | 0 | 0 | 0 |
| node **4** | | | | |
| **1** | | | | |

# Felsenstein's pruning algorithm: Tips



4, four: G

3, three: T

2, two: A

1, one: C

**if** n *is a tip* **then**
  **for** $i \leftarrow 1$ **to** N **do**
    $L_i(n, ) \leftarrow [0, 0, 0, 0]$;
    $L_i(n, A[n, i]) \leftarrow 1$;
  **end**
**else**
  ...;
**end**

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **3** | | | | |
| **1** | 1 | 0 | 0 | 0 |
| node **4** | | | | |
| **1** | 0 | 0 | 0 | **1** |

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Internal nodes



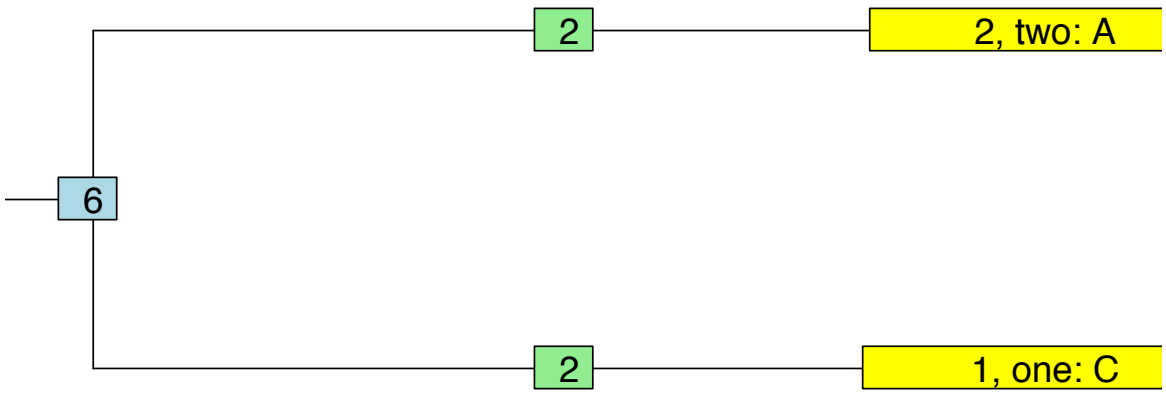
Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

**if** n *is a tip* **then**
  ...;
**else**
  **for** $i \leftarrow 1$ **to** N **do**
    **for** c *in* {T, C, A, G} **do**
      $L1 \leftarrow 0$;
      $L2 \leftarrow 0$;
      **for** $y$ *in* {T, C, A, G} **do**
        $L1 \leftarrow L1 + P(|n, child1|)[c, y] \cdot L_i(child1, y)$;
        $L2 \leftarrow L2 + P(|n, child2|)[c, y] \cdot L_i(child2, y)$;
      **end**
      $L_i(node, c) = L1 \cdot L2$;
    **end**
  **end**
**end**

# Internal node 6



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution




| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | | | | |

# Internal node 6, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | | | | |

$$L_1(6|T) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[T,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[T,Y]\Big)$$

# Internal node 6, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | | | | |

$$\mathbf{L}_1(6|\mathsf{T}) = \Big(\sum_{\mathsf{X}\in\{\mathsf{T},\mathsf{C},\mathsf{A},\mathsf{G}\}} \mathbf{L}_1(1|\mathsf{X})\mathbf{P}(2)[\mathsf{T},\mathsf{X}]\Big) \times \Big(\sum_{\mathsf{Y}\in\{\mathsf{T},\mathsf{C},\mathsf{A},\mathsf{G}\}} \mathbf{L}_1(2|\mathsf{Y})\mathbf{P}(2)[\mathsf{T},\mathsf{Y}]\Big)$$
$$= (0 \times \mathbf{P}(2)[\mathsf{T},\mathsf{T}] + 1 \times \mathbf{P}(2)[\mathsf{T},\mathsf{C}] + 0 \times \mathbf{P}(2)[\mathsf{T},\mathsf{A}] + 0 \times \mathbf{P}(2)[\mathsf{T},\mathsf{G}])$$
$$\times (0 \times \mathbf{P}(2)[\mathsf{T},\mathsf{T}] + 0 \times \mathbf{P}(2)[\mathsf{T},\mathsf{C}] + 1 \times \mathbf{P}(2)[\mathsf{T},\mathsf{A}] + 0 \times \mathbf{P}(2)[\mathsf{T},\mathsf{G}])$$

# Internal node 6, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | | | | |

$$\mathrm{L}_1(6|\mathrm{T}) = \Big(\sum_{\mathrm{X}\in\{\mathrm{T,C,A,G}\}} \mathrm{L}_1(1|\mathrm{X})\mathrm{P}(2)[\mathrm{T},\mathrm{X}]\Big) \times \Big(\sum_{\mathrm{Y}\in\{\mathrm{T,C,A,G}\}} \mathrm{L}_1(2|\mathrm{Y})\mathrm{P}(2)[\mathrm{T},\mathrm{Y}]\Big)$$

$$= (0 \times \mathrm{P}(2)[\mathrm{T},\mathrm{T}] + 1 \times \mathrm{P}(2)[\mathrm{T},\mathrm{C}] + 0 \times \mathrm{P}(2)[\mathrm{T},\mathrm{A}] + 0 \times \mathrm{P}(2)[\mathrm{T},\mathrm{G}])$$

$$\times (0 \times \mathrm{P}(2)[\mathrm{T},\mathrm{T}] + 0 \times \mathrm{P}(2)[\mathrm{T},\mathrm{C}] + 1 \times \mathrm{P}(2)[\mathrm{T},\mathrm{A}] + 0 \times \mathrm{P}(2)[\mathrm{T},\mathrm{G}])$$

$$= (0 \times 0.76 + 1 \times 0.24 + 0 \times 0.00089 + 0 \times 0.00051)$$

$$\times (0 \times 0.76 + 0 \times 0.24 + 1 \times 0.00089 + 0 \times 0.00051)$$

# Internal node 6, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | | | | |

$$
\begin{aligned}
\mathrm{L}_1(6|\mathrm{T}) &= \Big(\sum_{X\in\{\mathrm{T,C,A,G}\}} \mathrm{L}_1(1|X)\mathrm{P}(2)[\mathrm{T},X]\Big) \times \Big(\sum_{Y\in\{\mathrm{T,C,A,G}\}} \mathrm{L}_1(2|Y)\mathrm{P}(2)[\mathrm{T},Y]\Big) \\
&= (0\times \mathrm{P}(2)[\mathrm{T},\mathrm{T}] + 1\times \mathrm{P}(2)[\mathrm{T},\mathrm{C}] + 0\times \mathrm{P}(2)[\mathrm{T},\mathrm{A}] + 0\times \mathrm{P}(2)[\mathrm{T},\mathrm{G}]) \\
&\times (0\times \mathrm{P}(2)[\mathrm{T},\mathrm{T}] + 0\times \mathrm{P}(2)[\mathrm{T},\mathrm{C}] + 1\times \mathrm{P}(2)[\mathrm{T},\mathrm{A}] + 0\times \mathrm{P}(2)[\mathrm{T},\mathrm{G}]) \\
&= (0\times 0.76 + 1\times 0.24 + 0\times 0.00089 + 0\times 0.00051) \\
&\times (0\times 0.76 + 0\times 0.24 + 1\times 0.00089 + 0\times 0.00051) \\
&= 0.24\times 0.00089 \approx 0.00021
\end{aligned}
$$

# Internal node 6



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

# Internal node 6, nucleotides C, A, G



| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

$$\mathbf{L}_1(6|\mathsf{C}) = \Big(\sum_{\mathsf{X}\in\{\mathsf{T},\mathsf{C},\mathsf{A},\mathsf{G}\}} \mathbf{L}_1(1|\mathsf{X})\mathbf{P}(2)[\mathsf{C},\mathsf{X}]\Big) \times \Big(\sum_{\mathsf{Y}\in\{\mathsf{T},\mathsf{C},\mathsf{A},\mathsf{G}\}} \mathbf{L}_1(2|\mathsf{Y})\mathbf{P}(2)[\mathsf{C},\mathsf{Y}]\Big)$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Internal node 6, nucleotides C, A, G



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

$$L_1(6|C) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[C,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[C,Y]\Big)$$
$$= (1 \times P(2)[C,C]) \times (1 \times P(2)[C,A]) = 0.8 \times 0.00089 \approx 0.00071$$

# Internal node 6, nucleotides C, A, G



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

$$L_1(6|C) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[C,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[C,Y]\Big)$$

$$= (1 \times P(2)[C,C]) \times (1 \times P(2)[C,A]) = 0.8 \times 0.00089 \approx 0.00071$$

$$L_1(6|A) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[A,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[A,Y]\Big)$$

# Internal node 6, nucleotides C, A, G



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

$$L_1(6|C) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[C,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[C,Y]\Big)$$
$$= (1 \times P(2)[C,C]) \times (1 \times P(2)[C,A]) = 0.8 \times 0.00089 \approx 0.00071$$
$$L_1(6|A) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[A,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[A,Y]\Big)$$
$$= (1 \times P(2)[A,C]) \times (1 \times P(2)[A,A]) = 0.0007 \times 0.9 \approx 0.00063$$

# Internal node 6, nucleotides C, A, G



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

$$L_1(6|C) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[C,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[C,Y]\Big)$$
$$= (1 \times P(2)[C,C]) \times (1 \times P(2)[C,A]) = 0.8 \times 0.00089 \approx 0.00071$$
$$L_1(6|A) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[A,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[A,Y]\Big)$$
$$= (1 \times P(2)[A,C]) \times (1 \times P(2)[A,A]) = 0.0007 \times 0.9 \approx 0.00063$$
$$L_1(6|G) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[G,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[G,Y]\Big)$$

# Internal node 6, nucleotides C, A, G



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | | | |

$$L_1(6|C) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[C,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[C,Y]\Big)$$
$$= (1 \times P(2)[C,C]) \times (1 \times P(2)[C,A]) = 0.8 \times 0.00089 \approx 0.00071$$
$$L_1(6|A) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[A,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[A,Y]\Big)$$
$$= (1 \times P(2)[A,C]) \times (1 \times P(2)[A,A]) = 0.0007 \times 0.9 \approx 0.00063$$
$$L_1(6|G) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(1|X)P(2)[G,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(2|Y)P(2)[G,Y]\Big)$$
$$= (1 \times P(2)[G,C]) \times (1 \times P(2)[G,A]) = 0.0007 \times 0.17 \approx 0.00012$$

# Internal node 6

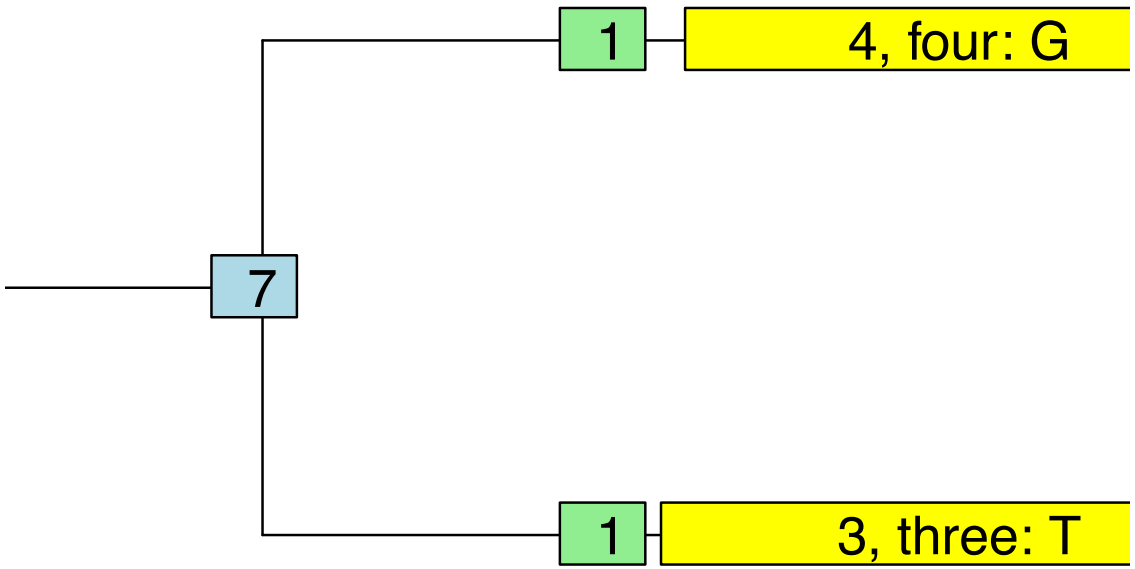


Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **1** | | | | |
| **1** | 0 | 1 | 0 | 0 |
| node **2** | | | | |
| **1** | 0 | 0 | 1 | 0 |
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |

# Internal node 7

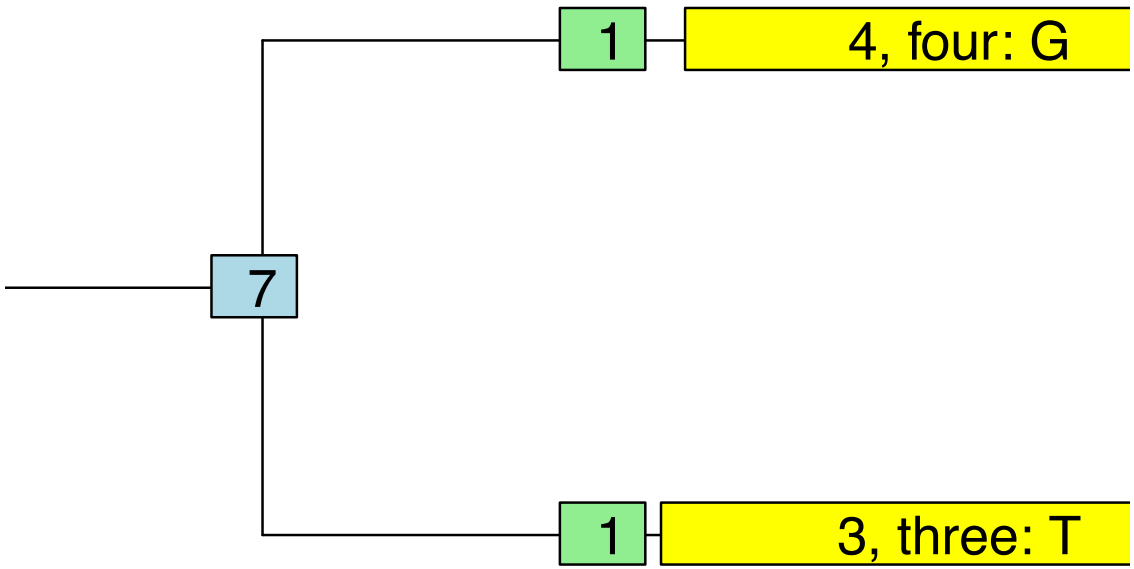





Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Internal node 7



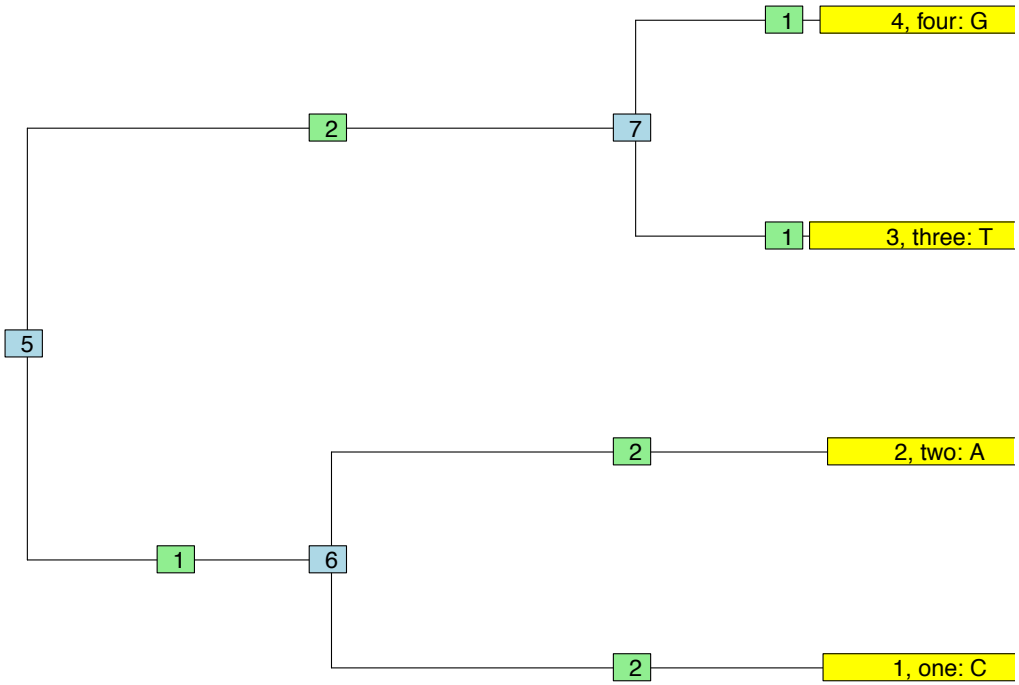



| | T | C | A | G |
|---|---|---|---|---|
| node **3** | | | | |
| **1** | 1 | 0 | 0 | 0 |
| node **4** | | | | |
| **1** | 0 | 0 | 0 | 1 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Internal node 5






| | T | C | A | G |
|---|---|---|---|---|
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |
| node **5** | | | | |
| **1** | | | | |

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Internal node 5, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |
| node **5** | | | | |
| **1** | | | | |

# Internal node 5, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |
| node **5** | | | | |
| **1** | | | | |

$$\mathbf{L}_1(5|\mathrm{T}) = \Big(\sum_{\mathrm{X}\in\{\mathrm{T,C,A,G}\}} \mathbf{L}_1(6|\mathrm{X})\mathbf{P}(1)[\mathrm{T},\mathrm{X}]\Big) \times \Big(\sum_{\mathrm{Y}\in\{\mathrm{T,C,A,G}\}} \mathbf{L}_1(7|\mathrm{Y})\mathbf{P}(2)[\mathrm{T},\mathrm{Y}]\Big)$$

# Internal node 5, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |
| node **5** | | | | |
| **1** | | | | |

$$L_1(5|T) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(6|X)P(1)[T,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(7|Y)P(2)[T,Y]\Big)$$

$$= (0.00021 \times 0.861 + 0.00071 \times 0.13 + 0.00063 \times 0.00045 + 0.00012 \times 0.00026)$$

$$\times (0.00022 \times 0.76 + 0.000029 \times 0.24 + 000015 \times 0.00089 + 0.00027 \times 0.00051)$$

# Internal node 5, nucleotide T



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |
| node **5** | | | | |
| **1** | | | | |

$$L_1(5|T) = \Big(\sum_{X\in\{T,C,A,G\}} L_1(6|X)P(1)[T,X]\Big) \times \Big(\sum_{Y\in\{T,C,A,G\}} L_1(7|Y)P(2)[T,Y]\Big)$$

$$= (0.00021 \times 0.861 + 0.00071 \times 0.13 + 0.00063 \times 0.00045 + 0.00012 \times 0.00026)$$

$$\times (0.00022 \times 0.76 + 0.000029 \times 0.24 + 000015 \times 0.00089 + 0.00027 \times 0.00051)$$

$$\approx 4.89e{-}08$$

# Internal node 5



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node **6** | | | | |
| **1** | 0.00021 | 0.00071 | 0.00063 | 0.00012 |
| node **7** | | | | |
| **1** | 0.00022 | 0.000029 | 0.000015 | 0.00027 |
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

# Felsenstein's pruning algorithm: Example



| | T | C | A | G |
|---|---|---|---|---|
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

$$\pi = (0.22, 0.26, 0.33, 0.19)$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Example



| | T | C | A | G |
|---|---|---|---|---|
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

$$\pi = (0.22, 0.26, 0.33, 0.19)$$

$$\mathrm{LogL} = \left( \sum_{i \in 1:N} \log\left( \sum_{X \in \{T,C,A,G\}} \pi_X \times L_i(5|X) \right) \right)$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Example



| | T | C | A | G |
|---|---|---|---|---|
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

$$\pi = (0.22, 0.26, 0.33, 0.19)$$

$$\mathrm{LogL} = \Big( \sum_{i \in 1:N} \log\Big( \sum_{X \in \{T,C,A,G\}} \pi_X \times L_i(5|X) \Big) \Big)$$

$$= \log\Big( \sum_{X \in \{T,C,A,G\}} \pi_X \times L_1(5|X) \Big)$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Example



| | T | C | A | G |
|---|---|---|---|---|
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

$$\pi = (0.22, 0.26, 0.33, 0.19)$$

$$\begin{aligned}
\mathrm{LogL} &= \Big(\sum_{i \in 1:N} \log\Big(\sum_{X \in \{T,C,A,G\}} \pi_X \times L_i(5|X)\Big)\Big) \\
&= \log\Big(\sum_{X \in \{T,C,A,G\}} \pi_X \times L_1(5|X)\Big) \\
&= \log(\pi_T \times L_1(5|T) + \pi_C \times L_1(5|C) \\
&+ \pi_A \times L_1(5|A) + \pi_G \times L_1(5|G))
\end{aligned}$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Example



| | T | C | A | G |
|---|---|---|---|---|
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

$$\pi = (0.22, 0.26, 0.33, 0.19)$$

$$\begin{aligned}
\mathrm{LogL} &= \Big(\sum_{i \in 1:N} \log\Big(\sum_{X \in \{T,C,A,G\}} \pi_X \times L_i(5|X)\Big)\Big) \\
&= \log\Big(\sum_{X \in \{T,C,A,G\}} \pi_X \times L_1(5|X)\Big) \\
&= \log(\pi_T \times L_1(5|T) + \pi_C \times L_1(5|C) \\
&+ \pi_A \times L_1(5|A) + \pi_G \times L_1(5|G)) \\
&= \log(0.22 \times 4.89e{-}08 + 0.26 \times 4.43e{-}08 \\
&+ 0.33 \times 2.42e{-}08 + 0.19 \times 3.72e{-}08)
\end{aligned}$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Felsenstein's pruning algorithm: Example



| | T | C | A | G |
|---|---|---|---|---|
| node **5** | | | | |
| **1** | 4.89e-08 | 4.43e-08 | 2.42e-08 | 3.72e-08 |

$$\pi = (0.22, 0.26, 0.33, 0.19)$$

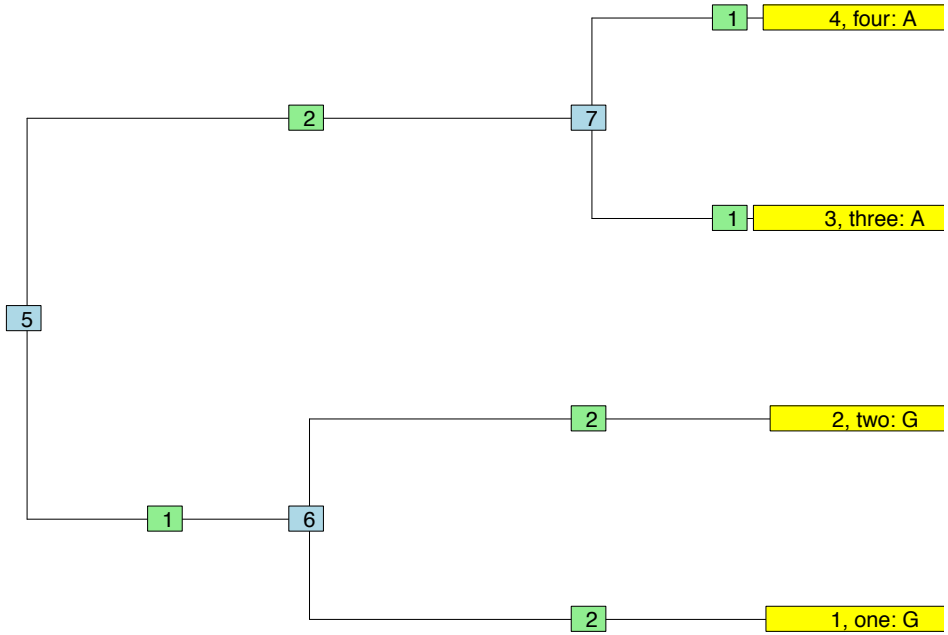
$$\begin{aligned}
\mathrm{LogL} &= \Big(\sum_{i \in 1:N} \log\Big(\sum_{X \in \{T,C,A,G\}} \pi_X \times L_i(5|X)\Big)\Big) \\
&= \log\Big(\sum_{X \in \{T,C,A,G\}} \pi_X \times L_1(5|X)\Big) \\
&= \log(\pi_T \times L_1(5|T) + \pi_C \times L_1(5|C) \\
&+ \pi_A \times L_1(5|A) + \pi_G \times L_1(5|G)) \\
&= \log(0.22 \times 4.89e{-}08 + 0.26 \times 4.43e{-}08 \\
&+ 0.33 \times 2.42e{-}08 + 0.19 \times 3.72e{-}08) \\
&= \log(3.73e{-}08) \approx -17.1
\end{aligned}$$

Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

# Pen'n'paper exercise



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution




$\pi = (0.22, 0.26, 0.33, 0.19)$

P(1):

| | T | C | A | G |
|---|---|---|---|---|
| T | 0.86 | 0.13 | 0.00045 | 0.00026 |
| C | 0.11 | 0.89 | 0.00045 | 0.00026 |
| A | 0.0003 | 0.00035 | 0.95 | 0.052 |
| G | 0.0003 | 0.00035 | 0.09 | 0.91 |

P(2):

| | T | C | A | G |
|---|---|---|---|---|
| T | 0.76 | 0.24 | 0.00089 | 0.00051 |
| C | 0.2 | 0.8 | 0.00089 | 0.00051 |
| A | 0.00059 | 0.0007 | 0.9 | 0.096 |
| G | 0.00059 | 0.0007 | 0.17 | 0.83 |

# Pen'n'paper solution



Tree (log-)likelihood computation
Felsenstein's tree-pruning algorithm
Pseudocode
Example
Calculation at tips
Calculation at internal nodes
Exercise
Solution

| | T | C | A | G |
|---|---|---|---|---|
| node 1 | | | | |
| 1 | 0 | 0 | 0 | 1 |
| node 2 | | | | |
| 1 | 0 | 0 | 0 | 1 |
| node 6 | | | | |
| 1 | 2.601e-07 | 2.601e-07 | 9.216e-03 | 0.6889 |
| node 3 | | | | |
| 1 | 0 | 0 | 1 | 0 |
| node 4 | | | | |
| 1 | 0 | 0 | 1 | 0 |
| node 7 | | | | |
| 1 | 2.025e-07 | 2.025e-07 | 0.9025 | 8.1e-03 |
| node 5 | | | | |
| 1 | 1.482021e-07 | 1.448554e-07 | 0.036 | 0.099 |

$$\begin{aligned}\mathrm{LogL} &= \log(\pi_\mathsf{T} \times \mathrm{L}_1(5|\mathsf{T}) + \pi_\mathsf{C} \times \mathrm{L}_1(5|\mathsf{C}) + \pi_\mathsf{A} \times \mathrm{L}_1(5|\mathsf{A}) + \pi_\mathsf{G} \times \mathrm{L}_1(5|\mathsf{G})) \\ &= \log(0.22 \times 1.482021e{-}07 + 0.26 \times 1.448554e{-}07 \\ &\quad + 0.33 \times 0.03624314450 + 0.19 \times 0.1005294544) \\ &= \log(0.0310609) \approx -3.472\end{aligned}$$