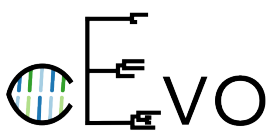
Computational Biology

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

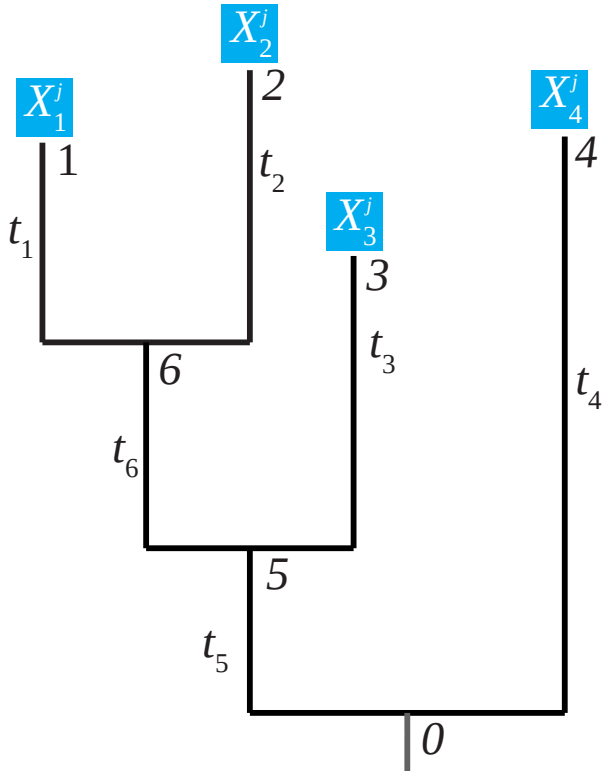


Modelling traits on a tree

Each of j traits:

- starts at a single root value,
- evolves independently as a Brownian Motion on distinct branches,
- finishes evolving upon reaching the tips.

The traits are correlated because they evolved along the same tree.

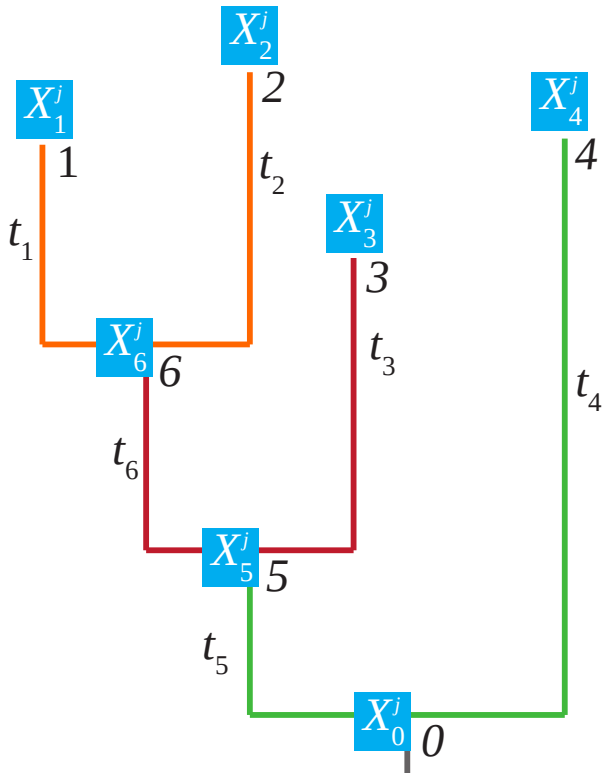


How can we do regression after accounting for this tree-driven correlation ?

Intuition

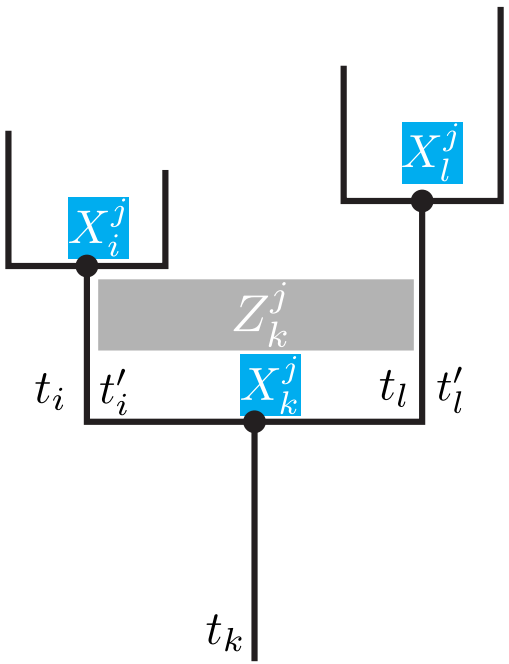
We look for alternative quantities (*contrasts*) that would be:

1. independent:
   - independent Brownian trajectories in subtrees.
   - use the tree structure to find independent variables.
2. identically distributed:
   - traits are normally distributed with variance that scales with branch lengths,
   - use branch lengths to define *contrasts* with equal variance.



Notation

- X_i^j: value of trait j at node i
- t_i: length of the branch leading to node i
- t_i': *corrected* branch length leading to node i
- Z_k^j: value of contrast corresponding to trait j on node k.



Algorithm - Overview

If k is a tip:

X_k is the observed value

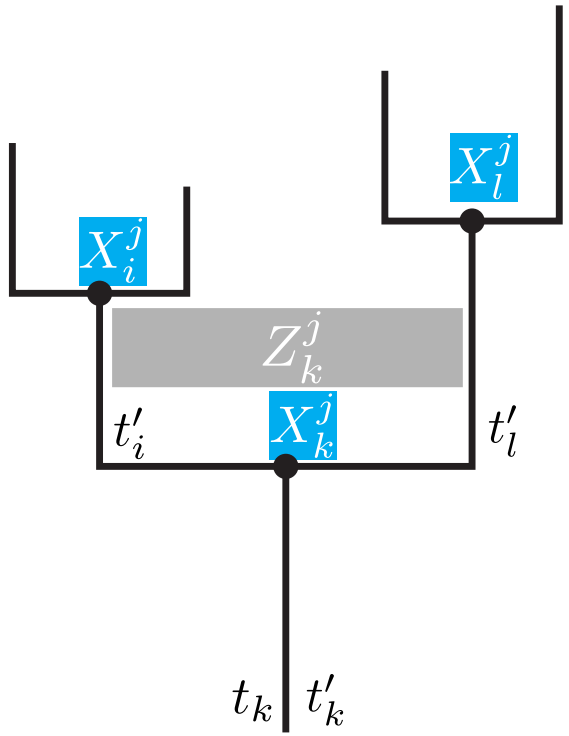
$t'_k = t_k$

Else if k is the parent of i, l:

$$t'_k = \frac{t'_i \cdot t'_l}{t'_i + t'_l} + t_k$$

$$X^j_k = X^j_i \frac{t'_l}{t'_i + t'_l} + X^j_l \frac{t'_i}{t'_i + t'_l}$$

$$Z^j_k = \frac{X^j_i - X^j_l}{\sqrt{t'_i + t'_l}}$$

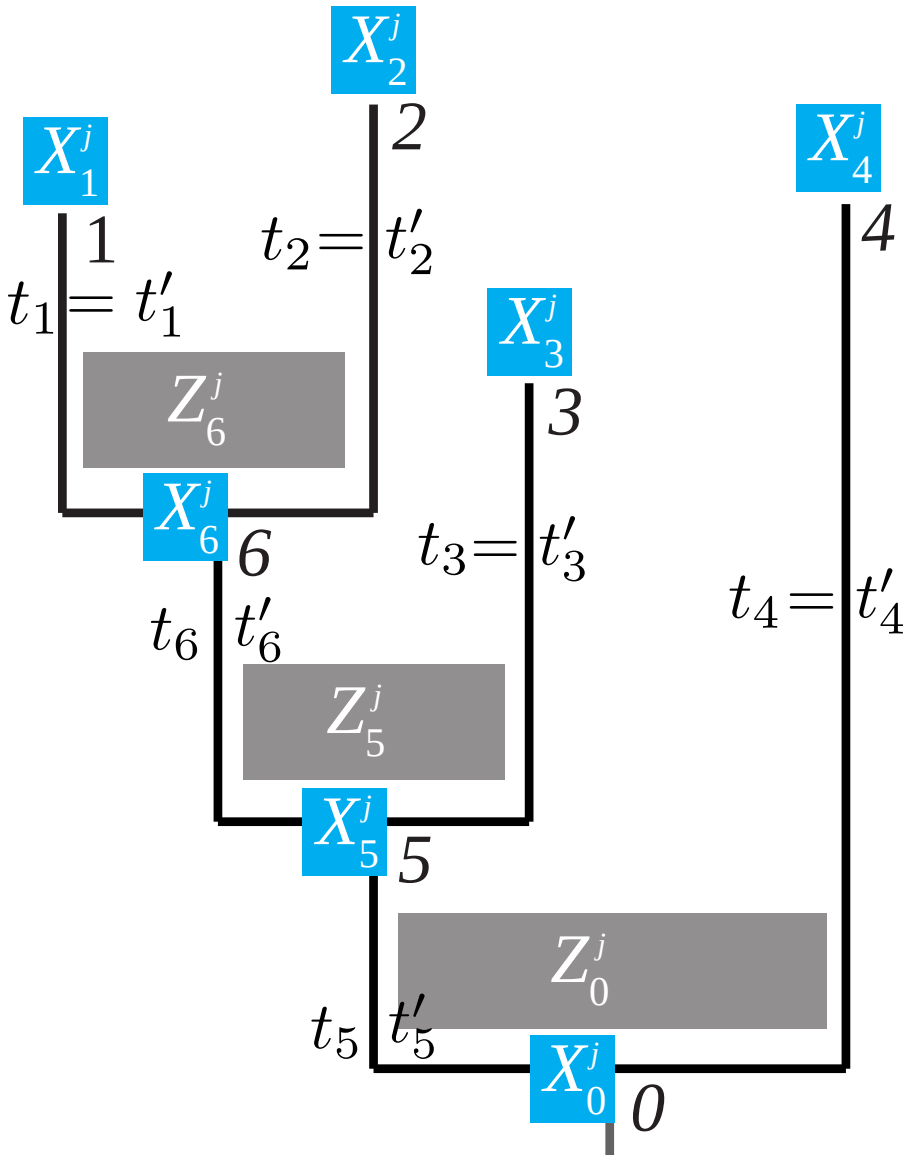


Always visit the parent node k after the two child nodes i, l.

Algorithm - Pseudo-code

Contrast computation for node k **and trait** j
Data: Node k**, tree** τ**, trait tips values** obs^j
Result: t'_k, X^j_k, Z^j_k
if k *is a tip* **then**
  $t'_k \leftarrow t_k$;
  $X^j_k \leftarrow obs^j_k$;
  $Z^j_k \leftarrow NA$;
else
  Compute t'_i, X^j_i, Z^j_i **where** i **is the first child of** k;
  Compute t'_l, X^j_l, Z^j_l **where** l **is the second child of** k;
  $t'_k \leftarrow \frac{t'_i t'_l}{t'_i + t'_l} + t_k$;
  $X^j_k \leftarrow X^j_i \frac{t'_l}{t'_i + t'_l} + X^j_l \frac{t'_i}{t'_i + t'_l}$;
  $Z^j_k \leftarrow \frac{X^j_i - X^j_l}{\sqrt{t'_i + t'_l}}$;
end

Illustration



Pen and Paper exercise

Calculate the normalised independent contrasts for both traits on this tree.

- Corrected branch lengths

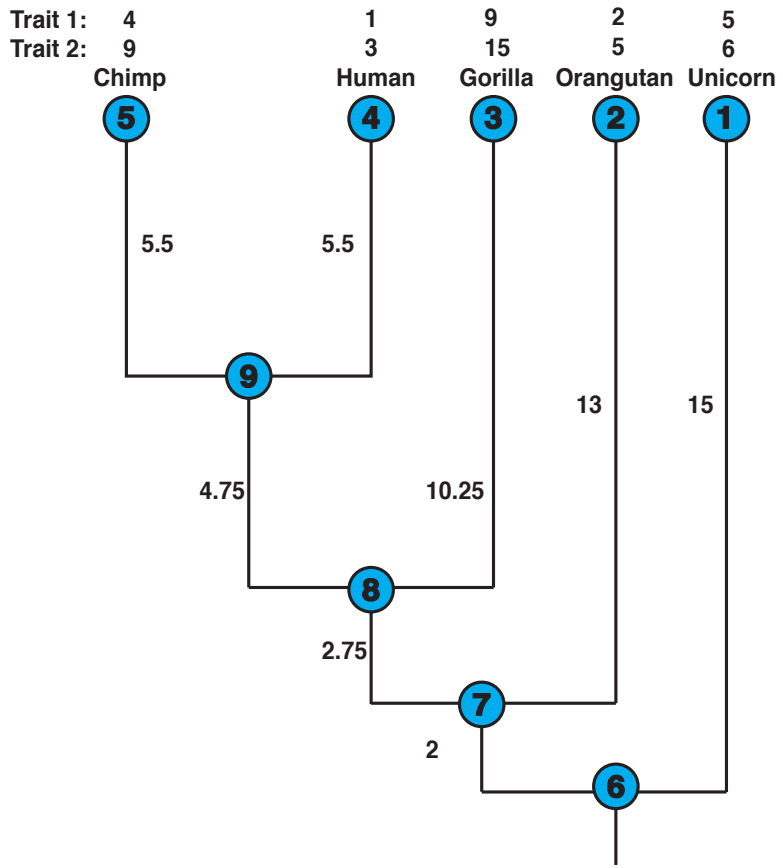
$$t'_k = \frac{t'_i \cdot t'_l}{t'_i + t'_l} + t_k$$

- Trait values of trait j

$$x^j_k = x^j_i \frac{t'_l}{t'_i + t'_l} + x^j_l \frac{t'_i}{t'_i + t'_l}$$

- Normalised contrasts of trait j

$$z^j_k = \frac{x^j_i - x^j_l}{\sqrt{t'_i + t'_l}}$$

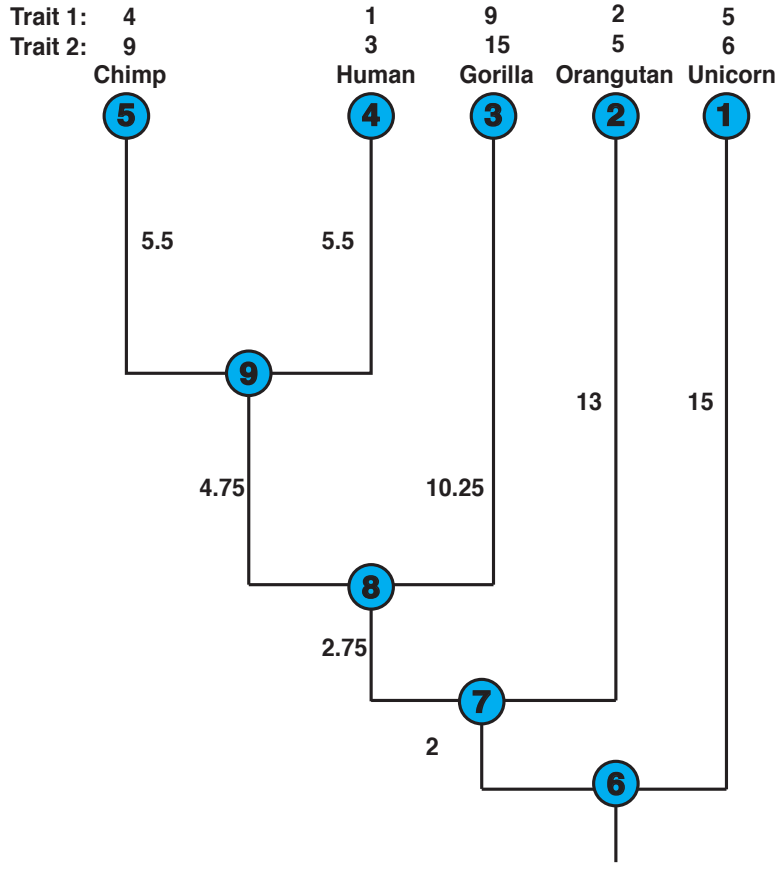


Pen and Paper exercise - solution

Calculate the normalised independent contrasts for both traits on this tree.

Trait 1:

$$Z_6^1 = 0.1933545$$
$$Z_7^1 = -0.7244694$$
$$Z_8^1 = 1.5428162$$
$$Z_9^1 = -0.9045340$$



Trait 2:

$$Z_6^2 = -0.4540037$$
$$Z_7^2 = -1.0717747$$
$$Z_8^2 = 2.1362070$$
$$Z_9^2 = -1.8090681$$