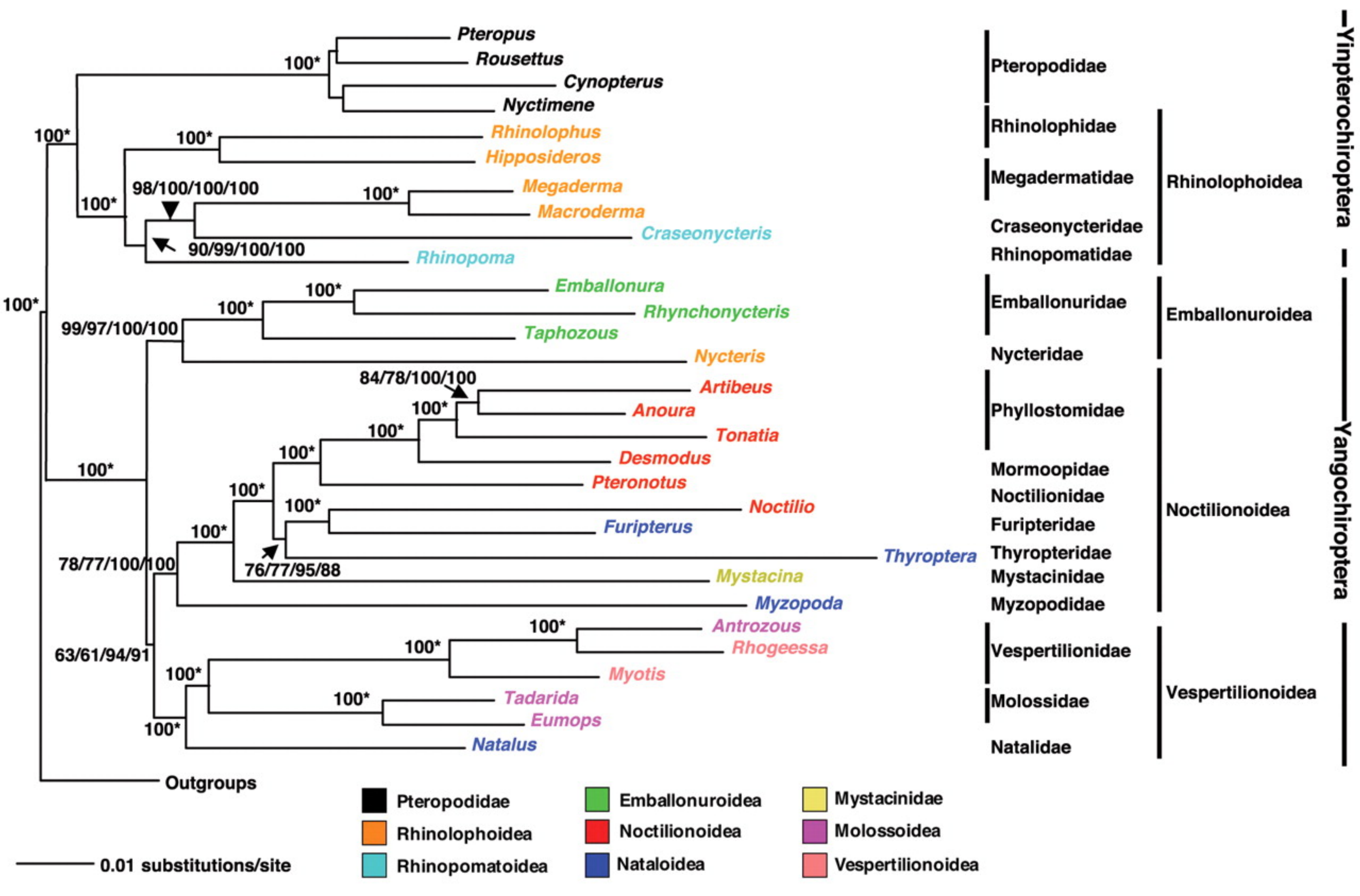
Bioinformatics debates: Phylogenetics?

Alicia, James and Thomas

guillert@tcd.ie @TGuillerme

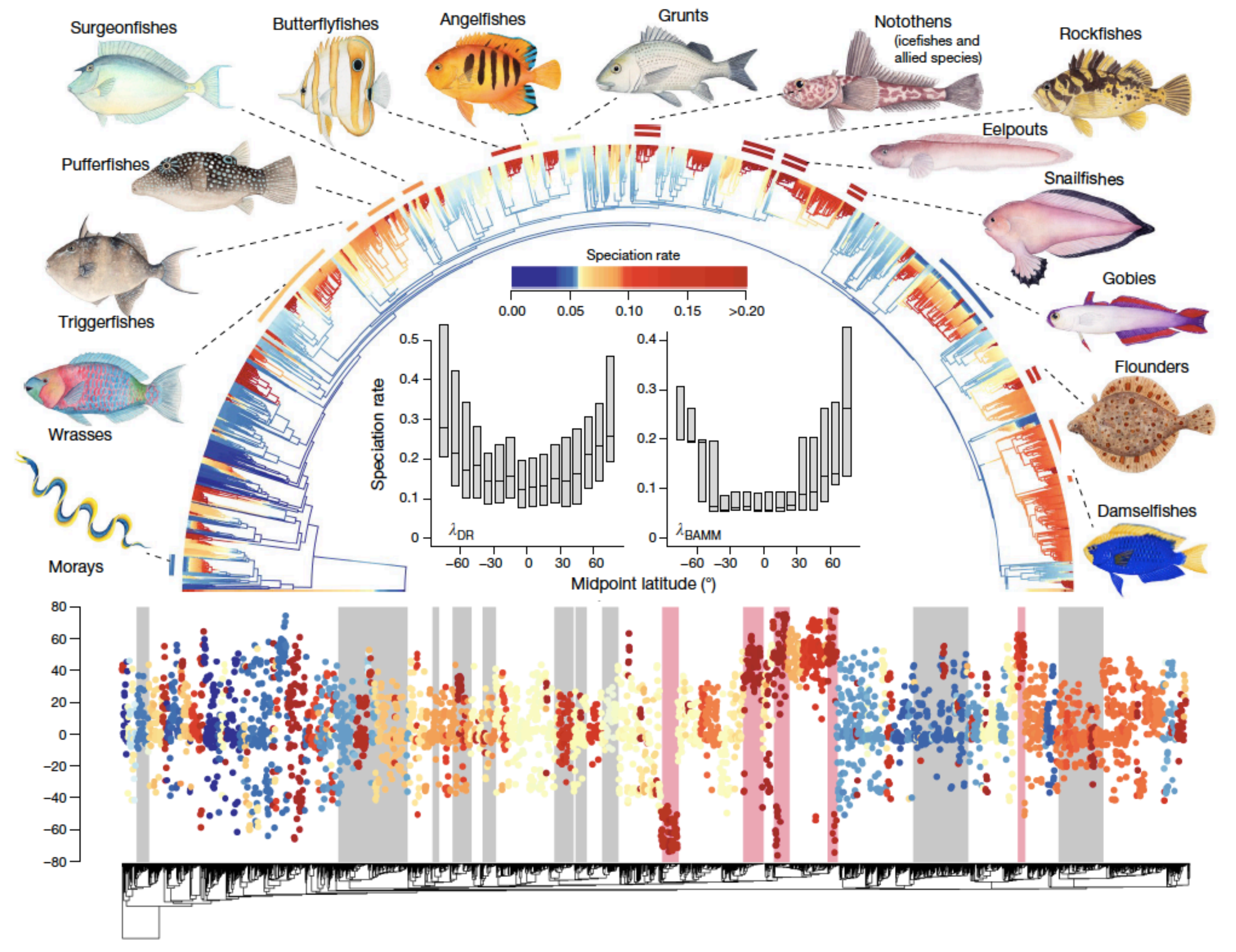
Why do we need trees?

Understanding evolutionary relations



Teeling et al. 2005 *Science*

Understanding evolutionary processes

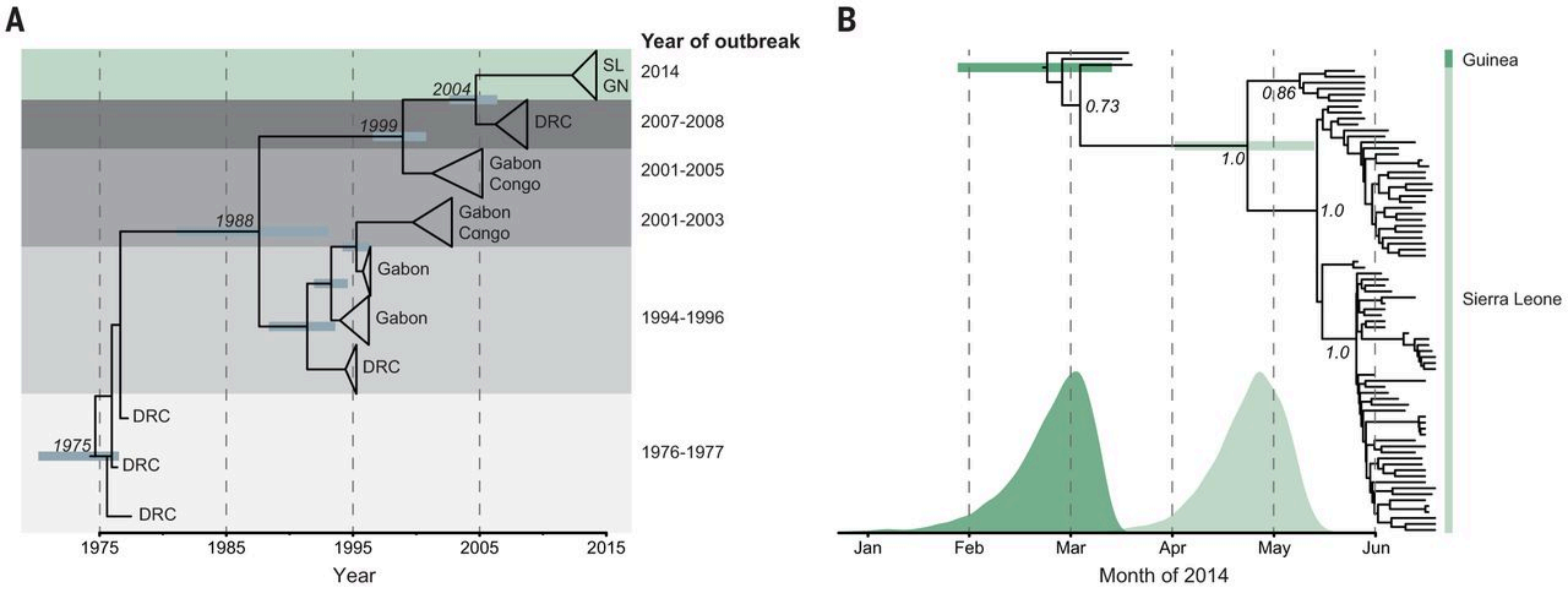


Rabosky et al. 2018 *Nature*

Why do we need trees?

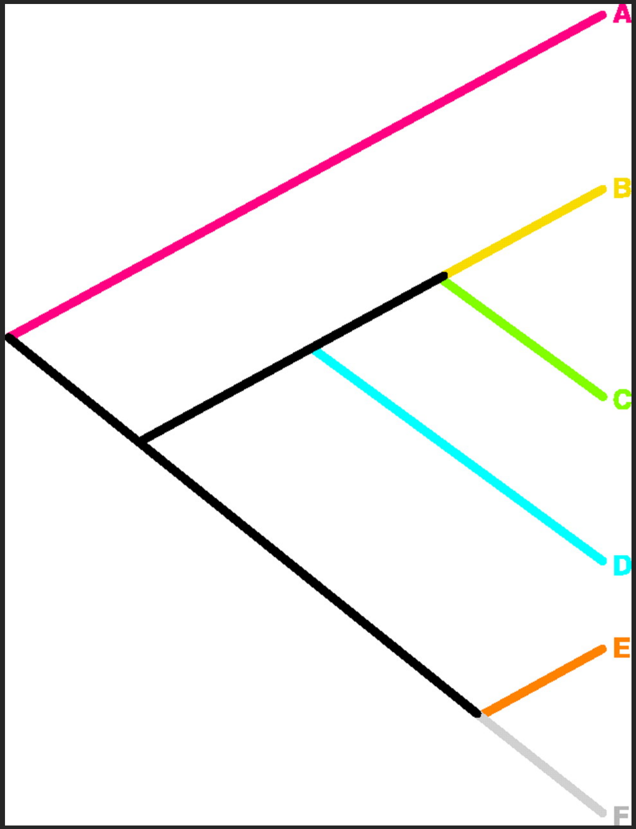
but also…

Public health & medicine



Gire et al 2014 *Science*

Conservation



Mace et al 2003 *Science*

"Nothing in biology makes sense except in the light of evolution"

Phylogeny = **THE** base tool in biology!

Phylogenetics

=

Using traits to solve relations among stuff (dead or alive) assuming descent with modification through time.

Phylogenetics

=

Using traits to solve relations among stuff (dead or alive) assuming descent with modification through time.

Using traits

- DNA
- Proteins

Using traits

- DNA
- Proteins
- Morphology

Using traits

- DNA
- Proteins
- Morphology
- Behavior
- Biogeography
- …

Using traits

- DNA
- Proteins
- Morphology

All of them:
Total Evidence
Method

Phylogenetics

=

Using traits to solve relations among stuff (dead or alive) assuming descent with modification through time.

Solving relations

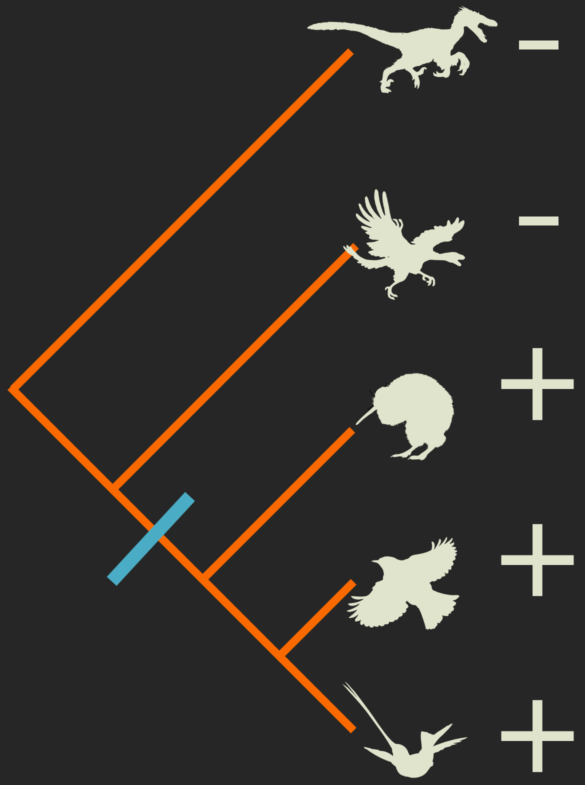
Topology theory +
combinatorics +
optimality criterion

Solving relations

Topology theory +
combinatorics +
optimality criterion

3 taxa -> 3 trees
10 taxa -> 3.4×10^{7} trees
30 taxa -> 4.9×10^{38} trees
54 taxa -> 2.8×10^{82} trees

Solving relations



1 – Propose a topology
2 – Count number of changes
3 – Rank topology

Maximum parsimony

Solving relations

A T T G A A
A T C G A A
A C C G A A

	A	C	G	T
A	p1	p2	p3	p4
C	p5	p6	p7	p8
G	p9	p1	p1	p2
T	p3	p4	p5	p6

hypothesis

data

Maximum likelihood

1 – Propose a topology
2 – Calculate the likelihood of this topology
3 – Rank topology

Solving relations

Bayesian!

1 – Propose a topology
2 – Calculate the likelihood of this topology including prior beliefs
3 – Rank topology
4 – Include uncertainty!

Phylogenetics

=

Using traits to solve relations among stuff (dead or alive) assuming descent with modification through time.

Assuming evolution

	A	C	G	T
A				
C				
G				
T		$p(TC)$		

$p(TC)$ = chances of getting from T to C

Assuming evolution

	A	C	G	T
A				
C				$p(CT)$
G				
T		$p(TC)$		

$p(TC) = p(CT)$?

Assuming evolution

	A	C	G	T
A				
C	p(CA)			p(CT)
G				
T		p(TC)		p(T T)

p(TC) = p(CT) ? = p(TT) ? = p(CA) ?

Assuming evolution

	A	C	G	T
A	*p1*	*p2*	*p3*	*p4*
C	*p5*	*p6*	*p7*	*p8*
G	*p9*	*p10*	*p11*	*p12*
T	*p13*	*p14*	*p15*	*p16*

All rates can be different (but 16 parameters)

Phylogenetics

=

Using traits to solve relations among stuff (dead or alive) assuming descent with modification through time.

The molecular* clock

Time = distance/speed

The molecular* clock

*and morphological, protein, biogeographic, etc...

Time = distance/speed

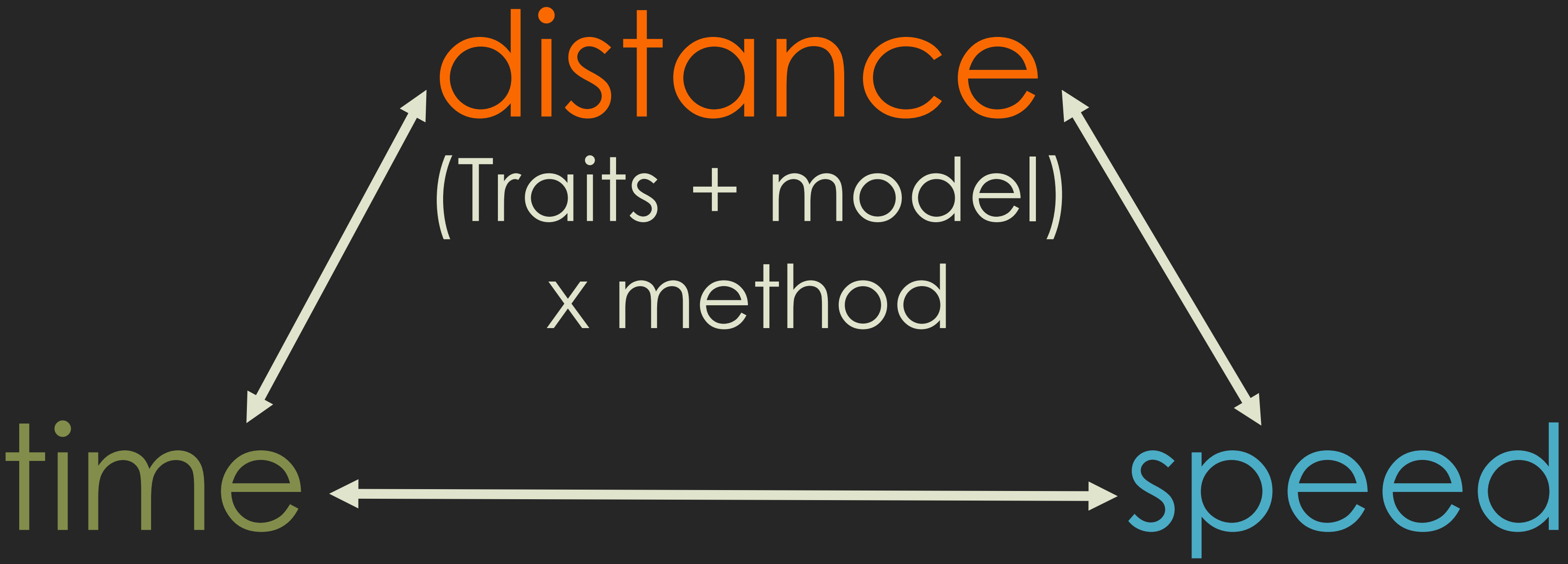
The molecular clock

distance

time

speed

The molecular clock



distance

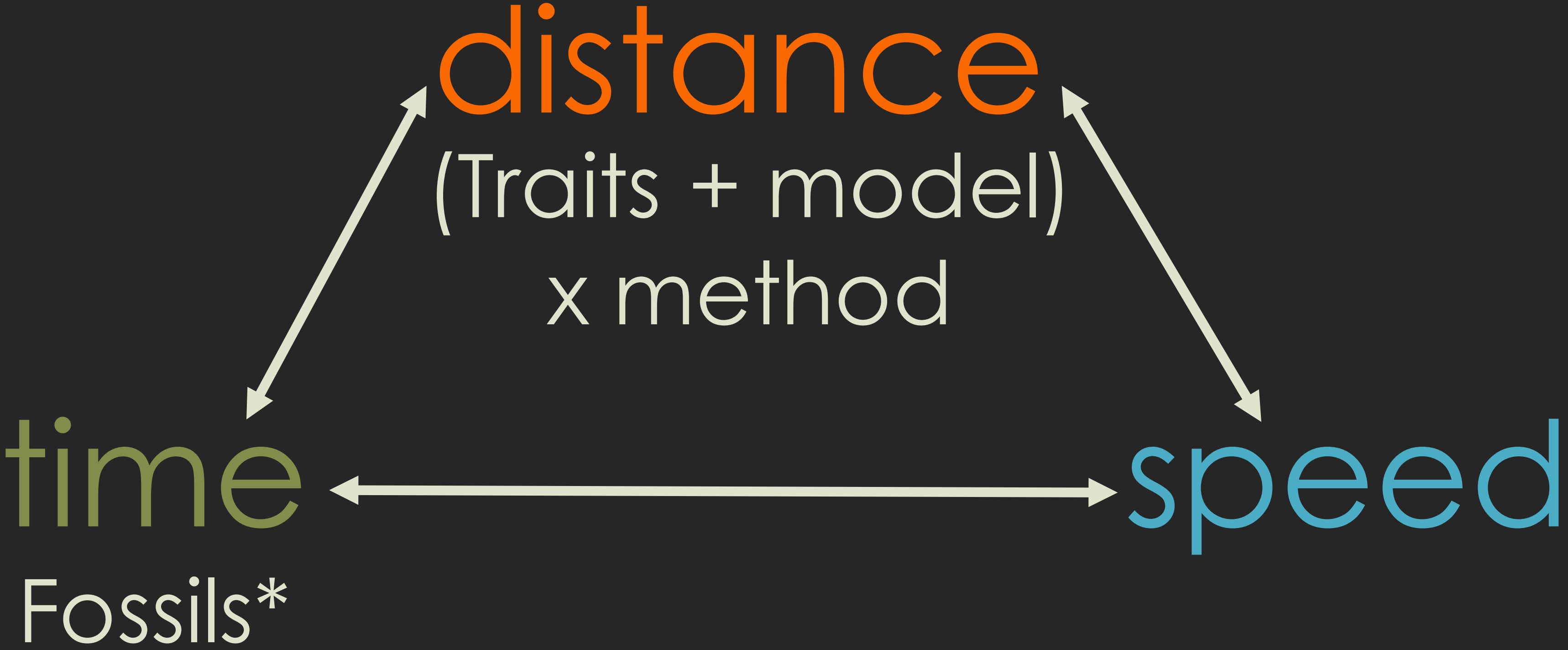
(Traits + model)

x method

time

speed

The molecular clock



distance

(Traits + model)

x method

time

Fossils*

speed

The molecular clock

distance

(Traits + model)

x method

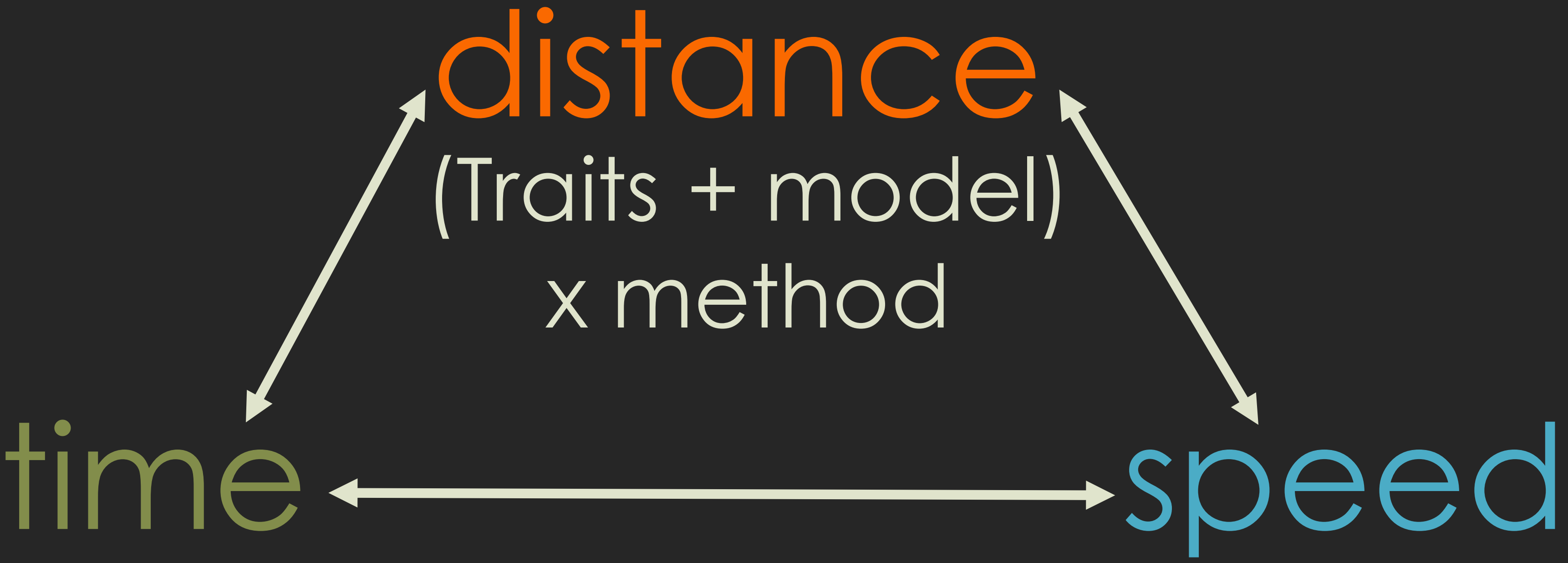
time

Fossils*

speed

*Unless some super rare cases

The molecular clock



distance

(Traits + model)

x method

time

Fossils*

speed

Inferred

(constant or

variable?)

*Unless some super rare cases

Beast2

Bayesian evolutionary analysis by sampling trees

Tutorial with Alicia
29th of October
14:00-15:00
Goddard 501

Bioinformatics debates:

Phylogenetics?

Using traits

Are all traits equivalent?

Are traits really independent?

Solving relations

Parsimony vs. ML/Bayesian?

Do we need models or methods (processes, e.g. distances) to solve trees?

Assuming evolution

What are our assumptions on the process (e.g. Yule, coalescent, birth-death)?

How can we differentiate models (e.g. AIC, number of parameters)?

Assuming evolution

Do we need to estimate the whole process or estimate it step by step (topology, models, dating)?

Timing evolution

How do we solve the equation speed = time/distance (without knowing speed and with uncertain time)?