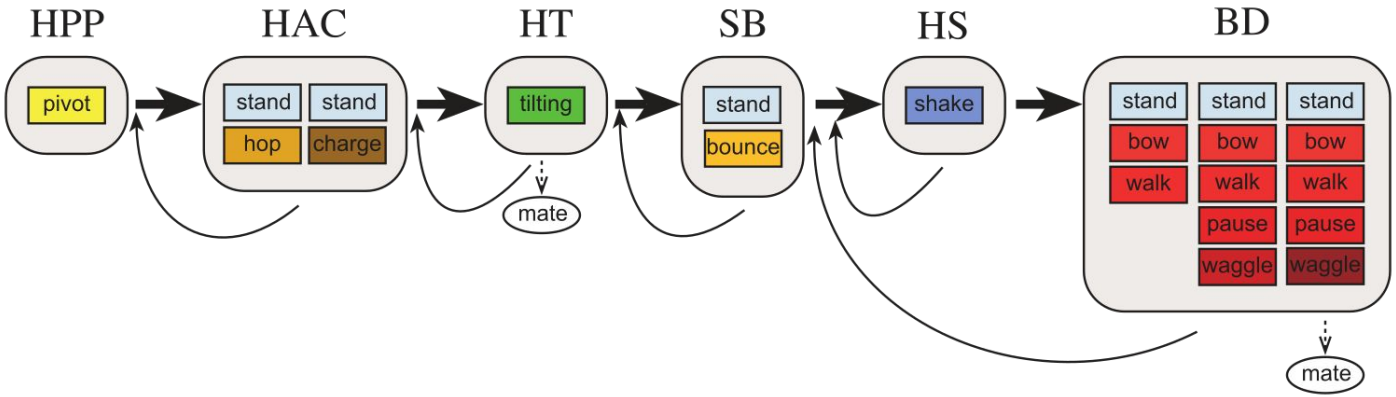
Comparative analyses of phenotypic sequences: a case study using cricket songs

Daniel Caetano and Jeremy Beaulieu
2019

UNIVERSITY OF ARKANSAS

What is a phenotypic sequence?

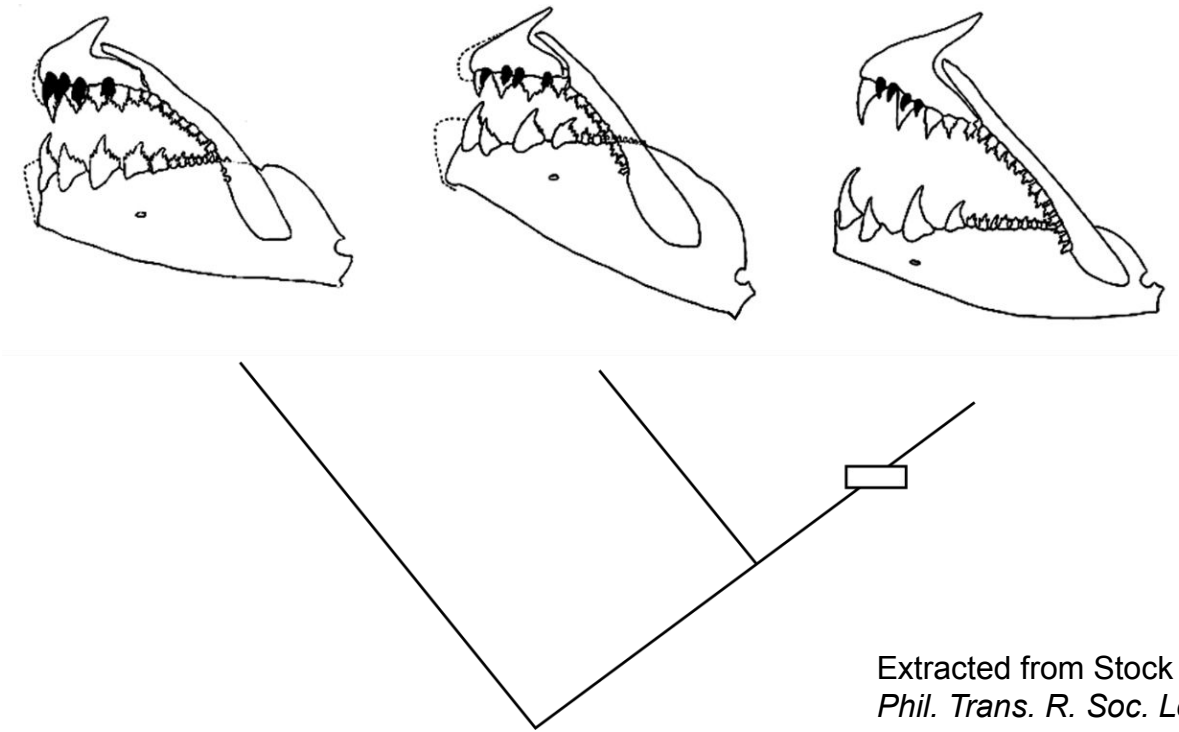


Biological Journal of the Linnean Society, 2008, **94**, 491–504. With 3 figures

Evolution of the courtship phenotype in the bird of paradise genus *Parotia* (Aves: Paradisaeidae): homology, phylogeny, and modularity

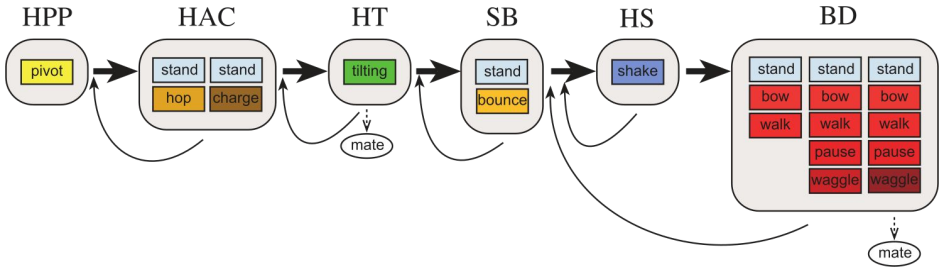
EDWIN SCHOLES III*
Cornell Laboratory of Ornithology, Cornell University, Ithaca, NY 14850, USA

What is a phenotypic sequence?

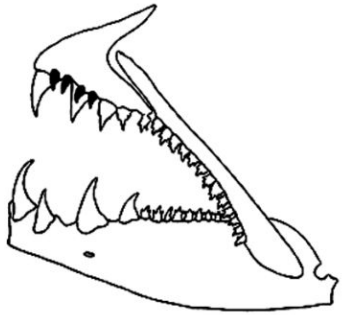


Extracted from Stock D. W. (2001)
Phil. Trans. R. Soc. Lond. B

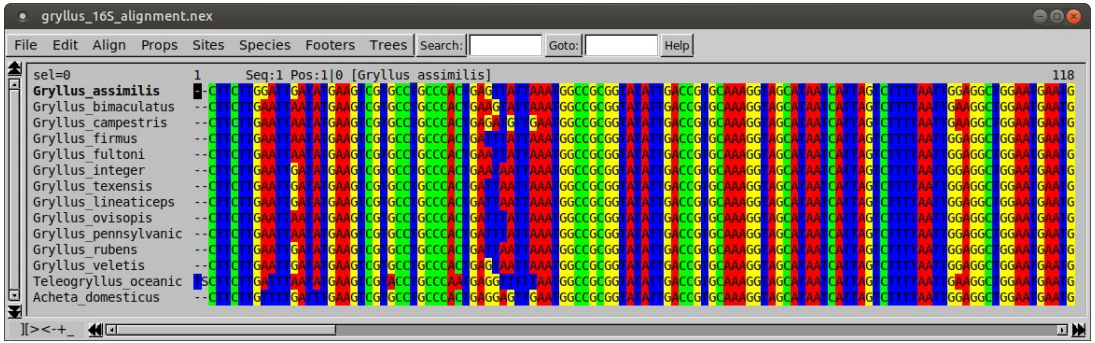
Caetano and Beaulieu caetanods1@gmail.com @ds_caetano



time sequence

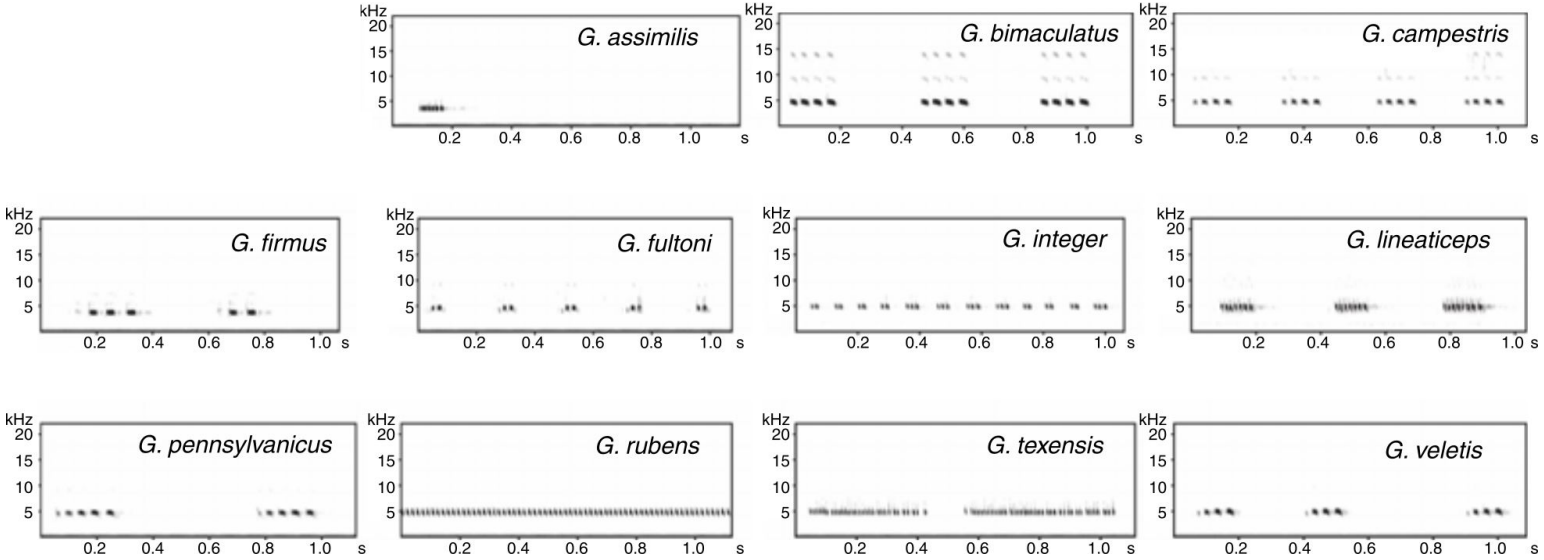


morphological position sequence



molecular sequence

Calling songs of *Gryllus* crickets (data from Robillard et al. 2006 *Cladistics*)



Questions

How rates of evolution vary along the sequence trait?

Are rates autocorrelated?

Do silent and chirp regions show similar rates of evolution?

Methods

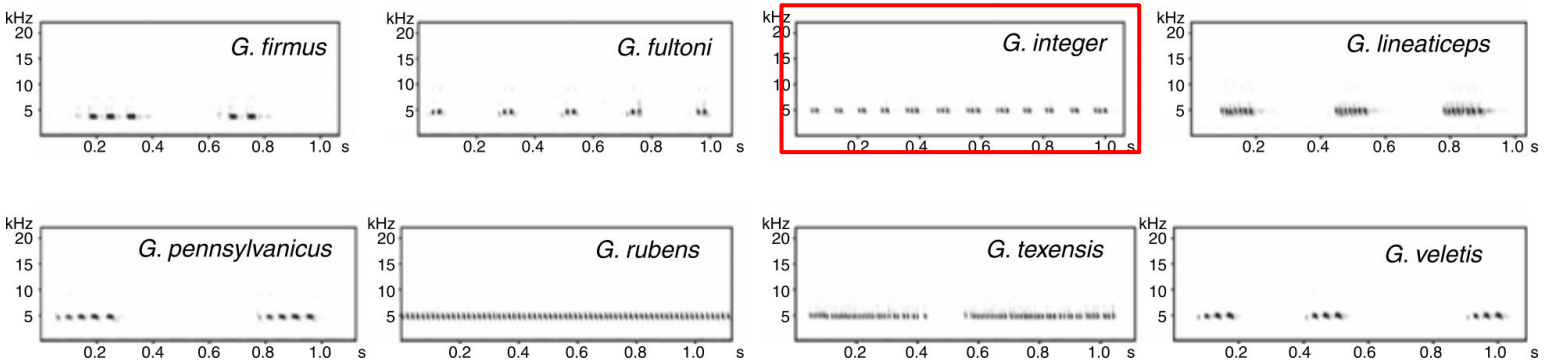
Estimate phylogeny **using molecular data**.

Estimate sequence alignment **conditioned on the phylogeny**.

Estimate rates of evolution along the sequence positions.

Contrast independent rates with autocorrelated rates.

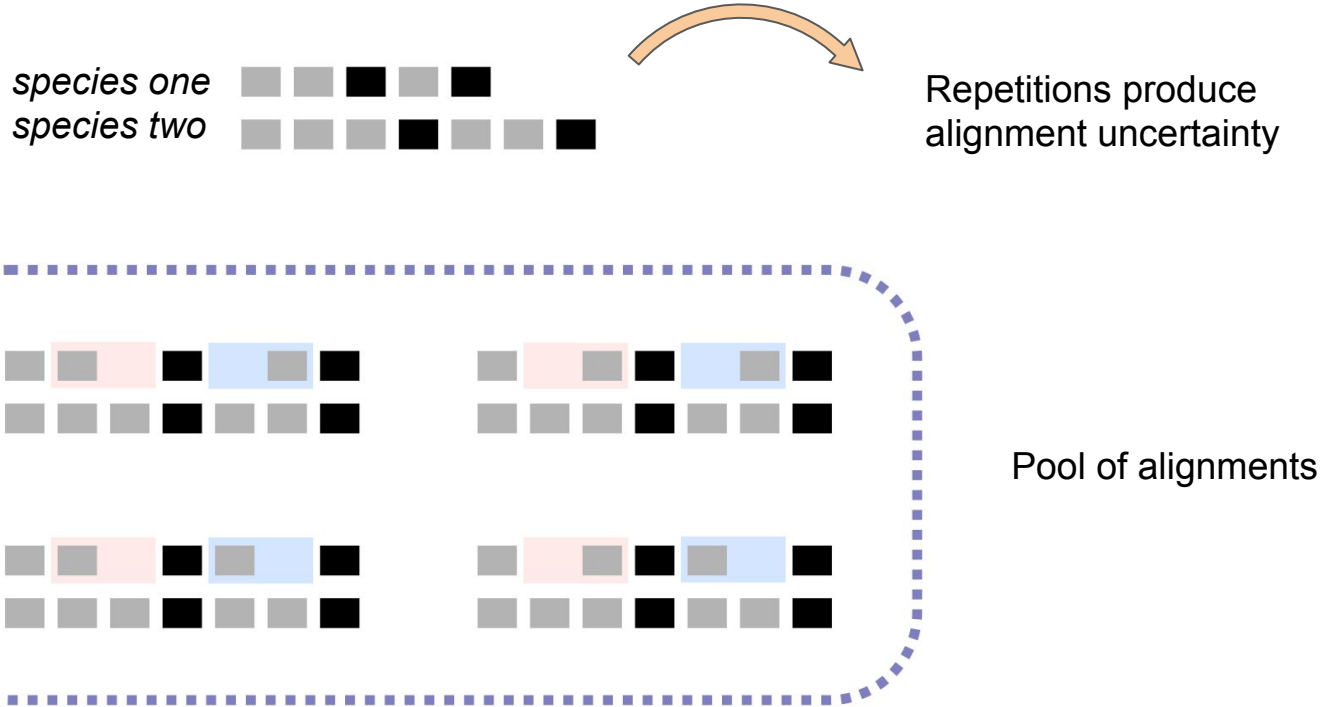
Calling songs of *Gryllus* crickets (data from Robillard et al. 2006 *Cladistics*)

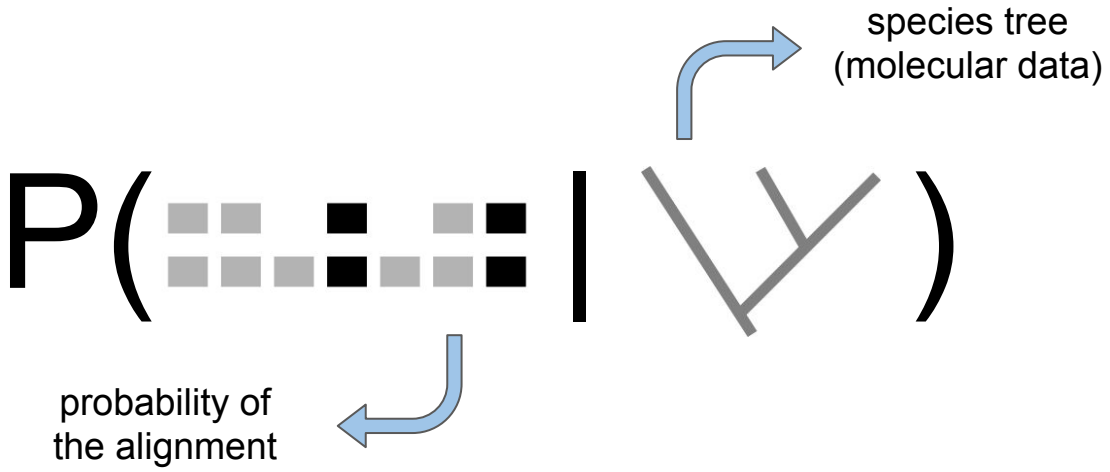


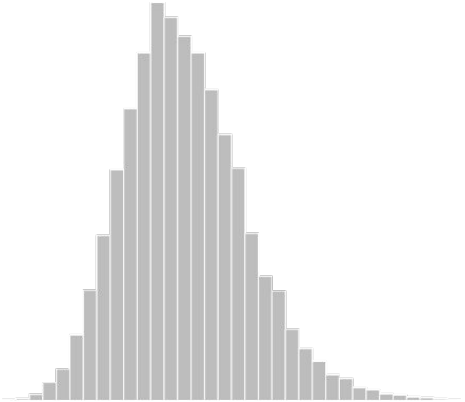
interchirp silence chirp (sound) unknown state

G. integer

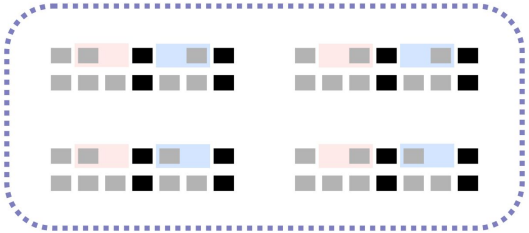
Alignment of non-molecular sequences







posterior distribution
of alignments



BAli---Phy

Suchard and Redelings (2006)

Distance-based alignment with custom distance matrices

MAFFT

Katoh and Standley (2013)

MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences

This feature is supported in versions ≥7.120.

Non-biological sequences

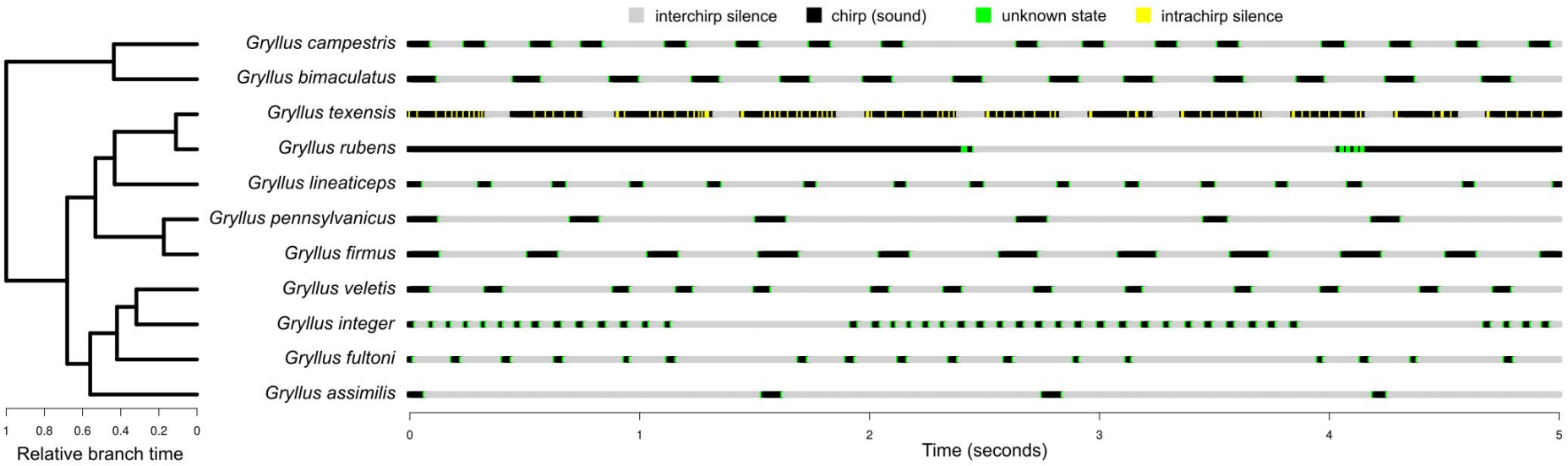
Non-biological sequences, or texts consisting of printable characters, can be aligned in the **--text** mode.

Input:

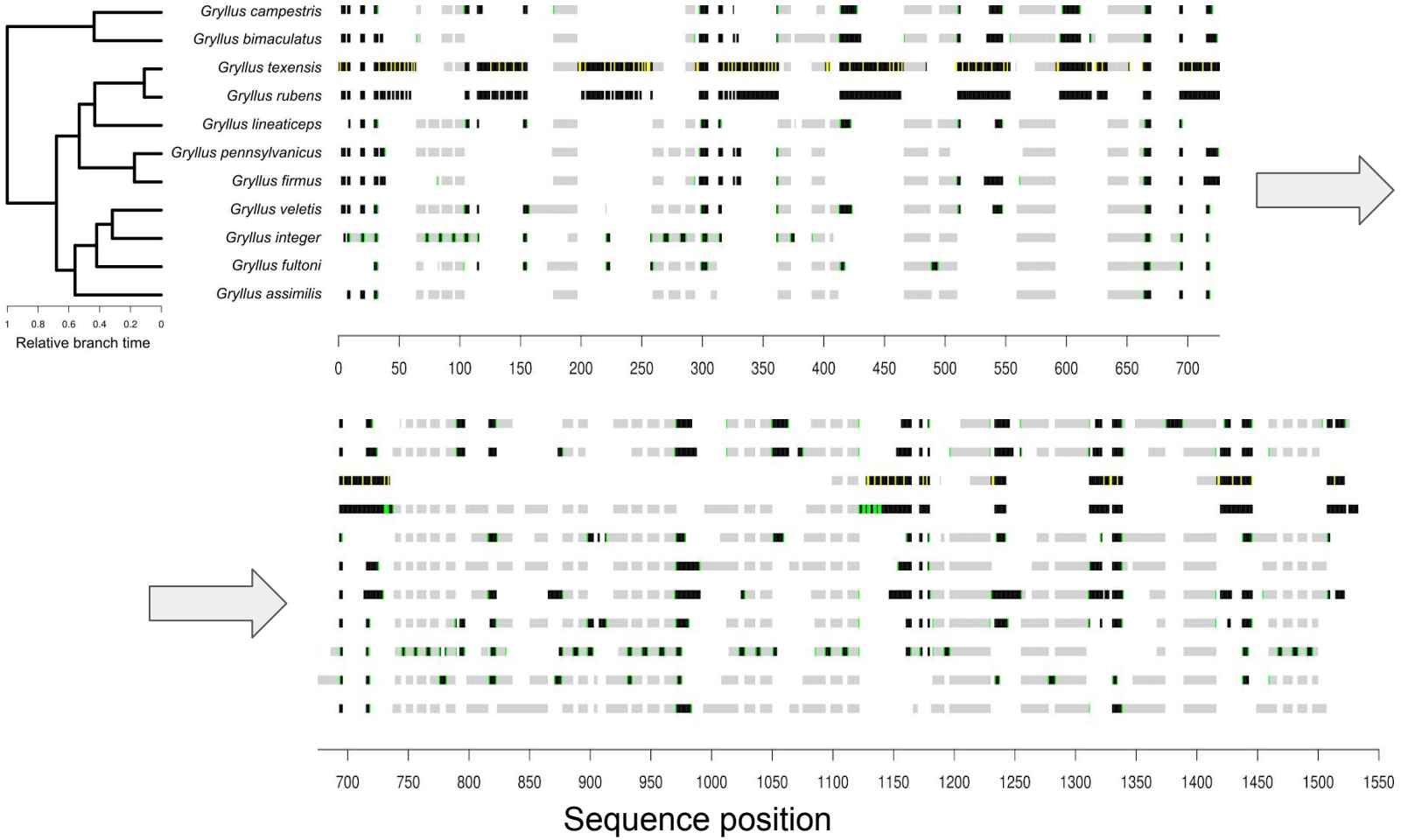
```
> text 1
2008~KATO~Toh
> text 2
2005~Katoh~Kuma~MIYATA~Toh
```

- Most printable characters, except for =<>, are accepted in the input sequences.
- Minus (-) is used to represent a gap in the output. If the input data has -s, they are removed.
- Spaces in the input data are also removed. They have to be converted to another symbol (~ in the

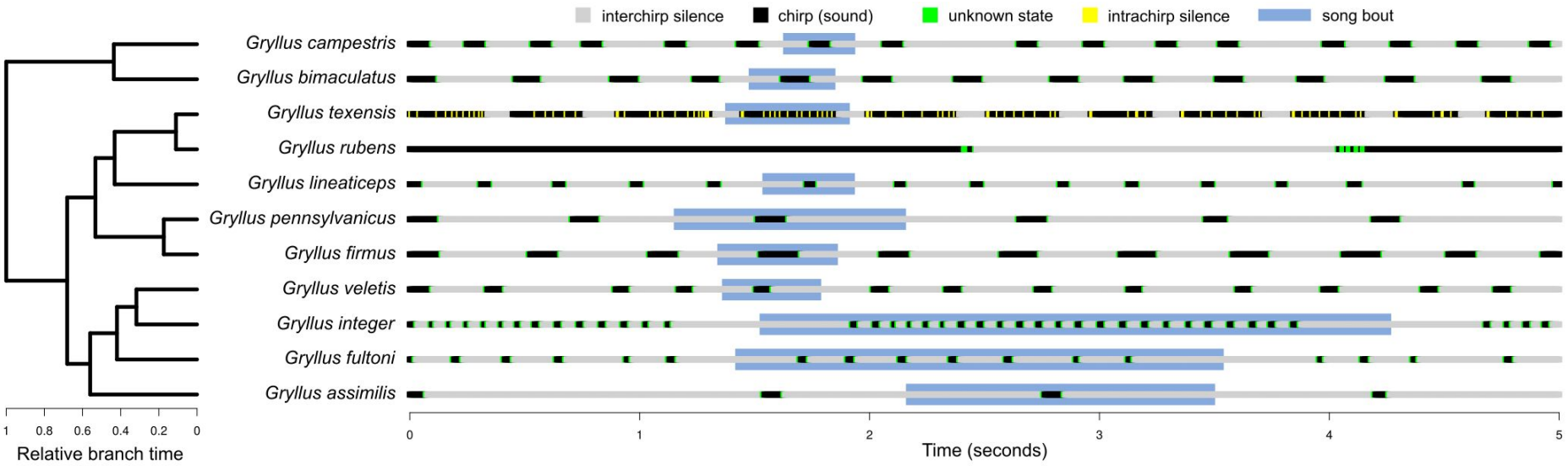
Calling songs of *Gryllus* crickets (data from Robillard et al. 2006 *Cladistics*)



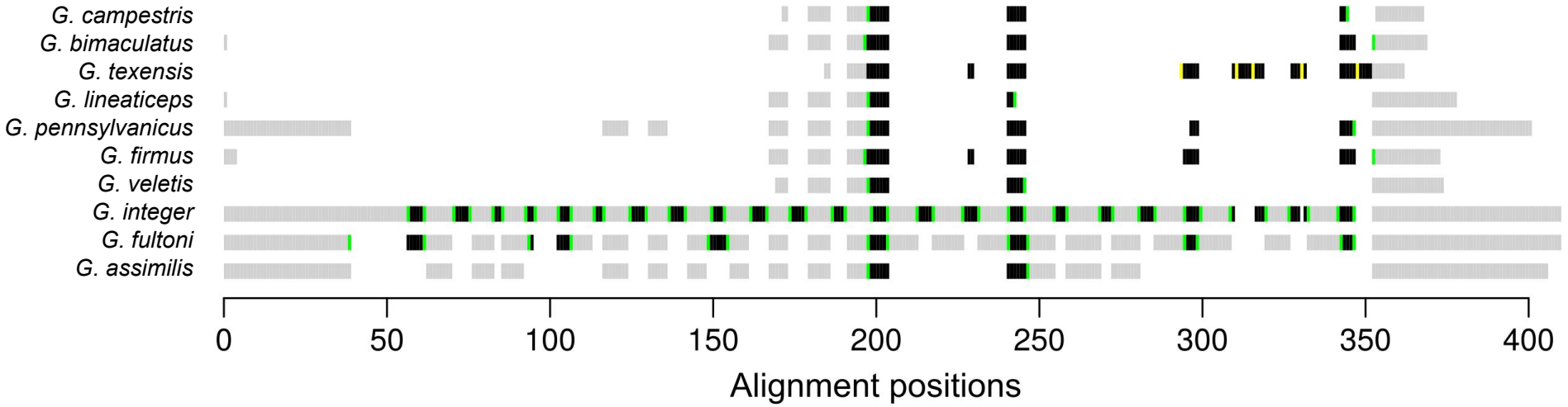
Caetano and Beaulieu caetanods1@gmail.com @ds_caetano



Calling songs of *Gryllus* crickets (data from Robillard et al. 2006 *Cladistics*)



Alignment of song bout sequences



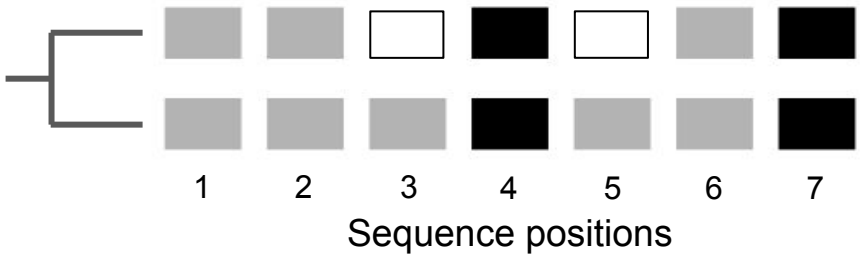
The model

The model

gap state (loss of trait)

sound (chirp)

silence

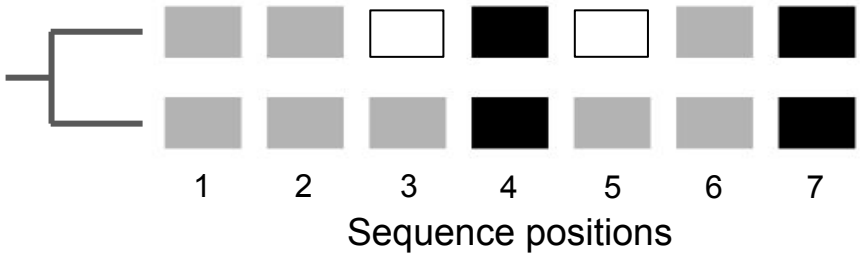


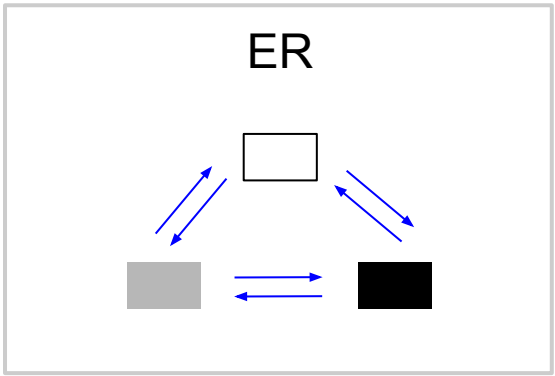
The model

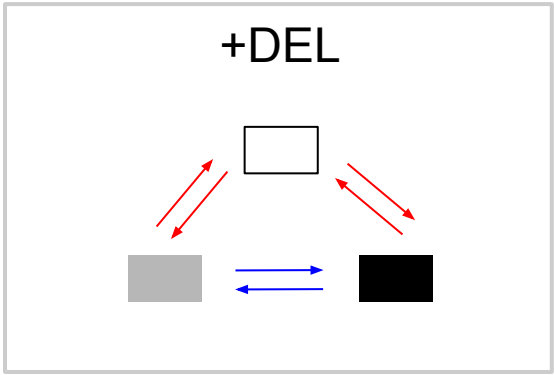
gap state (loss of trait)

sound (chirp)

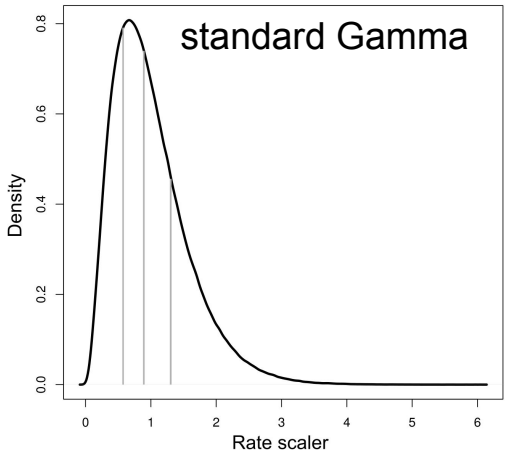
silence

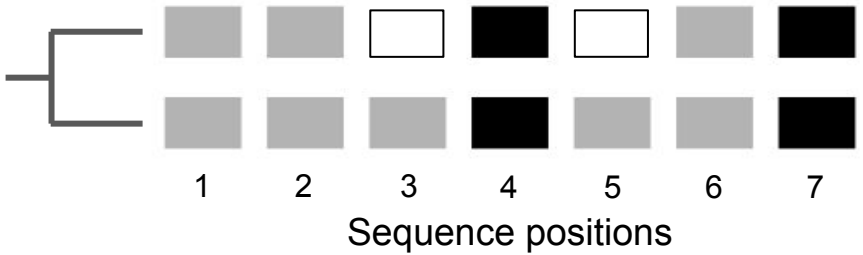


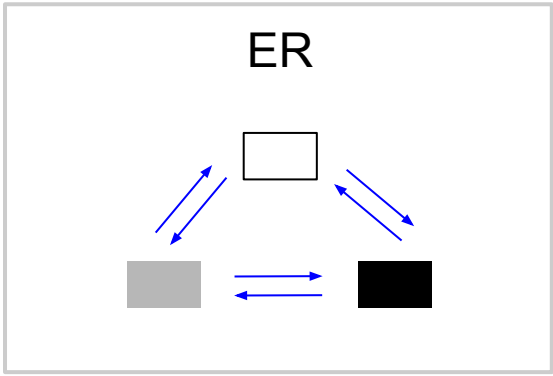


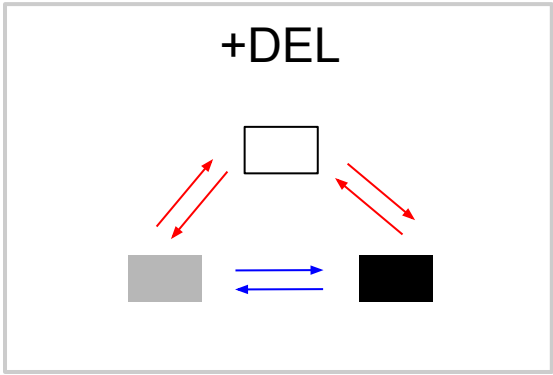


The model

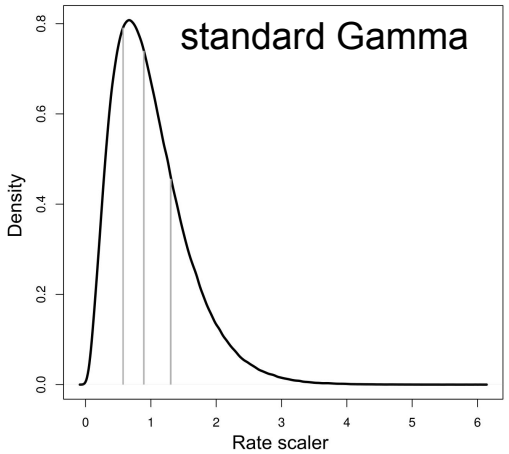

standard Gamma
Density
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
Rate scaler


1
2
3
4
5
6
7
Sequence positions


ER

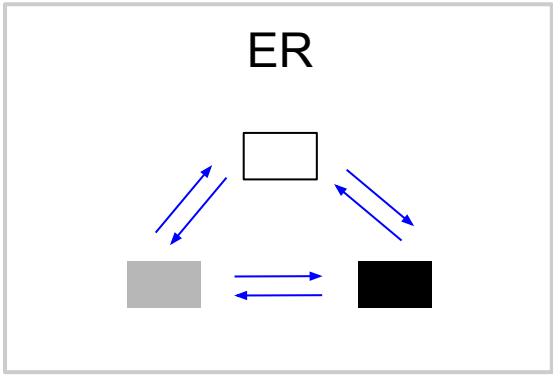

+DEL

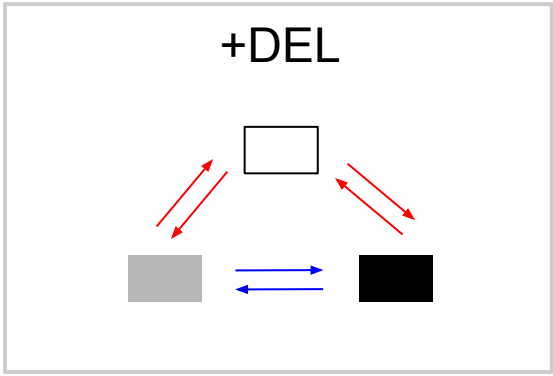
The model

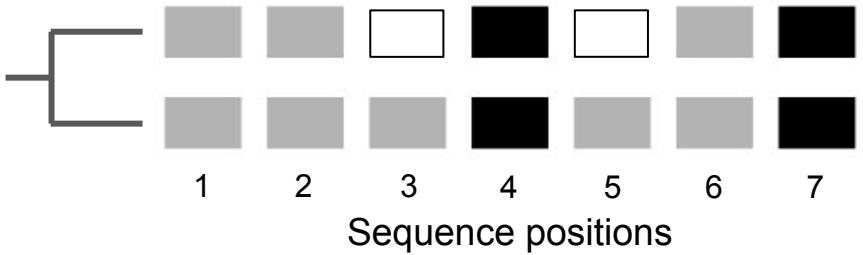


$r \cdot Q$

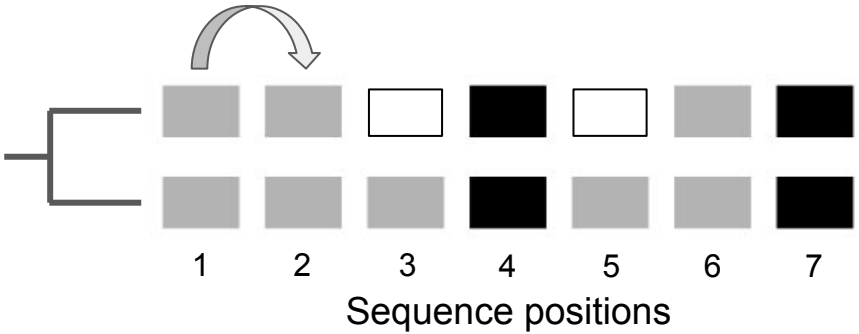
Yang (1994)



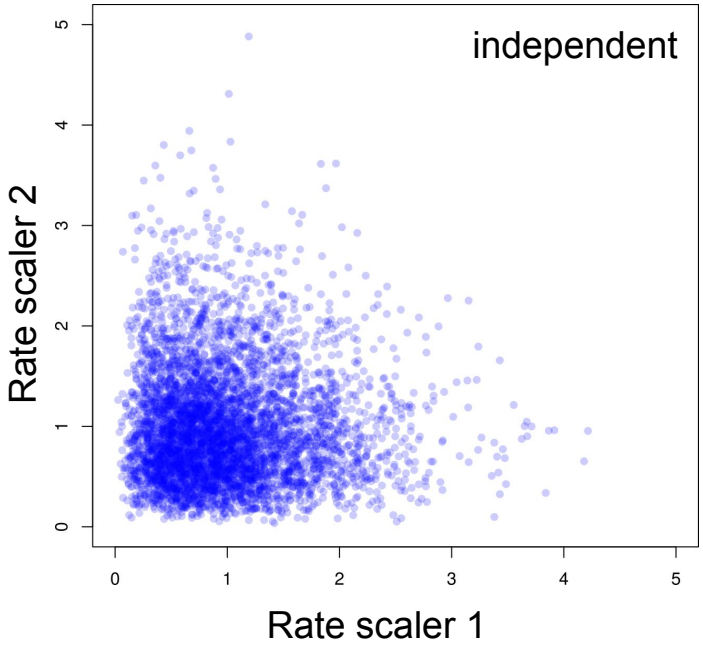


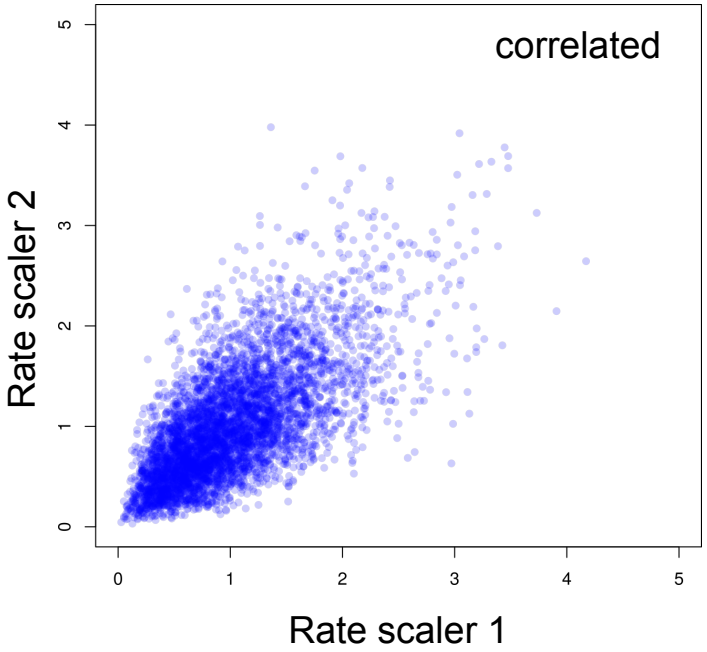


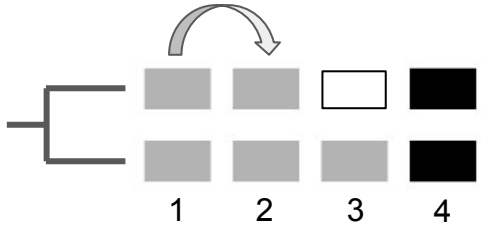
Including autocorrelation of rates



Including autocorrelation of rates

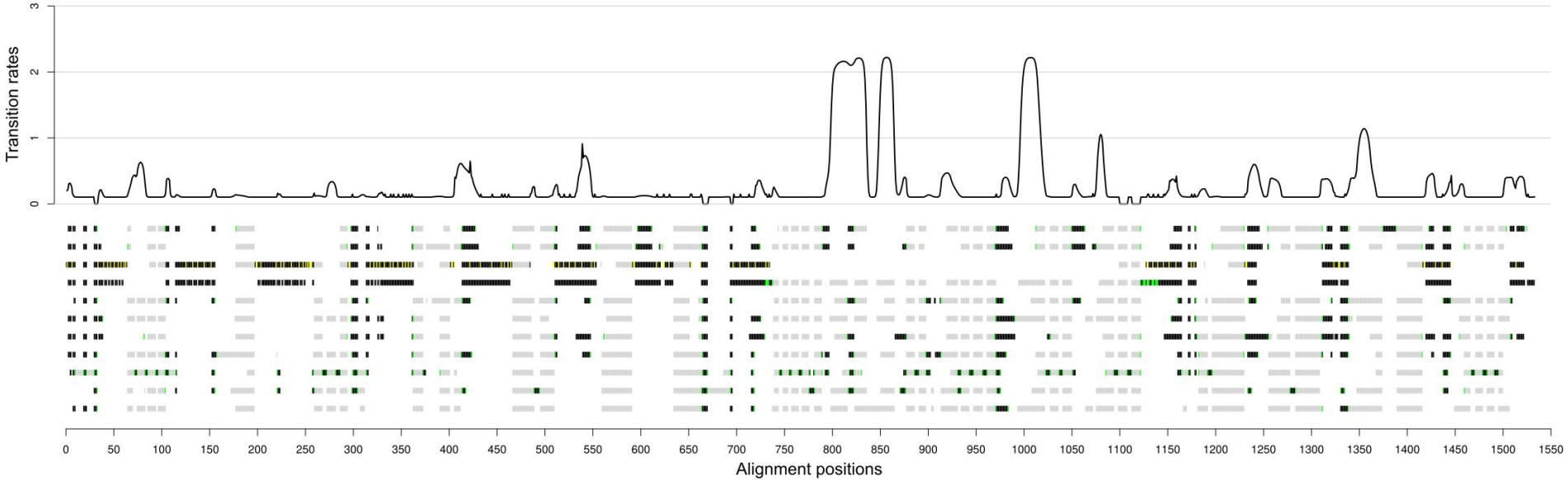




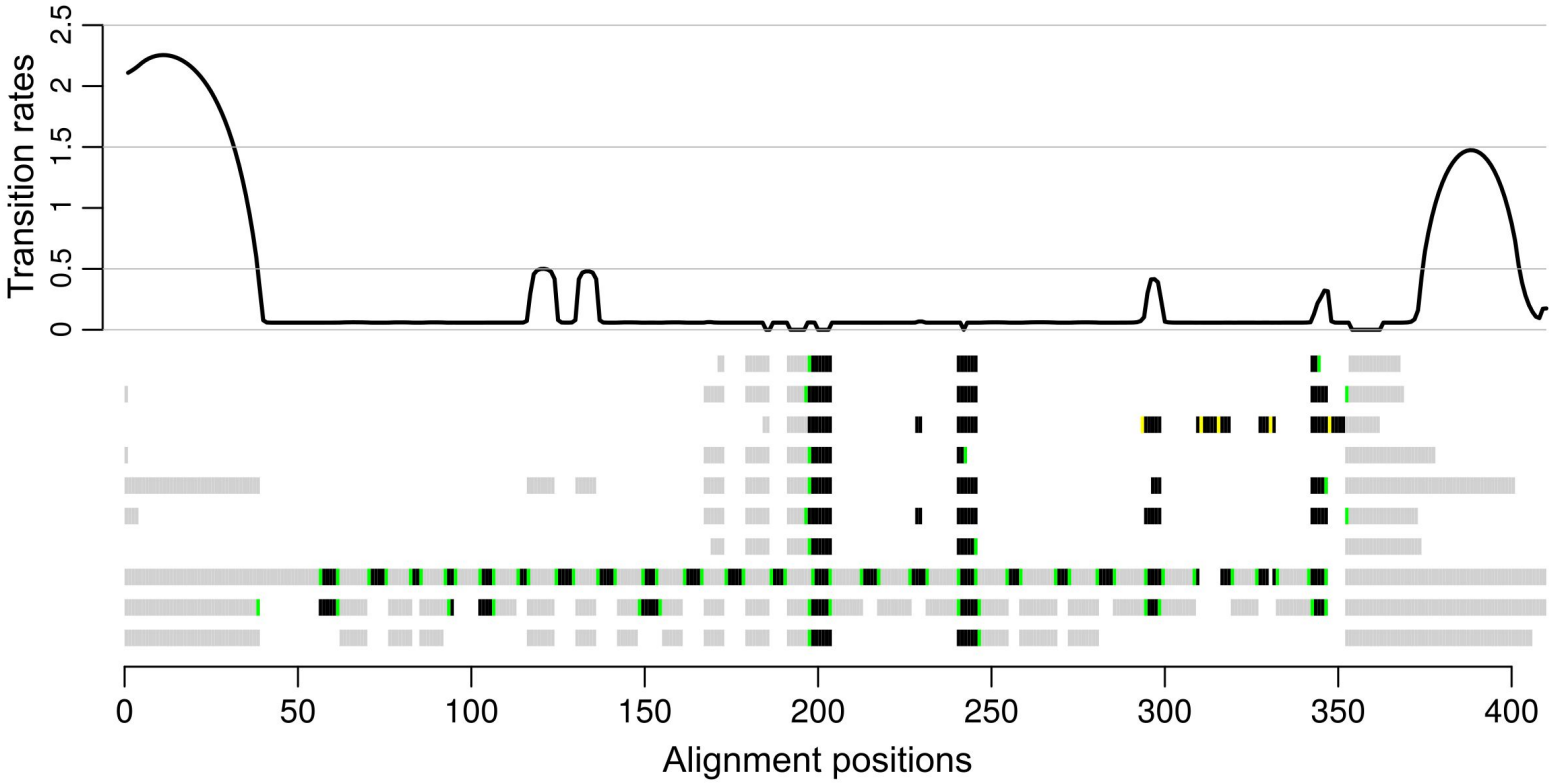


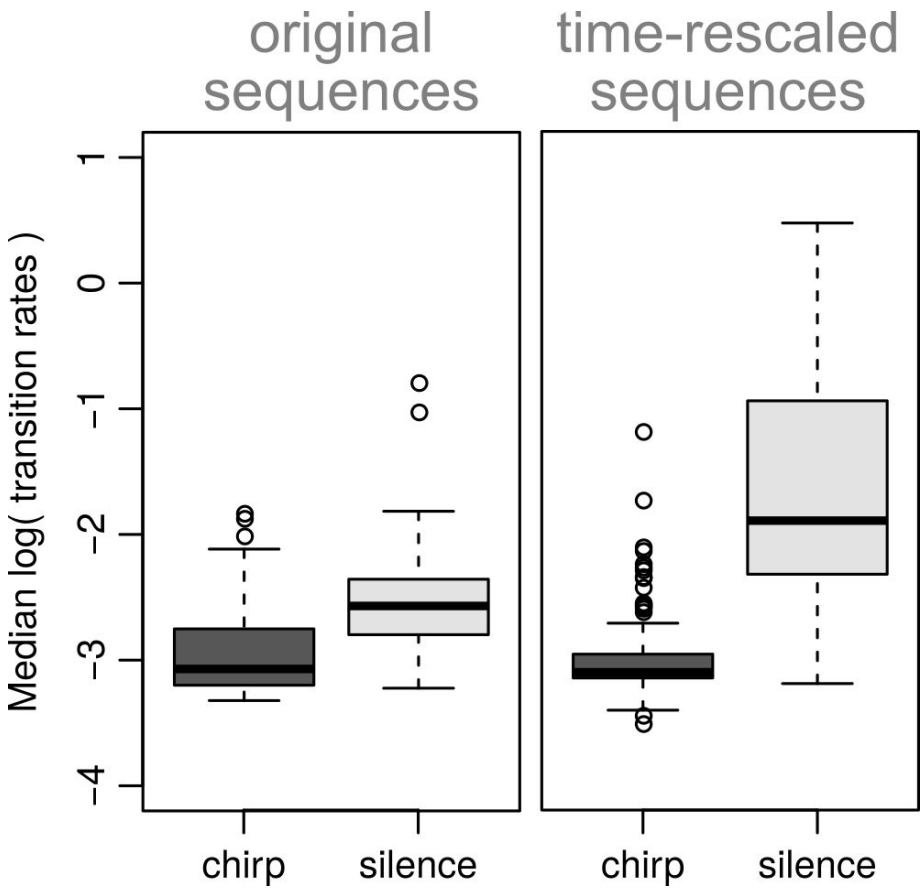
Yang 1995 - *Genetics*

Evolution of temporal pattern is autocorrelated



Evolution of temporal pattern is autocorrelated





Wilcoxon signed-rank test ($p < 0.001$)

Conclusion

Sequence organization allows for context-dependent evolution

An underexplored axis of variation of multivariate traits

Faster rates of evolution for silent region suggests less constraints associated with predation and parasitism

Jeremy Beaulieu
University of Arkansas

Teo Nakov
University of Arkansas

Tony Robillard
Muséum national
d'Histoire naturelle