

# Phylogenetic Gaussian processes for the Ancestral Reconstruction of Bat Echolocation Calls

J.P. Meagher



July 4, 2017

# Acknowledgements

The logo for EPSRC (Engineering and Physical Sciences Research Council) features the letters 'EPSRC' in a bold, purple, sans-serif font. The text is centered between two horizontal green lines.The word 'engage' is written in a lowercase, sans-serif font. Each letter is composed of a cluster of small green dots of varying sizes, creating a textured, bubbly effect.The logo for The Alan Turing Institute consists of the words 'THE ALAN TURING INSTITUTE' in a bold, black, sans-serif font, stacked vertically within a black rectangular border.The logo for GPAO features the letters 'GPAO' in a bold, black, sans-serif font. The letter 'O' is replaced by a black circle containing a white silhouette of a person in a batting stance, holding a bat and a ball.

Mark Girolami  
Imperial College London

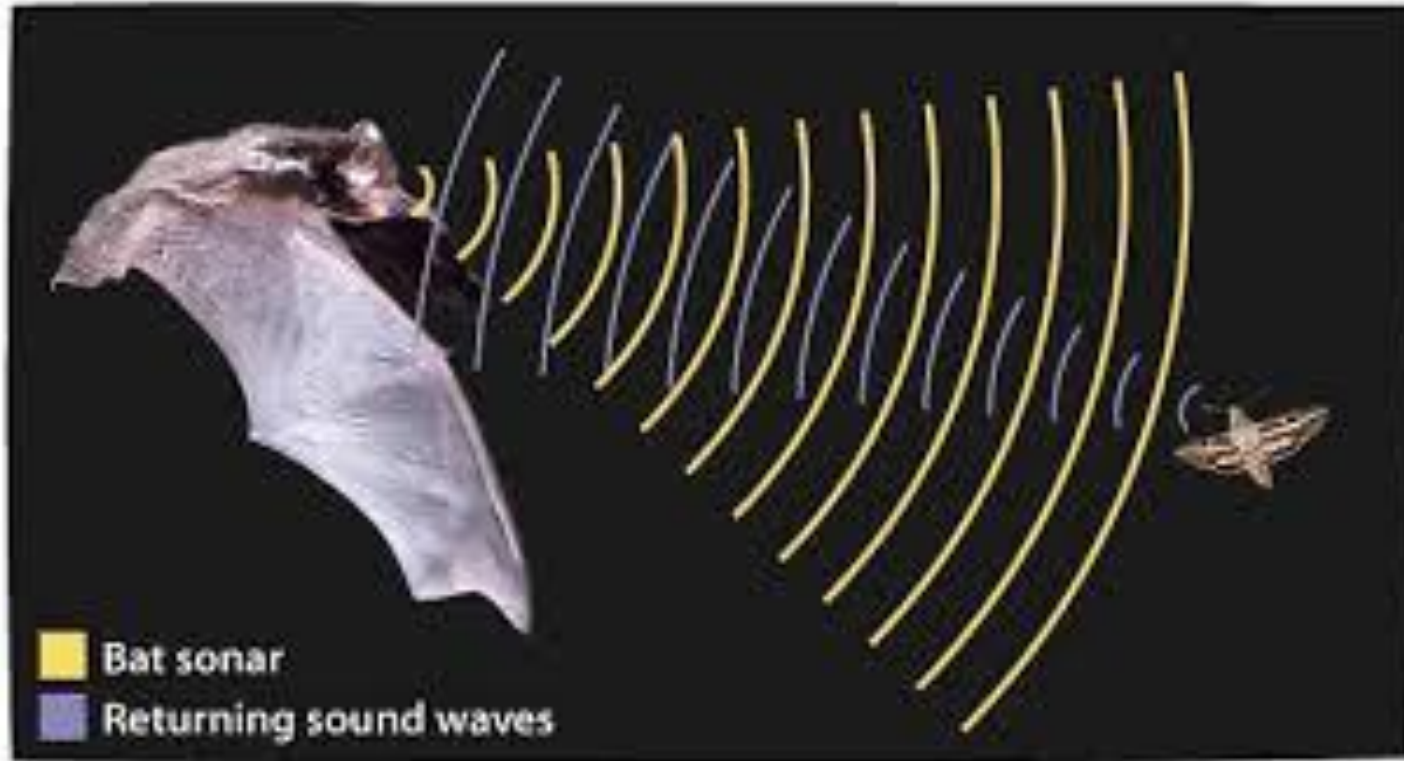


Theo Damoulas  
University of Warwick



Kate Jones  
University College London

# Bats: Some Background



# Bats: Engaging the Public

## Science & Environment

### Bat-sound library tracks biodiversity

By Helen Briggs  
BBC News

14 April 2016 | Science & Environment



El Pinacate reserve is one of the driest places on Earth and home to many bats

Scientists have compiled the biggest known library of bat sounds in an effort to identify and conserve rare species.

#### Top Stories

##### MP swaps sides over 'untrue' Leave claims

Tory MP Sarah Wollaston quits the campaign to leave the EU saying its claim that £350m a week could be saved for the NHS 'simply isn't true'.

48 minutes ago

##### Government criticised over flood plans

1 hour ago

##### Body of missing man found in Vietnam

34 minutes ago

#### Features



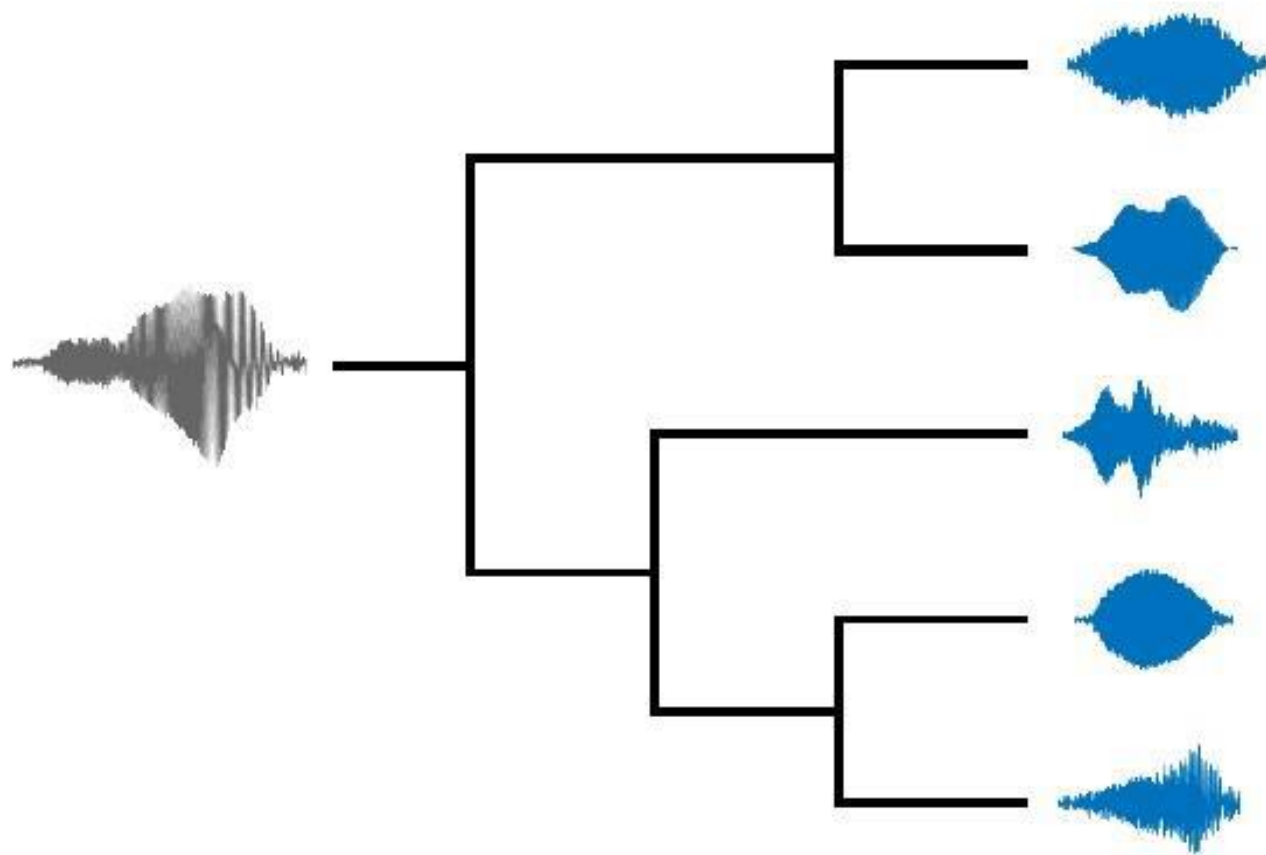
##### Political gaffe

Russian PM comments to pensioner spark

# Bats: Non-invasive Monitoring



# Research Project Summary

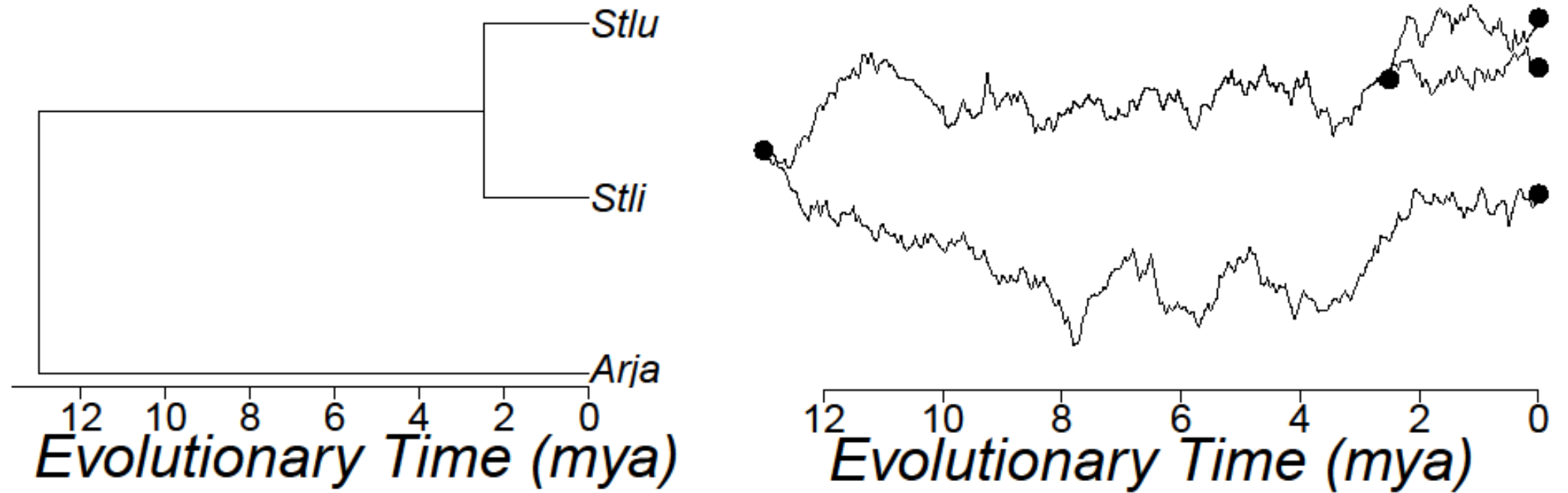


# Ancestral Reconstruction: A Definition

The extrapolation back in time from measured characteristics of individuals (or populations) to their common ancestors.

# Continuous Character Trait Evolution

Phylogenetic Ornstein-Uhlenbeck Process





# Function-Valued Traits: A Definition

A trait that is repeatedly measured, along some continuous scale, where measurements can represent points on a curve, and both means and covariances along the trajectory can change, gradually and continually.

# Gaussian Processes

A collection of random variables, any finite number of which have a joint Gaussian distribution.

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

$$m(\mathbf{x}) = \mathbf{E}(f(\mathbf{x}))$$

$$k(\mathbf{x}, \mathbf{x}') = \text{cov}(\mathbf{x}, \mathbf{x}')$$

$$\mathbf{x} \in \mathbf{R}^p$$

# Phylogenetic Gaussian Processes: Assumptions

- 1. Conditional on their common ancestors in the phylogenetic tree  $\mathbf{T}$  any two traits are statistically independent.*
- 2. The statistical relationship between a trait and any of its descendants in  $T$  is independent of the topology of  $\mathbf{T}$*
- 3. The marginal GP along each branch of the phylogeny is space-time separable.*

$$\Sigma((f, t), (f', t')) = K(f, f')k(t, t')$$

# Phylogenetic Gaussian Processes: Theoretical Result

For a phylogenetic Gaussian Process  $Y$  with time-space separable covariance function, when  $K$  is a degenerate Mercer kernel, there exists a set of  $n$  deterministic basis functions  $\phi_i : F \rightarrow \mathbf{R}$  and univariate GPs  $X_i$  for  $i = 1, \dots, n$  such that

$$g(f, \mathbf{t}) = \sum_{i=1}^n \phi_i(f) X_i(\mathbf{t})$$

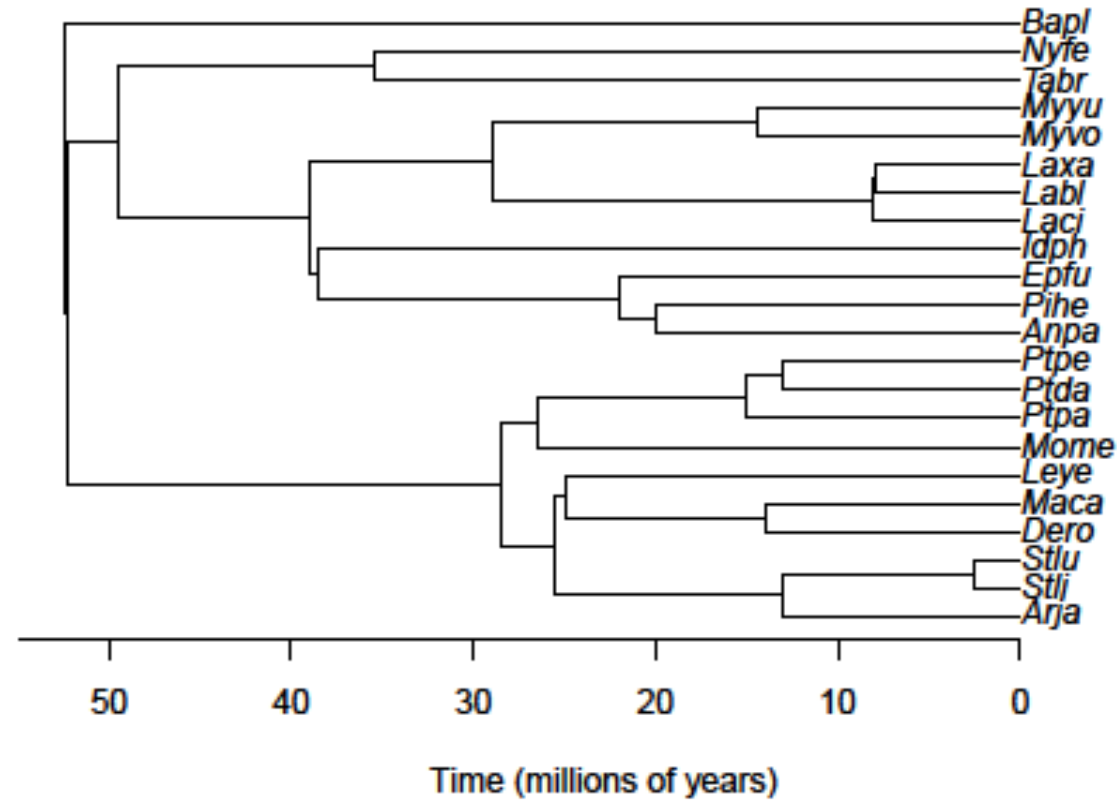
has the same distribution as  $Y$ . The full phylogenetic covariance function of this phylogenetic GP is

$$\Sigma_{\mathbf{T}}((f, \mathbf{t}), (f', \mathbf{t}')) = \sum_{i=1}^n k_{\mathbf{T}}^i(\mathbf{t}, \mathbf{t}') \phi_i(f) \phi_i(f'),$$

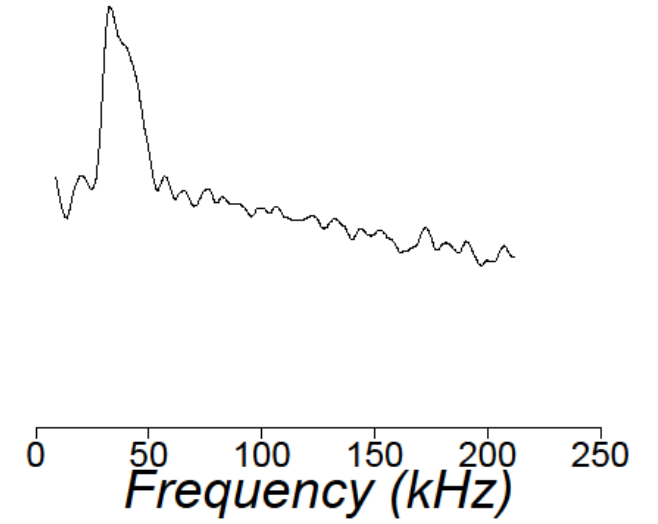
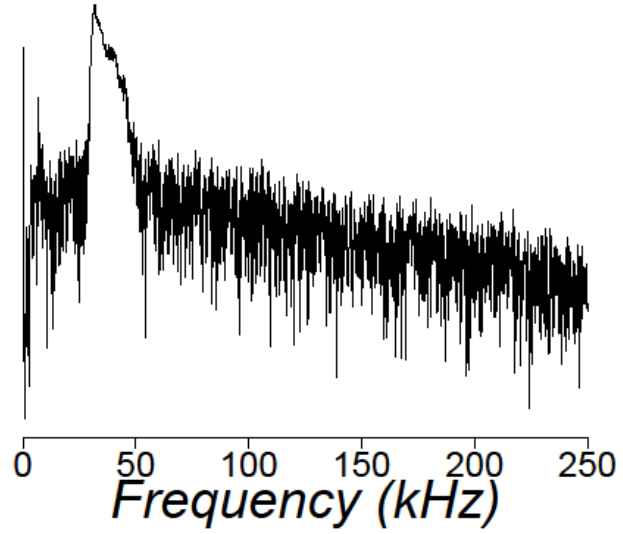
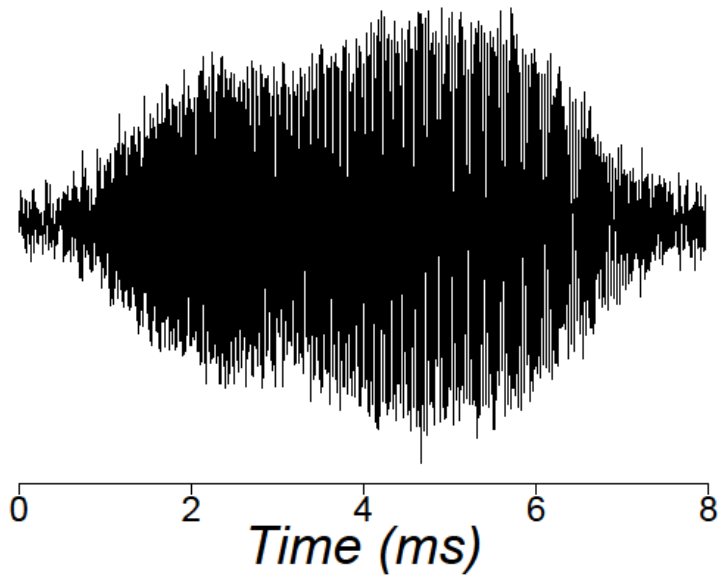
where  $\int \phi_i(f) \phi_j(f) df = \delta_{ij}$ ,  $\delta$  being the Kronecker delta, and so the phylogenetic covariance function depends only on  $\mathbf{t}, \mathbf{t}' \in \mathbf{T}$ .

# Bat Echolocation Calls: The Phylogenetic Tree

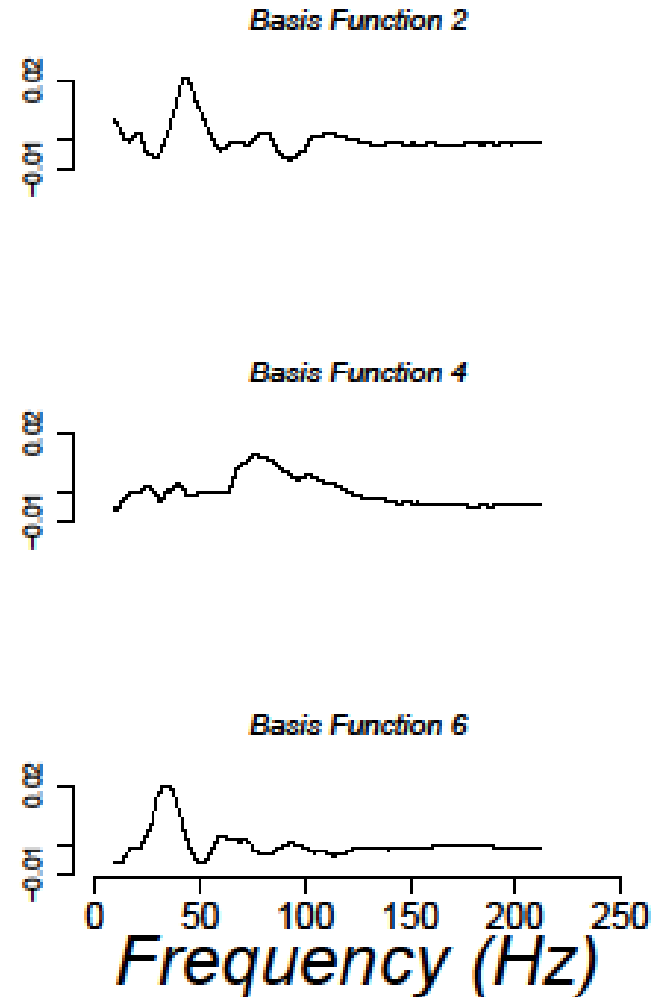
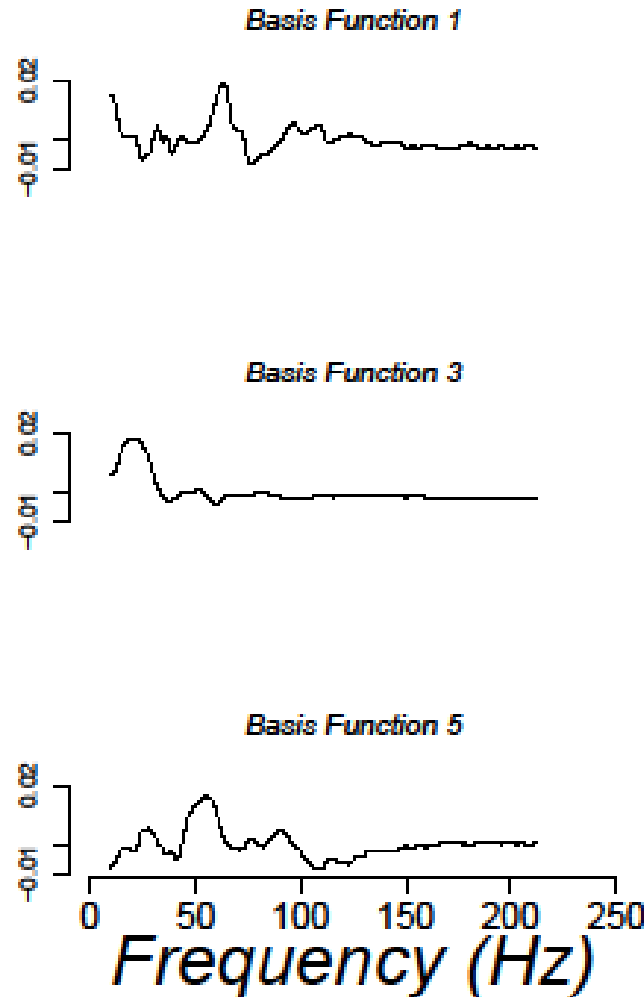
Mexican Bat Phylogeny



# Bat Echolocation Calls: A Function-Valued Trait



# Bat Echolocation Calls: Independent Basis



# Results: Phylogenetic Ornstein-Uhlenbeck Process

$$k_{\mathbf{T}}^i(\mathbf{t}, \mathbf{t}') = (\sigma_p^i)^2 \exp\left(\frac{-d_{\mathbf{T}}(\mathbf{t}, \mathbf{t}')}{\ell^i}\right) + (\sigma_n^i)^2 \delta_{\mathbf{t}, \mathbf{t}'}$$



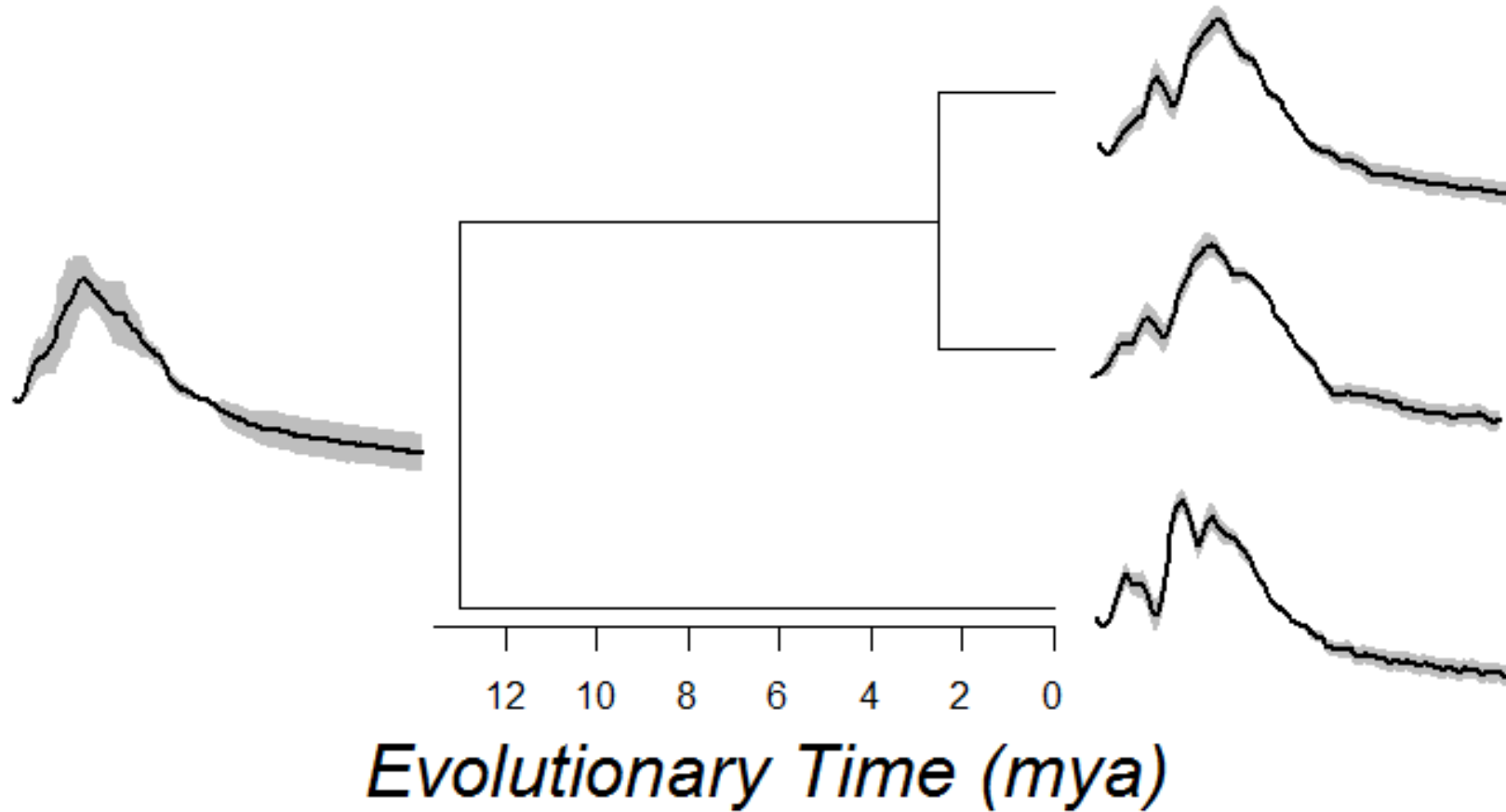
# Results: Hyperparameter Estimates

Basis	$\hat{\sigma}_p$	$\hat{\ell}$	$\hat{\sigma}_n$
1	2.30 (0.11)	12.27 ( 4.18)	1.18 (0.11)
2	3.17 (0.11)	27.63 ( 3.70)	1.26 (0.13)
3	4.05 (0.32)	70.50 (20.31)	1.19 (0.12)
4	3.32 (0.17)	22.86 ( 8.95)	1.96 (0.19)
5	3.00 (0.13)	26.93 ( 2.85)	1.21 (0.11)
6	3.70 (0.14)	12.82 ( 4.52)	1.28 (0.15)

# Results: Hyperparameter Estimates

Basis	$\hat{\sigma}_p$	$\hat{\ell}$	$\hat{\sigma}_n$
1	2.30 (0.11)	12.27 ( 4.18)	1.18 (0.11)
2	3.17 (0.11)	27.63 ( 3.70)	1.26 (0.13)
3	4.05 (0.32)	70.50 (20.31)	1.19 (0.12)
4	3.32 (0.17)	22.86 ( 8.95)	1.96 (0.19)
5	3.00 (0.13)	26.93 ( 2.85)	1.21 (0.11)
6	3.70 (0.14)	12.82 ( 4.52)	1.28 (0.15)

# Results: Posterior Predictive Distributions



# Conclusions

- Phylogenetic Signal Identified
- Strong Phylogenetic signal at low frequencies.

## Future Work

- Implement the Phylogenetic Gaussian Process Regression Model for Echolocation Call Spectrograms.
- Consider Alternative Echolocation Call Representations.
- Consider Alternative models for the evolutionary dynamics.

