

# A Flexible Bayesian Framework to Study Viral Trait Evolution

Paul Bastide<sup>1</sup>, G. Baele<sup>1</sup>, M. A. Suchard<sup>2,3</sup> and P. Lemey<sup>1</sup>

KU LEUVEN

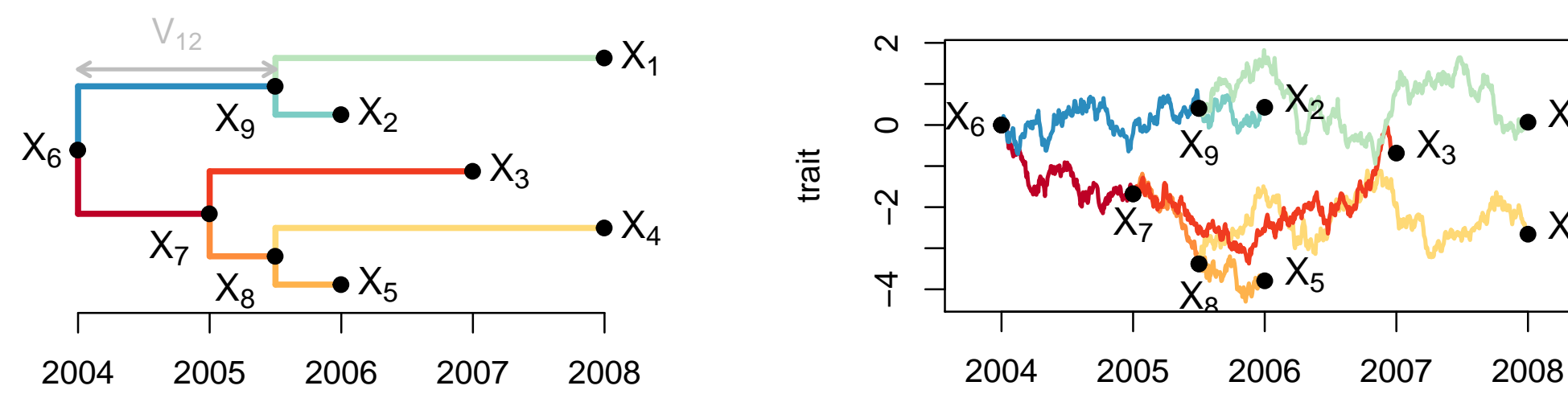
UCLA

<sup>1</sup> Evolutionary and Computational Virology, Rega Institute, KU Leuven, Belgium

<sup>2</sup> Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, UCLA, USA

<sup>3</sup> Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, USA

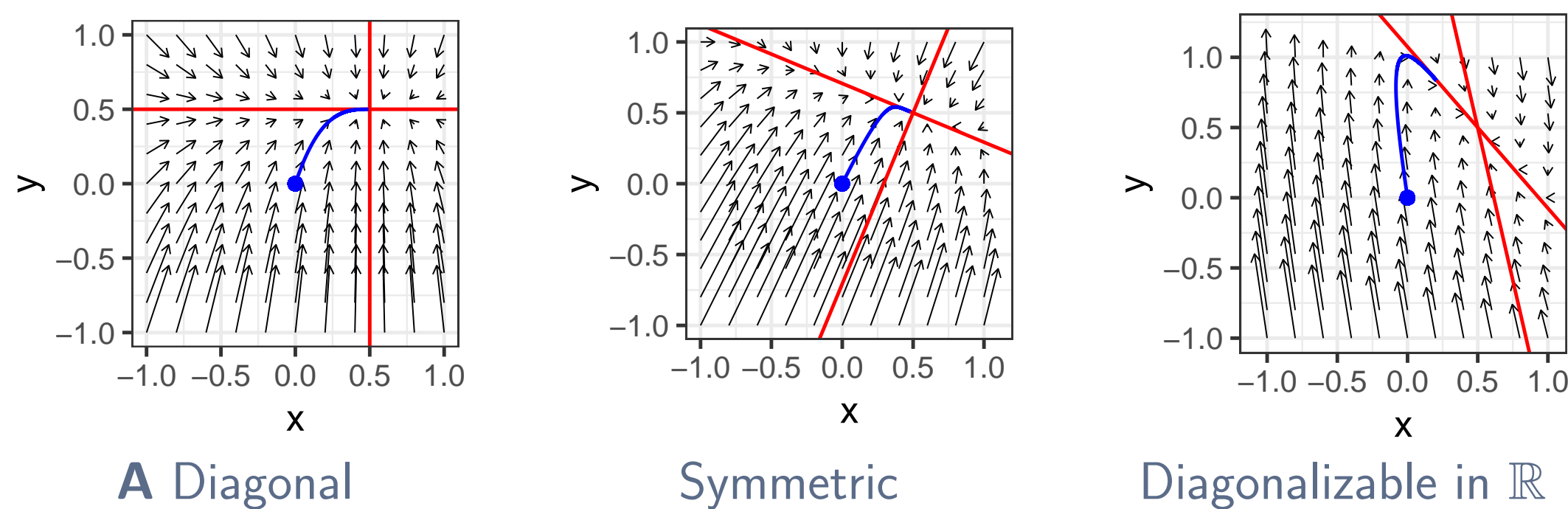
## Model: Stochastic Process on a Tree



Brownian Motion tip distribution:  $\mathbf{X} \sim \mathcal{MN}(\mathbf{1}_n \mu^T, \mathbf{V}, \mathbf{R})$

## Multivariate Ornstein-Uhlenbeck

$$d\mathbf{W}(t) = \mathbf{A}[\boldsymbol{\beta}(t) - \mathbf{W}(t)]dt + \boldsymbol{\Sigma}d\mathbf{B}(t)$$



## General Model

[2]

$$\begin{aligned} \mathbf{X}_1 &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) && \text{root} \\ \mathbf{X}_j \mid \mathbf{X}_{\text{pa}(j)} &\sim \mathcal{N}(\mathbf{q}_j \mathbf{X}_{\text{pa}(j)} + \mathbf{r}_j, \boldsymbol{\Sigma}_j) && \text{latent tips and nodes} \\ \mathbf{Y}_i^o \mid \mathbf{X}_{\text{pa}(i)} &\sim \mathcal{N}(\mathbf{X}_{\text{pa}(i)}, \mathbf{S}) && \text{observations} \end{aligned}$$

Includes: BM, OU, Drift, Integrated OU, ...

## Bayesian Inference Framework

Assumption:  $\mathbf{Y}$  and  $\mathbf{S}$  independent conditionally on  $\mathcal{T}$ .

$$\begin{aligned} p(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi} \mid \mathbf{Y}, \mathbf{S}) &\propto p(\mathbf{Y}, \mathbf{S} \mid \boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}) \\ &\propto p(\mathbf{Y} \mid \boldsymbol{\theta}, \mathcal{T}) p(\mathbf{S} \mid \mathcal{T}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}) \end{aligned}$$

This poster:  $\mathcal{T}$  fixed.

## Sampling the Correlation and Attenuation [3]

- ▶ Cholesky  $\mathbf{C} = \mathbf{W}^T \mathbf{W}$ 
  - ▷  $W_{1k}^2 + \dots + W_{kk}^2 = 1$  (cor)
  - ▷  $W_{kk} > 0$  (identifiability)

↔ Each column  $\mathbf{W}_{\cdot k}$  is in the half euclidean unit sphere  $\mathcal{S}_k^h$ .

- ▶ Diag.  $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ 
  - ▷  $\lambda_k \in \mathbb{R}_+^*$  (model)
  - ▷  $\lambda_1 < \dots < \lambda_p$  (identifiability)

↔ Each column  $\mathbf{P}_{\cdot k}$  is in the half euclidean unit sphere  $\mathcal{S}_p^h$ .

## Efficient Likelihood Computation

- ▶  $\log p(\mathbf{Y}^o)$ : one post-order traversal of the tree.
- ▶  $\frac{\partial}{\partial \phi_j} [\log p(\mathbf{Y}^o)]$ : one pre-order traversal of the tree.

↔  $O(np^3)$ .

## Implementation: BEAST [4]

- 
- ▶ Tree and Trait estimation.
  - ▶ MCMC, HMC, Marginal likelihood.

## HIV Virulence Heritability

[1, 5]

- ▶ Measures (correlated):
  - ▷ Set point viral load
  - ▷ CD4+ T cell decline rate
- ▶ Heritability:

$$h_k^2 = \frac{V(\mathbf{X}_{1:n,k})}{V(\mathbf{Y}_{1:n,k}^o)}$$

- ▶ Results
  - $h^2$  VL: 17% [0.007, 82.5]% (95% CI)
  - $h^2$  CD4: 0.02% [0.0024, 0.16]%
- ↔ “Consistent” with previous estimates.

OU  
BM

## Conclusion and Perspectives

A general statistical framework for trait evolution models with dated tips.

- ▶ Perspectives:
  - ▷ Develop HMC.
  - ▷ Identifiability of  $\mathbf{A}$ .
  - ▷ Applications: geography, paleontology.



pbastide.github.io

## References

- [1] Alizon, von Wyl, Stadler, et al. 2010. *PLoS Pathog.* 6.
- [2] Bastide, Ané, Robin, et al. 2018. *Syst. Biol.* 67:662–680.
- [3] Lewandowski, Kurowicka, Joe. 2009. *J. Multivar. Anal.* 100:1989–2001.
- [4] Suchard, Lemey, Baele, et al. 2018. *Virus Evol.* 4:1–5.
- [5] Vrancken, Lemey, Rambaut, et al. 2015. *Methods Ecol. Evol.* 6:67–82.