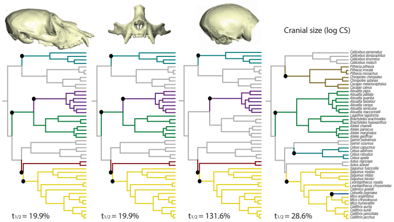
# A Flexible Bayesian Framework to Study Viral Trait Evolution

Paul Bastide[1], LST Ho[2], G Baele[3], P Lemey[3] and MA Suchard[4]

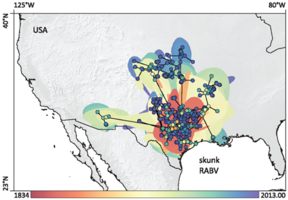
[1] IMAG, Université de Montpellier, CNRS
[2] Department of Mathematics and Statistics, Dalhousie University, Canada
[3] Evolutionary and Computational Virology, Rega Institute, KU Leuven, Belgium
[4] Department of Biostatistics, Biomathematics and Human Genetics, UCLA, USA

20 May 2021

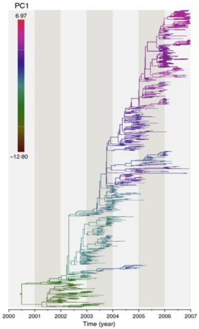
# Phylogenetic Comparative Methods




Aristide et al. (2018)




Dellicour et al. (2017)



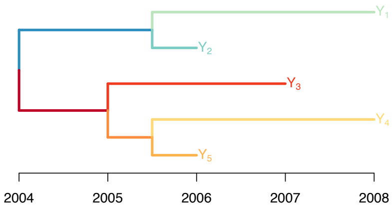

Vrancken et al. (2015)

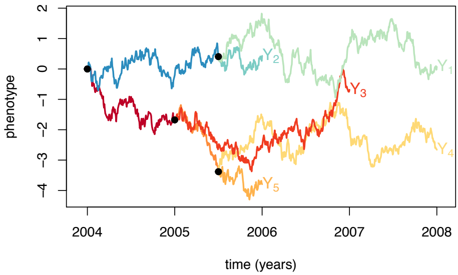
- Various time scales: Myr – decade.
- Various traits: morpho, geo, viral.

Question: Trait dynamics for an evolving organism ?

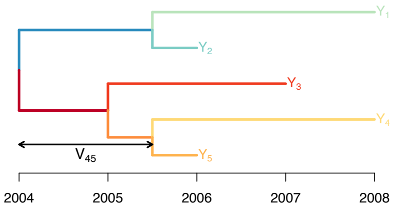
# Outline

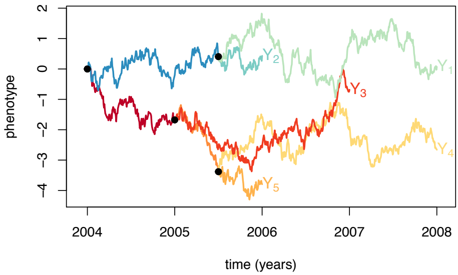
# BM on a Tree


Y1
Y2
Y3
Y4
Y5
2004
2005
2006
2007
2008


phenotype
2
1
0
−1
−2
−3
−4
Y1
Y2
Y3
Y4
Y5
2004
2005
2006
2007
2008
time (years)

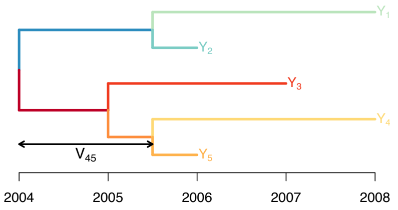
# BM on a Tree

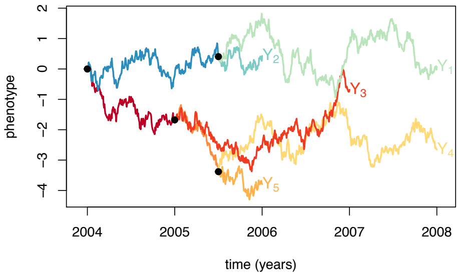



| | | |
|---|---|---|
| SDE: | $\mathrm{d}\, X_t = \sigma \,\mathrm{d}\, B_t$ | |
| Variance: | $\mathbb{C}\mathrm{ov}\left[Y_4; Y_5\right] = \sigma^2 \times V_{45}$ | shared evolution time |
| Expectation: | $\mathbb{E}\left[Y_i\right] = \mu$ | ancestral root value |

# BM on a Tree

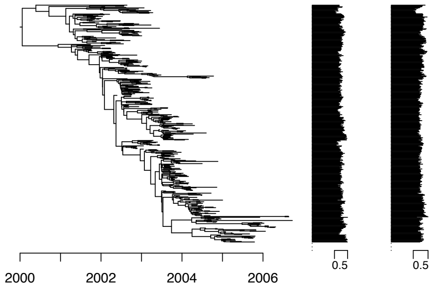



Distribution: Normal

$$\mathbf{Y} \sim \mathcal{N}(\mu \mathbf{1}_n, \sigma^2 \mathbf{V})$$

# Multivariate BM

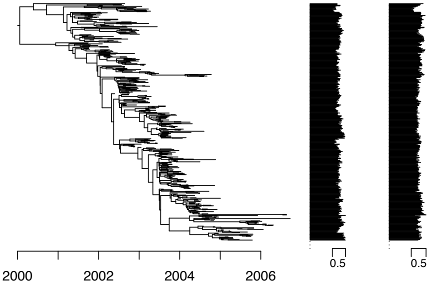



Data: Vectors of $p$ traits

$$\mathbf{Y}_i^T = (Y_{i1}, \dots, Y_{ip})$$

Tree: Influenza H3N2 (Lemey et al., 2014)

# Multivariate BM




Data: Vectors of $p$ traits

$$\mathbf{Y}_i^T = (Y_{i1}, \dots, Y_{ip})$$

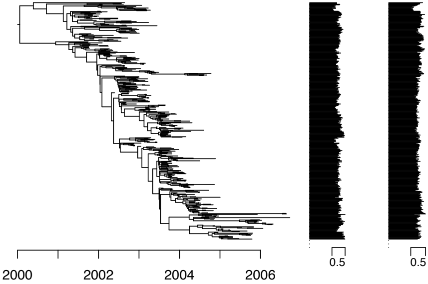
SDE: $\mathrm{d}\,\mathbf{X}_t = \boldsymbol{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$ $\qquad \mathbf{R} = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$

Variance: $\mathbb{C}\mathrm{ov}\left[Y_{ik}; Y_{jl}\right] = R_{kl} \times V_{ij}$ shared evolution time

Expectation: $\mathbb{E}\left[\mathbf{Y}_{\cdot k}\right] = \mu_k$ ancestral root value

Tree: Influenza H3N2 (Lemey et al., 2014)

# Multivariate BM




Data: Vectors of $p$ traits

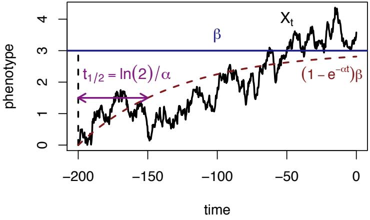
$$\mathbf{Y}_i^T = (Y_{i1}, \dots, Y_{ip})$$

Distribution: Matrix Normal

$$\mathbf{Y} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R}) \qquad \mathbb{V}\mathrm{ar}\left[\mathrm{vec}(\mathbf{Y})\right] = \mathbf{R} \otimes \mathbf{V}$$

Tree: Influenza H3N2 (Lemey et al., 2014)

# Ornstein-Uhlenbeck Modeling (Hansen, 1997)




$$\mathrm{d}\,X_t = \alpha[\beta - X_t]\,\mathrm{d}\,t + \sigma\,\mathrm{d}\,B_t$$

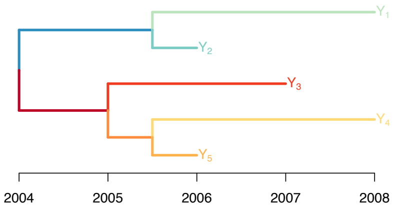
Deterministic part:

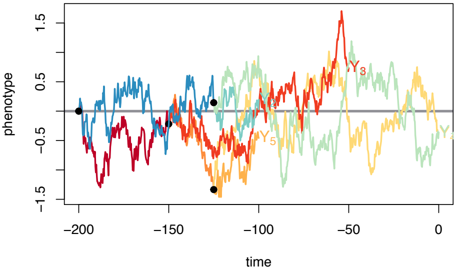
- $\beta$: primary optimum (mechanistically defined).
- $\ln(2)/\alpha$: phylogenetic half live.

Stochastic part:

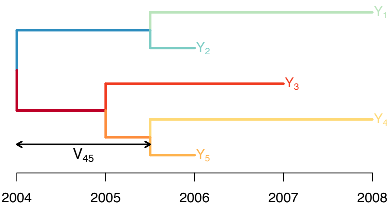
- $X_t$: trait value (actual optimum).
- $\sigma dB(t)$: Brownian fluctuations.

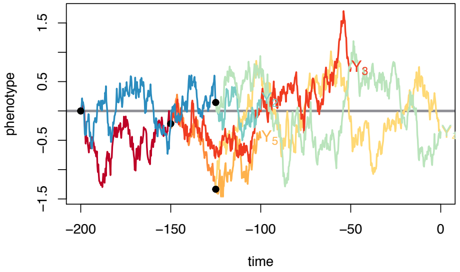
# OU on a Tree


Y1
Y2
Y3
Y4
Y5
2004
2005
2006
2007
2008


phenotype
1.5
0.5
−0.5
−1.5
Y3
Y2
Y5
−200
−150
−100
−50
0
time

# OU on a Tree




SDE: $\mathrm{d}\,X_t = \alpha[\beta - X_t]\,\mathrm{d}\,t + \sigma\,\mathrm{d}\,B_t$

Variance: $\mathbb{C}\mathrm{ov}\left[Y_4; Y_5\right] = \frac{\sigma^2}{2\alpha} e^{-\alpha(V_4+V_5)}(e^{2\alpha V_{45}} - 1)$

Expectation: $\mathbb{E}\left[Y_i\right] = \mu e^{-\alpha V_i} + \beta(1 - e^{-\alpha V_i})$

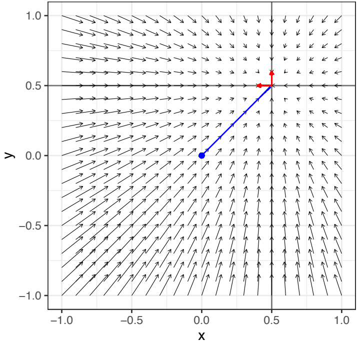
# Multivariate OU Modeling

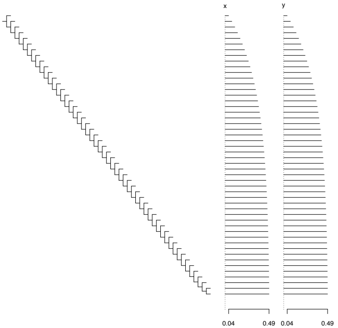
$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

# Multivariate OU Modeling

$$d\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,d\,t + \mathbf{\Sigma}\,d\,\mathbf{B}_t$$

Scalar: $\mathbf{A} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

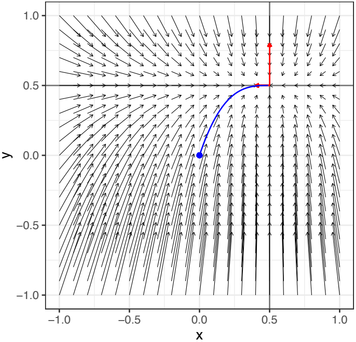


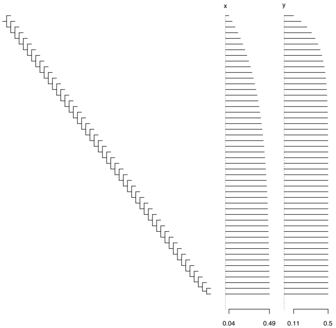



# Multivariate OU Modeling

$$d\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,d\,t + \mathbf{\Sigma}\,d\,\mathbf{B}_t$$

Diagonal: $\mathbf{A} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.3 \end{pmatrix} \quad \beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

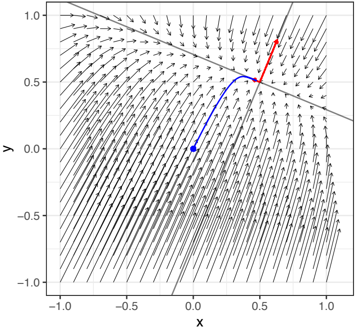


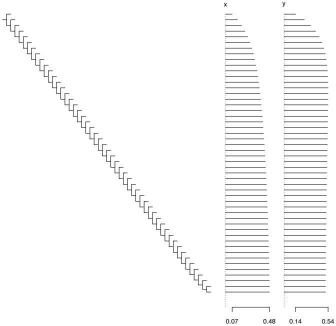



# Multivariate OU Modeling

$$d\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,d\,t + \mathbf{\Sigma}\,d\,\mathbf{B}_t$$

Symmetric: $\mathbf{A} = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

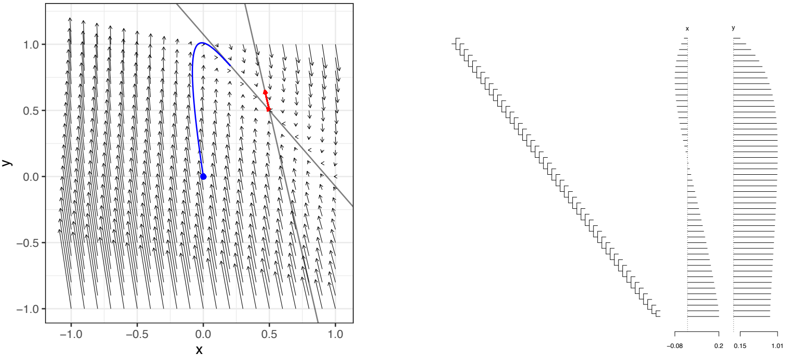


# Multivariate OU Modeling

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable in $\mathbb{R}$: $\mathbf{A} = \begin{pmatrix} -0.02 & -0.04 \\ 0.2 & 0.2 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$



# Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

# Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Expectation: $\mathbb{E}[\mathbf{Y}_i] = \boldsymbol{\mu} e^{-\mathbf{A}V_i} + \beta(1 - e^{-\mathbf{A}V_i})$

# Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Expectation: $\mathbb{E}\left[\mathbf{Y}_i\right] = \boldsymbol{\mu} e^{-\mathbf{A}V_i} + \beta(1 - e^{-\mathbf{A}V_i})$

Variance: $\mathbb{C}\mathrm{ov}\left[\mathbf{Y}_i; \mathbf{Y}_j\right] = \mathbf{P}\left[\mathbf{W}_{ij} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right]\mathbf{P}^T$

$$\mathbf{W}_{ij} = \left[\frac{1}{\lambda_q + \lambda_r} e^{-\lambda_q V_i} e^{-\lambda_r V_j}\left(e^{(\lambda_q+\lambda_r)V_{ij}} - 1\right)\right]_{1 \leq q, r \leq p}$$

# Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Expectation: $\mathbb{E}\left[\mathbf{Y}_i\right] = \boldsymbol{\mu} e^{-\mathbf{A}V_i} + \beta(1 - e^{-\mathbf{A}V_i})$

Variance: $\mathbb{C}\mathrm{ov}\left[\mathbf{Y}_i;\mathbf{Y}_j\right] = \mathbf{P}\left[\mathbf{W}_{ij} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right]\mathbf{P}^T$

$$\mathbf{W}_{ij} = \left[\frac{1}{\lambda_q + \lambda_r} e^{-\lambda_q V_i} e^{-\lambda_r V_j}\left(e^{(\lambda_q+\lambda_r)V_{ij}} - 1\right)\right]_{1 \leq q,r \leq p}$$

Distribution: Still Gaussian.

No nice Kronecker product.

# Bayesian Phylogenetics

Goal:

$$p\left(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi} \mid \mathbf{Y}, \mathbf{S}\right)$$

# Bayesian Phylogenetics

Goal:

$$p\left(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi} \mid \mathbf{Y}, \mathbf{S}\right) \propto p\left(\mathbf{Y}, \mathbf{S} \mid \boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}\right) p\left(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}\right)$$

# Bayesian Phylogenetics

Goal:

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

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

# Bayesian Phylogenetics

Goal:

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

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

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

# BM: Gibbs with Conjugate Priors

Goal: sample from the posterior $p(\mathbf{R}, \boldsymbol{\mu} | \mathbf{Y})$

# BM: Gibbs with Conjugate Priors

Goal: sample from the posterior $p(\mathbf{R}, \boldsymbol{\mu}|\mathbf{Y})$

Likelihood:

$$\mathbf{Y}|\mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

# BM: Gibbs with Conjugate Priors

Goal: sample from the posterior $p(\mathbf{R}, \boldsymbol{\mu} | \mathbf{Y})$

Likelihood:

$$\mathbf{Y} | \mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu} | \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1} \mathbf{R})$$

# BM: Gibbs with Conjugate Priors

Goal: sample from the posterior $p(\mathbf{R}, \boldsymbol{\mu} | \mathbf{Y})$

Likelihood:

$$\mathbf{Y} | \mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu} | \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1} \mathbf{R})$$

Gibbs:

+

$$\mathbf{R} | \mathbf{Y}, \boldsymbol{\mu} \sim \mathcal{IW}(\mathbf{R}_n, \nu_n) \quad \text{with} \quad \mathbf{R}_n = f(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{V})$$

# BM: Gibbs with Conjugate Priors

Goal: sample from the posterior $p(\mathbf{R}, \boldsymbol{\mu} | \mathbf{Y})$

Likelihood:

$$\mathbf{Y} | \mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu} | \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1} \mathbf{R})$$

Gibbs:

+

$$\mathbf{R} | \mathbf{Y}, \boldsymbol{\mu} \sim \mathcal{IW}(\mathbf{R}_n, \nu_n) \quad \text{with} \quad \mathbf{R}_n = f(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{V})$$

$\hookrightarrow$ Automatic sampling in the space of variance matrices.

# BM: Gibbs with Conjugate Priors

Goal: sample from the posterior $p(\mathbf{R}, \boldsymbol{\mu} | \mathbf{Y})$

Likelihood:

$$\mathbf{Y} | \mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu} | \mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1} \mathbf{R})$$

Gibbs:

+

$$\mathbf{R} | \mathbf{Y}, \boldsymbol{\mu} \sim \mathcal{IW}(\mathbf{R}_n, \nu_n) \quad \text{with} \quad \mathbf{R}_n = f(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{V})$$

$\hookrightarrow$ Automatic sampling in the space of variance matrices. :-)

# OU: No Gibbs

Likelihood:

$$\mathbf{Y} | \mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

# OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

# OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

Gibbs: Not possible.

# OU: No Gibbs

Likelihood:

$$\mathbf{Y} | \mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

Gibbs: Not possible.

Metropolis - Hasting:

Need to sample in constrained spaces ($\mathbf{A}$, $\mathbf{R}$).

# OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

Gibbs: Not possible.

Metropolis - Hasting:

Need to sample in constrained spaces (**A**, **R**).

$\hookrightarrow$ Find a transformation $f : \mathcal{C}_q \to \mathbb{R}^q$

# OU: MH in constrained space

Goal: sample from the posterior $p(\mathbf{R}, \mathbf{A}, \boldsymbol{\mu} | \mathbf{Y})$

# OU: MH in constrained space

Goal: sample from the posterior $p(\mathbf{R}, \mathbf{A}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{R} = \boldsymbol{\sigma}\mathbf{C}\boldsymbol{\sigma}$: Correlation / Variance decomposition
- $\mathbf{C}$: LKJ transformation

# OU: MH in constrained space

Goal: sample from the posterior $p(\mathbf{R}, \mathbf{A}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{R} = \boldsymbol{\sigma} \mathbf{C} \boldsymbol{\sigma}$: Correlation / Variance decomposition +
- $\mathbf{C}$: LKJ transformation
- $\mathbf{A} = \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}^{-1}$ positive, real eigenvalues +
- $\mathbf{P}$: Unitary eigen-vectors
- $\boldsymbol{\Lambda}$, $\boldsymbol{\sigma}$: Positive (ordered) vectors

# OU: MH in constrained space

Goal: sample from the posterior $p(\mathbf{R}, \mathbf{A}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{R} = \boldsymbol{\sigma}\mathbf{C}\boldsymbol{\sigma}$: Correlation / Variance decomposition
- $\mathbf{C}$: LKJ transformation
- $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ positive, real eigenvalues
- $\mathbf{P}$: Unitary eigen-vectors
- $\boldsymbol{\Lambda}$, $\sigma$: Positive (ordered) vectors

Priors:

- $\mathbf{C}$: LKJ distribution
- $\mathbf{P}$: Spherical beta
- $\boldsymbol{\Lambda}$, $\sigma$: half-normal

# OU: MH in constrained space

Goal: sample from the posterior $p(\mathbf{R}, \mathbf{A}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{R} = \boldsymbol{\sigma}\mathbf{C}\boldsymbol{\sigma}$: Correlation / Variance decomposition +
- $\mathbf{C}$: LKJ transformation
- $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ positive, real eigenvalues +
- $\mathbf{P}$: Unitary eigen-vectors
- $\boldsymbol{\Lambda}$, $\boldsymbol{\sigma}$: Positive (ordered) vectors

Priors:

- $\mathbf{C}$: LKJ distribution +
- $\mathbf{P}$: Spherical beta
- $\boldsymbol{\Lambda}$, $\boldsymbol{\sigma}$: half-normal

$\hookrightarrow$ We have a running random walk MCMC.

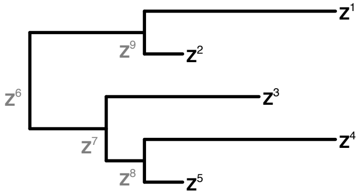
**Question:** Can we use the tree ?

# General Model

BM, OU: Instance of a general Gaussian propagation model.

# General Model

BM, OU: Instance of a general Gaussian propagation model.



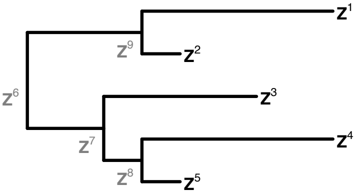

$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

# General Model

BM, OU: Instance of a general Gaussian propagation model.




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

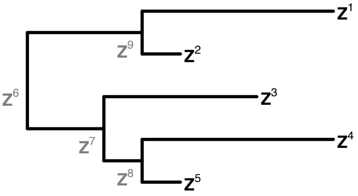
$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\boldsymbol{\beta}_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

# General Model

BM, OU: Instance of a general Gaussian propagation model.




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \,\middle|\, \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

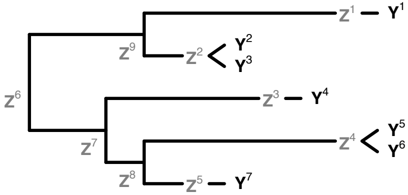
BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\boldsymbol{\beta}_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

Drift, shifts, Integrated OU...

# General Model

BM, OU: Instance of a general Gaussian propagation model.




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \quad \text{obs}$$

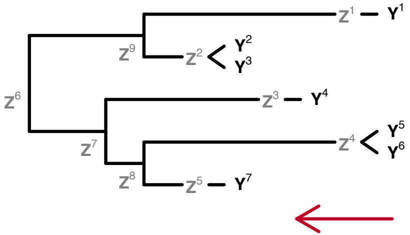
BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\beta_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

Drift, shifts, Integrated OU...

Error Model, "Heritability".

# Efficient Computations: Likelihood

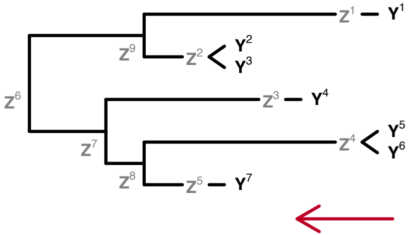



$$\begin{aligned} \mathbf{Z}^r &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) && \text{root} \\ \mathbf{Z}^j \mid \mathbf{Z}^{\text{pa}(j)} &\sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) && \text{nodes} \\ \mathbf{Y}^i \mid \mathbf{Z}^{\text{pa}(i)} &\sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, \mathbf{S}_i\right) && \text{obs} \end{aligned}$$

Likelihood: $\log p(\mathbf{Y})$ in one post-order traversal.

$\hookrightarrow O(N)$.

# Efficient Computations: Likelihood




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

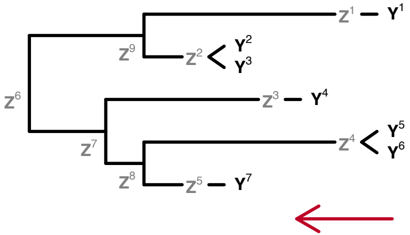
$$\mathbf{Y}^i \mid \mathbf{Z}^{\text{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, \mathbf{S}_i\right) \quad \text{obs}$$

Likelihood: $\log p(\mathbf{Y})$ in one post-order traversal.

↪ $O(N)$.

↪ "Pruning", "Gaussian elimination", "Phylogenetic Kalman filter", ...

# Efficient Computations: Likelihood




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \,\Big|\, \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

$$\mathbf{Y}^i \,\Big|\, \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \quad \text{obs}$$

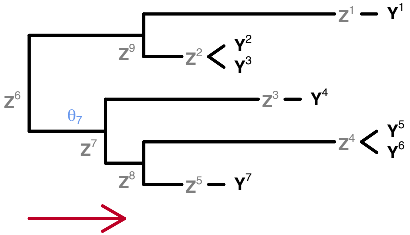
Likelihood: $\log p(\mathbf{Y})$ in one post-order traversal.

$\hookrightarrow$ $O(N)$.

$\hookrightarrow$ "Pruning", "Gaussian elimination", "Phylogenetic Kalman filter", ...

Difficulty: Numerical robustness.

# Efficient Computations: Gradient

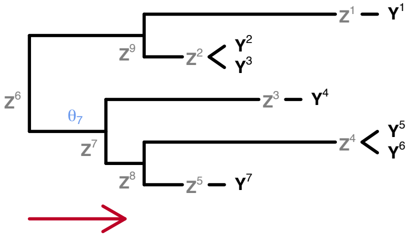



$$\begin{aligned} \mathbf{Z}^r &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) && \text{root} \\ \mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} &\sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) && \text{nodes} \\ \mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} &\sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) && \text{obs} \end{aligned}$$

Branch-specific Gradient: in one pre-order traversal.

$$\begin{aligned} \nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Y} \mid \boldsymbol{\theta}_k\right)\right] &= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Z}^k, \mathbf{Y} \mid \boldsymbol{\theta}_k\right)\right] \mid \mathbf{Y}\right] \\ &= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Z}^k \mid \mathbf{Y}_{\lceil k \rceil}, \boldsymbol{\theta}_k\right)\right] \mid \mathbf{Y}\right]. \end{aligned}$$

# Efficient Computations: Gradient



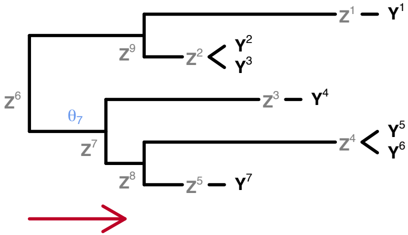

$$\begin{aligned}
\mathbf{Z}^r &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) && \text{root} \\
\mathbf{Z}^j \mid \mathbf{Z}^{\text{pa}(j)} &\sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) && \text{nodes} \\
\mathbf{Y}^i \mid \mathbf{Z}^{\text{pa}(i)} &\sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, \mathbf{S}_i\right) && \text{obs}
\end{aligned}$$

Branch-specific Gradient: in one pre-order traversal.

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}_k} [\log p(\mathbf{Y} \mid \boldsymbol{\theta}_k)] &= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Z}^k, \mathbf{Y} \mid \boldsymbol{\theta}_k\right)\right] \mid \mathbf{Y}\right] \\
&= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Z}^k \mid \mathbf{Y}_{\lceil k \rceil}, \boldsymbol{\theta}_k\right)\right] \mid \mathbf{Y}\right].
\end{aligned}$$

Chain rule: Gradient w.r.t. any parameter in $O(N)$.

# Efficient Computations: Gradient




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \quad \text{obs}$$

Branch-specific Gradient: in one pre-order traversal.

$$\begin{aligned}\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Y} \mid \boldsymbol{\theta}_k\right)\right] &= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Z}^k, \mathbf{Y} \mid \boldsymbol{\theta}_k\right)\right] \mid \mathbf{Y}\right] \\ &= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}_k}\left[\log p\left(\mathbf{Z}^k \mid \mathbf{Y}_{\lceil k \rceil}, \boldsymbol{\theta}_k\right)\right] \mid \mathbf{Y}\right].\end{aligned}$$

Chain rule: Gradient w.r.t. any parameter in $O(N)$. $\rightarrow$ HMC

# Implementation

(Suchard et al., 2018)

BEAST
Bayesian Evolutionary Analysis Sampling Trees

- MCMC for tree estimation
- Comprehensive set of tools:
  - Factor model
  - Marginal Likelihood
  - ...
- Developed in `Java` since 2002.
- This is `BEAST` 1.10.

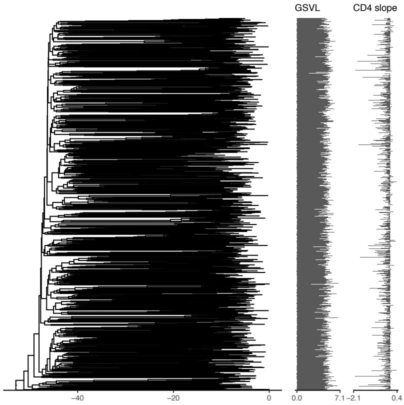
What's new:

- Flexible sampling of variance
- Efficient HMC

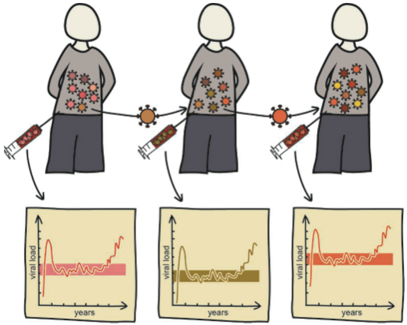
Limitations:

- Diagonal OU

# HIV Virulence Heritability (Alizon et al., 2010; Blanquart et al., 2017)



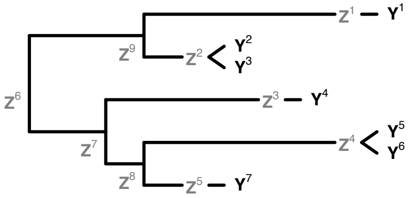

CD4: T cells decline rate
GSVL: Set point viral load




Fraser et al. (2014)

Questions: Is virulence "heritable"? Which model of trait evolution?

# Phylogenetic Heritability




$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \quad \text{obs}$$

$$\mathbf{V}(\mathbf{Z}^{\mathrm{tip}}) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}}[\mathbf{Z}^j], \qquad \mathbf{V}(\mathbf{Y}) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}}[\mathbf{Y}^i],$$

$$H_{kl} = \frac{V_{kl}(\mathbf{Z}^{\mathrm{tip}})}{\sqrt{V_{kk}(\mathbf{Y}) V_{ll}(\mathbf{Y})}}.$$

# Models

(Alizon et al., 2010)

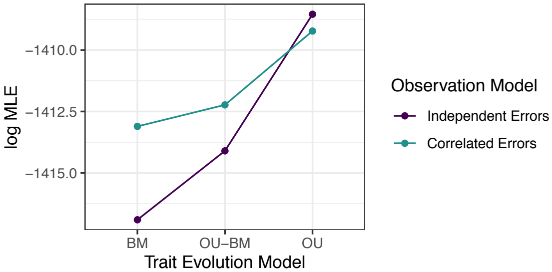
We use three different models:

- **BM** no selection on the traits.
- **OU-BM** selection on VL, not on CD4.
- **OU** selection on both traits.

With independent or correlated errors.

- Each model is fitted using a HMC
- Estimated Marginal likelihood is used to compare them.

# Results



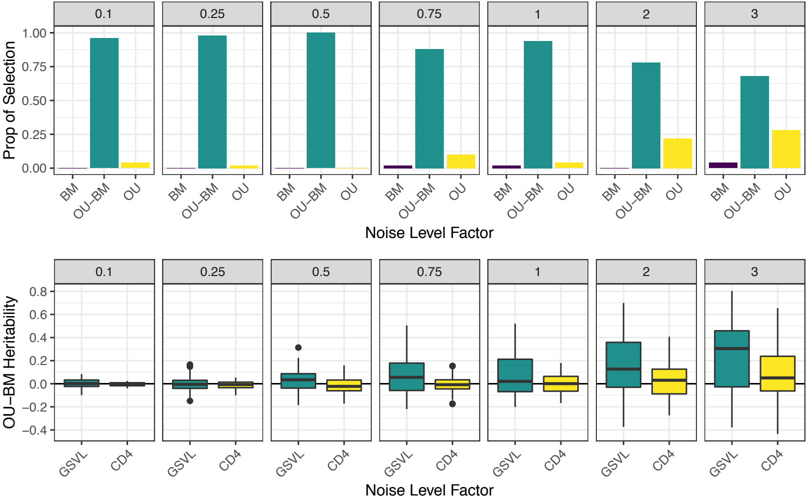

Heritability (using OU-BM):

VL $h^2 = 0.35 \quad [0.22, 0.51]$ (95% CI)

CD4 $h^2 = 0.21 \quad [0.12, 0.32]$ (95% CI)

Consistent with previous estimates.

# OU-BM Simulations



# Conclusion and Perspectives

A general framework for trait evolution with dated tips.

Main Features:

- Flexible models and implementation
- Efficient algorithms
- Applicable to virology, ecology, paleontology.

weasels

Preprint: `arxiv:2003.10336` (in print, AoAS)

# Conclusion and Perspectives

A general framework for trait evolution with dated tips.

Main Features:

- Flexible models and implementation
- Efficient algorithms
- Applicable to virology, ecology, paleontology. weasels

Perspectives:

- Identifiability full OU ?
- Shifting parameters
- Other questions: geographical spread, ...

Preprint: `arxiv:2003.10336` (in print, AoAS)

# Bibliography


Alizon, von Wyl, Stadler, et al. 2010. *PLoS Pathogens*. 6.

Aristide, Bastide, dos Reis, et al. 2018. *Evolution*. 72:2697–2711.

Betancourt. 2017. *arXiv e-print*. p. 170102434.

Blanquart, Wymant, Cornelissen, et al. 2017. *PLoS Biology*. 15:1–26.

Dellicour, Rose, Faria, et al. 2017. *Molecular Biology and Evolution*. 34:2563–2571.

Fraser, Lythgoe, Leventhal, et al. 2014. *Science*. 343:1243727–1243727.

Hansen. 1997. *Evolution*. 51:1341.

Lemey, Rambaut, Bedford, et al. 2014. *PLoS Pathogens*. 10:e1003932.

Lewandowski, Kurowicka, Joe. 2009. *Journal of Multivariate Analysis*. 100:1989–2001.

Schnitzler, Theis, Polly, et al. 2017. *Evolutionary Ecology Research*. 18:187–200.

Suchard, Lemey, Baele, et al. 2018. *Virus Evolution*. 4:1–5.

Vrancken, Lemey, Rambaut, et al. 2015. *Methods in Ecology and Evolution*. 6:67–82.

# Thank you for listening

IMAG

Rega Instituut

Reagan Medical Center

# Appendices

# MCMC

Goal: Sample from $p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$

# MCMC

Goal: Sample from $p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$

Metropolis - Hasting: Iterate:

- Draw $\boldsymbol{\theta}^*$ in $q\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\theta}^t\right)$.
- Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$ with probability:

$$r_t = \min\left\{1, \frac{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^*\right)}{p\left(\mathbf{Y} \,\middle|\, \boldsymbol{\theta}^t\right)} \frac{p\left(\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^t\right)} \frac{q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})}\right\}.$$

# MCMC

Goal: Sample from $p\left(\boldsymbol{\theta} \mid \mathbf{Y}\right) \propto p\left(\mathbf{Y} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$

Metropolis - Hasting: Iterate:

- Draw $\boldsymbol{\theta}^*$ in $q\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\theta}^t\right)$.
- Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$ with probability:

$$r_t = \min\left\{1, \frac{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^*\right)}{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^t\right)} \frac{p\left(\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^t\right)} \frac{q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})}\right\}.$$

Gibbs:

- Split $\boldsymbol{\theta} = (\boldsymbol{\theta}_{[1]}, \dots, \boldsymbol{\theta}_{[K]})$.
- Draw $\boldsymbol{\theta}^*$ in $p(\boldsymbol{\theta}_{[k]}|\boldsymbol{\theta}^{(t)}_{[-k]}, \mathbf{Y})$ so that $r_t = 1$.

back

# MH in constrained space

Transformation:

$$f : \begin{cases} \mathcal{C}_q \to \mathbb{R}^q \\ \boldsymbol{\theta} \mapsto \boldsymbol{\nu} = f(\boldsymbol{\theta}) \end{cases}$$

# MH in constrained space

Transformation:

$$f : \begin{cases} \mathcal{C}_q \to \mathbb{R}^q \\ \boldsymbol{\theta} \mapsto \boldsymbol{\nu} = f(\boldsymbol{\theta}) \end{cases}$$

Distribution: For a distribution $\pi$:

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \pi_{\boldsymbol{\nu}}(f(\boldsymbol{\theta})) \times |J_f(\boldsymbol{\theta})|$$

# MH in constrained space

Transformation:

$$f : \begin{cases} \mathcal{C}_q \to \mathbb{R}^q \\ \boldsymbol{\theta} \mapsto \boldsymbol{\nu} = f(\boldsymbol{\theta}) \end{cases}$$

Distribution: For a distribution $\pi$:

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \pi_{\boldsymbol{\nu}}(f(\boldsymbol{\theta})) \times |J_f(\boldsymbol{\theta})|$$

Metropolis - Hasting: Iterate:

- Draw $\boldsymbol{\nu}^*$ in $q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^t)$.
- Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^* = f^{-1}(\boldsymbol{\nu}^*)$ with probability

$$r_t = \min\left\{1, \frac{p(\mathbf{Y} \mid \boldsymbol{\theta}^*)}{p(\mathbf{Y} \mid \boldsymbol{\theta}^t)} \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^t)} \frac{q(\boldsymbol{\nu}^{(t)}|\boldsymbol{\nu}^*)}{q(\boldsymbol{\nu}^*|\boldsymbol{\nu}^{(t)})} \frac{|J_f(\boldsymbol{\theta}^{(t)})|}{|J_f(\boldsymbol{\theta}^*)|}\right\}.$$

back

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

Decomposition: Use correlation matrix **C**

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

Decomposition: Use correlation matrix **C**

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

Decomposition: Use correlation matrix **C**

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

**C**: Correlation matrix :-(

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

Decomposition: Use correlation matrix **C**

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

**C**: Correlation matrix :-(

LKJ: Transformation on the space of correlation matrices.

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

Decomposition: Use correlation matrix **C**

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

**C**: Correlation matrix :-(

LKJ: Transformation on the space of correlation matrices.
- ↪ Use "vine" theory.
- ↪ Easier and more efficient: Cholesky representation.

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

$$\mathbf{C} = \mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}^T \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}$$

With:

- $W_{1k}^2 + \cdots + W_{kk}^2 = 1$ (correlation)
- $W_{kk} > 0$ (identifiability)

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

$$\mathbf{C} = \mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}^T \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}$$

With:

- $W_{1k}^2 + \cdots + W_{kk}^2 = 1$ (correlation)
- $W_{kk} > 0$ (identifiability)

$\hookrightarrow$ Each column $k$ is in the half euclidian unit sphere $\mathcal{S}_k^{\mathrm{h}}(\mathbb{R})$.

# Variance: LKJ transformation (Lewandowski, Kurowicka, and Joe, 2009)

$$\mathbf{C} = \mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}^T \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}$$

With:

- $W_{1k}^2 + \cdots + W_{kk}^2 = 1$ (correlation)
- $W_{kk} > 0$ (identifiability)

$\hookrightarrow$ Each column $k$ is in the half euclidian unit sphere $\mathcal{S}_k^{\mathrm{h}}(\mathbb{R})$.

$\hookrightarrow$ :-)

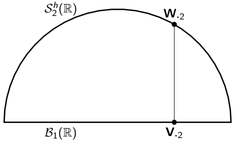
+

back

# Sampling in the Sphere

→ Sphere $\mathcal{S}_k^{\mathrm{h}}(\mathbb{R})$ to euclidean ball $\mathcal{B}_{k-1}(\mathbb{R})$:

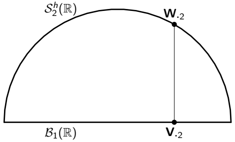
$$\mathbf{F} : \begin{cases} \mathcal{B}_{k-1}(\mathbb{R}) & \to \mathcal{S}_k^{\mathrm{h}}(\mathbb{R}) \\ \mathbf{V}_{\cdot k} & \mapsto \mathbf{W}_{\cdot k} = \left(\mathbf{V}_{\cdot k},\ \sqrt{1 - \left\|\mathbf{V}_{\cdot k}\right\|^2}\right) \end{cases}$$



# Sampling in the Sphere

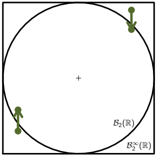
→ Sphere $\mathcal{S}_k^h(\mathbb{R})$ to euclidean ball $\mathcal{B}_{k-1}(\mathbb{R})$:

$$\mathbf{F} : \begin{cases} \mathcal{B}_{k-1}(\mathbb{R}) & \to \mathcal{S}_k^h(\mathbb{R}) \\ \mathbf{V}_{\cdot k} & \mapsto \mathbf{W}_{\cdot k} = \left(\mathbf{V}_{\cdot k}, \ \sqrt{1 - \|\mathbf{V}_{\cdot k}\|^2}\right) \end{cases}$$




→ $\mathcal{B}_{k-1}(\mathbb{R})$ to infinite-norm ball $\mathcal{B}_{k-1}^{\infty}(\mathbb{R})$ : "LKJ"

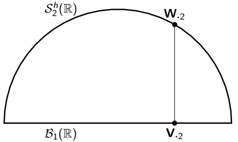
$$\mathbf{LKJ}_i(\mathbf{z}) = \begin{cases} z_i & \text{if } i = 1 \\ z_i \prod_{k=1}^{i-1}(1 - z_k^2)^{1/2} & \text{if } 1 < i \leq k-1. \end{cases}$$



# Sampling in the Sphere

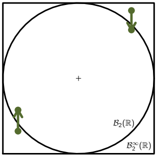
→ Sphere $\mathcal{S}_k^h(\mathbb{R})$ to euclidean ball $\mathcal{B}_{k-1}(\mathbb{R})$:

$$\mathbf{F} : \begin{cases} \mathcal{B}_{k-1}(\mathbb{R}) & \to \mathcal{S}_k^h(\mathbb{R}) \\ \mathbf{V}_{\cdot k} & \mapsto \mathbf{W}_{\cdot k} = \left(\mathbf{V}_{\cdot k}, \sqrt{1 - \|\mathbf{V}_{\cdot k}\|^2}\right) \end{cases}$$




→ $\mathcal{B}_{k-1}(\mathbb{R})$ to infinite-norm ball $\mathcal{B}_{k-1}^{\infty}(\mathbb{R})$ : "LKJ"

$$\mathbf{LKJ}_i(\mathbf{z}) = \begin{cases} z_i & \text{if } i = 1 \\ z_i \prod_{k=1}^{i-1}(1 - z_k^2)^{1/2} & \text{if } 1 < i \leq k-1. \end{cases}$$




→ $\mathcal{B}_{k-1}^{\infty}(\mathbb{R})$ to $\mathbb{R}^{k-1}$: "Fisher Z" transform

$$\tanh^{-1} : \begin{cases} ]-1, 1[ \to \mathbb{R} \\ x \mapsto \frac{1}{2} \ln\left(\frac{1+x}{1-x}\right) \end{cases}$$

back

# Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0$$

# Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathsf{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

# Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P\Lambda P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathrm{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

$\hookrightarrow$ Decomposition is unique.

# Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathrm{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

$\hookrightarrow$ Decomposition is unique.

Sampling:

$\mathbf{P}_{\cdot k}$: Same as $\mathbf{W}_{\cdot k}$

# Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathrm{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

$\hookrightarrow$ Decomposition is unique.

Sampling:

$\mathbf{P}_{\cdot k}$: Same as $\mathbf{W}_{\cdot k}$
$\mathbf{\Lambda}$: Use $\log(\lambda_i) - \log(\lambda_{i-1})$

back

# LKJ and Spherical Beta Distributions

LKJ distribution:

$$\mathrm{LKJ}(\mathbf{C} \mid \eta) = c_p(\eta)\, |\mathbf{C}|^{\eta-1}$$

| | |
|---|---|
| $\eta = 1$: | Uniform. |
| $\eta > 1$: | Peak around identity matrix. |
| $0 < \eta < 1$: | Trough around identity matrix. |

Spherical Beta distribution:

$$\mathrm{SBeta}(\mathbf{V}|\beta) \propto (1 - \|\mathbf{V}\|^2)^{\beta}$$

$\hookrightarrow$ LKJ $\iff$ spherical beta on each of the $\mathbf{V}_{\cdot k} = \mathbf{W}_{\cdot k}^{-k}$ (with $\beta_k = \eta + (p-k)/2$).

back

# Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

# Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) = \text{Potential energy} + \text{Kinetic energy}$$

# Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\begin{aligned}\text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) &= \text{Potential energy} + \text{Kinetic energy} \\ &= U(\mathbf{q}) + K(\mathbf{p}) \\ &= -\log \underbrace{p(\mathbf{Y}|\mathbf{q})p(\mathbf{q})}_{\text{posterior}} - \log \underbrace{\phi(\mathbf{p})}_{\text{Gaussian}}\end{aligned}$$

# Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\begin{aligned}\text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) &= \text{Potential energy} + \text{Kinetic energy} \\ &= U(\mathbf{q}) + K(\mathbf{p}) \\ &= -\log \underbrace{p(\mathbf{Y}|\mathbf{q})p(\mathbf{q})}_{\text{posterior}} - \log \underbrace{\phi(\mathbf{p})}_{\text{Gaussian}}\end{aligned}$$

$H$ **total energy** invariant by Hamiltonian dynamic: $\begin{cases} \frac{d\mathbf{p}}{dt} = \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \frac{d\mathbf{q}}{dt} = -\nabla_{\mathbf{p}} K(\mathbf{p}) \end{cases}$

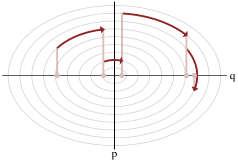
# Hamiltonian Monte Carlo

(Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\begin{aligned} \text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) &= \text{Potential energy} + \text{Kinetic energy} \\ &= U(\mathbf{q}) + K(\mathbf{p}) \\ &= -\log \underbrace{p(\mathbf{Y}|\mathbf{q})p(\mathbf{q})}_{\text{posterior}} - \log \underbrace{\phi(\mathbf{p})}_{\text{Gaussian}} \end{aligned}$$

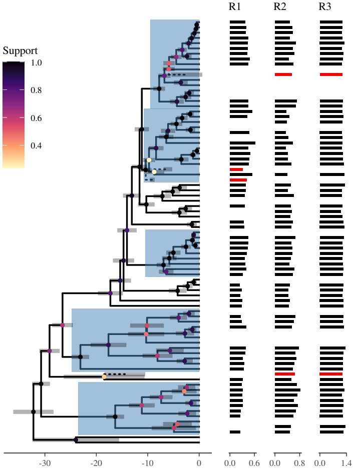
$H$ **total energy** invariant by Hamiltonian dynamic: $\begin{cases} \dfrac{d\mathbf{p}}{dt} = \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \dfrac{d\mathbf{q}}{dt} = -\nabla_{\mathbf{p}} K(\mathbf{p}) \end{cases}$

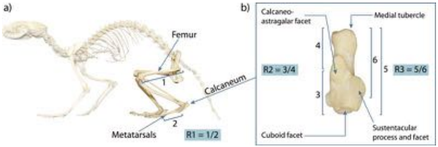



1. Draw random moments $\mathbf{p}$.
2. Propose a new $\mathbf{q}$ from the dynamic.

back

# Musteloidea Superfamily




Schnitzler et al. (2017)

- Weasels and allies
- Total evidence approach
- Fossils provide a temporal signal

Raccoon and skunk in a Hollywood back yard feasting on cat food.

# Total evidence approach

Question: Does the continuous trait helps placing the fossils ?

Measure: Entropy of the branch attachment frequency vector.

*The smaller the better*

Results:

| | *Pannonictis* | *Sivaonyx beyi* | *Teruelictis riparius* | *Trocharion albanense* |
|---|---|---|---|---|
| No Traits | 2.84 | 1.80 | 1.24 | 2.28 |
| BM | **2.81** | 1.22 | 1.22 | 2.01 |
| OU | 2.92 | **1.20** | **1.10** | **1.86** |

The signal is not strong.

# Model Selection (fixed tree)

Question: Does including fossils change the selected model ?

Schnitzler et al. (2017):

- Univariate analyses.
- OU or trend favored *when fossils are included*.

Results: The signal is not strong.

back