

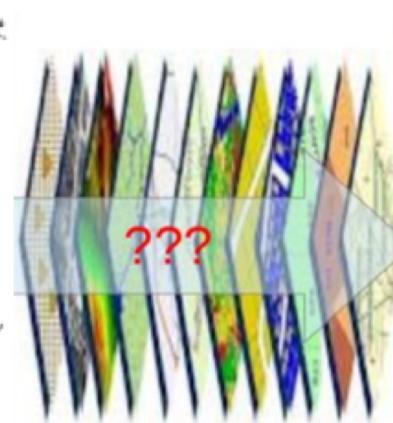
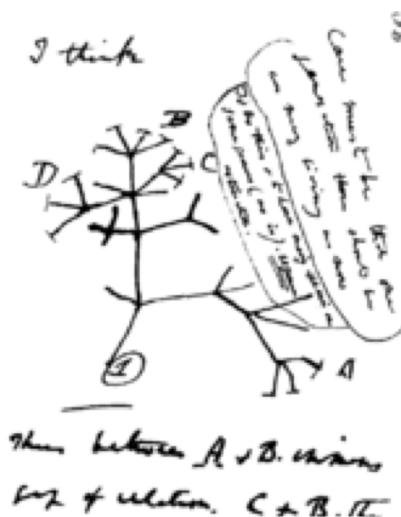
Phylogenetic Comparative Methods in the Era of Big Data

PhD Dissertation Defence
Venelin Mitov
Computational Evolution Group

Committee:
Tanja Stadler (advisor)
Samuel Alizon (University of Montpellier)
Roland Regös (ETH Zurich)
Niko Beerenwinkel (ETH Zurich, chair)

Genotype → Phenotype

Ch. Darwin (1837)



*If we are to relate genotype to phenotype,
we need such representations of the objects
and interactions that allow finding the mappings
between levels.*

Michael Savageau

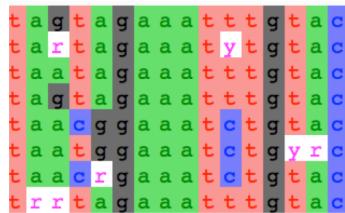
Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

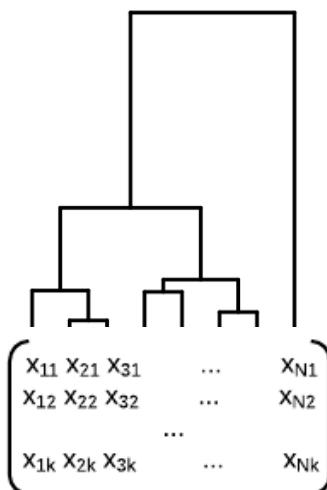
Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Phylogenetic comparative methods

A grid of DNA sequence data for multiple taxa. The columns represent positions in the sequence, and the rows represent different taxa. Colored boxes highlight specific mutations or changes across the taxa.

1.
Phylogenetic
inference



2. Trait-
evolution
inference

Model of
trait evolution



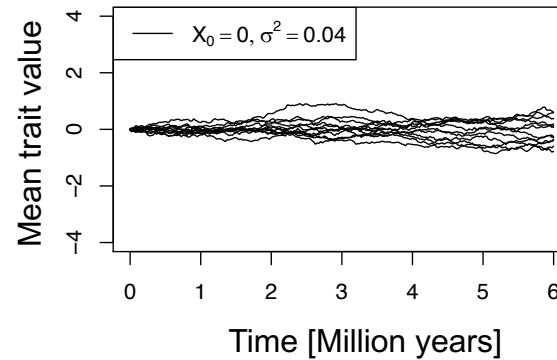
Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

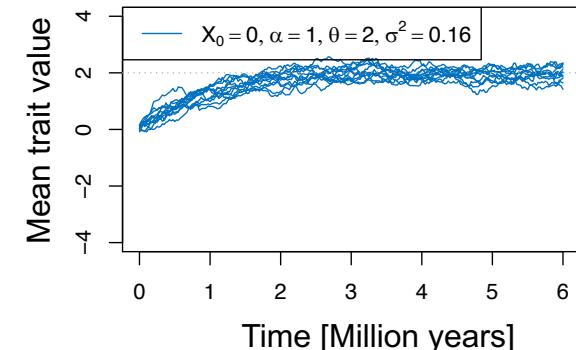
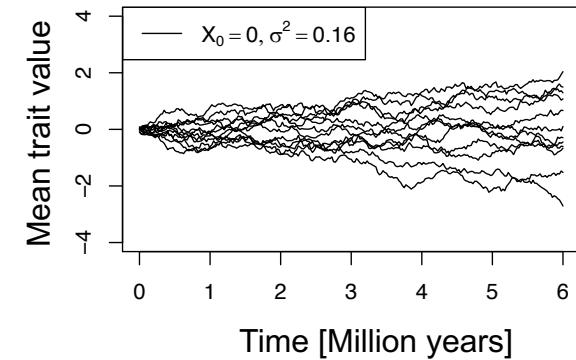
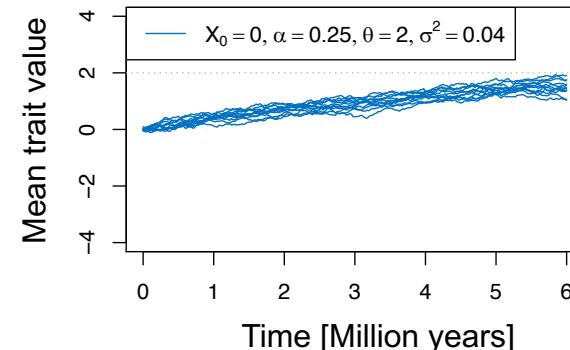
Microevolutionary forces → Continuous-time random processes

Lande, R. (1976). Natural-Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution*, 30(2), 314–334.

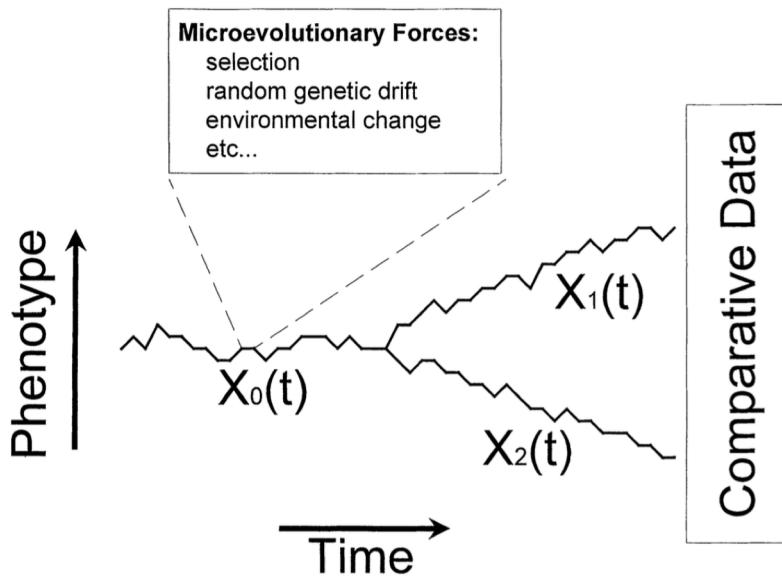
Random drift →
Brownian motion (BM)
parameters: X_0 , σ^2



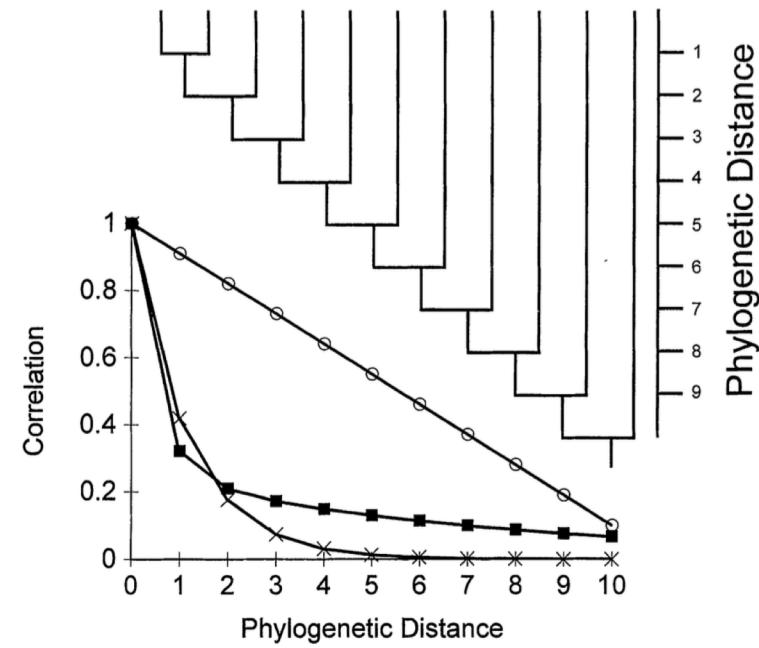
Random drift +
stabilizing selection →
Ornstein-Uhlenbeck
(OU)
parameters: X_0 , α , θ , σ^2



Branching BM and OU processes



Comparative Data



Key idea: It is possible to infer the parameters of the branching process based on the patterns of correlation in the data forming at the tips.

Branching BM and OU processes → Multivariate Gaussian distributions

- Denote the tree by \mathcal{T} and the model parameters by Θ . Under the model, the data is a sampled point from a N-dimensional Gaussian distribution:
 - mean-vector: $\vec{\mu}(\Theta, \mathcal{T})$
 - variance-covariance matrix: $\Sigma(\Theta, \mathcal{T})$
 - Model likelihood:

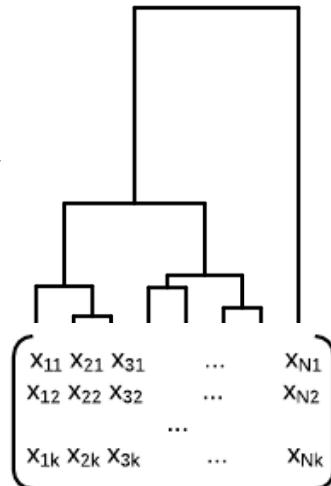
$$pdf(\vec{x}|\Theta, \mathcal{T}) = \frac{1}{\sqrt{\det(2\pi\Sigma(\Theta, \mathcal{T}))}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}(\Theta, \mathcal{T}))' \Sigma(\Theta, \mathcal{T})^{-1} (\vec{x} - \vec{\mu}(\Theta, \mathcal{T})) \right]$$

- Use standard statistical model inference (e.g. likelihood maximization or MCMC sampling) to infer Θ .

Summary: Phylogenetic comparative methods

A grid of DNA sequence data showing multiple taxa across several sites. The grid consists of 10 rows (taxa) and 15 columns (sites). The sequence characters are color-coded: A (green), T (red), C (blue), and G (purple).

1.
Phylogenetic
inference



2. Trait-
evolution
inference

Model of
trait evolution



Phylogenetic comparative methods (PCMs) combine **two types of input**:

- A phylogenetic tree;
- A set of measurements for a number of traits observed at each tip in the tree.

in order to infer a **branching (Gaussian) process** that models:

- the association between the traits;
- the association between the traits and the phylogenetic tree;
- the biological processes governing the evolution of the traits.

Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Challenges in the application of PCMs to big phylogenetic trees

Technical challenges

The existing tools for phylogenetic model inference are limited to small ultrametric trees – the inference on big/non-ultrametric trees is limited to BM models only
→ fast algorithms for model inference are needed.

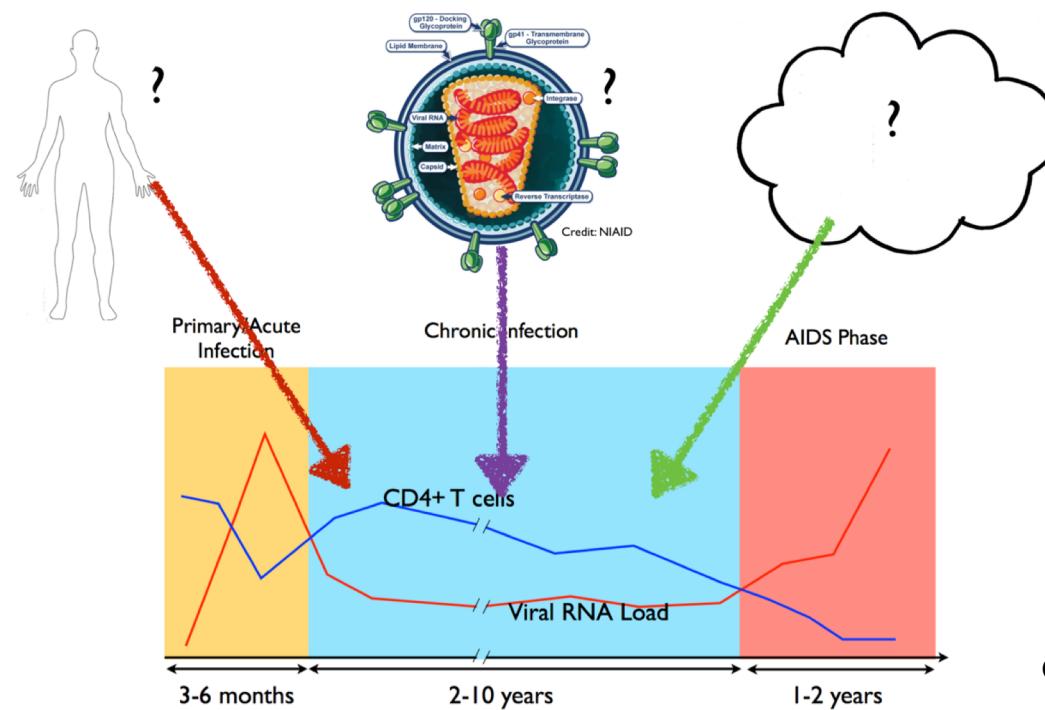
Conceptual challenges

Big phylogenetic trees imply heterogeneous evolutionary processes. Such a process cannot be described by a single phylogenetic model of evolution.
→ new phylogenetic models that incorporate heterogeneous evolutionary processes along the tree are needed.

Agenda

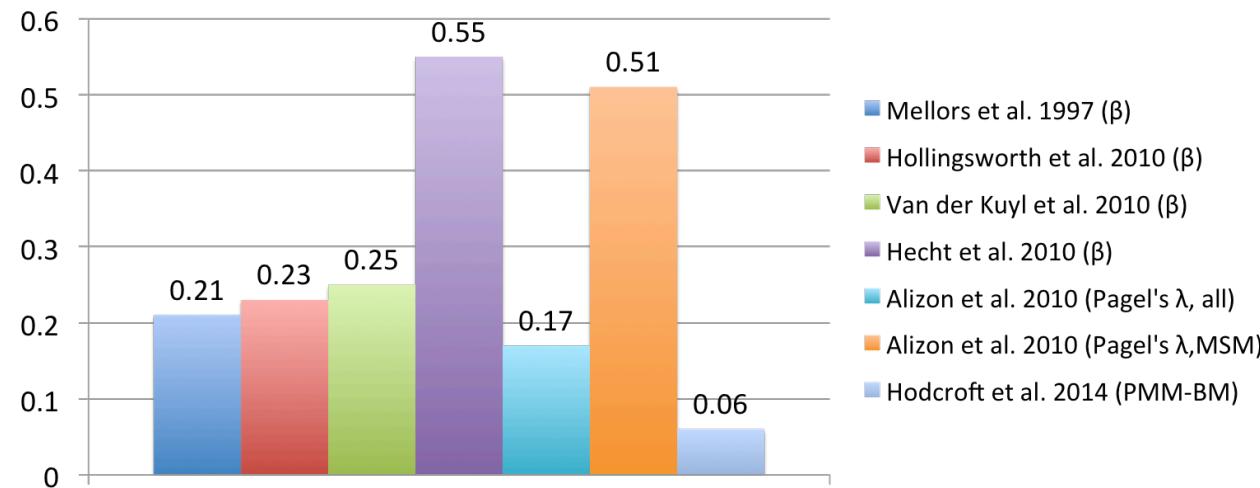
- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Virulence of an HIV-1 Infection



- “Virulence” is the inverse time to AIDS of a patient.
- The viral load during the chronic phase, known as set-point viral load (spVL) provides a proxy measure for virulence.

Virus control of HIV virulence is discussed controversially



We show that the differences in the estimates result from method artifacts, not differences in the data.

Approaches to determine the importance of the virus

Quantification measure

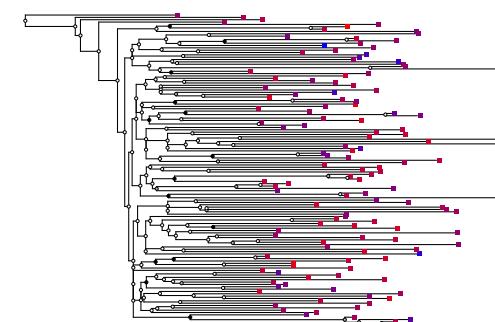
Broad-sense heritability H^2 : Amount of variation in a trait explained by the virus genotype

Estimator

Resemblance-based estimators: measuring the relative trait-similarity within groups of transmission-related patients
Tool: Donor-recipient regression (DR).



Phylogenetic comparative methods: measuring the association between observed trait values from patients and their (approximate) transmission tree
Tool: Phylogenetic mixed model (PMM).



H^2 comes from quantitative genetics for sexual reproducing populations

Transfer to
pathogens

Within-host evolution
Partial quasi-species transmission

DR

Within-host evolution is ignored if trait is measured late in infection:
→ **negative bias** (as difference due to evolution is treated as noise)

PMM

Within-host evolution is accounted for but the assumed Brownian motion model fits poorly to the data:
→ **negative bias caused by inflated parameter σ^2**

H^2 estimators overcoming the biases

Resemblance
-based

Phylogenetic-
based

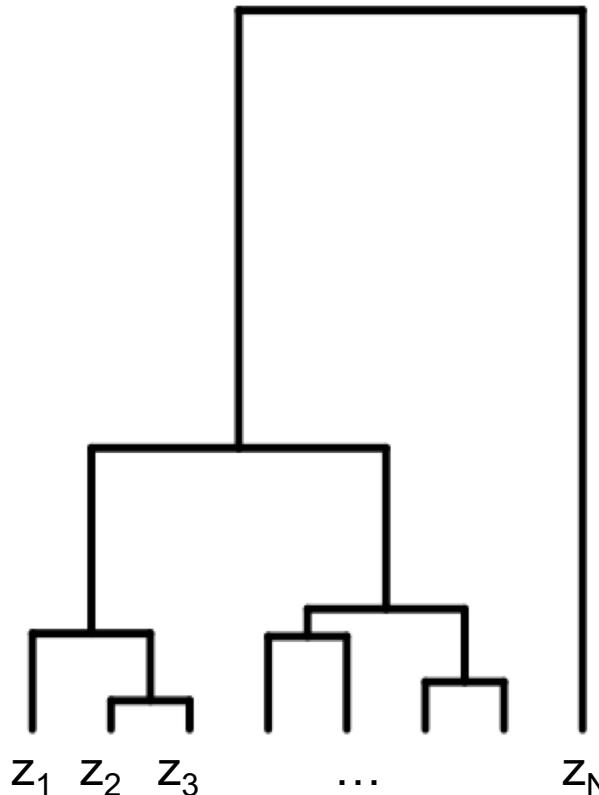
Anova-CPP (*extending PP by Shirreff et al., 2012*)

- Input is a phylogenetic tree; closest phylogenetic pairs (CPP) are determined;
- Anova on CPPs to determine how much more similar they are to each other than across pairs

POUMM (*generalizing PMM to selection*)

- Input is a phylogenetic tree
- Assumptions:
 - Ornstein-Uhlenbeck process for genotypic trait evolution
 - Contribution from host is drawn from a normal distribution
- Maximum-likelihood estimation of relative contribution of genotype

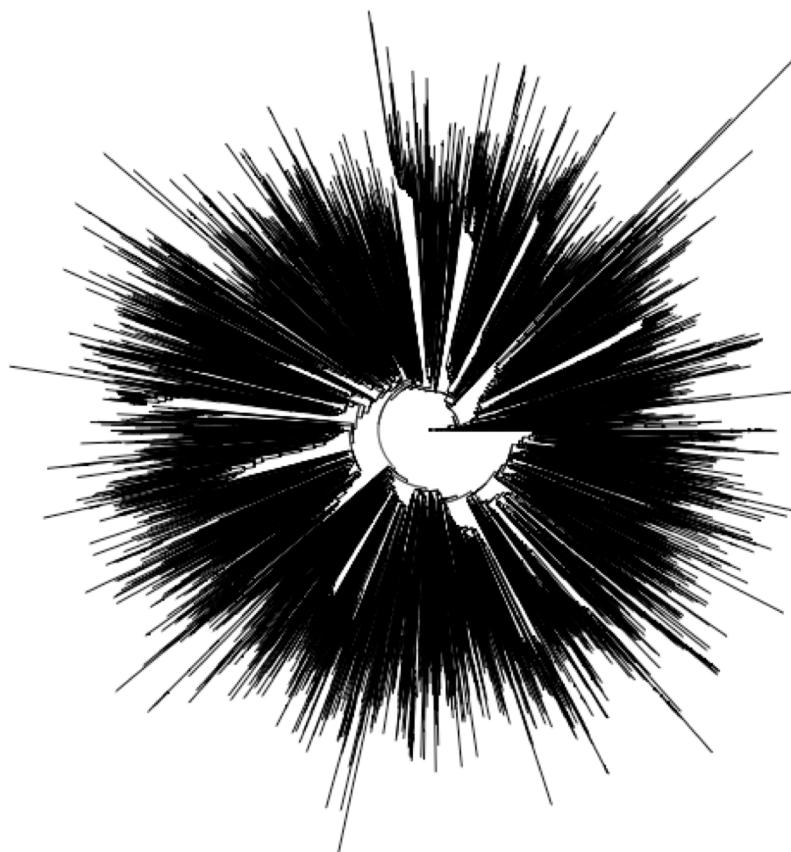
The phylogenetic Ornstein-Uhlenbeck mixed model (POUMM)



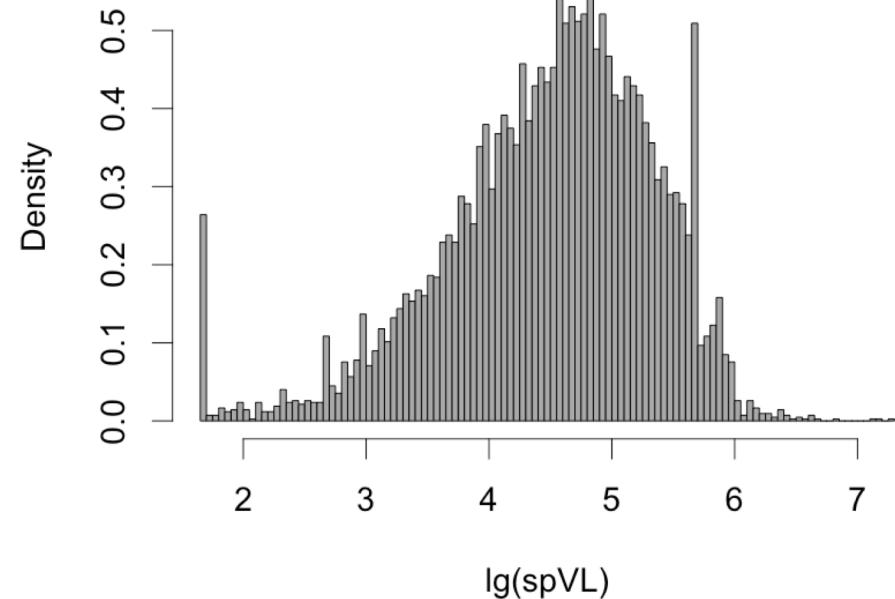
- Model the trait as: $z = G + e$
 - G : genotypic contribution, evolves according to an Ornstein-Uhlenbeck process with parameters $(\Theta, \alpha, \sigma^2)$
 - e : host contribution, assumed to be drawn from a normal distribution $N(0, \sigma_e^2)$
- Estimate the model parameters $(\Theta, \alpha, \sigma^2, \sigma_e^2)$;
- Estimate H^2 as:
$$H_{OUe}^2 = 1 - \sigma_e^2 / \sigma^2(z)$$
- Implemented in the R-package POUMM

HIV Data from the UK Drug Resistance Database

N=8483 patients

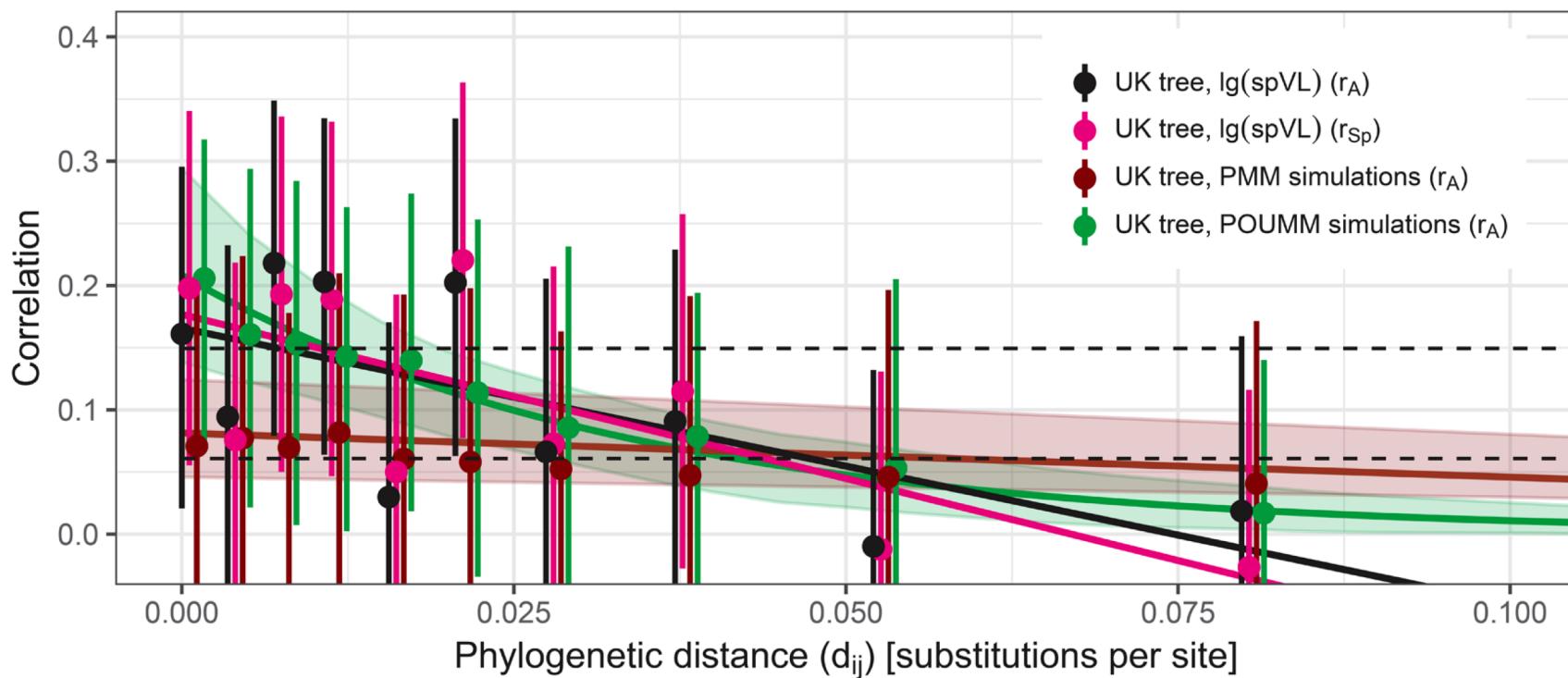


PMM: $H^2=0.06$, CI: [0.02, 0.1]



Hodcroft, E., et al. (2014). The Contribution of Viral Genotype to Plasma Viral Set-Point in HIV Infection. *PLoS Pathogens*, 10(5)

The correlation between phylogenetic pairs decreases with phylogenetic distance



Mitov, V., & Stadler, T. (2018a). A Practical Guide to Estimating the Heritability of Pathogen Traits. *Molecular Biology and Evolution*, 6(9). (original publication included as thesis Chapter 3)

Applications of the POUMM to other datasets and traits

- Swiss HIV cohort study (SHCS), N=2014

	PMM	POUMM
ΔCD4 (unadjusted)	25% (9%–40%)	17% (6%–29%)
ΔCD4 (adjusted)	24% (7%–39%)	17% (5%–30%)
spVL (unadjusted)	12% (2%–28%)	26% (8%–43%)
spVL (adjusted)	8% (0%–26%)	29% (12%–46%)
ppp	22% (5%–39%)	17% (4%–29%)

Bertels, F., Marzel, A., Leventhal, G., Mitov, V., Fellay, J., Günthard, H. F., et al. (2017). Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T Cell Decline and Per-Parasite Pathogenicity. *Molecular Biology and Evolution*, 35(1)(1), 27–37. (original publication included as thesis Chapter 4)

Applications of the POUMM to other datasets and traits

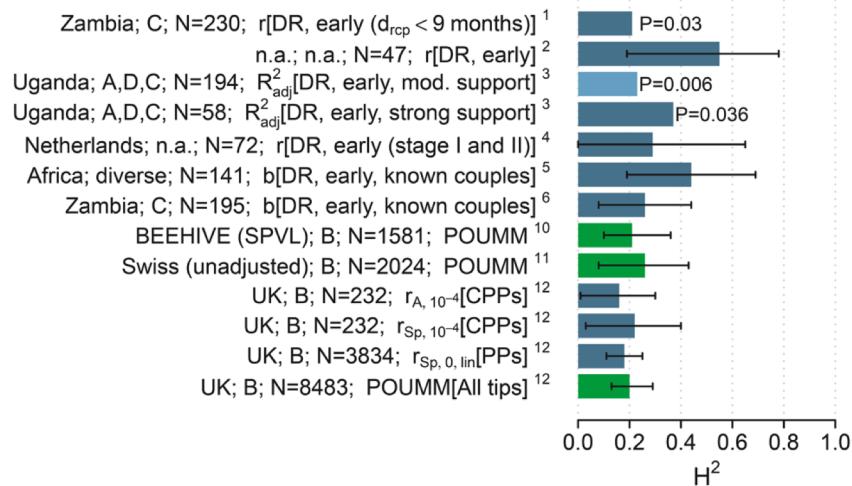
- BEEHIVE dataset

Model	Measure	N	h^2
NULL	GSQL	1,581	0 (0–0)
BM	GSQL	1,581	0.17 (0.08–0.26)
OU	GSQL	1,581	0.31 (0.15–0.43)
NULL	SPVL	1,581	0 (0–0)
BM	SPVL	1,581	0.13 (0.05–0.2)
OU	SPVL	1,581	0.21 (0.1–0.36)
NULL	CD4 slope	1,170	0 (0–0)
BM	CD4 slope	1,170	0.11 (0–0.19)
OU	CD4 slope	1,170	0.1 (0.01–0.27)

Blanquart, F., Wymant, C., Cornelissen, M., Gall, A., Bakker, M., Bezemer, D., et al. (2017). Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *Plos Biology*, 15(6)

Summary: an agreement between cohorts and methods accounting for within-host evolution

A Cohort; Subtype; N; Method[data, filter]



Resemblance-based

Phylogenetic-based (Brownian motion)

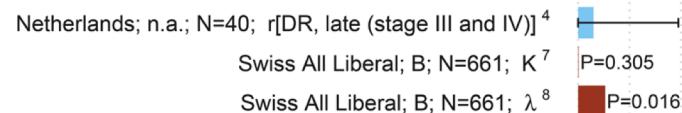
Phylogenetic-based (Ornstein-Uhlenbeck)

References in superscript:

- 1: Tang et al. (2004);
- 2: Hecht et al. (2010);
- 3: Hollingsworth et al. (2010);
- 4: van der Kuyl et al. (2010);
- 5: Lingappa et al. (2013);
- 6: Yue et al. (2013);
- 7: Alizon et al. (2010);
- 8: Shirreff et al. (2013);

Mitov, V., & Stadler, T. (2018a).

B Cohort; Subtype; N; Method[data, filter]



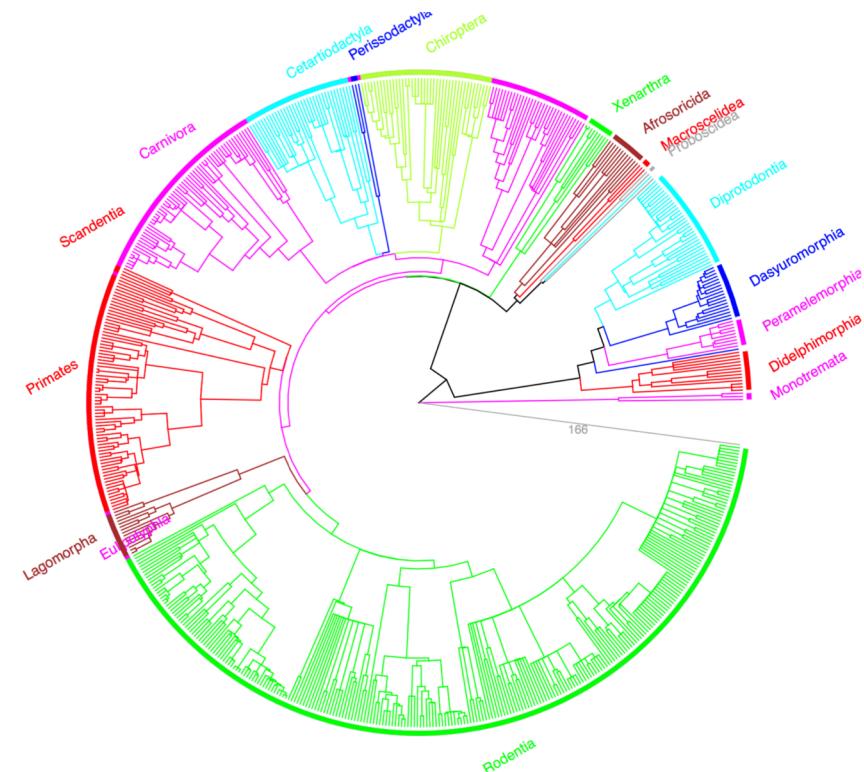
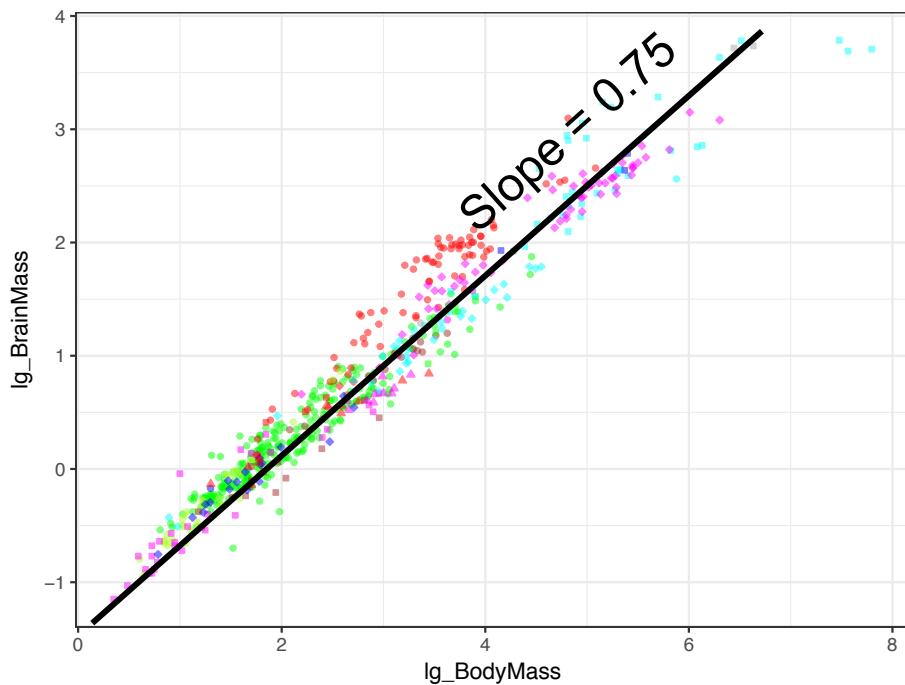
SED

The viral contribution to the variance in HIV-1 virulence is at least 20%.

Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

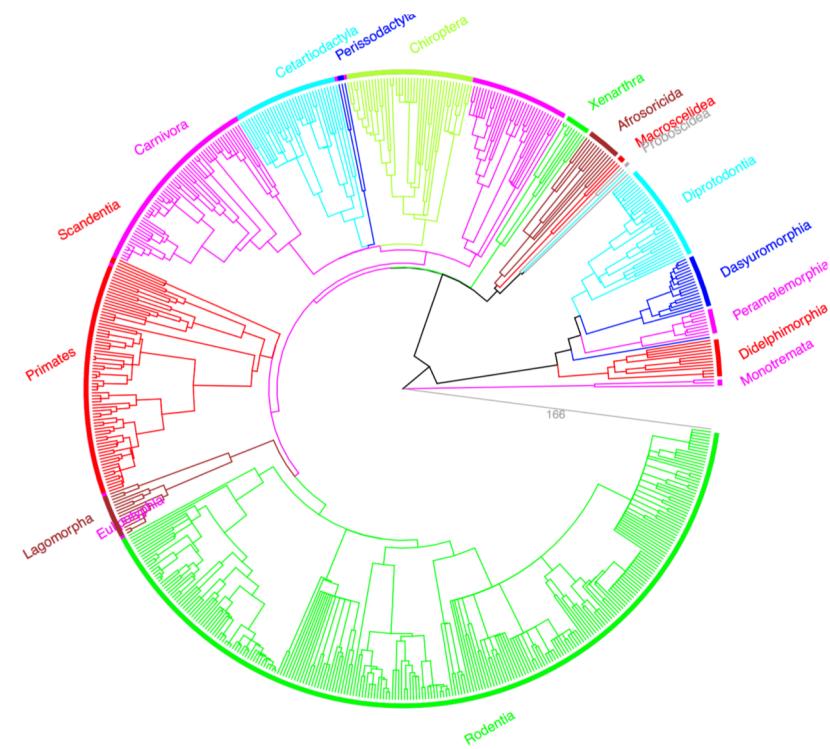
The allometry between brain-mass and body-mass in mammals



Boddy et al. (2012). Comparative analysis of encephalization in mammals reveals relaxed constraints on anthropoid primate and cetacean brain. Rininda-Emonds et al. (2007). The delayed rise of

How did the brain – body allometry evolve through time?

How did the brain – body allometry in mammals evolve through time?



Problem

- Heterogenous evolutionary process through time;
- Existing phylogenetic models assume a homogenous evolutionary process.

Approach: the \mathcal{G}_{LInv} family of models

Definition 1 (The \mathcal{G}_{LInv} family). *We say that a trait evolutionary model belongs to the \mathcal{G}_{LInv} family if it satisfies the following*

1. *after branching the traits evolve independently in all descending lineages,*
2. *the distribution of the trait vector at time t , $\vec{x}(t)$, conditional on the trait vector at time $s < t$, $\vec{x}(s)$, is Gaussian with the mean and variance satisfying*

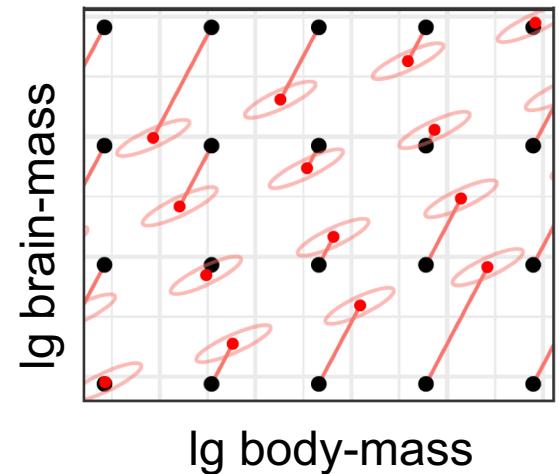
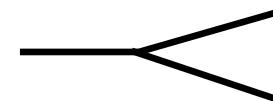
(2.a) $E[\vec{x}(t)|\vec{x}(s)] = \vec{\omega} + \Phi\vec{x}(s)$

(the expectation is a linear function of the ancestral value),

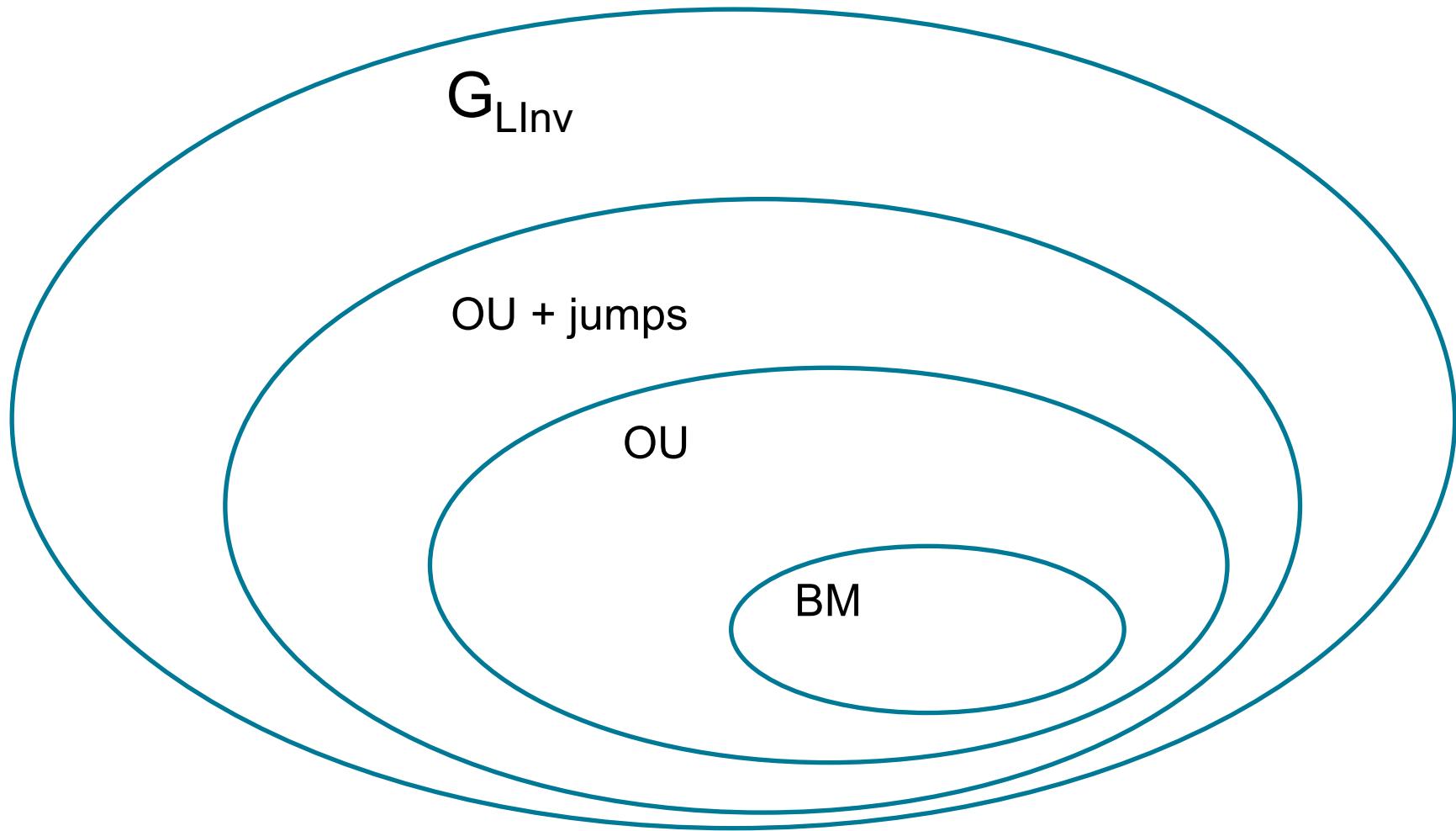
(2.b) $\text{Var}[\vec{x}(t)|\vec{x}(s)] = \mathbf{V}$

(variance is invariant with respect to the ancestral value),

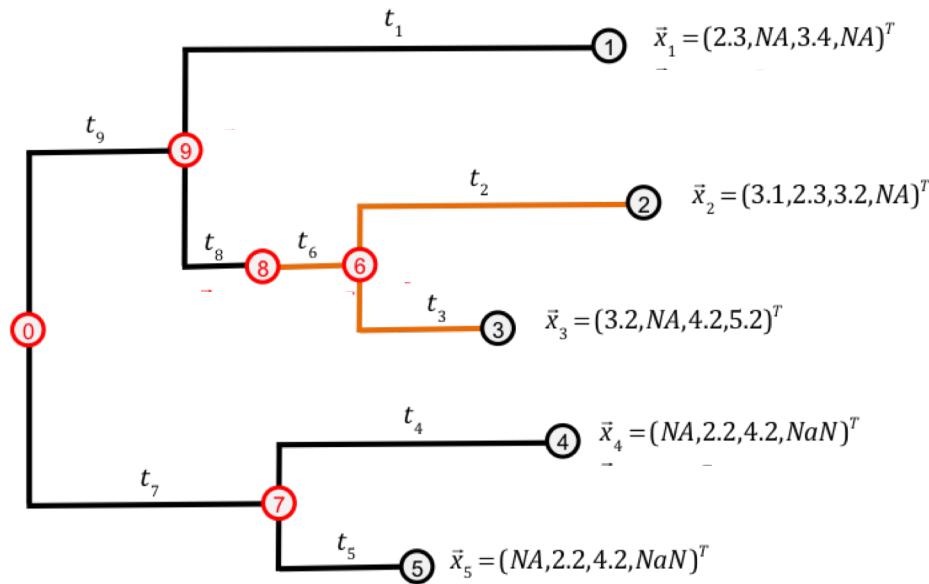
for some vector $\vec{\omega}$ and matrices Φ , \mathbf{V} which may depend on s and t but do not depend on the trait trajectory $\vec{x}(\cdot)$.



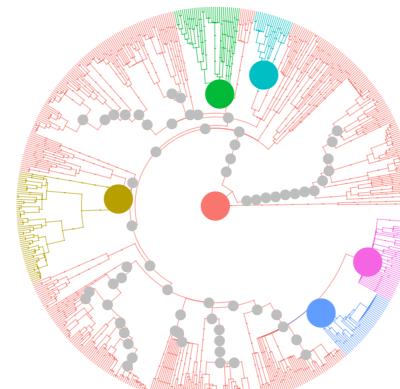
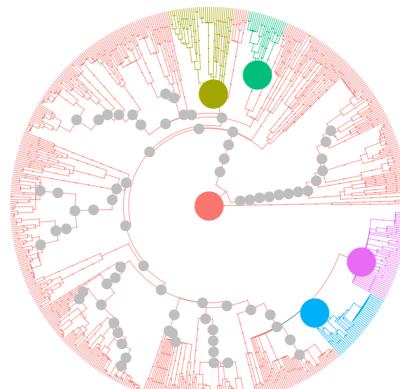
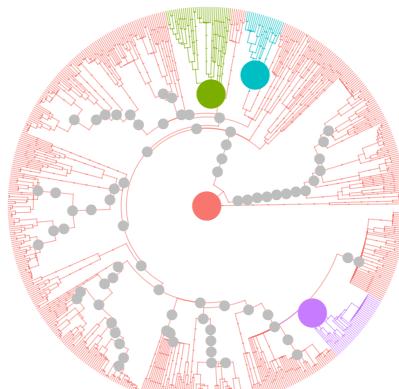
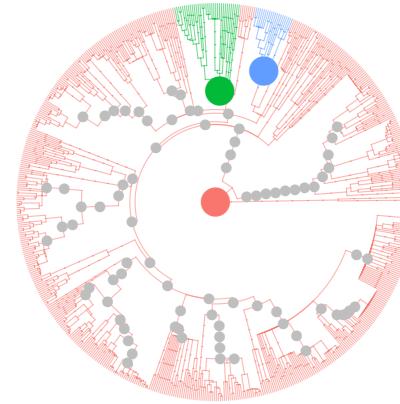
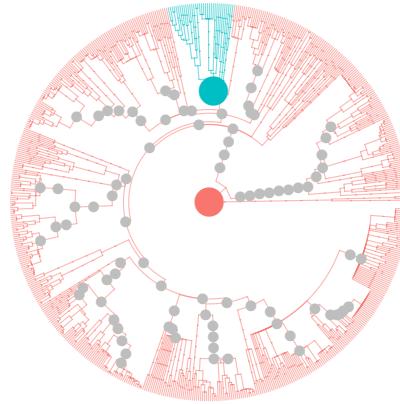
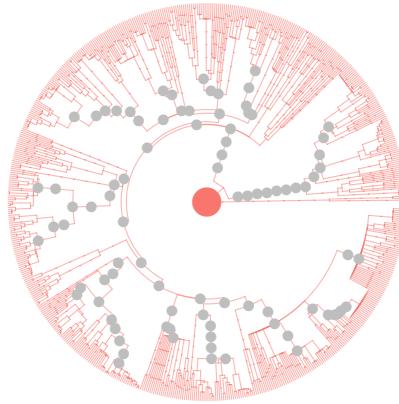
Approach: the G_{LInv} family of models



Approach: Use a mixed Gaussian phylogenetic model (MGPM) over the G_{LInv} family



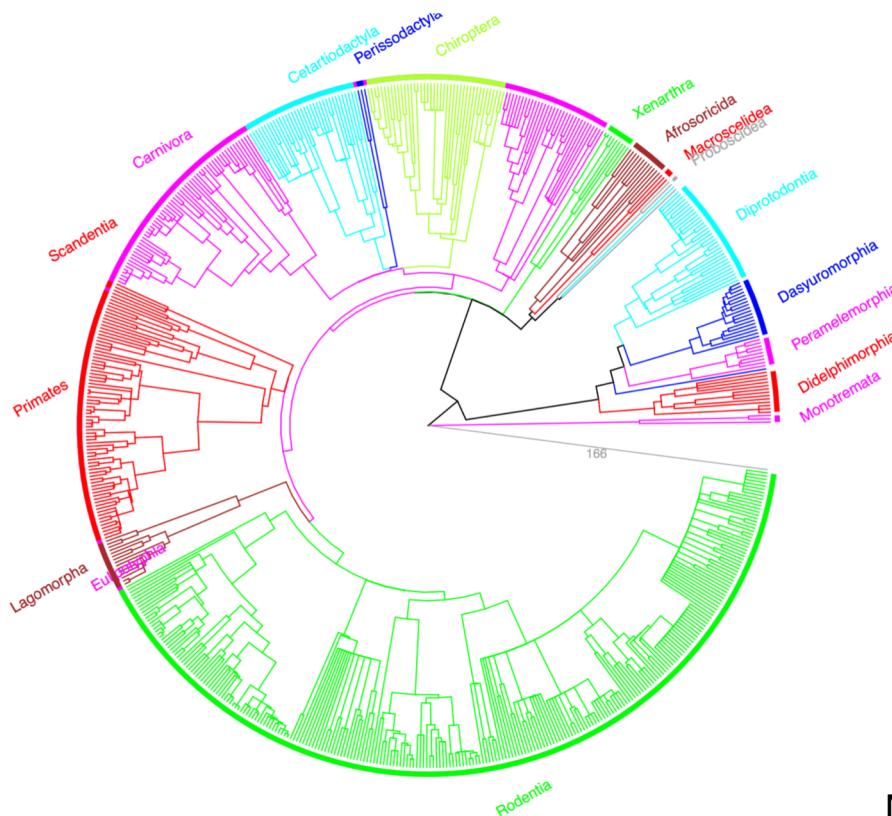
Stepwise AIC search for the “best” MGPM



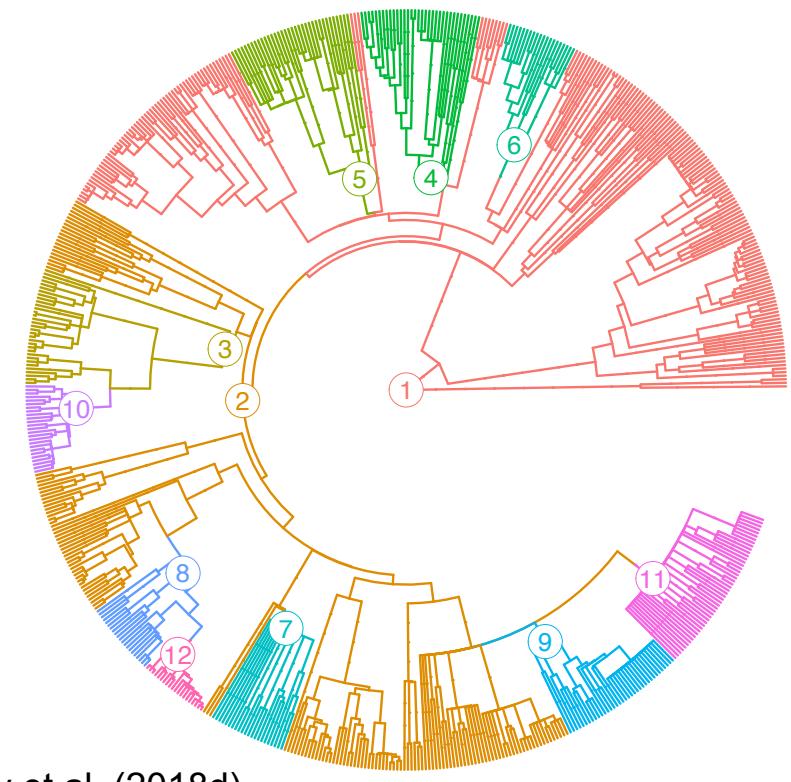
Mitov, V. & Bartoszek, K. & Stadler, T. (2018d). Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models, *Proc. Natl. Acad. Of Sciences* (in review, included as thesis Chapter 7).

An MGPM fit to the mammal data

21 mammal orders

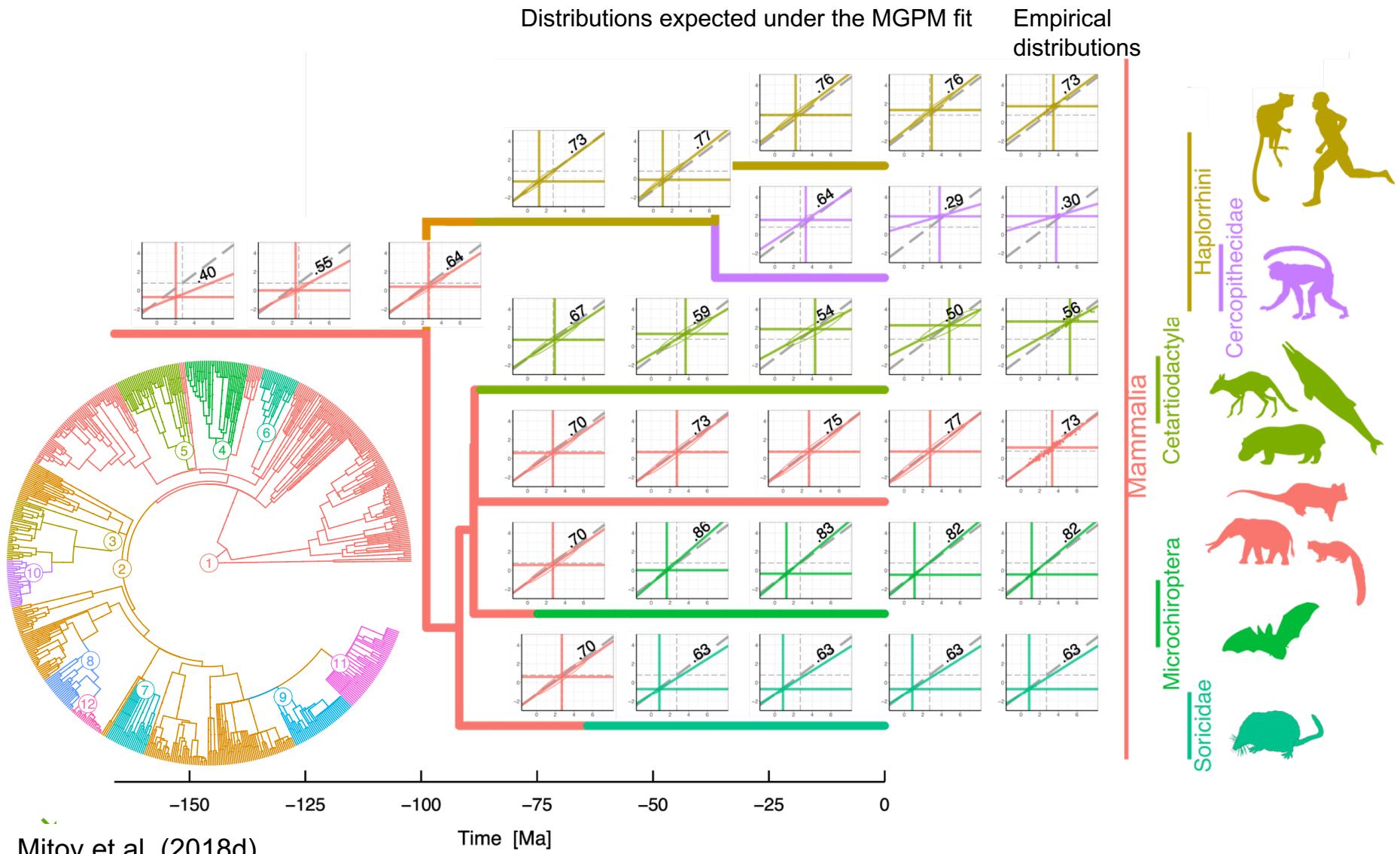


12 inferred evolutionary regimes
Best AIC=-241



Mitov et al. (2018d)

Evolution of the brain-body mass allometry



Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- **Fast likelihood calculation of Gaussian phylogenetic models**
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Software

TOYEPIDEMIC

<R-package>

Simulation of within and
between-host viral
dynamics during epidemics

PATHERIT

<R-package>

Estimators of pathogen
trait heritability

PCMFit

<R-package>

Fitting phylogenetic
comparative models

PCMBASE

<R-package>

Fast likelihood calculation
and trait simulation of
multivariate phylogenetic
models

SPLITT

<C++ template library>

Serial and Parallel Lineage Traversal of Trees

Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Software

TOYEPIDEMIC

<R-package>

Simulation of within and
between-host viral
dynamics during epidemics

PATHERIT

<R-package>

Estimators of pathogen
trait heritability

PCMFit

<R-package>

Fitting phylogenetic
comparative models

PCMBASE

<R-package>

Fast likelihood calculation
and trait simulation of
multivariate phylogenetic
models

POUMM

<R-package>

Phylogenetic Ornstein-
Uhlenbeck Mixed Model

SPLITT

<C++ template library>

Serial and Parallel Lineage Traversal of Trees

The likelihood of the POUMM model can be calculated in O(N) time

The likelihood function is equivalent to N-variate Gaussian probability density function:

$$pdf(\vec{\mathbf{x}}|\Theta, \mathcal{T}) = \frac{1}{\sqrt{\det(2\pi\Sigma(\Theta, \mathcal{T}))}} \exp \left[-\frac{1}{2} (\vec{\mathbf{x}} - \vec{\mu}(\Theta, \mathcal{T}))' \Sigma(\Theta, \mathcal{T})^{-1} (\vec{\mathbf{x}} - \vec{\mu}(\Theta, \mathcal{T})) \right]$$

	POUMM	PMM ($\alpha \rightarrow 0$)
Θ :	$\langle g_0, \alpha, \theta, \sigma, \sigma_e \rangle$	$\langle g_0, \sigma, \sigma_e \rangle$
$\mu_i(\Theta, \mathcal{T})$:	$e^{-\alpha t_{0i}} g_0 + (1 - e^{-\alpha t_{0i}}) \theta$	g_0
$\Sigma_{ii}(\Theta, \mathcal{T})$:	$\sigma^2 \frac{(1 - e^{-2\alpha t_{0i}})}{2\alpha} + \sigma_e^2$	$\sigma^2 t_{0i} + \sigma_e^2$

Calculating the matrix Σ is an $O(N^2)$ operation

Inverting the matrix Σ is an $O(N^3)$ operation

=> For big N, the likelihood calculation is slow.

The likelihood of the POUMM model can be calculated in O(N) time

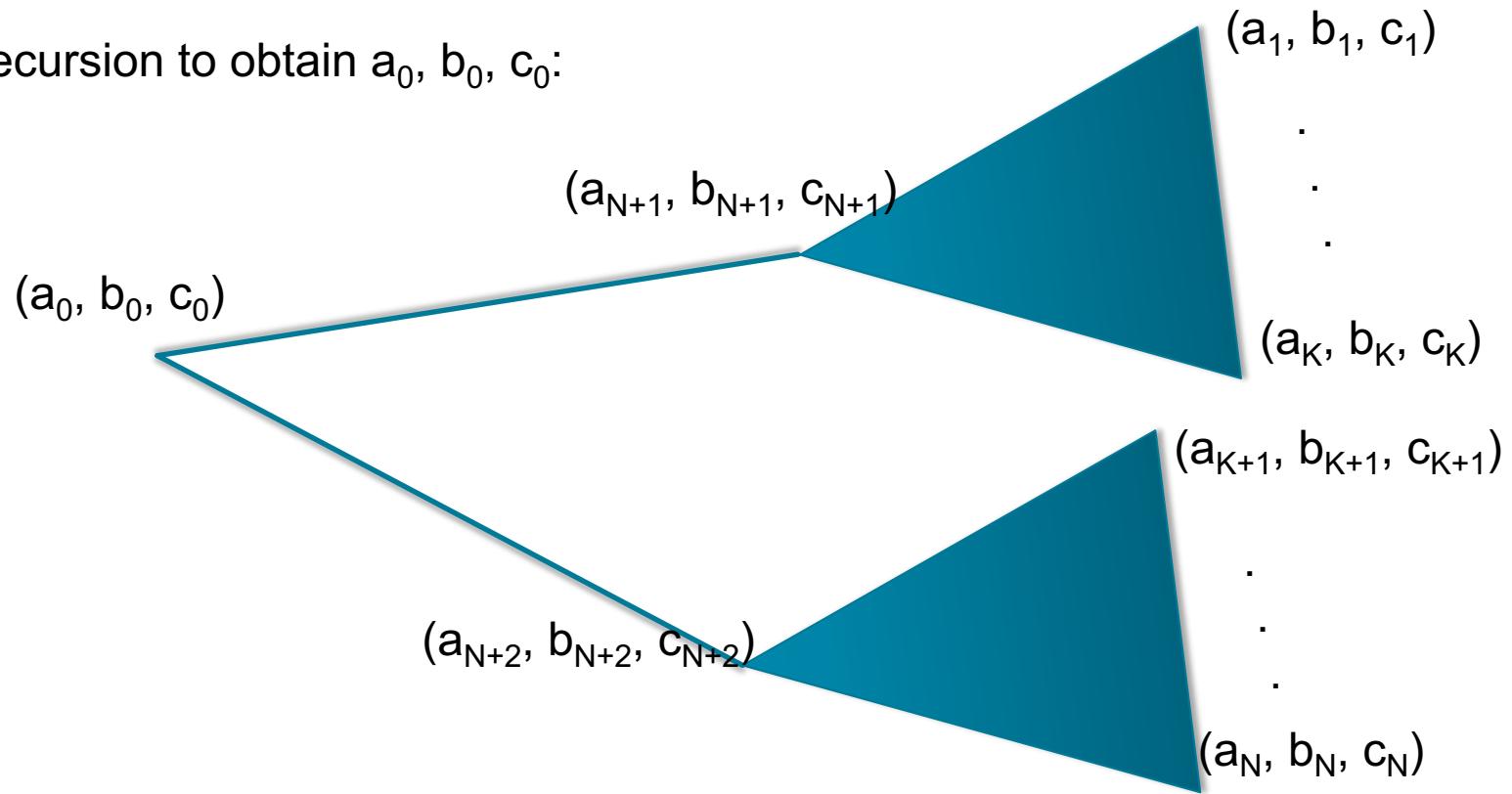
The likelihood function can be expressed as a quadratic polynomial of the root-value:

$$\ln(\text{pdf}(\vec{X}|\Theta, \mathcal{T})) = \ell\ell(\Theta, \mathcal{T}) = a_0 x_0^2 + b_0 x_0 + c_0$$

Mitov, V. & Stadler, T. (2018b). Fast Bayesian Inference of Gaussian Phylogenetic Models Using Parallel Likelihood Calculation. (manuscript currently in revision for Methods in Ecology and Evolution, included as **thesis Chapter 5**).

Pruning from the tips to the root of the tree to calculate the coefficients a , b , c for each node

Use recursion to obtain a_0 , b_0 , c_0 :



Mitov, V. & Stadler, T. (2018b).

Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Software

TOYEPIDEMIC

<R-package>

Simulation of within and
between-host viral
dynamics during epidemics

PATHERIT

<R-package>

Estimators of pathogen
trait heritability

PCMFit

<R-package>

Fitting phylogenetic
comparative models

PCMBASE

<R-package>

Fast likelihood calculation
and trait simulation of
multivariate phylogenetic
models

SPLITT

<C++ template library>

Serial and Parallel Lineage Traversal of Trees

Quadratic polynomial representation of the log-likelihood of \mathbf{G}_{LInv} models

$$pdf(\vec{\mathbf{X}}|\Theta, \mathcal{T}) = \frac{1}{\sqrt{\det(2\pi\Sigma(\Theta, \mathcal{T}))}} \exp \left[-\frac{1}{2} (\vec{\mathbf{X}} - \vec{\mu}(\Theta, \mathcal{T}))' \Sigma(\Theta, \mathcal{T})^{-1} (\vec{\mathbf{X}} - \vec{\mu}(\Theta, \mathcal{T})) \right]$$

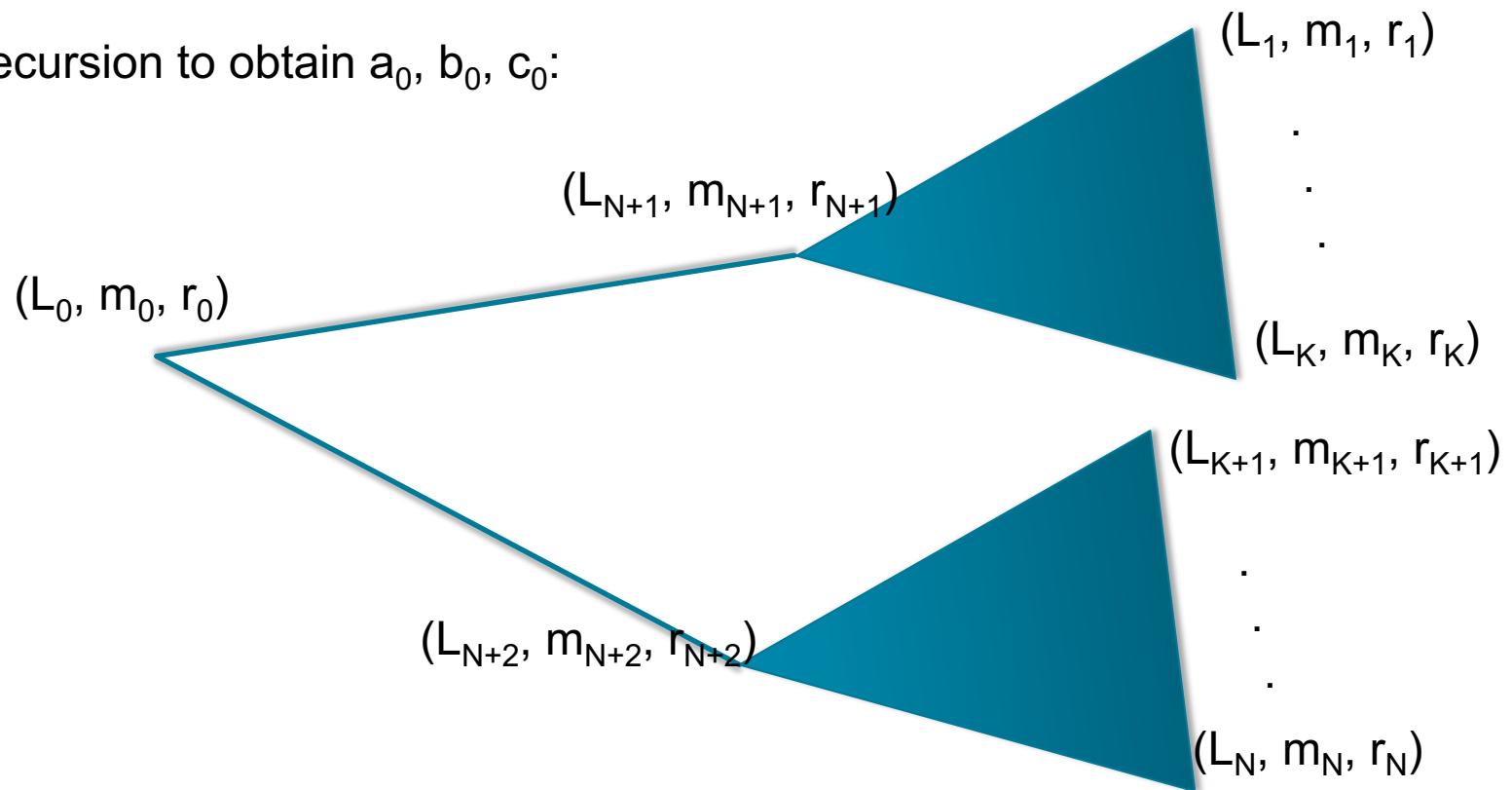
Theorem 2. Let \mathcal{M} be a trait evolutionary model from the $\mathcal{G}_{\text{LInv}}$ -family and \mathbb{T} be a phylogenetic tree. Let Θ be the parameters of \mathcal{M} . For the root (0) or any internal node j in \mathbb{T} , there exists a $k_j \times k_j$ matrix \mathbf{L}_j , a k_j -vector \vec{m}_j and a scalar r_j , such that the likelihood of \mathcal{M} for the data \mathbf{X}_j , conditioned on $\vec{x}_j \in \mathbb{R}^{k_j}$ and \mathbb{T} is expressed as:

$$pdf(\mathbf{X}_j | \vec{x}_j, \mathbb{T}, \Theta) = \exp \left(\vec{x}_j^T \mathbf{L}_j \vec{x}_j + \vec{x}_j^T \vec{m}_j + r_j \right). \quad (5)$$

The parameters \mathbf{L}_j , \vec{m}_j , r_j are functions of Θ , the observed data \mathbf{X}_j , and the tree \mathbb{T} , namely, equations 8, 9, and 10.

Pruning from the tips to the root of the tree to calculate the coefficients L , m , r for each node

Use recursion to obtain a_0 , b_0 , c_0 :



Mitov, V. & Stadler, T. (2018b).

Agenda

- Phylogenetic comparative methods
 - Phylogenetic models of trait evolution
 - Challenges in the application of PCMs to big phylogenetic trees
- Example applications
 - Estimating the heritability of virulence in HIV infections
 - Quantifying the brain-body-mass allometry in mammals
- Fast likelihood calculation of Gaussian phylogenetic models
 - The univariate POUMM model
 - Generalization to multivariate mixed Gaussian phylogenetic models
 - Parallel likelihood calculation

Software

TOYEPIDEMIC

<R-package>

Simulation of within and
between-host viral
dynamics during epidemics

PATHERIT

<R-package>

Estimators of pathogen
trait heritability

PCMFit

<R-package>

Fitting phylogenetic
comparative models

PCMBASE

<R-package>

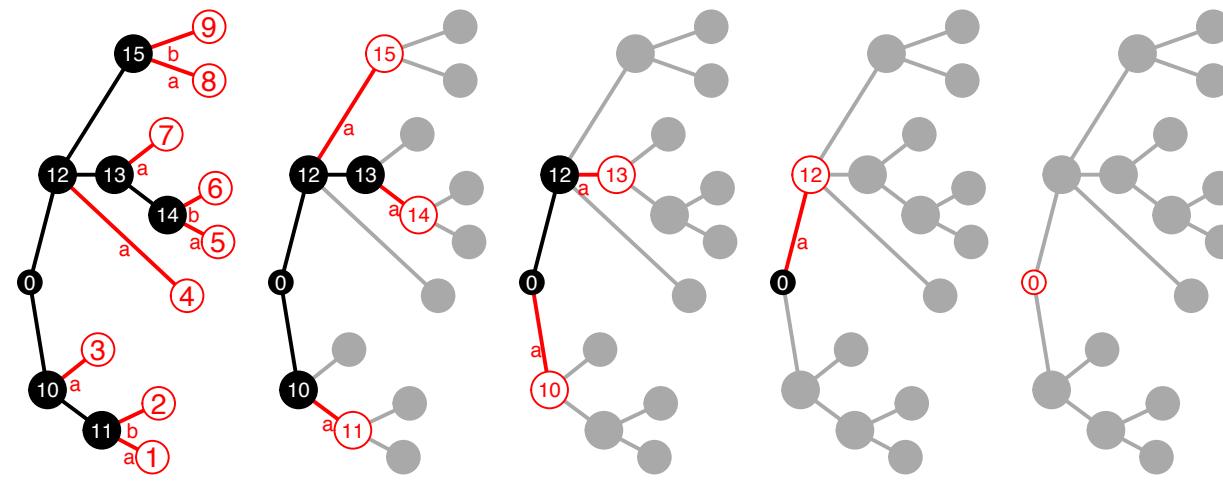
Fast likelihood calculation
and trait simulation of
multivariate phylogenetic
models

SPLITT

<C++ template library>

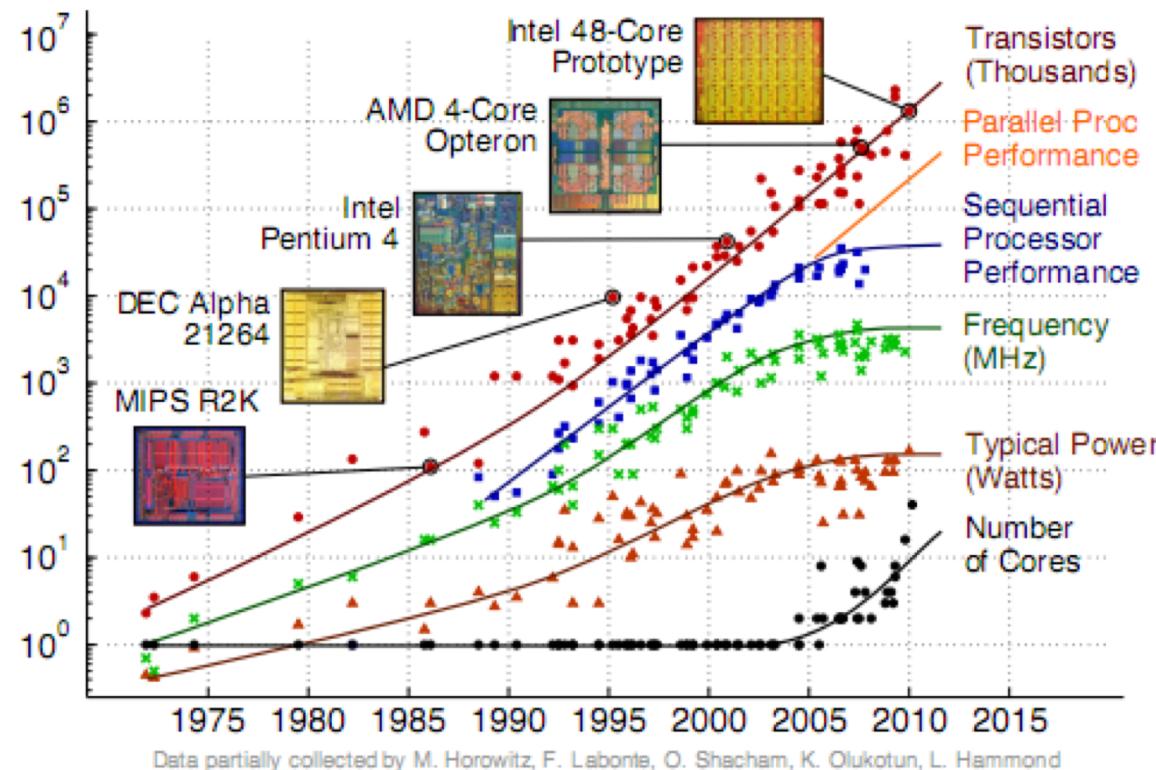
Serial and Parallel Lineage Traversal of Trees

Parallel tree traversal: the SPLITT library



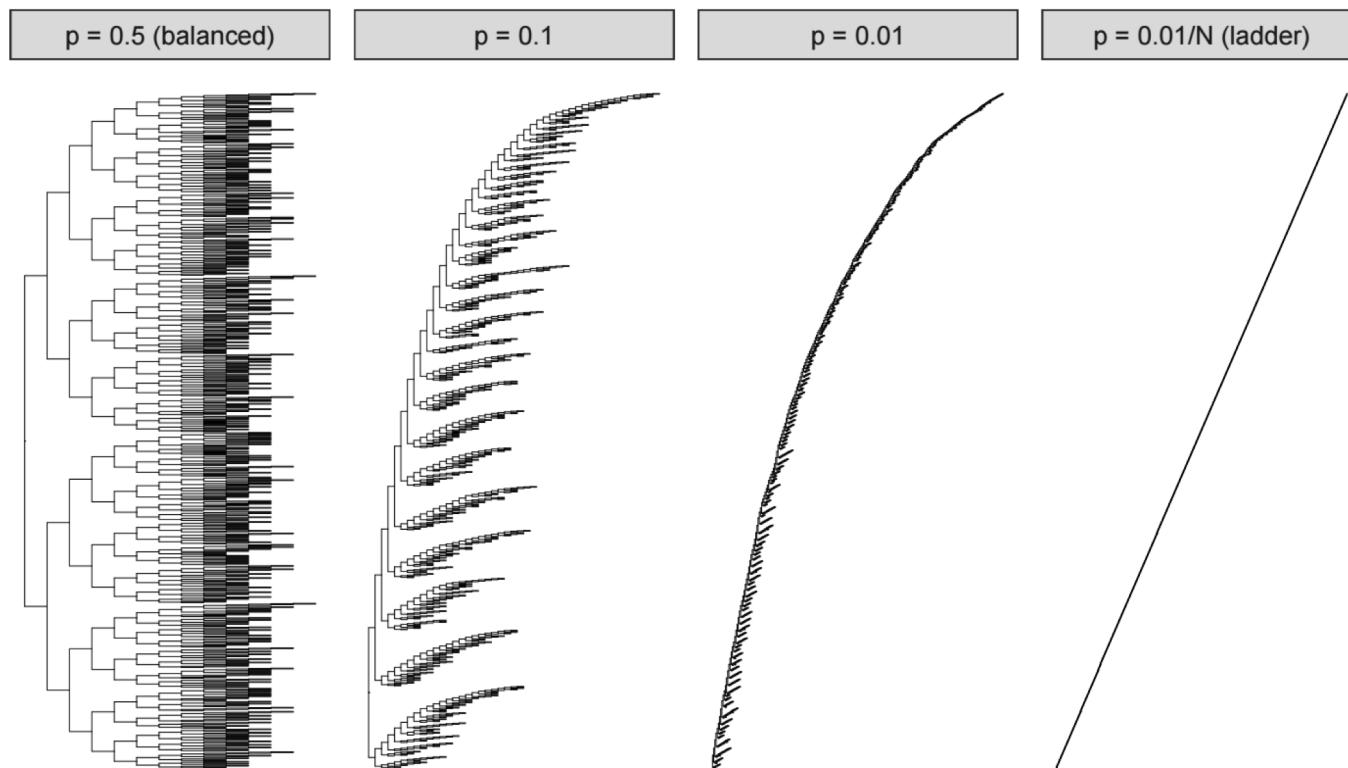
Mitov, V. & Stadler, T. (2018b).

Clock rate vs number of cores in CPUs



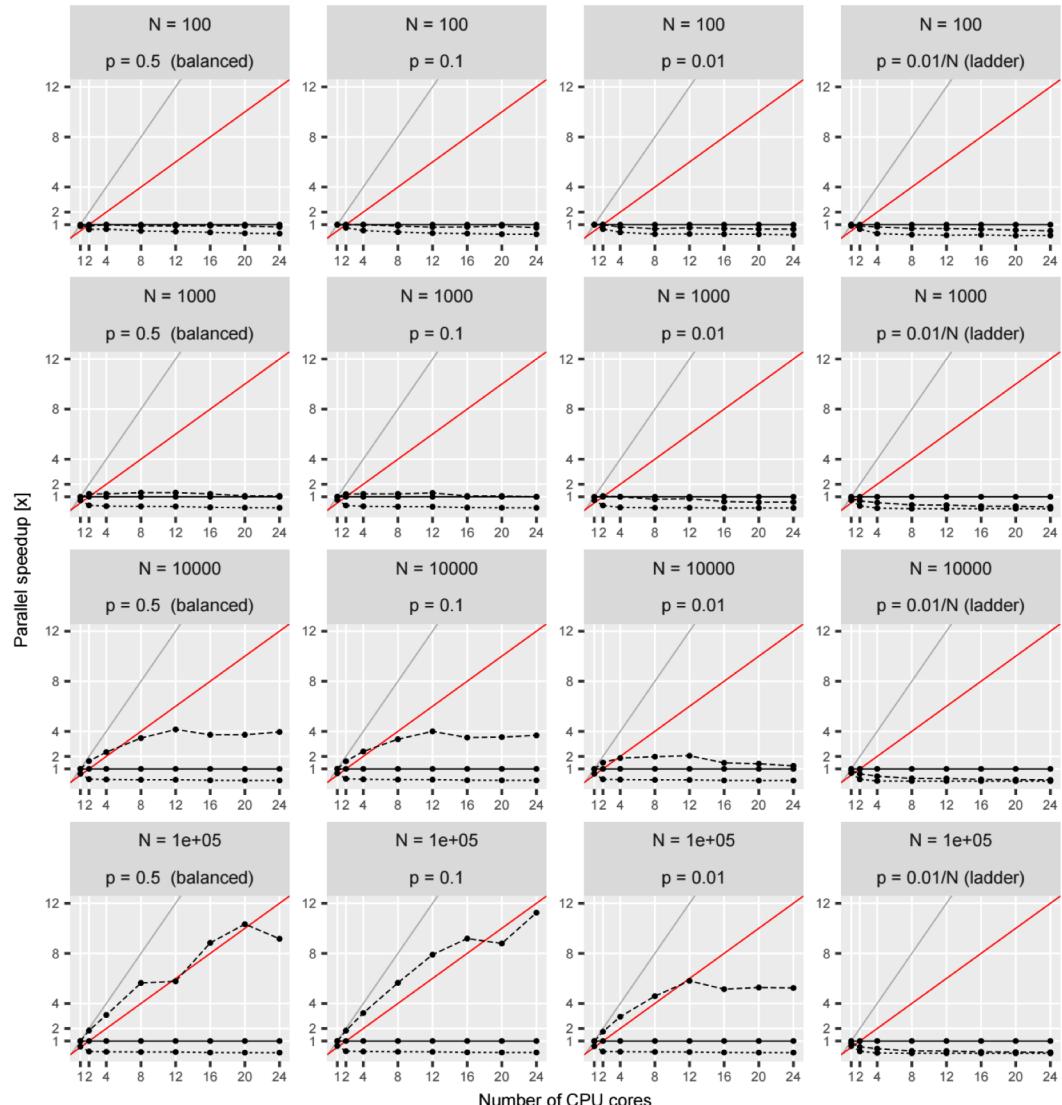
Prepared by C. Batten - School of Electrical and Computer Engineering - Cornell University - 2005 - retrieved Dec 12 2012 -
<http://www.csl.cornell.edu/courses/ece5950/handouts/ece5950-overview.pdf>

Performance benchmark



Mitov, V. & Stadler, T. (2018b).

Up to 10x parallel speed-up of the POUMM likelihood calculation with 20 CPU cores



Mitov, V. & Stadler, T. (2018b).

Conclusions

- Phylogenetic comparative methods provide an evolutionary perspective to the analysis of comparative data. However, model misspecifications are potential pitfalls that can lead to false biological conclusions. Therefore, results should be treated as hypotheses to be validated against competing methods and other types of data.
- The fast likelihood calculation for MGPMs enables data-driven hypothesis generation. Yet, at present only approximate maximum likelihood inference is possible.

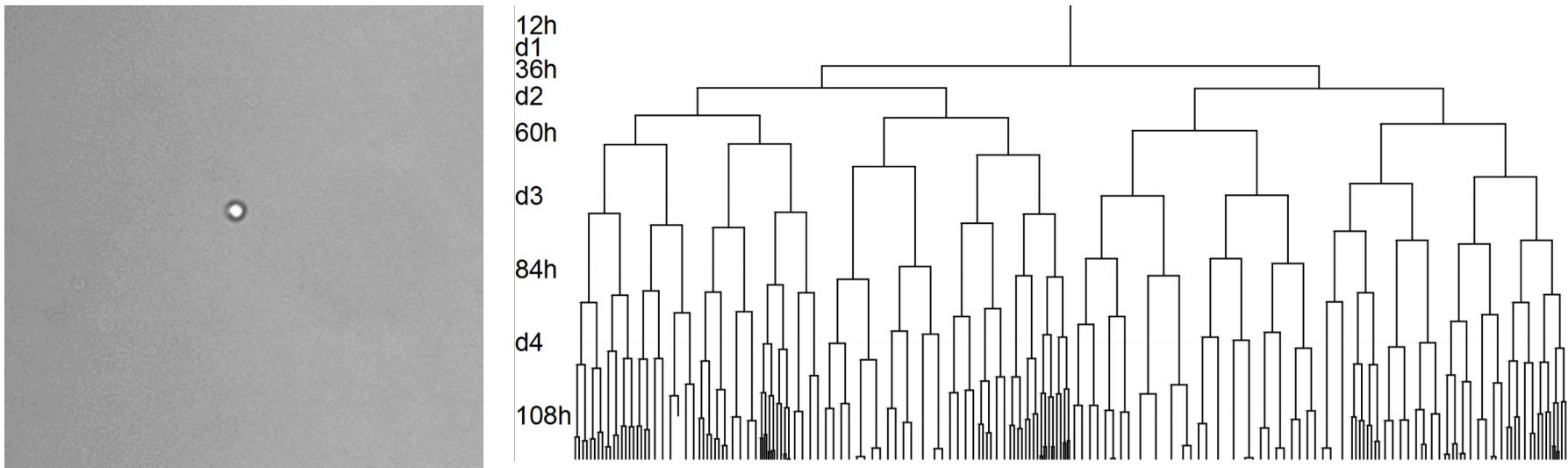
Future directions – methods

- Inclusion of new G_{LInv} model-types in the MGPM framework;
 - Jump-enabled models of punctuated equilibrium;
 - Acceleration/deceleration models of adaptive radiation;
- Bayesian MGPM inference;
- Integrating PCMs with demographic models

Future directions – data analysis

- MGPM search for stem-cell differentiation patterns

Visual tracking of hematopoietic stem cells (HSC)



Morphological data (e.g. Cell area and perimeter) +
Cell signaling data (e.g. Sca1 signal, CD41 signal)

Courtesy: Timm Schröder's laboratory, D-BSSE

Future directions – data analysis

- MGPM search for stem-cell differentiation patterns
- Multiple-trait analysis of HIV epidemiological data, e.g. performing an MGPM analysis jointly for spVL and CD4 decline.
- MGPM based cluster identification in HIV transmission trees.

Acknowledgements

- Tanja Stadler
- My collaborators: Krzysztof Bartoszek, Georgios Asimomitis
- My co-authors: Denise Kühnert, Carsten Magnus, Nicola Müller, Julija Pecerska, David Rasmussen, Chi Zhang, Alexei Drummond, Tracy Heath, Oliver Pybus, Timothy Vaughan, Frederic Bertels, Alex Maryel, Gabriel Leventhal, Jacques Fellay, Huldrych Günthard, Jürg Boni, Sabine Yerli, Thomas Klimkait, Vincent Aubert, Manuel Battegay, Andri Rauch, Mattias Cavassini, Alexandra Calmy, Enos Bernasconi, Patrick Schmid, Alexandra Scherrer, Viktor Müller, Sebastian Bonhoeffer and Roger Kouyos
- Co-referees of the thesis: Roland Regös, Samuel Alizon
- All colleagues from the cEvo group
- My friends, family and my parents

Mammal fossil record

Juramaia

From Wikipedia, the free encyclopedia

Juramaia is an extinct genus of very basal eutherian mammal from the Late Jurassic (Oxfordian stage) deposits of western Liaoning, China; it is a small shrew-like mammal of body length approximately 70–100 mm.^{[1][2]}

Juramaia is known from the holotype BMNH PM1343, an articulated and nearly complete skeleton including incomplete skull preserved with full dentition.

Contents [hide]

- 1 Discovery
- 2 Evolution
- 3 Classification
- 4 See also
- 5 References
- 6 External links

Discovery [edit]

It was collected in the Daxigou site, Jianchang, from the Tiaojishan Formation dated at about 160 million years ago.^[3] It was first named by Zhe-Xi Luo, Chong-Xi Yuan, Qing-Jin Meng and Qiang Ji in 2011 and the type species is *Juramaia sinensis*.^[4]

Evolution [edit]

The discovery of *Juramaia* provides new insight into the evolution of placental mammals by showing that their lineage diverged from that of the marsupials 35 million years earlier than previously thought.^[4] Furthermore, its discovery fills gaps in the fossil record and helps to calibrate modern, DNA-based methods of dating the evolution.^{[5][6]} Based on climbing adaptations found in the forelimb bones, it has been suggested that the basal stock of Eutheria was arboreal,^[4] in a manner resembling that of modern rats.^[7]



Large brains in mammals evolved for better sense of smell



Hadrocodium wui

Photograph by: Klingler and Luo, Carnegie Museum of Natural History

The fossil of the Jurassic mammal *Hadrocodium wui* -- its skull is only 12 millimeters (less than 0.5 inch) long and estimated to weigh only 2 grams. The CT scan of its braincase reveals new information about the sense of smell in early mammals. Credit: Photo: Mark A. Klingler/Carnegie Museum of Natural History

An MGPM fit to the mammal dataset

