

EVOLUTION 2019

June 21-25

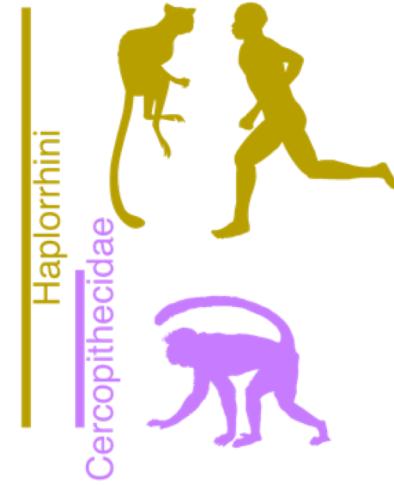
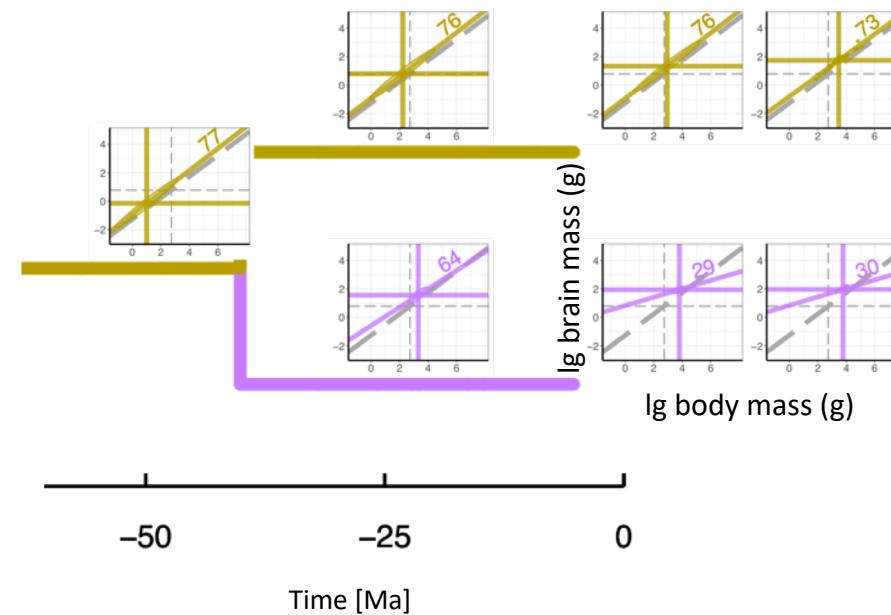
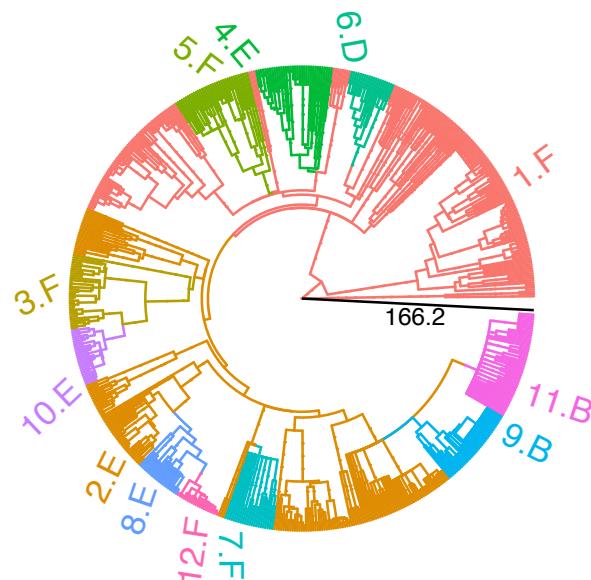
Providence, Rhode Island



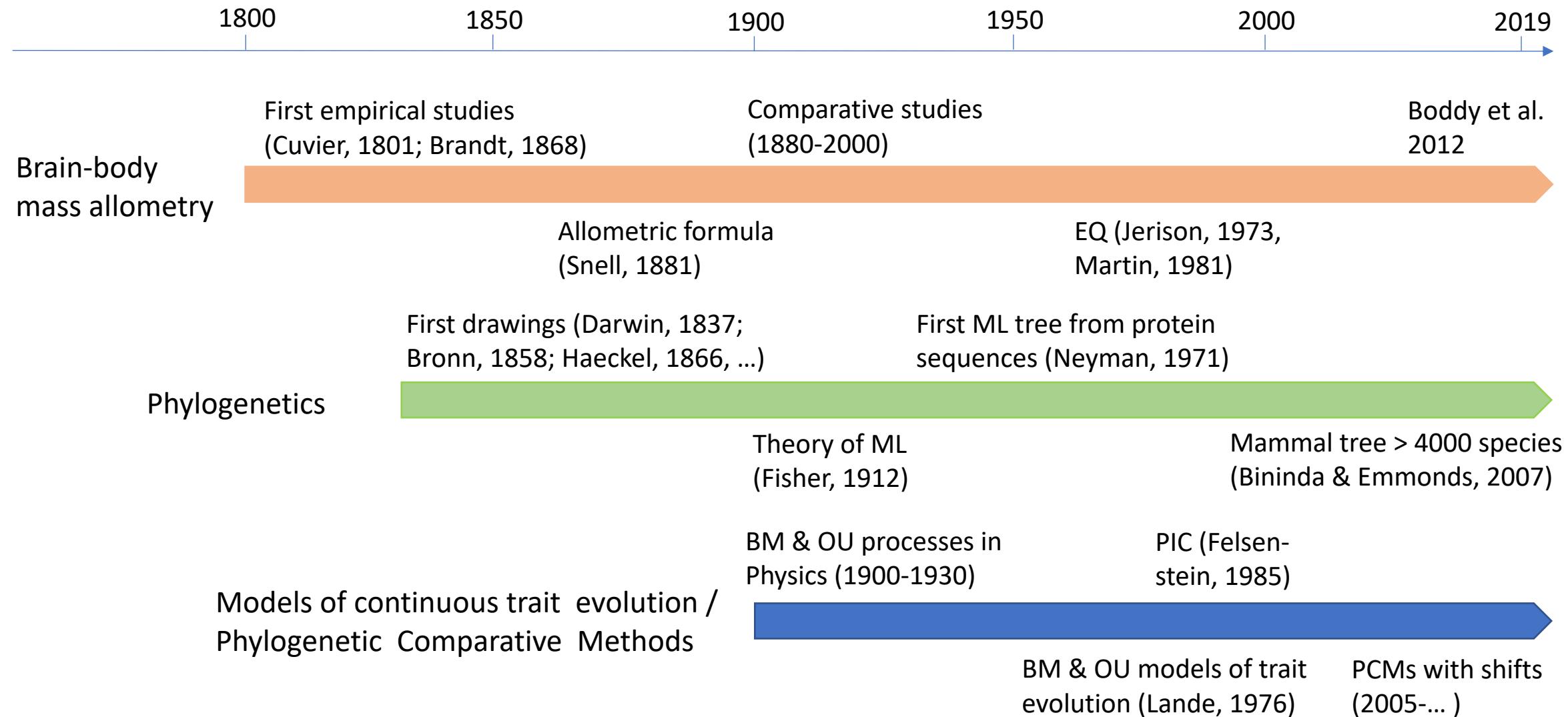
American Society of Naturalists
Society of Systematic Biologists
Society for the Study of Evolution

Automatic Generation of Evolutionary Hypotheses Using Mixed Gaussian Phylogenetic Models

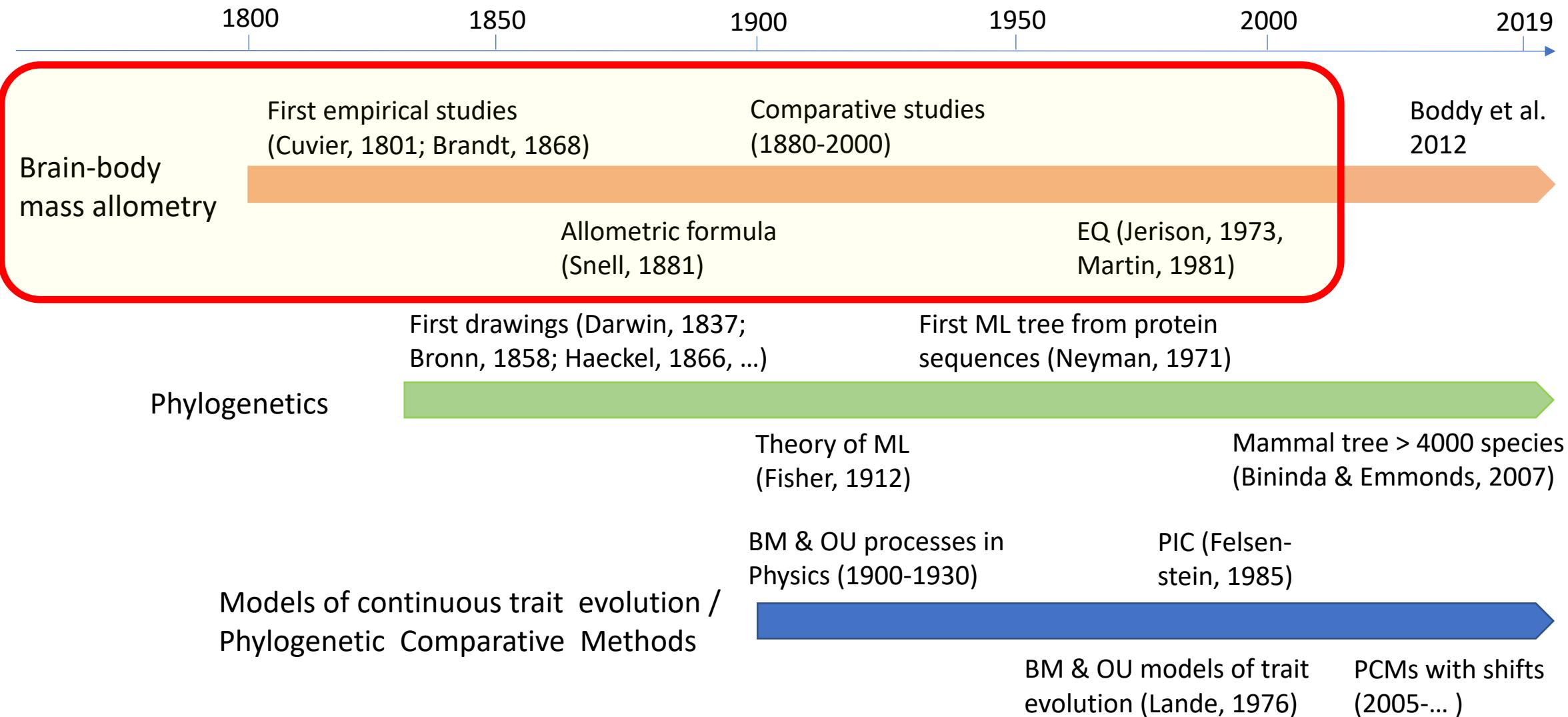
Venelin Mitov, Krzysztof Bartoszek, Tanja Stadler



Timeline



Timeline



The brain-body mass allometry in mammals



Image by Charles J Sharp



Image by Charles J Sharp



Image by Macinate

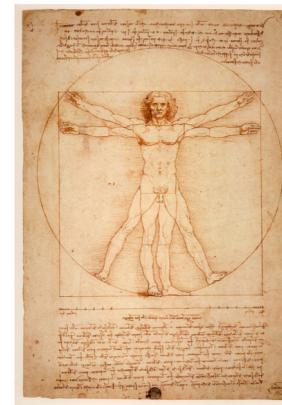


Image by Leonardo Da Vinci



Image by Bernard DUPONT

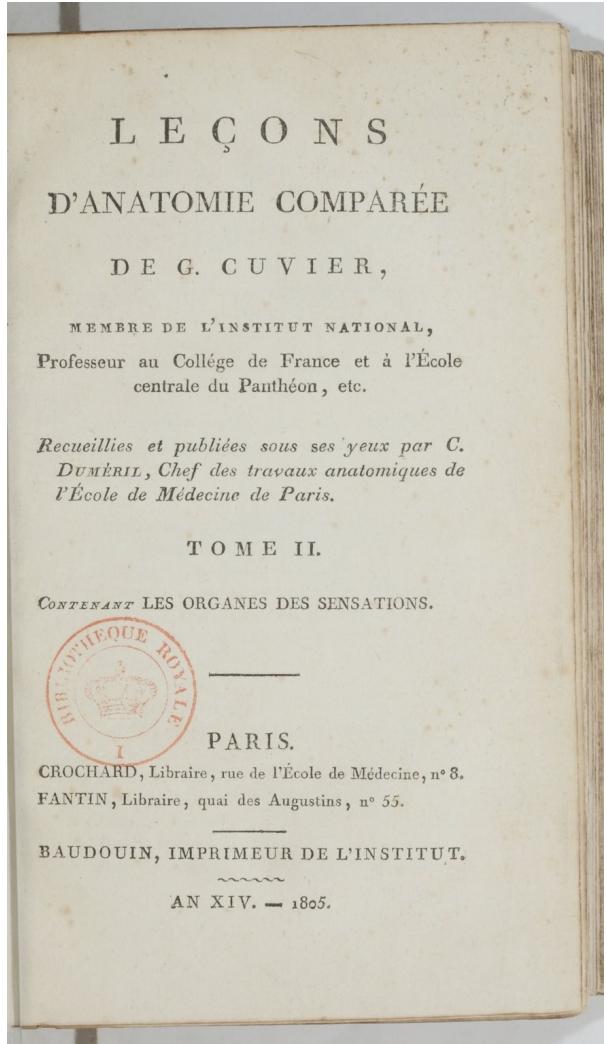


Brain images downloaded from <https://www.quantamagazine.org/how-humans-evolved-supersize-brains-20151110/>



	Capybara	Rhesus Macaque	Western Gorilla	Human	African Bush Elephant
Mean body mass:	28.5 kg	4.6 kg	121 kg	65 kg	4301 kg
Mean brain mass:	75 g (0.26%)	88 g (1.9%)	455 g (0.38%)	1250 g (1.9%)	5436 g (0.13%)

Georges Cuvier, 1801



<p>ART. V. <i>Du cerveau des mammifères.</i> 149</p> <p>soit l'ont énormément plus petit que ceux à sang chaud, etc.</p> <p>Homme. $\frac{1}{22} \cdot \frac{1}{25} \cdot \frac{1}{30} \cdot \frac{1}{35}$</p> <p>Selon qu'il est jeune ou vieux.</p>	<p>O R A N G E S.</p> <p>Gibbon $\frac{1}{45}$</p> <p>S A P A J O U S.</p> <p>Saïmiri $\frac{1}{25}$ Saï $\frac{1}{25}$ Ouïstiti $\frac{1}{25}$ Coaïta. $\frac{1}{41}$</p> <p>G U E N O N S.</p> <p>Malbrouc jeune. $\frac{1}{21}$ Callitriches et Patas $\frac{1}{41}$ Mone $\frac{1}{41}$ Mangabey $\frac{1}{48}$</p> <p>M A G O T S et M A C A Q U E S.</p> <p>Mataque. $\frac{1}{45}$ Magot. $\frac{1}{105}$ Papion $\frac{1}{105}$</p> <p>M A K I S.</p> <p>Mococo jeune $\frac{1}{61}$ Vari. $\frac{1}{84}$</p>
--	--

K 5

<p>150 IX^e LEÇON. <i>Du cerv. des an. vertébrés.</i></p> <p>C H E İ R O P T È R E S.</p> <p>Noctule $\frac{1}{96}$</p> <p>P L A N T I G R A D E S.</p> <p>Taupe. $\frac{1}{86}$ Ours $\frac{1}{265}$ Hérisson. $\frac{1}{168}$</p> <p>C A R N I V O R E S.</p> <p>Chien $\frac{1}{47} \cdot \frac{1}{30} \cdot \frac{1}{57} \cdot \frac{1}{154} \cdot \frac{1}{161} \cdot \frac{1}{305}$ Renard $\frac{1}{205}$ Loup $\frac{1}{210}$ Chat $\frac{1}{82} \cdot \frac{1}{94} \cdot \frac{1}{156}$ Panthere $\frac{1}{247}$ Marte. $\frac{1}{565}$ Furet. $\frac{1}{158}$</p> <p>R O N G E U R S.</p> <p>Castor. $\frac{7}{290}$ Lièvre $\frac{1}{220}$ Lapin. $\frac{1}{140} \cdot \frac{1}{152}$ Ondatra. $\frac{1}{124}$ Rat $\frac{1}{76}$ Souris. $\frac{1}{45}$ Mulot. $\frac{1}{51}$</p>



Image from the National Library of Medicine, Bethesda, Maryland

Alexandre Brandt, 1868

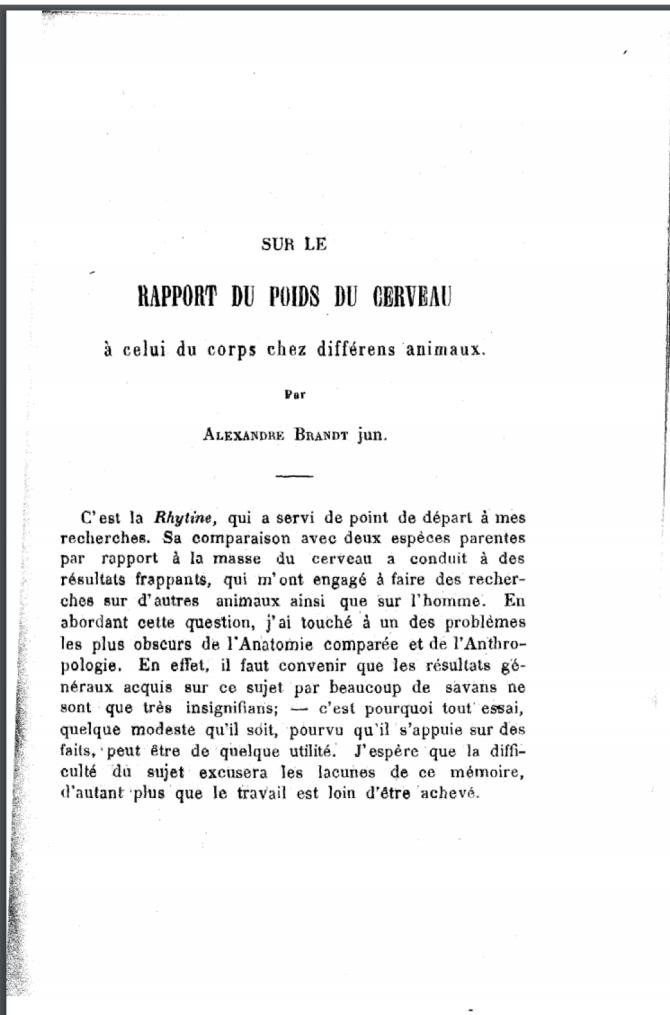
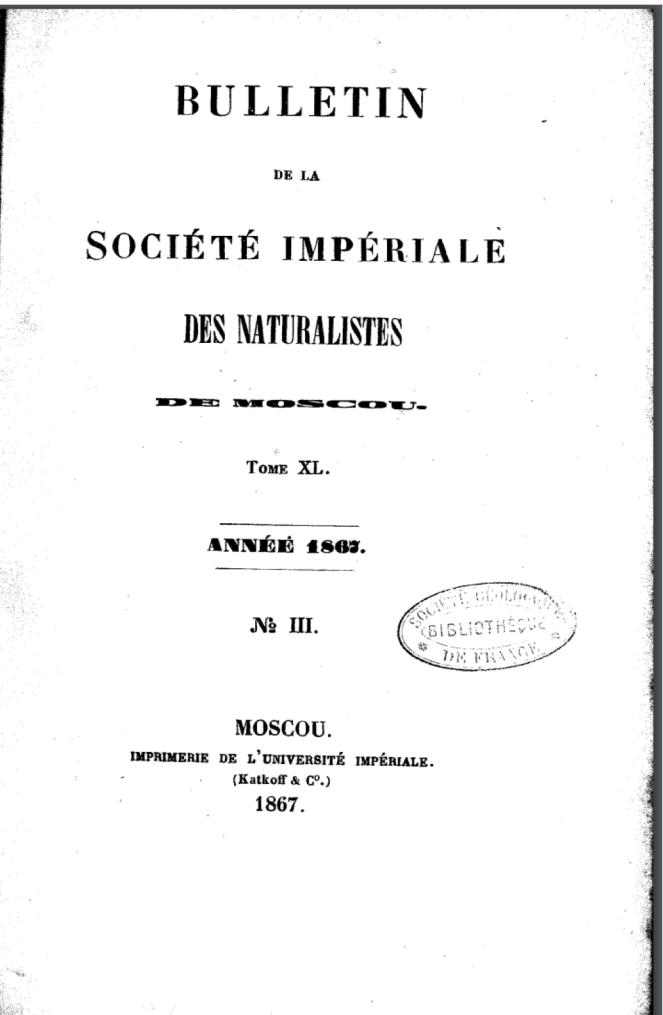


Image by Viktor Kornienko, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=8019484>

Alexandre Brandt, 1868

1. *Plus un animal est petit, plus il a relativement de cerveau.*

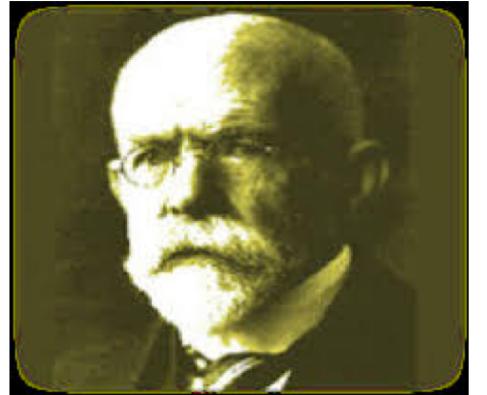
2. *Plus un animal est petit, plus les procès physiologiques sont relativement actifs chez lui.*



Image by Viktor Kornienko, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=8019484>

1. *The smaller an animal is, the more brain matter it has relative to its size.*
2. *The smaller an animal is, the more active its physiological processes are.*

Otto Snell's allometric formula, 1881

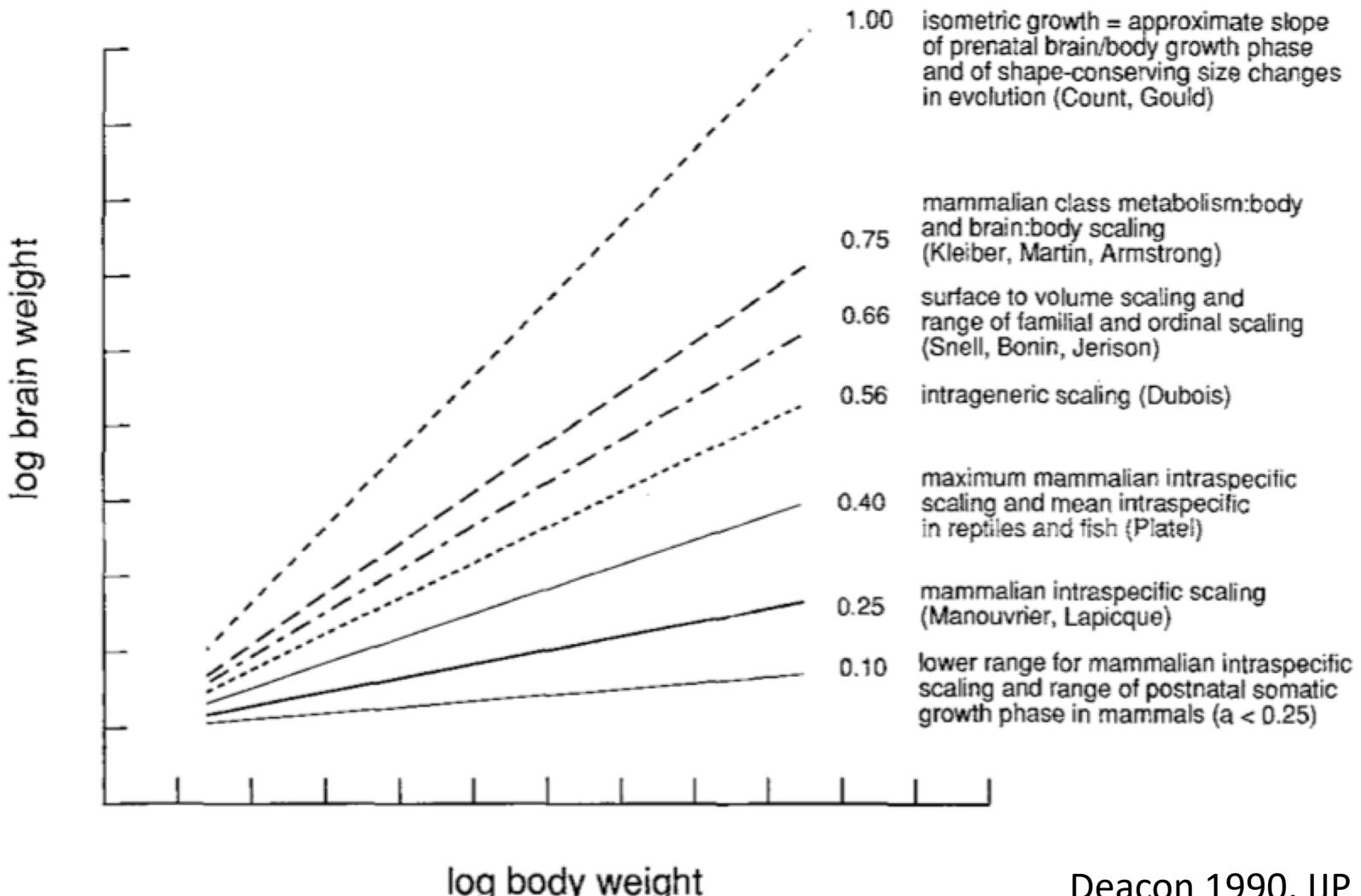


Source: see at the end of the presentation

$$(\text{Brain mass}) = k (\text{Body mass})^a$$

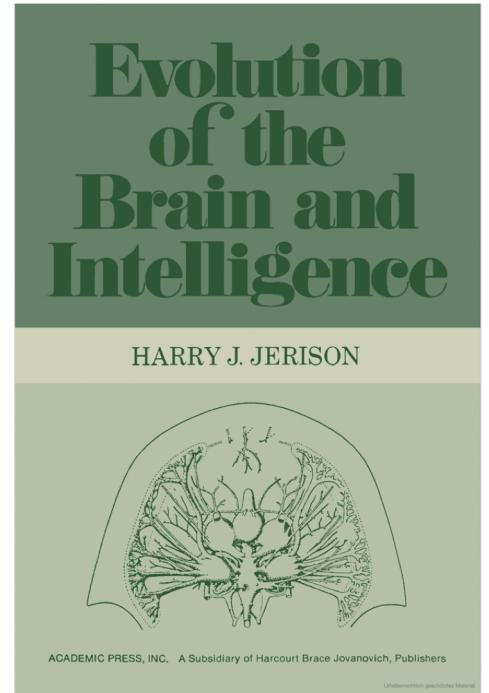
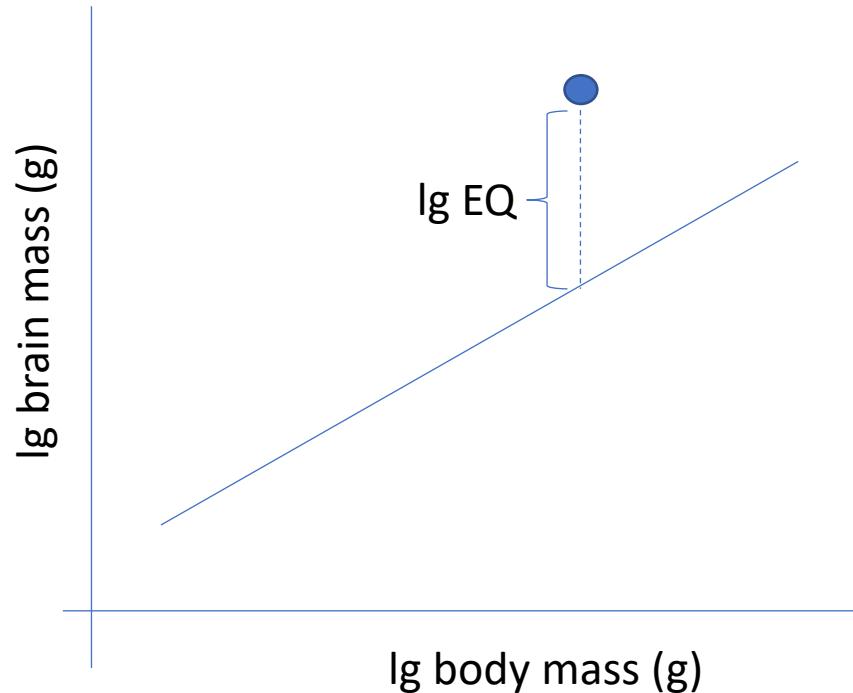
$$\lg(\text{Brain mass}) = a \lg(\text{Body mass}) + \lg k$$

Comparative studies (1880-1990)

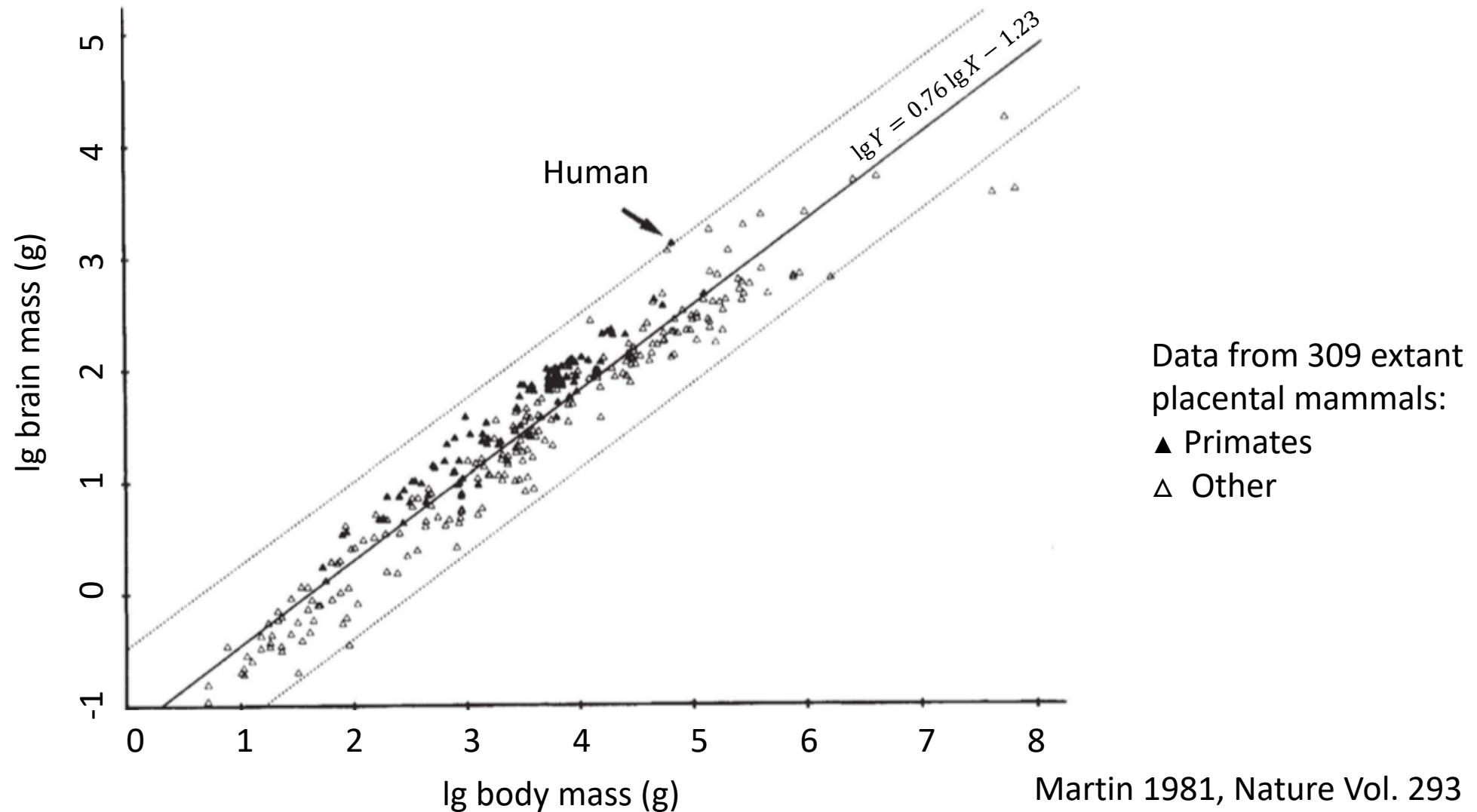


EQ, Jerison 1973

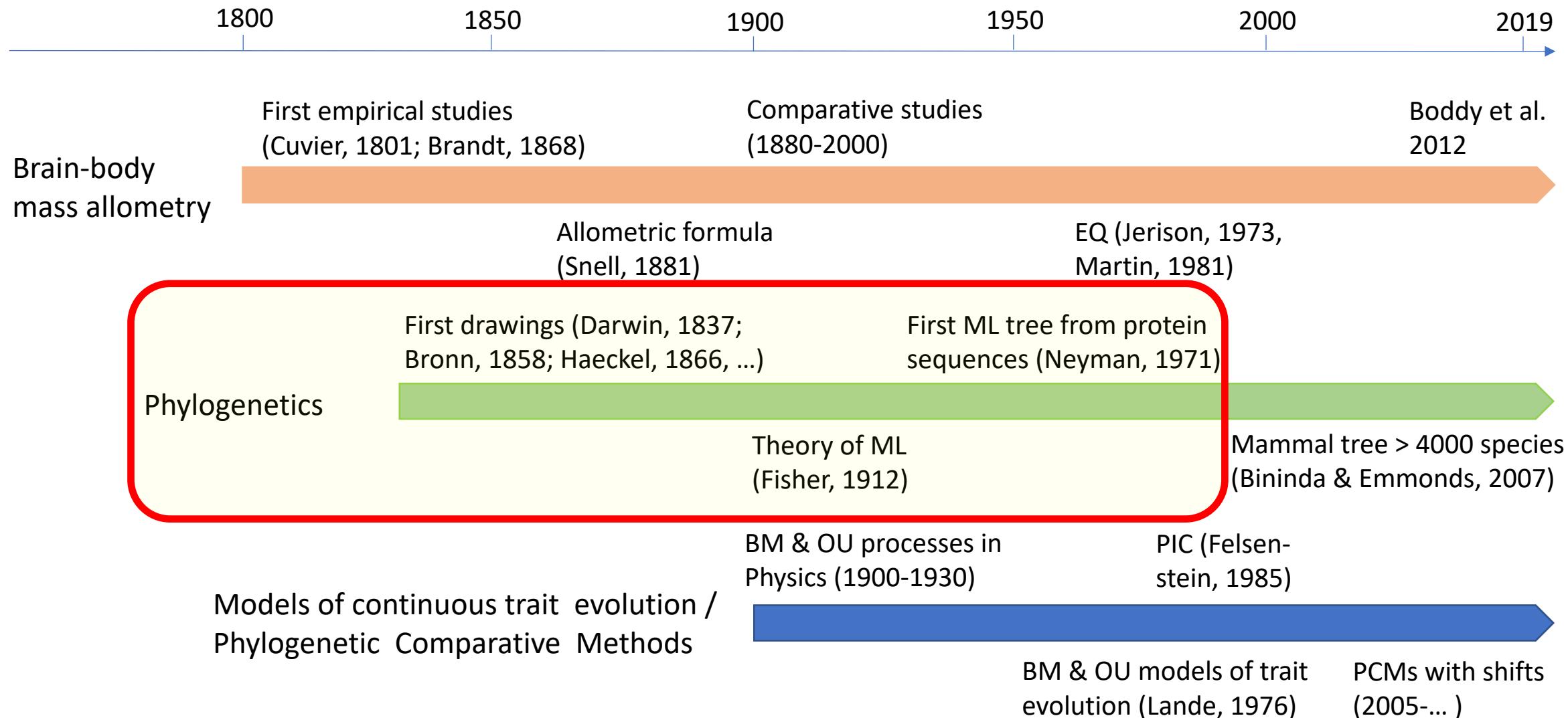
Definition “*Encephalization Quotient*”: the number of times a species’ brain mass exceeds the expectation relative to its body mass.



EQ → IQ, Martin 1981



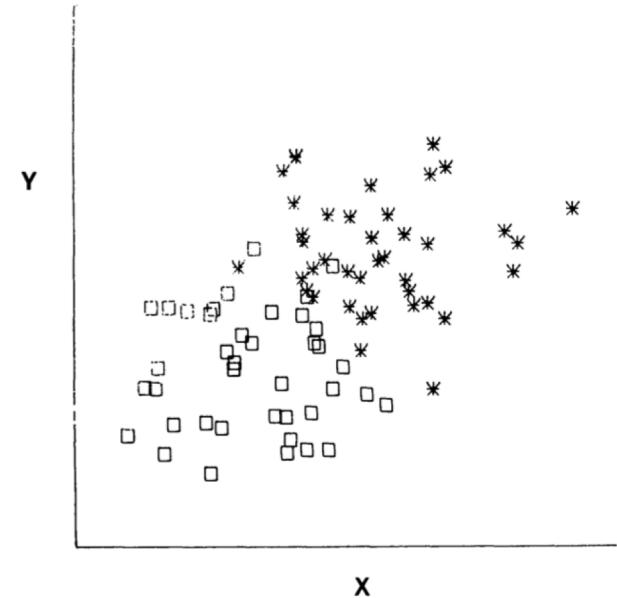
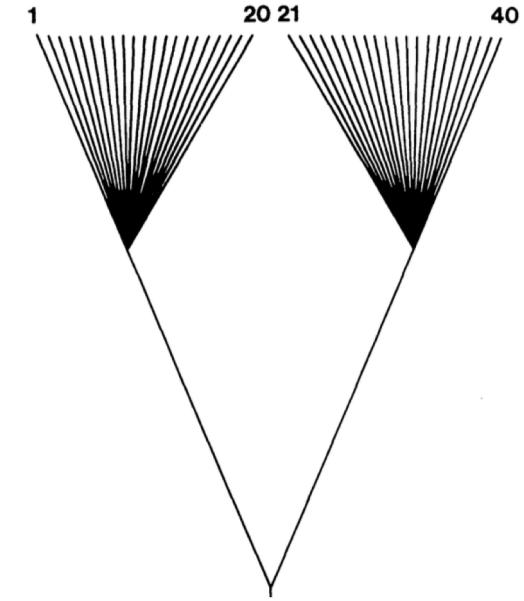
Timeline



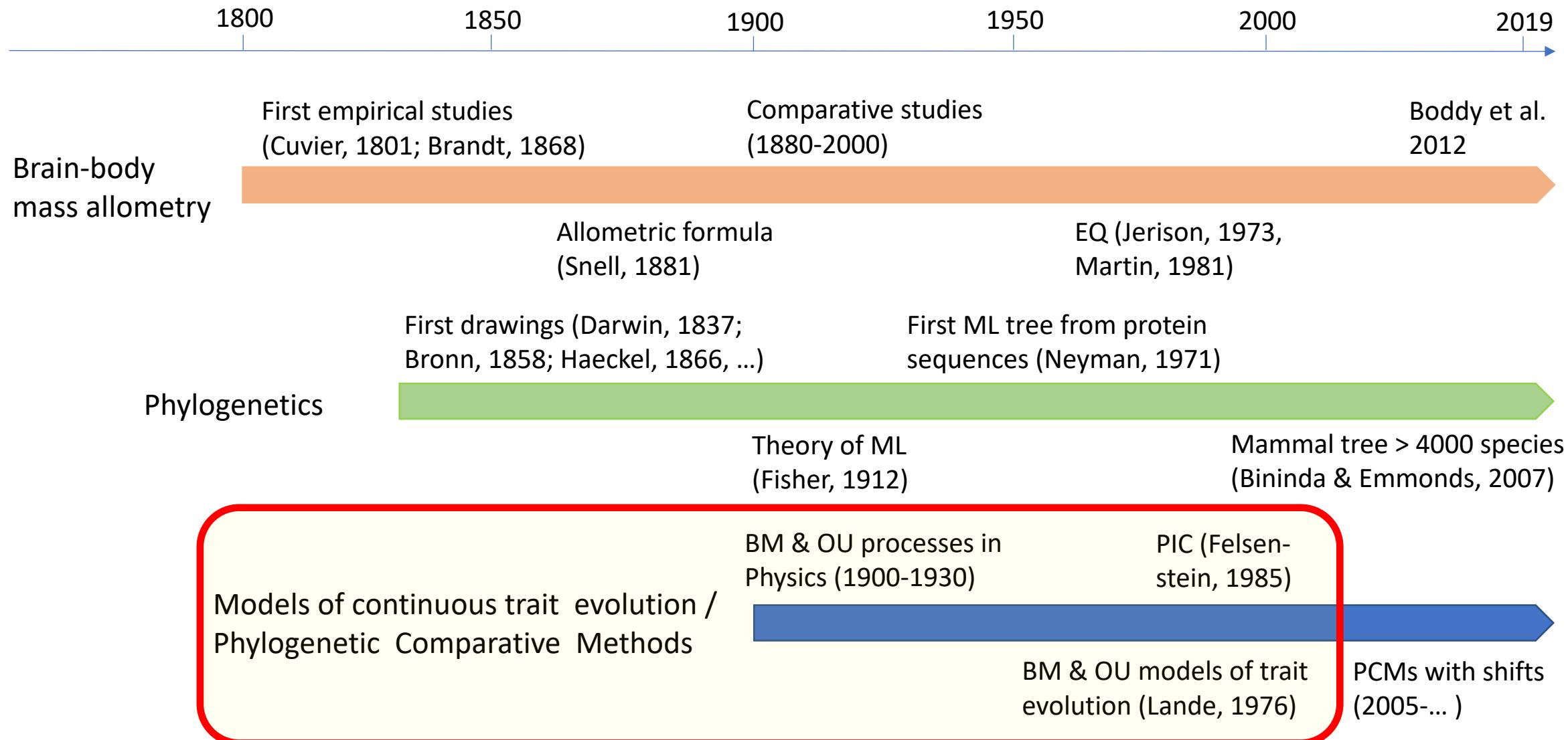
Felsenstein, 1985

- Possible bias in the regression slope due to ignoring the common history;
- A tree inferred from sequence data could be used to correct for this bias.

=> Phylogenetic Independent Contrasts (Felsenstein, 1985)

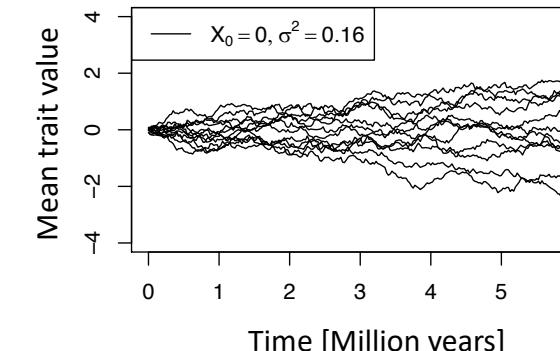
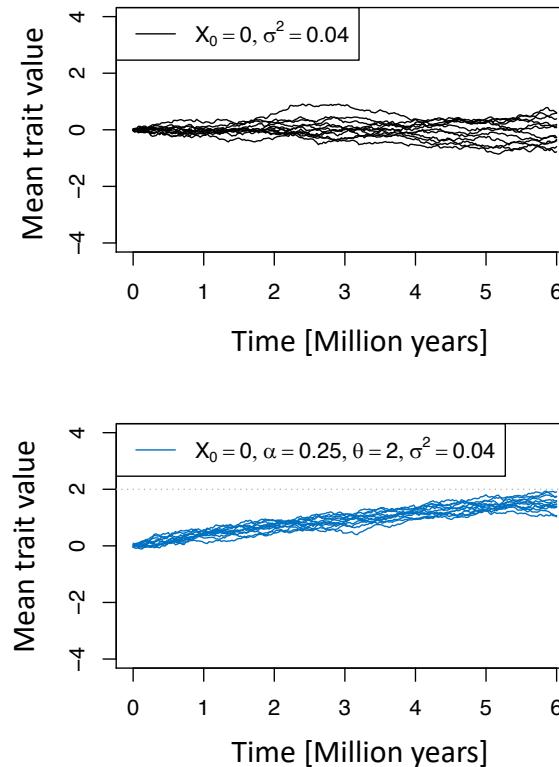


Timeline

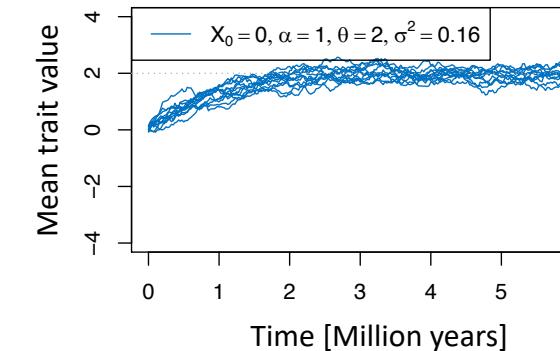
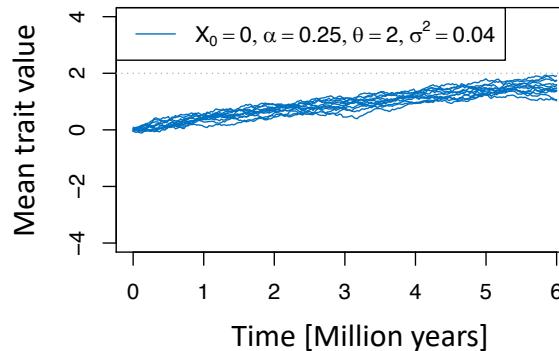


Models of trait evolution at the macro-evolutionary time scale

Random drift →
Brownian motion (BM)

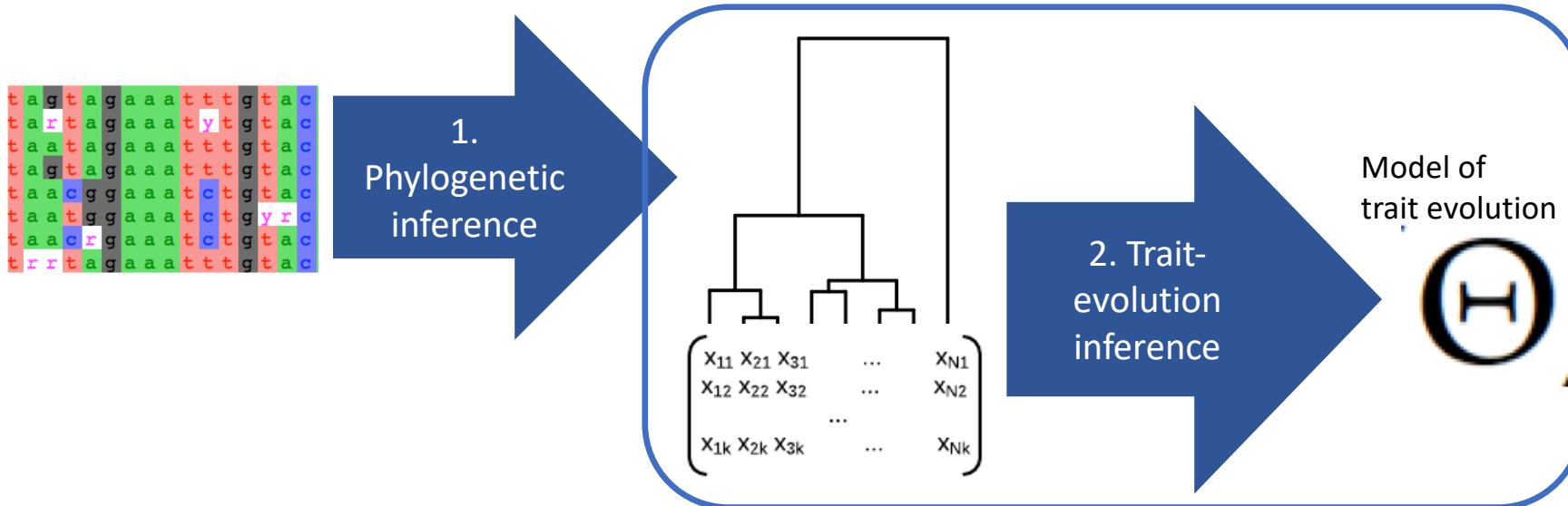


Random drift + stabilizing selection → Ornstein-Uhlenbeck (OU)



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

Phylogenetic comparative methods (PCMs)



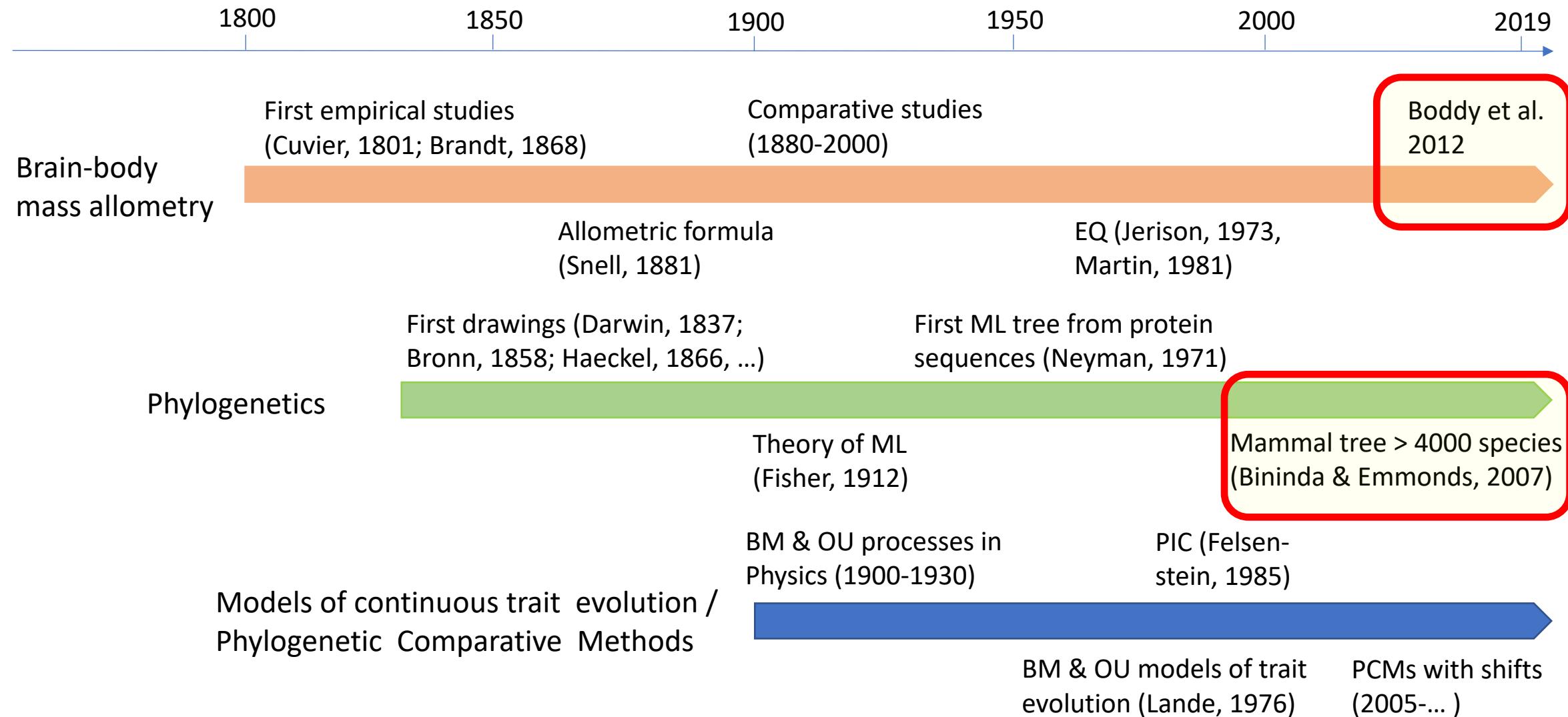
Phylogenetic comparative methods (PCMs) combine **two types of input**:

- A phylogenetic tree;
- A set of trait measurements for each tip in the tree.

in order to infer a **branching random process** that:

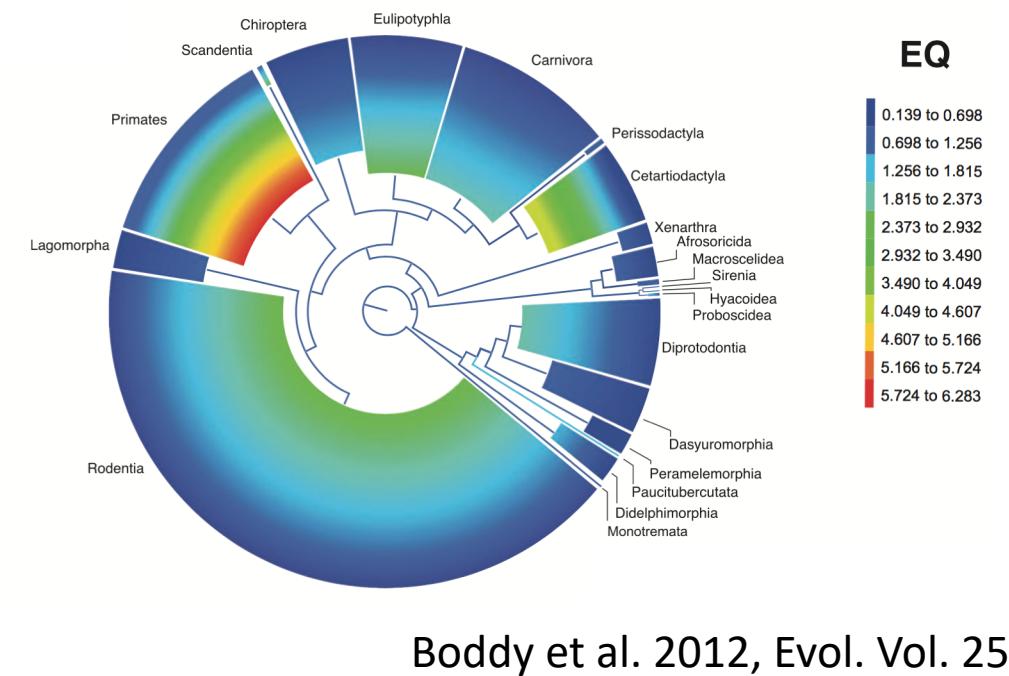
- models the (co)evolution of the traits along the phylogenetic tree;
- can be used to correct for the phylogenetic correlation.

Timeline



Boddy et al. 2012

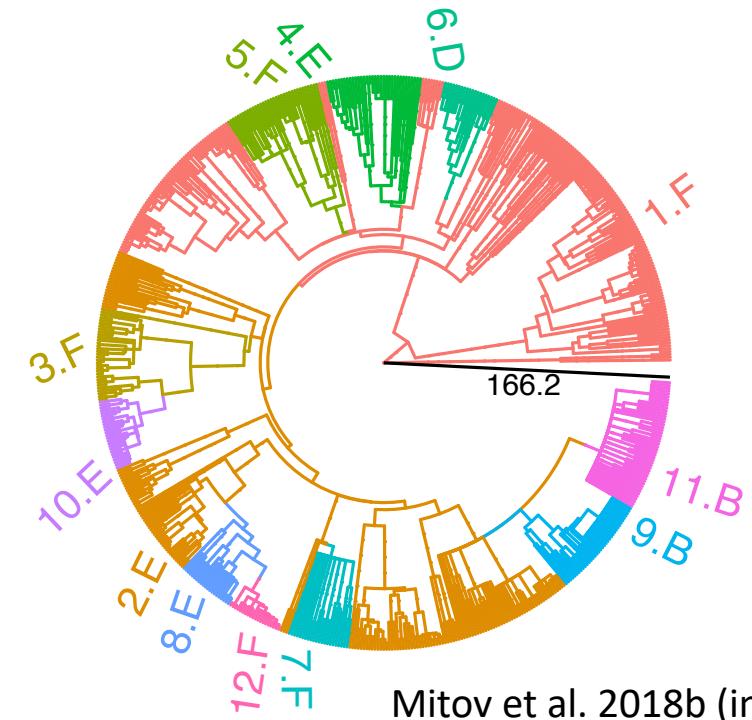
- Tree of 630 mammal species (subset of the tree from Bininda & Emmonds 2007);
- Estimation of EQ for *Homo Sapiens*
 - Standard regression (ignores phylogeny): EQ = 5.72
 - Brownian motion: EQ = 1.16
 - Ornstein-Uhlenbeck: EQ = 12.6
- *“To retain the same biological context as other encephalization studies, we chose to mainly report EQ values using the*



Is the assumption of a global BM or OU process adequate?

A Mixed Gaussian Phylogenetic Model (MGPM)

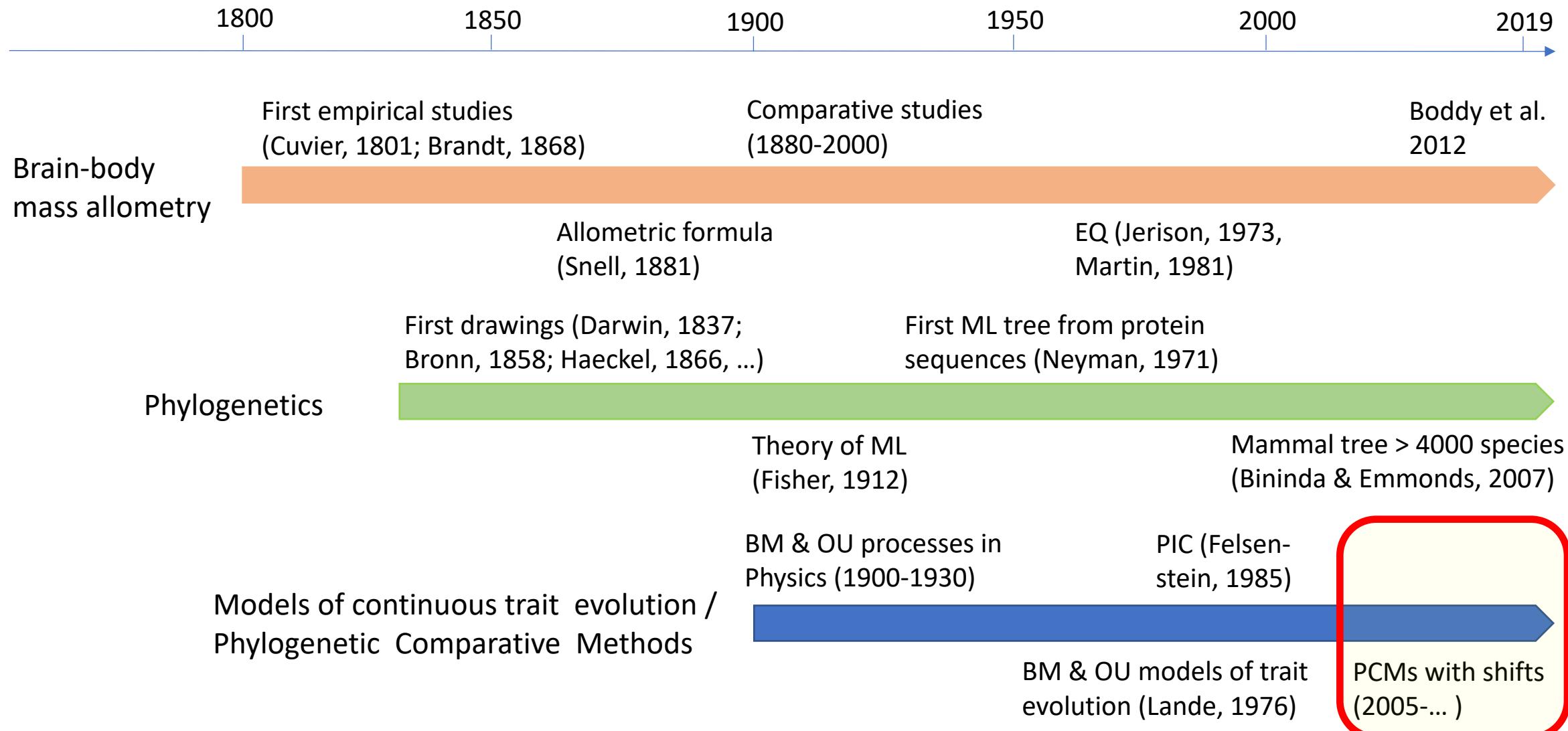
- A shift point configuration partitioning the three into regimes (colors 1, 2, ...);
- A mapping of model types (A, B, ..., F) to the regimes;
- Values for the model parameters.



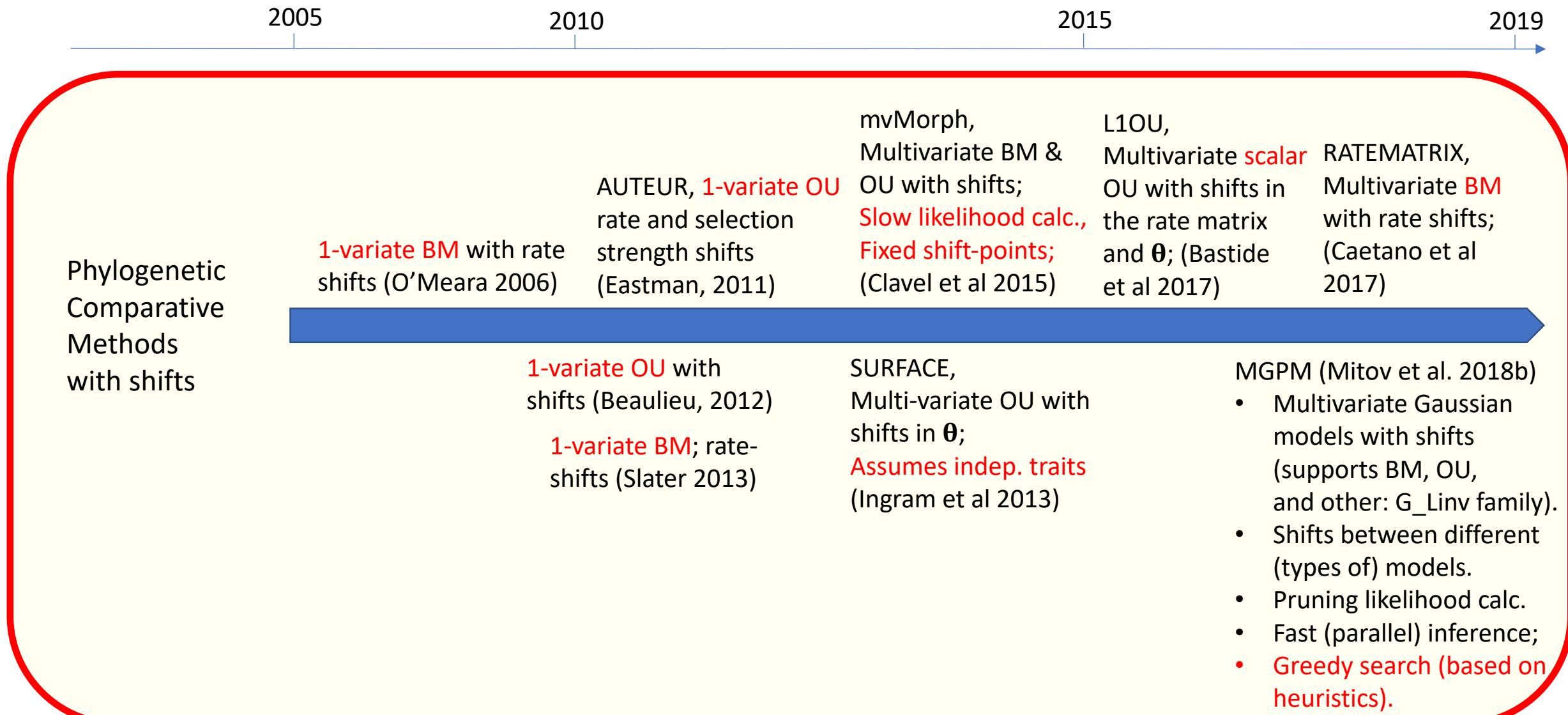
Mitov et al. 2018b (in review)

Goal: Infer an optimal MGPM selecting between BM and OU model types for subsets of species in the mammal tree.

Timeline



Timeline



An MGPM over 6 candidate model types

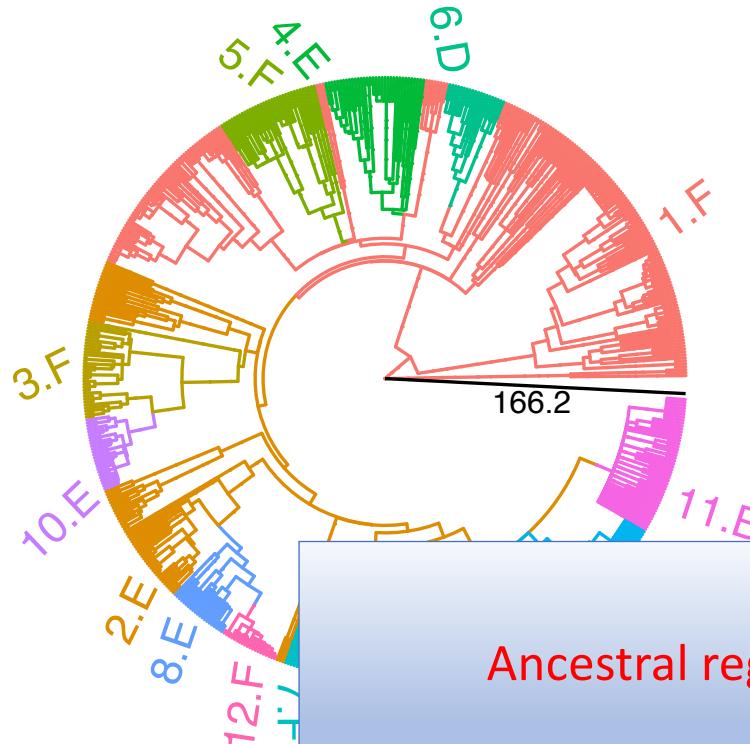
Model type	Parameters	Assumptions
A (BM)	$\Sigma = \begin{pmatrix} a & b \\ b & b \end{pmatrix}$	Brownian motion (BM), uncorrelated traits
B (BM)	$\Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	BM, correlated traits
C (OU)	$H = \begin{pmatrix} f & g \\ g & g \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & b \\ b & b \end{pmatrix}$	Ornstein-Uhlenbeck (OU), Uncorrelated traits
D (OU)	$H = \begin{pmatrix} f & g \\ g & g \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	OU, correlated traits, diagonal H
E (OU)	$H = \begin{pmatrix} f & h \\ h & g \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	OU, Correlated traits, symmetric H
F (OU)	$H = \begin{pmatrix} f & h \\ i & g \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	OU, correlated traits, asymmetric H

Model complexity increases 

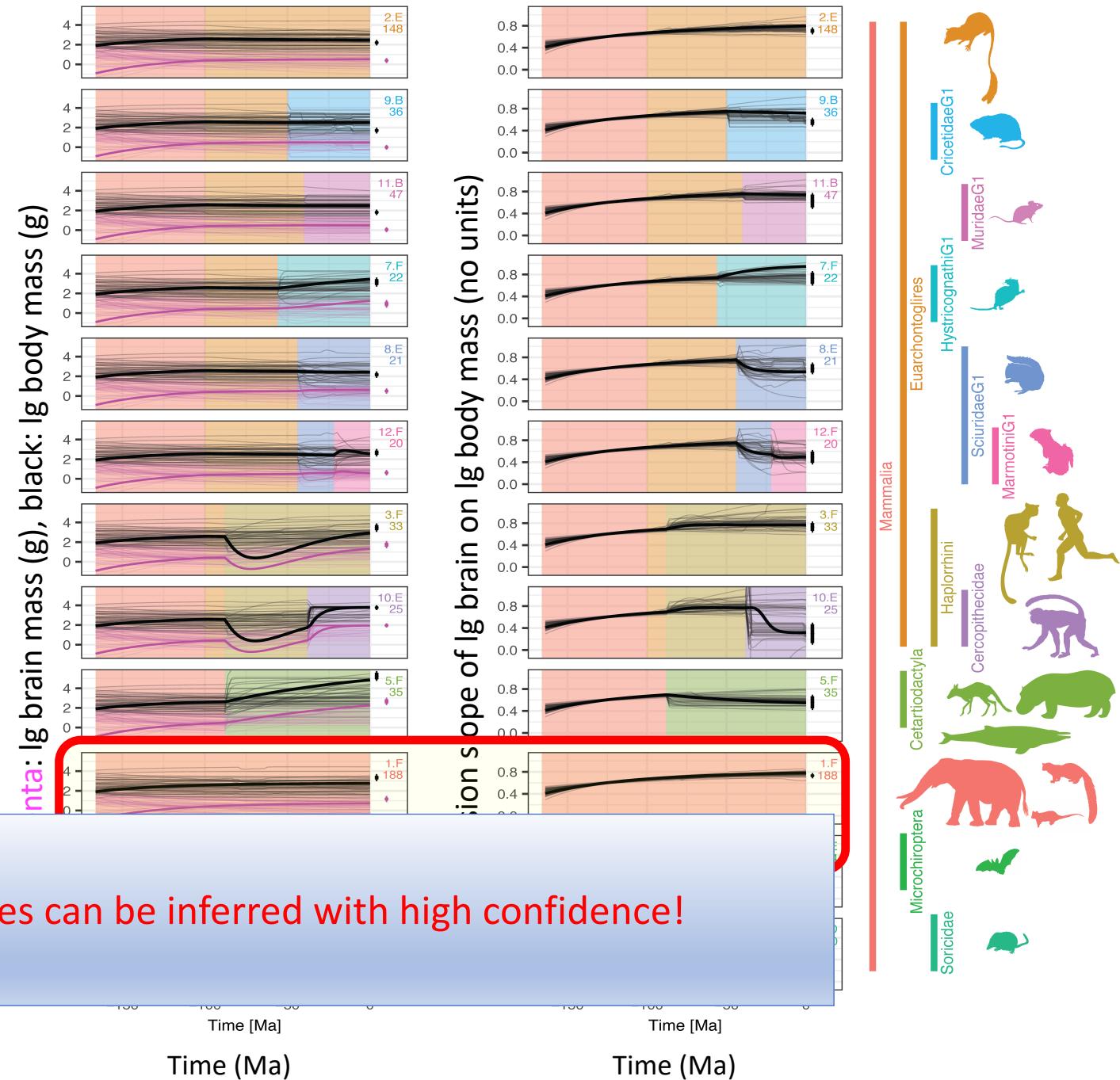
Σ : Evolutionary rate matrix parameter (BM and OU) H : Selection strength matrix parameter (OU)

$\vec{\theta}$: Long term optimum trait vector parameter (OU)

MGPM*: Parametric bootstrap



Mitov et al. 2018b (in review)



Conclusions

- The MGPM
 - provides a unified computationally efficient and extensible framework for a large family of models and for any type of tree;
 - enables a data-driven choice of groups for analysis.
- The mammal data has a strong statistical support for an MGPM over BM and OU models of evolution.
- The ancestral levels of brain-body mass allometry can be inferred with high confidence.

EVOLUTION 2019

June 21-25

Providence, Rhode Island

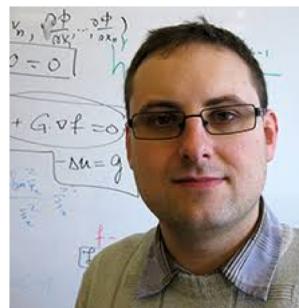


American Society of Naturalists
Society of Systematic Biologists
Society for the Study of Evolution

Prof. Dr. Tanja Stadler



Prof. Dr. Krzysztof Bartoszek



Georgios Asimomitis



The cEvo group at D-BSSE Basel



ETH zürich



Swiss Institute of
Bioinformatics



Source: 123rf.com

Used images' sources and licenses

- Capybara: By Charles J Sharp - Own work, from Sharp Photography, sharpphotography, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=4425429>
- Rhesus Macaque By Charles J Sharp - Own work, from Sharp Photography, sharpphotography.co.uk, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=65645730>
- Western Gorilla By Macinate - <https://www.flickr.com/photos/macinate/2445153914/>, CC BY 2.0, <https://commons.wikimedia.org/w/index.php?curid=4308155>
- African Bush Elephant By Bernard DUPONT from FRANCE - African Elephant (*Loxodonta africana*) bull, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=53675212>
- Vitruvian man: By Leonardo da Vinci - Leonardo Da Vinci - Photo from www.lucnix.be. 2007-09-08 (photograph).
- Otto Snell's foto: Dowloaded from the presentation <https://ecol586.files.wordpress.com/2018/05/alsmith-evolution-of-allometry.pdf>

Brain images: <https://www.quantamagazine.org/how-humans-evolved-supersize-brains-20151110/>

Used images' sources and licenses 2

- Cricetidae by Natasha Vitek: 81930c02-5f26-43f7-9c19-e9831e780e53, PDD 1.0;
- Hystricognathi by Zimices: 8d4497e7-a1b4-49e8-9e02-23db5f9aa37c, CCASAU 3.0;
- Muridae by Daniel Jaron: 92989e35-4e68-4a2d-b3a2-191ba9da671a, PDD 1.0;
- Haplorrhini (1) by Gareth Monger: 24230275-1bfa-4ec2-a946-ca1ececdf216, CCASAU 3.0;
- Haplorrhini (2, *Homo sapiens sapiens*) by T. Michael Keesey 2b4c32f6-99d0-43ba-9180-8013aa5bccd2, PDD 1.0;
- Cercopithecidae (uncredited), eccbb404-c99f-41f9-8785-01a7f57f1269, PDD 1.0;
- Marmotini by T. Michael Keesey 61440e34-7d24-4607-8479-2708ac45663f, PDD 1.0;
- Cetartiodactyla (1, Artiodactyla) by T. Michael Keesey 407f51d5-aa40-4e71-a5a7-7a6d6f328b5d, PDD 1.0;
- Cetartiodactyla (2, Cetacea) by Scott Hartman e68270c1-3091-4aee-92ae-51341a40e94a, PDD 1.0;
- Cetartiodactyla (3, *Hippopotamus amphibius*) by Jan A. Venter, Herbert H. T. Prins, David A. Balfour \& Rob Slotow (vectorized by T. Michael Keesey) , 6336f90c-8f02-48f5-94d1-1d85c0100473, CCASAU 3.0;
- Microchiroptera by Yan Wong, 18bfd2fc-f184-4c3a-b511-796aafcc70f6, PDD 1.0;
- Soricidae by Becky Barnes, 822c549b-b29b-47eb-9fe3-dc5bbb0abccb, PDD 1.0;
- Sciuridae by Catherine Yasuda, 5ebe5f2c-2407-4245-a8fe-397466bb06da, PDD 1.0;
- Feliformia (uncredited), ec56fa32-947b-4f0c-976b-c456132f2d6e, PDD 1.0;
- Diprotodontia by Michael Scroggie, f5592cab-cc61-4aab-b1dd-fba7cd2df7c9, PDD 1.0;
- Euarchonta by T. Michael Keesey (after Joseph Wolf), 88a07585-846a-405d-9195-c15c010e7443, PDD 1.0;
- Elephantidae by T. Michael Keesey, a15244a4-ecaa-4891-b870-31e5c8d9b5b3, PDD 1.0;

MGPM* vs other models

Model	#regimes	Maximum log-lik	ΔAIC (bigger is worse)
Global regime A (BM)	1	-541	1321
Global regime B (BM)	1	31	181
Global regime C (OU)	1	-541	1329
Global regime D (OU)	1	31	189
Global regime E (OU)	1	48	156
Global regime F (OU)	1	63	128
SURFACE OU	1	-541	1329
SCALAR OU	6	98	111
RATEMATRIX BM	9	116	73
MGPM* (A-F)	12	230	0

Software

PCMFit

<R-package>
MGPM inference

Mitov et al., 2018b (in review)

PCMBase

<R-package>

Fast likelihood calculation and trait simulation of multivariate phylogenetic
models

Mitov et al., 2018a (in review)

PCMBaseCpp

<R-package>

Fast likelihood calculation and trait simulation of multivariate phylogenetic
models

Mitov & Stadler, 2019, MEE

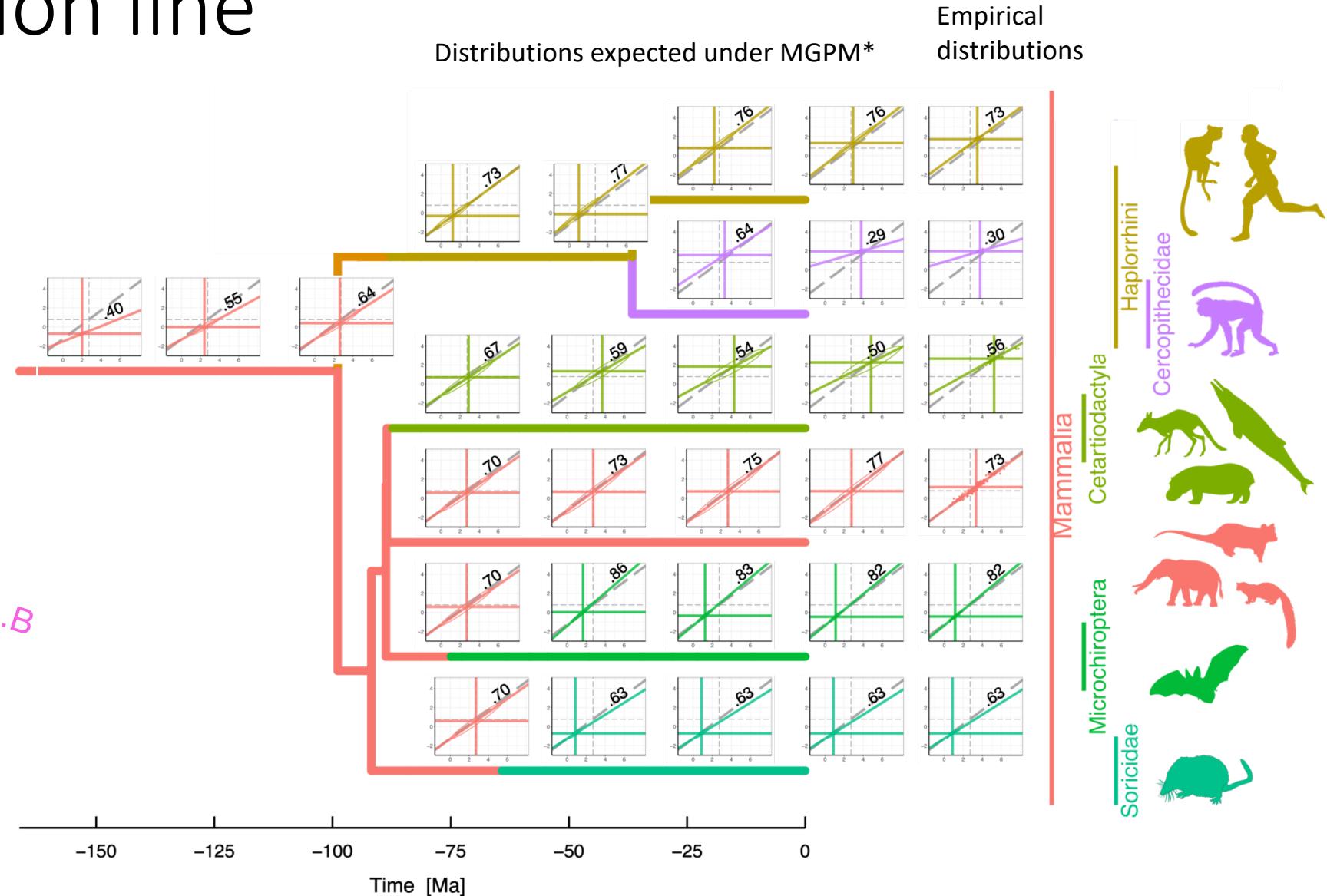
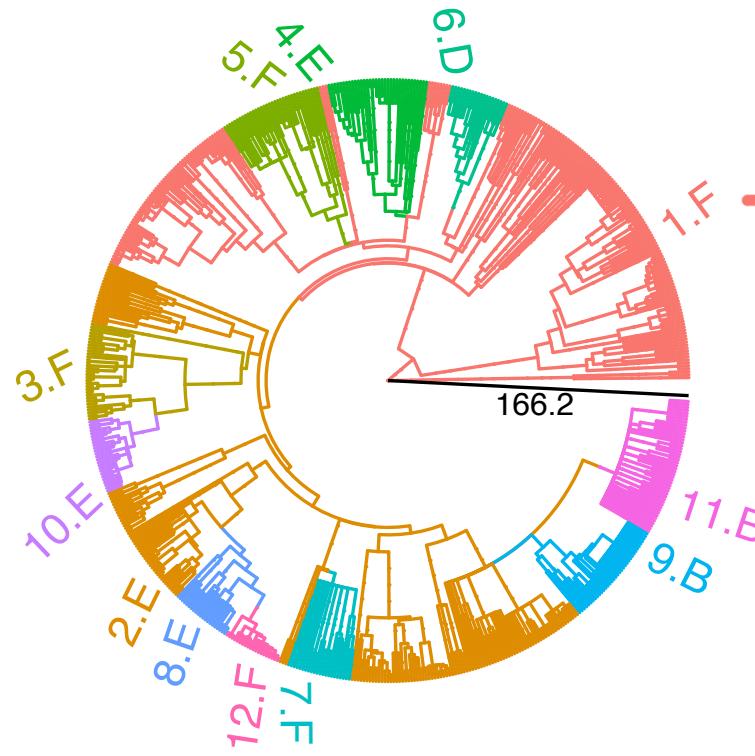
SPLITT

<C++ template library>

Serial and Parallel Lineage Traversal of Trees

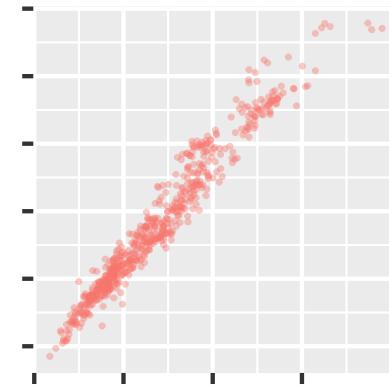
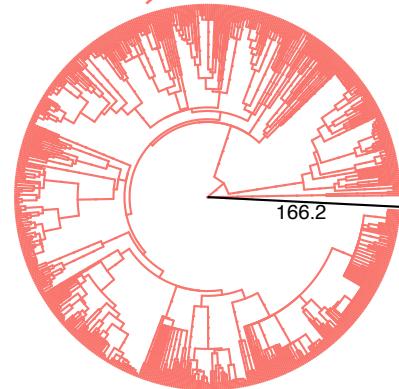
Mitov & Stadler, 2019, MEE

MGPM*: Inferred evolution of the brain-body mass regression line

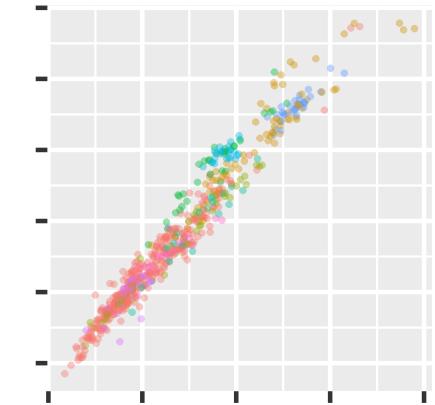
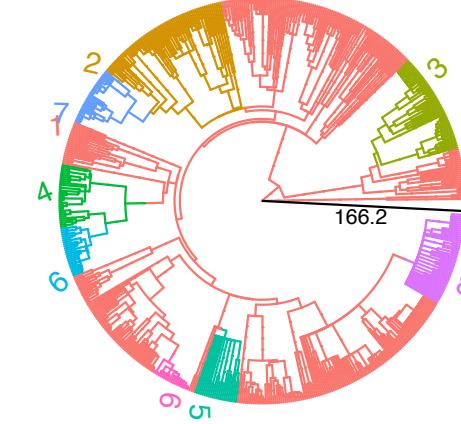


Shift point configurations of different models

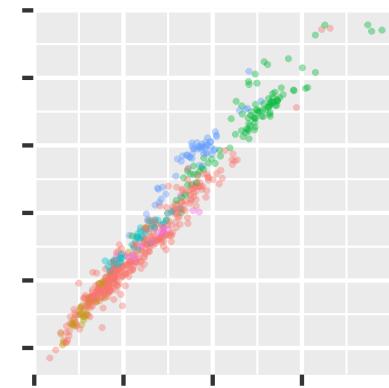
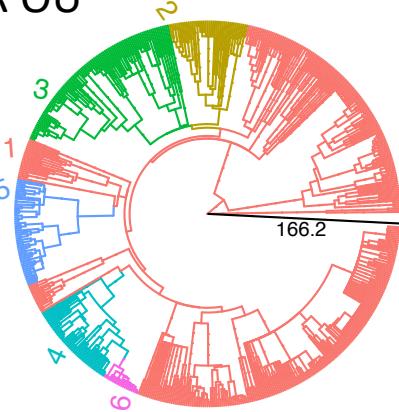
SURFACE OU



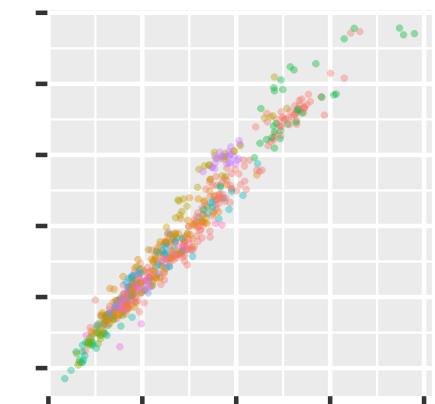
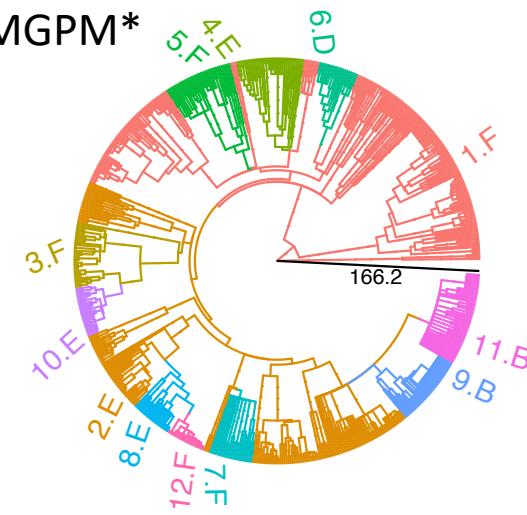
RATEMATRIX BM



SCALAR OU



MGPM*



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

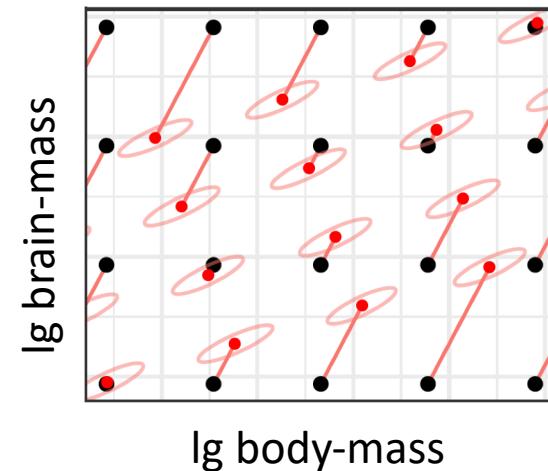
$$(2.a) \quad E[\vec{x}(t)|\vec{x}(s)] = \vec{\omega} + \Phi \vec{x}(s)$$

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

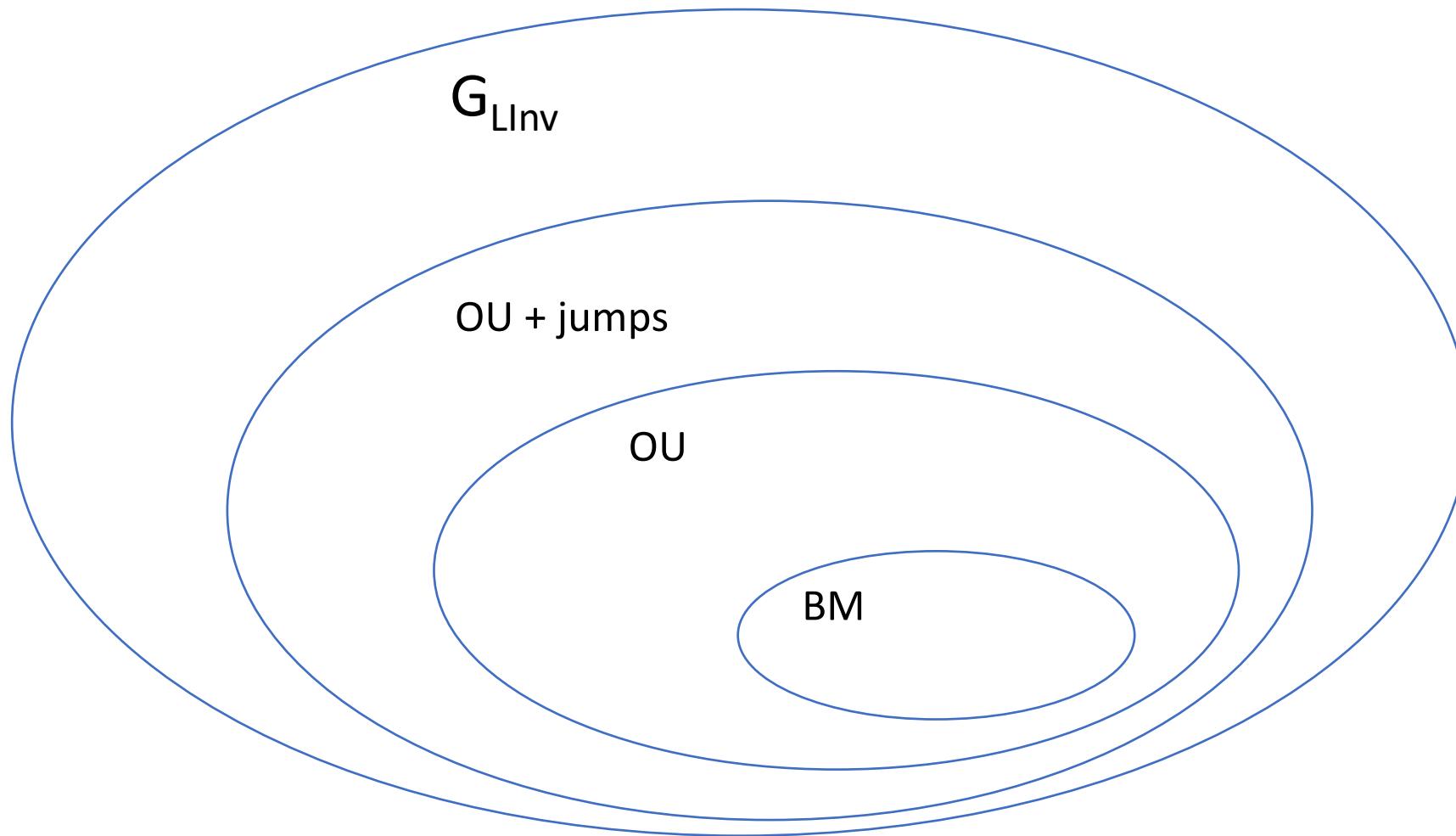
$$(2.b) \quad \text{Var}[\vec{x}(t)|\vec{x}(s)] = \mathbf{V}$$

(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 $\vec{x}(\cdot)$.



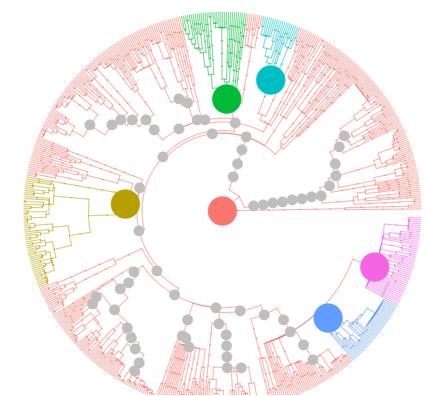
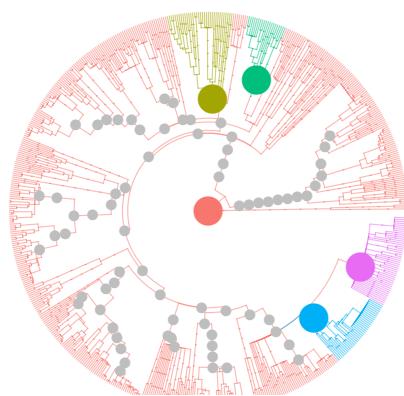
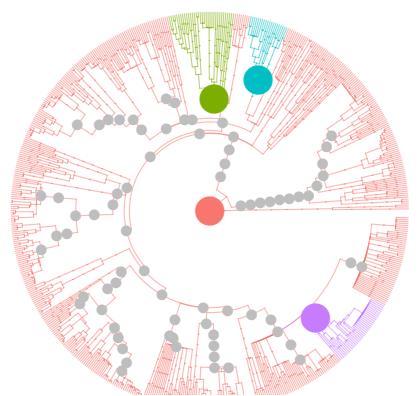
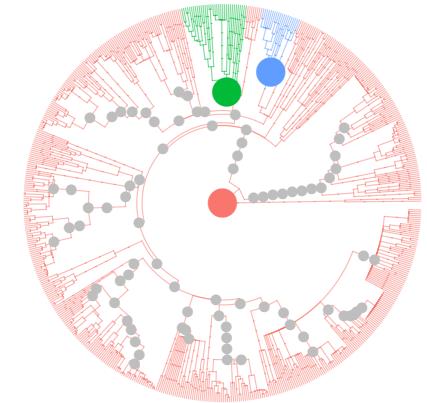
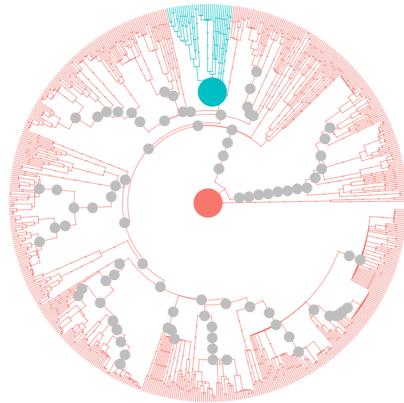
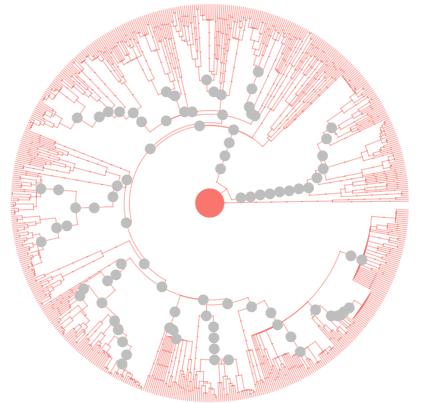
Approach: the G_{LInv} family of models



Models in G_{LInv} with shifts

Model (impl.)	Parameters	Shifts in	Assumptions
BM _B with shifts (R-package ratematrix)	$\Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	Σ	BM process with shifts, correlated traits
SURFACE OU (R-package surface)	$H = \begin{pmatrix} f & \\ & g \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & \\ & b \end{pmatrix}$	$\vec{\theta}$	Ornstein-Uhlenbeck, Uncorrelated traits
SCALAR OU (R-package l1OU)	$H = \begin{pmatrix} f & \\ & f \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	$\vec{\theta}, \Sigma$	Ornstein-Uhlenbeck, Correlated traits, Scalar H
MGPM (A-F) (R-package PCMBase + PCMFit)	$H = \begin{pmatrix} f & h \\ i & g \end{pmatrix}, \vec{\theta} = \begin{pmatrix} d \\ e \end{pmatrix}, \Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$	$H, \vec{\theta}, \Sigma$ & Model type	Shifts between BM _A -OU _F model types

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**).