

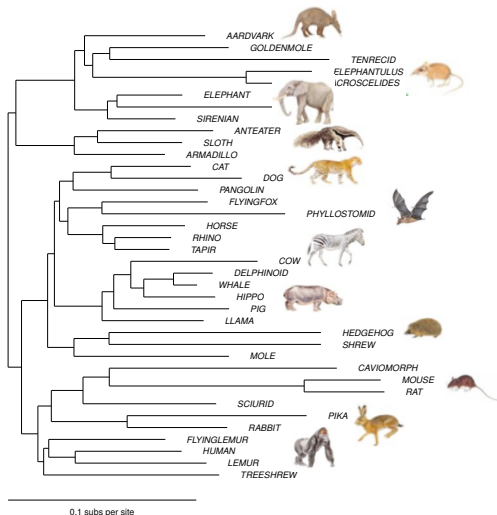
Patterns and processes

Bayesian models for macroevolutionary studies

Nicolas Lartillot, Raphael Poujol, Frederic Delsuc

October 2011

Variation of the substitution rate among lineages

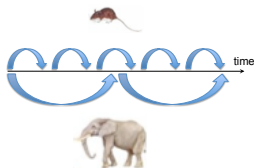


concatenation of 13 nuclear genes, 38 placentals

Variation of the substitution rate among lineages

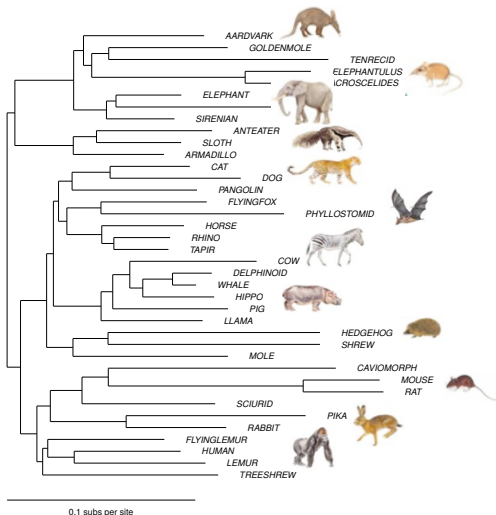
Possible causes

- generation-time effect



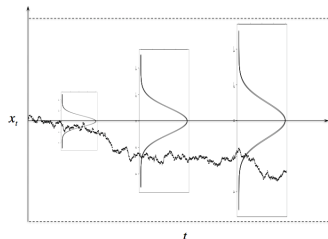
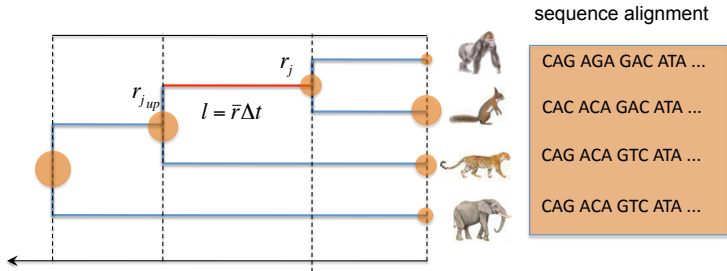
- metabolic rate effects
- selection for longevity

(reviewed in Lanfear et al, 2010)



concatenation of 13 nuclear genes, 38 placentals

Estimating divergence times: the relaxed clock model



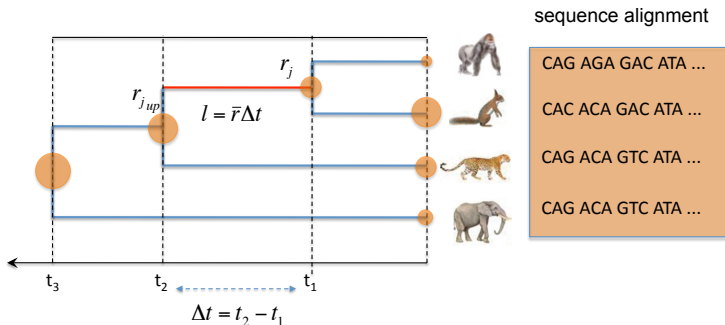
Brownian process

$$x_t = \ln r_t$$

$$x_t \sim N(x_0, \nu t)$$

(Thorne et al 1998, Lepage et al 2007, Rannala and Yang 2007)

Estimating divergence times: the relaxed clock model



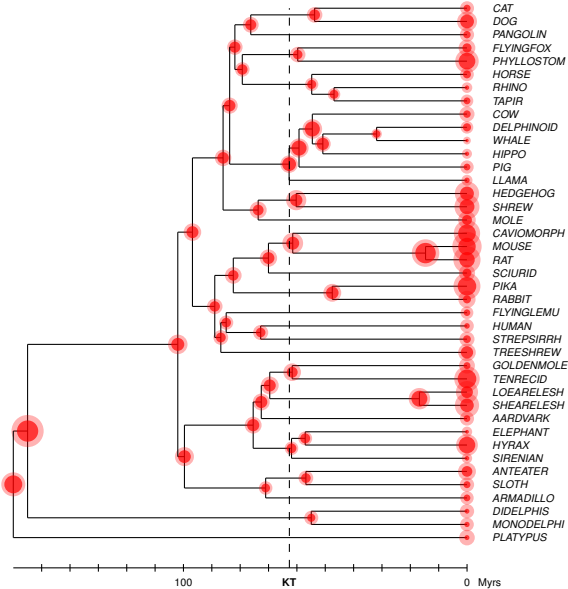
Sampling posterior density by MCMC

parameter vector: $\theta = (\nu, r, t, Q)$

$$p(D | r, t, Q) p(r | t, \nu) p(t) p(\nu) p(Q)$$

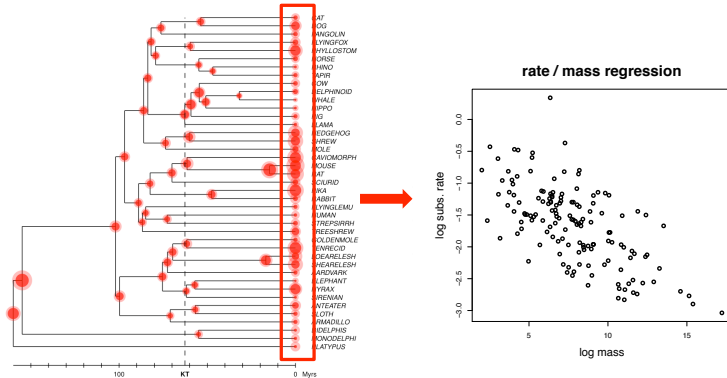
(Thorne et al 1998, Lepage et al 2007, Rannala and Yang 2007)

Divergence times and substitution rates



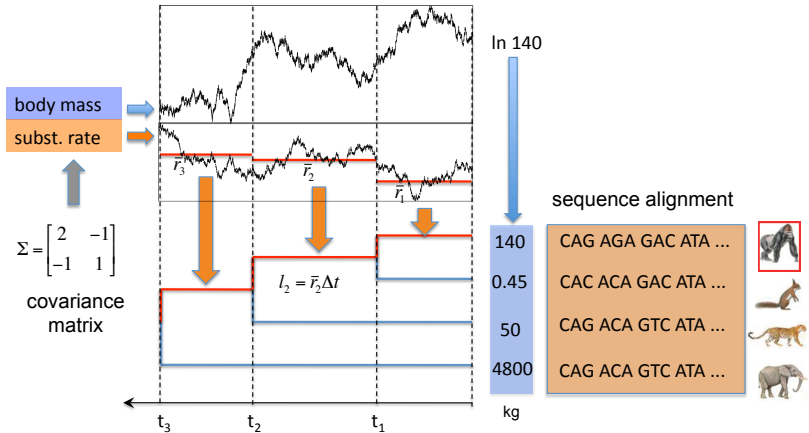
- carnivores
- chiropteres
- perissodactyls
- cetartiodactyls
- eulipotyphlans
- rodents
- lagomorphs
- primates
- afrotherians
- xenarthrans
- marsupials
- monotremes

Correlating rates and life-history traits



- correcting for phylogenetic inertia (independent contrasts)
- adaptation to trait/rate correlations (Welch 2011).
- sequential method: error propagation
- no feedback of rate variations on life-history evolution

Coupling life-history and substitution rate variations



Joint estimation (Bayesian MCMC)

divergence times, covariances, rates, and life-history evolution
(Lartillot and Poujol, 2011, Molecular Biology and Evolution)

Generalization

substitution parameters

- rate of synonymous substitution
- non-synonymous / synonymous ratio
- equilibrium GC

codon model (Goldman Yang, Muse Gaut 1994)

life-history traits

- sexual maturity
- mass
- maximum lifespan
- metabolic rate

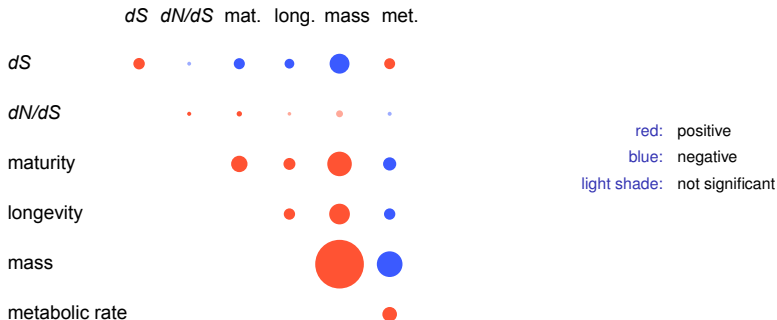
Priors

- uniform or birth death on divergence times
- fossil calibrations (Springer et al, 2003, Benton 2009)

Data

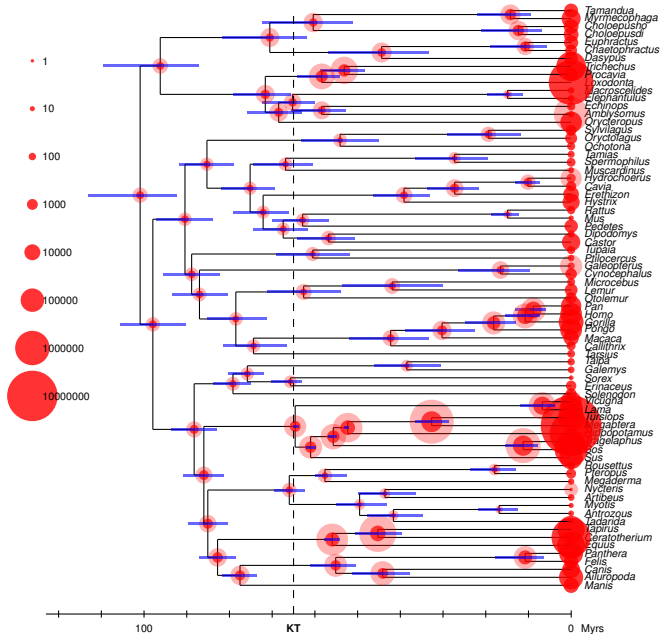
- nuclear data: 16 genes in 73 mammals
- nuclear data: 115 genes in 33 mammals
- mitochondrial data: cytochrome b in 100 mammals

1. Nuclear data: correlates of synonymous rate

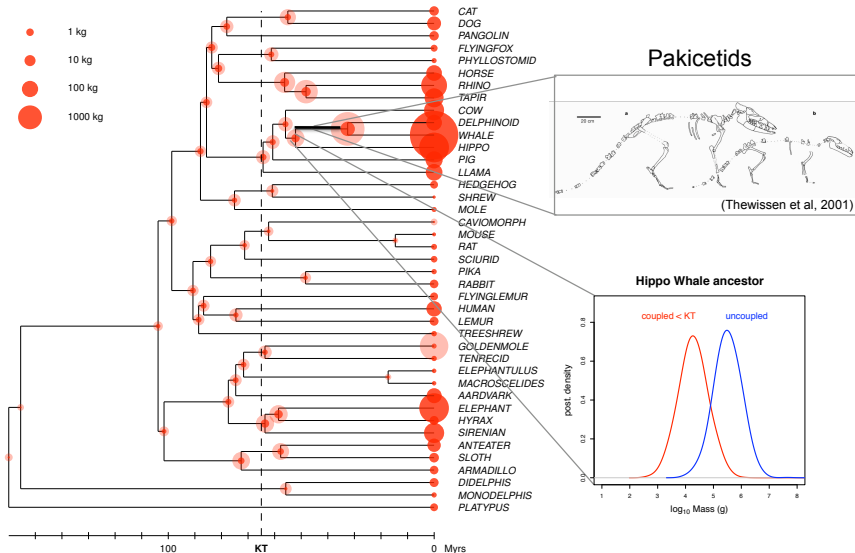


- strong correlations between life-history traits
- *dS* correlates negatively with body mass, gen. time and longevity
- R^2 : life-history variations explain $\sim 35\%$ of synonymous rate.
- partial correlations: longevity; generation time effect ?

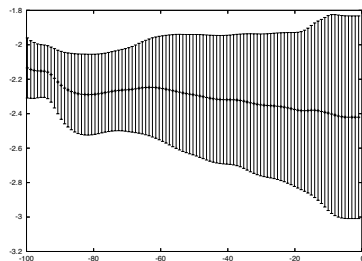
Inferring divergence times and body size evolution



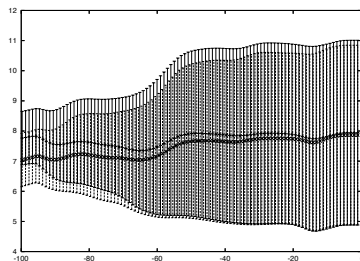
The evolution of body size



Systematic trends



substitution rate

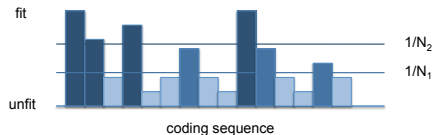
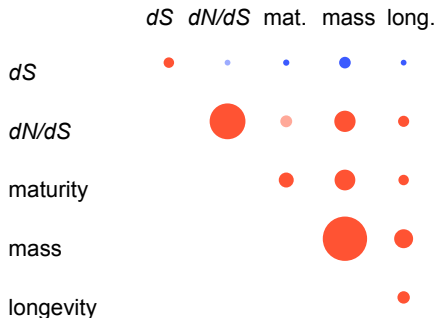


body mass

Cope's or Stanley's rule

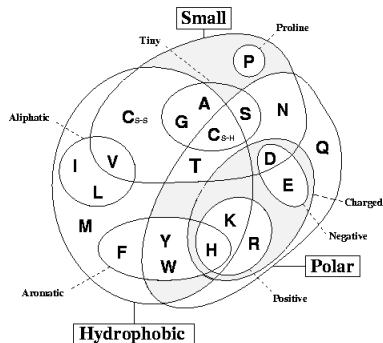
- intra-lineage drive towards larger body size
- more frequent extinction of large-bodied mammals
- needs to be explicitly modeled (directed Brownian motion)
- possible impact in estimated divergence times (Welch 2008)
- connections with mass-dependent extinction (FitzJohn, 2010).

2. Mitochondrial data: correlates of dN/dS



- positive correlation between dN/dS and body size
- compatible with a nearly-neutral interpretation
- via negative correlation body size population size (N)
- (Ohta, 1972, Kimura, 1979, Popadin, 2007)

Radical-conservative amino-acid replacement model



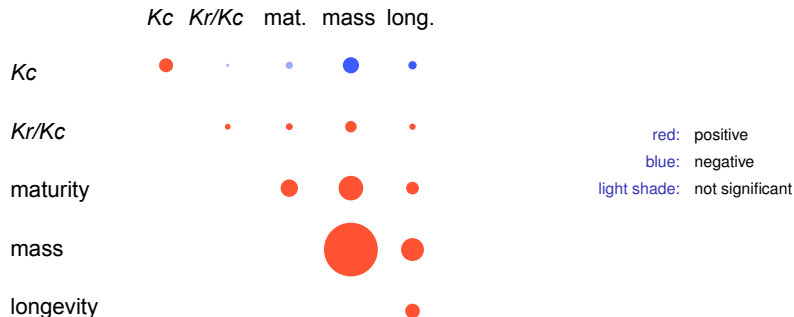
(adapted from Livingston and Barton, 1993)

$$\omega = K_r/K_c$$

$$Q_{ab} = R_{ab} \quad \text{if } a \rightarrow b \text{ conservative,}$$
$$Q_{ab} = R_{ab} \omega \quad \text{if } a \rightarrow b \text{ radical.}$$

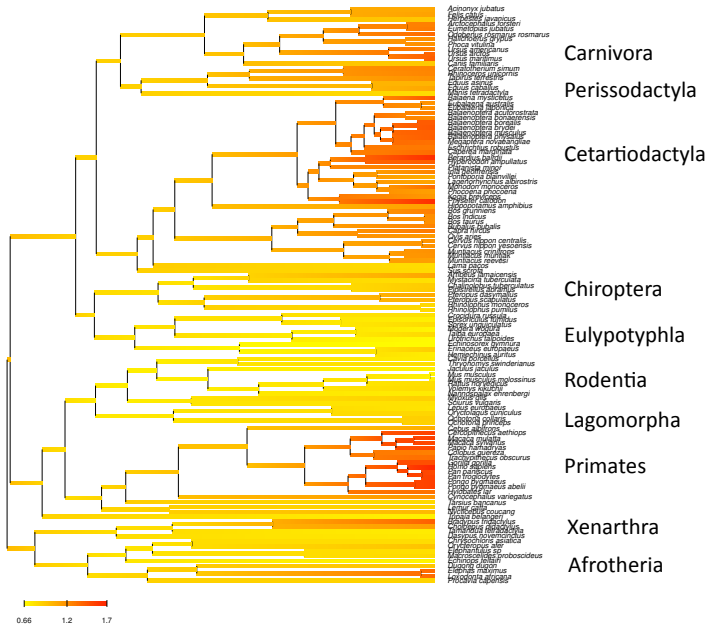
- R_{ab} : a general time reversible 20x20 process.
- conservative = conserving volume and/or polarity (and/or charge)

Mitochondrial data K_r/K_c (volume + polarity)



- positive correlation between K_r/K_c and body size
- similar to that observed for dN/dS (but higher R^2)
- charge: no significant effect
- polarity + volume : strongest correlation (highest R^2)

Reconstructed variations of K_r/K_c



Carnivora

Perissodactyla

Cetartiodactyla

Chiroptera

Eulypotyphla

Rodentia

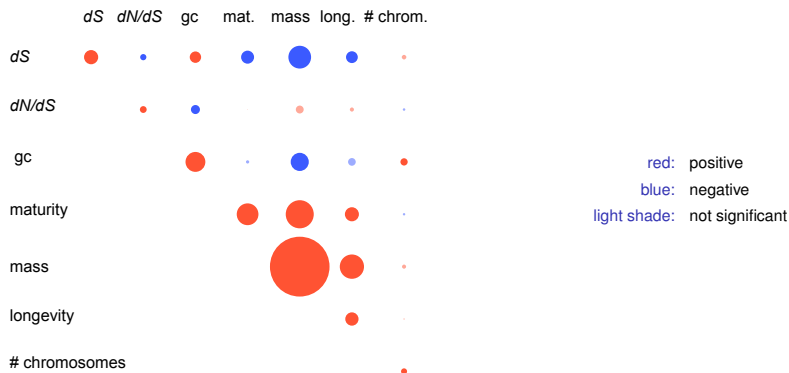
Lagomorpha

Primates

Xenarthra

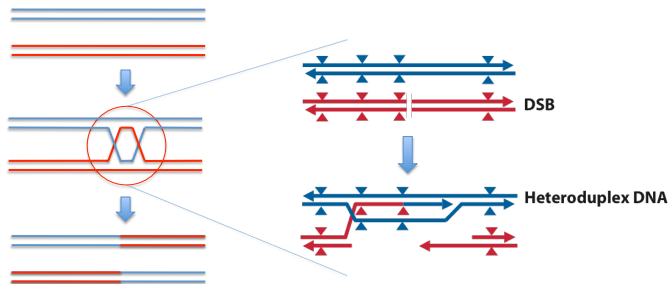
Afrotheria

3. Nuclear genes GC^*



- no correlation between dN/dS and body size
- negative correlation between GC^* and body size
- positive correlation between GC^* and number of chromosomes

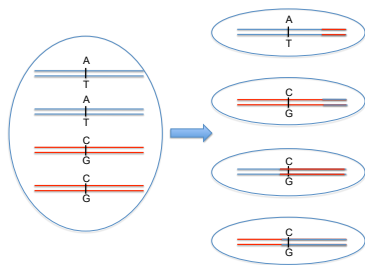
Biased gene conversion (BGC) during meiosis



adapted from Duret and Galtier 2009

- mismatches in heteroduplex sometimes repaired
- mutation biased towards AT
- repair pathways have evolved a bias towards GC

The population genetics of GC biased gene conversion



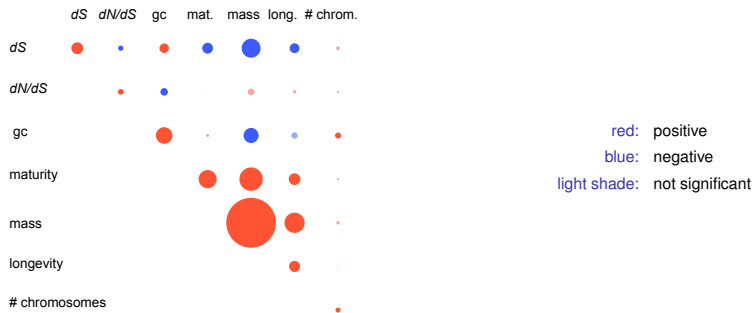
transmitted proportions

$$x_{GC} = \frac{1 + b}{2}$$

$$x_{AT} = \frac{1 - b}{2}$$

- mismatches in heteroduplex repaired towards GC
- overtransmission of GC compared to AT allele in heterozygotes
- equivalent to positive selection in favor of GC
- apparent selection coefficient: b (strength of the bias)
- b proportional to local recombination rate ($b = b_0 r$).

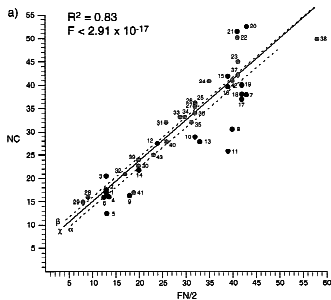
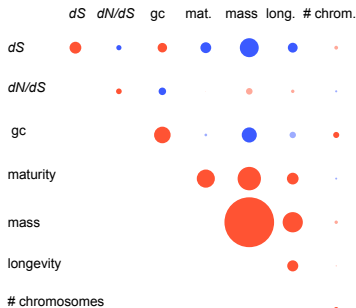
Biased gene conversion explains variations of GC^*



Negative correlation GC^* / body size

- larger animals = smaller population = less efficient selection
- also less efficient BGC (lower GC^*)

Biased gene conversion explains variations of GC^*

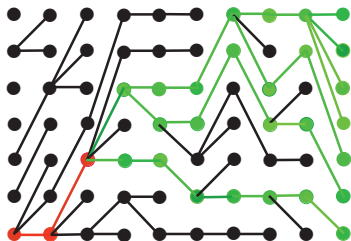


de Villena and Sapienza, 2001

Positive correlation GC^* / chromosome number

- conversion bias proportional to recombination rate
- ~ 1 recombination event per chromosome arm per meiosis
- recombination rate inversely proportional to chromosome size
- stronger gene conversion bias in more fragmented karyotypes

Population-genetics derivation



Fixation probability

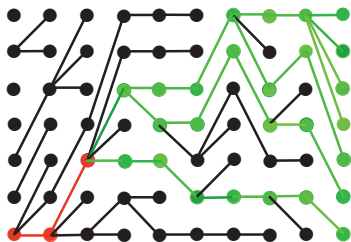
neutral case

$$p_0 = \frac{1}{2N}$$

general case

$$p = \frac{2s}{1 - e^{-4Ns}}$$

Population-genetics derivation



Fixation probability for biased gene conversion

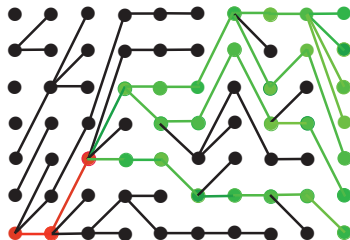
neutral case

$$p_0 = \frac{1}{2N}$$

general case

$$p = \frac{2b}{1 - e^{-4Nb}}$$

Population-genetics derivation

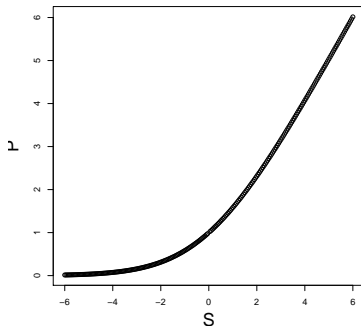


Scaled fixation probability P

$$P = p/p_0 = 2Np = \frac{4Nb}{1 - e^{-4Nb}} = \frac{B}{1 - e^{-B}}$$

with $B = 4Nb$ the *scaled* selection coefficient.

Scaled fixation probability as a function of $S = B = 4Nb$



- neutral case $S = 0$: $P = 1$
- deleterious $S < 0$: $P < 1$
- advantageous $S > 0$: $P > 1$

A mechanistic phylogenetic covariance model

$$\begin{pmatrix} - & \mu_{AC} & \mu_{AG} & \mu_{AT} \\ \mu_{CA} & - & \mu_{CG} & \mu_{CT} \\ \mu_{GA} & \mu_{GC} & - & \mu_{GT} \\ \mu_{TA} & \mu_{TC} & \mu_{TG} & - \end{pmatrix} + B \Rightarrow \begin{pmatrix} - & \mu_{AC} \frac{B}{1-e^{-B}} & \mu_{AG} \frac{B}{1-e^{-B}} & \mu_{AT} \\ \mu_{CA} \frac{-B}{1-e^{-B}} & - & \mu_{CG} & \mu_{CT} \frac{-B}{1-e^{-B}} \\ \mu_{GA} \frac{-B}{1-e^{-B}} & \mu_{GC} & - & \mu_{GT} \frac{-B}{1-e^{-B}} \\ \mu_{TA} & \mu_{TC} \frac{B}{1-e^{-B}} & \mu_{TG} \frac{B}{1-e^{-B}} & - \end{pmatrix}$$

Substitution rate (low mutation approx.)

Substitution rate = mutation rate x fixation probability

$$\begin{aligned} \rho &= 2N\mu p_{fix} = \mu 2Np_{fix} = \mu P \\ &= \mu \frac{B}{1-e^{-B}} \end{aligned}$$

A mechanistic phylogenetic covariance model

$$\begin{pmatrix} - & \mu_{AC} & \mu_{AG} & \mu_{AT} \\ \mu_{CA} & - & \mu_{CG} & \mu_{CT} \\ \mu_{GA} & \mu_{GC} & - & \mu_{GT} \\ \mu_{TA} & \mu_{TC} & \mu_{TG} & - \end{pmatrix} + B \implies \begin{pmatrix} - & \mu_{AC} \frac{B}{1-e^{-B}} & \mu_{AG} \frac{B}{1-e^{-B}} & \mu_{AT} \\ \mu_{CA} \frac{-B}{1-e^{-B}} & - & \mu_{CG} & \mu_{CT} \frac{-B}{1-e^{-B}} \\ \mu_{GA} \frac{-B}{1-e^{-B}} & \mu_{GC} & - & \mu_{GT} \frac{-B}{1-e^{-B}} \\ \mu_{TA} & \mu_{TC} \frac{B}{1-e^{-B}} & \mu_{TG} \frac{B}{1-e^{-B}} & - \end{pmatrix}$$

$$B = 4N_e b$$

- only 4-fold degenerate third codon positions
- modeling joint variations of B , body mass (M) and karyotype ($2n$)
- modeling variations among genes (local recombination rates)

Life-history and karyotypic covariates of BGC

Predicted allometric scaling of $B = 4N_e b$

$$N_e \sim M^{\gamma_M}, (\gamma_M < 0)$$

$$b = b_0 r$$

$$r \sim 2n$$

therefore, $B \sim M^{\gamma_M} 2n^{\gamma_n}, (\gamma_M < 0, \gamma_n = 1)$.

Estimated scaling coefficients and mutation bias ($\lambda = AT^*/GC^*$)

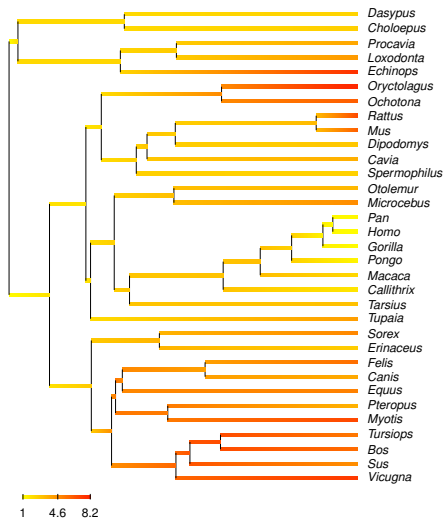
	γ_M	γ_n	λ
73 taxa 17 genes	-0.11** (-0.19, -0.03)	1.28** (0.54, 2.03)	1.38 (1.27, 1.50)
33 taxa 115 genes	-0.28* (-0.52, -0.01)	0.21 (-1.20, 1.56)	2.09 (2.04, 2.14)

A history of biased gene conversion in placentals

reconstruction of
 $B = 4N_e b$

$B < 1$: effectively neutral

$B > 1$: selective regime



- BGC above the nearly neutral threshold ($B > 1$) in some taxa
- significant force, deleterious effects (Galtier et al 2009, Berglund et al 2009)

Perspectives on biased gene conversion

BGC and recombination landscapes

- joint reconstruction of GC^* and genome rearrangements
- teasing out population size, recombination rate, and repair bias
- modeling overdispersion due to recombination hotspots turnover

understanding the (mal)adaptive value of BGC

- population genetics models (modifier theory)
- is there a selective regulation (buffering) of BGC intensity?

Conclusions

- integrative approach for correlating substitution patterns and quantitative traits
- can yield mechanistic insights about causes of molecular evolution
- potential source of information for reconstructing evolution of life-history, population size, karyotype, and genetic systems

Perspectives

- further into mechanistic modeling (dN/dS , BGC)
- including data about body size of fossil taxa
- modeling bursts (punctuated equilibria) and trends (Cope's rule)
- including diversification models (as priors on divergence times)
- modeling trait-dependent speciation and extinction
- modeling correlation with discrete characters

Acknowledgments

- Raphael Poujol
- Nicole Uwimana
- Frédéric Delsuc
- Nicolas Rodrigue
- Hervé Philippe
- many others...

Software availability (*coevol*)

- www.phylobayes.org

A mechanistic phylogenetic covariance model

$$\begin{pmatrix} - & \mu_{AC} & \mu_{AG} & \mu_{AT} \\ \mu_{CA} & - & \mu_{CG} & \mu_{CT} \\ \mu_{GA} & \mu_{GC} & - & \mu_{GT} \\ \mu_{TA} & \mu_{TC} & \mu_{TG} & - \end{pmatrix} + B \Rightarrow \begin{pmatrix} - & \mu_{AC} \frac{B}{1-e^{-B}} & \mu_{AG} \frac{B}{1-e^{-B}} & \mu_{AT} \\ \mu_{CA} \frac{-B}{1-e^{-B}} & - & \mu_{CG} & \mu_{CT} \frac{-B}{1-e^{-B}} \\ \mu_{GA} \frac{-B}{1-e^{-B}} & \mu_{GC} & - & \mu_{GT} \frac{-B}{1-e^{-B}} \\ \mu_{TA} & \mu_{TC} \frac{B}{1-e^{-B}} & \mu_{TG} \frac{B}{1-e^{-B}} & - \end{pmatrix}$$

$$B = 4N_e b$$

- only 4-fold degenerate third codon positions
- substitution rate = mutation rate \times fixation prob. (depends on B)
- modeling joint variations of B , body mass (M) and karyotype ($2n$)
- modeling variations among genes (local recombination rates)

Life-history and karyotypic covariates of BGC

Predicted allometric scaling of $B = 4N_e b$

$$N_e \sim M^{\gamma_M}, (\gamma_M < 0)$$

$$b = b_0 r$$

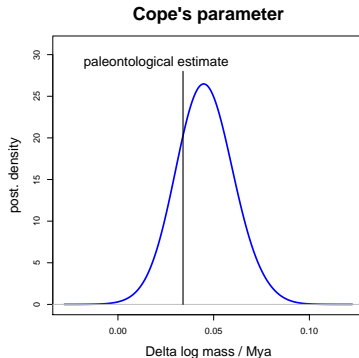
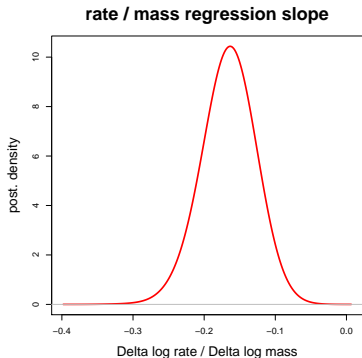
$$r \sim 2n$$

therefore, $B \sim M^{\gamma_M} 2n^{\gamma_n}, (\gamma_M < 0, \gamma_n = 1)$.

Estimated scaling coefficients and mutation bias ($\lambda = AT^*/GC^*$)

	γ_M	γ_n	λ
73 taxa 17 genes	-0.11** (-0.19, -0.03)	1.28** (0.54, 2.03)	1.38 (1.27, 1.50)
33 taxa 115 genes	-0.28* (-0.52, -0.01)	0.21 (-1.20, 1.56)	2.09 (2.04, 2.14)

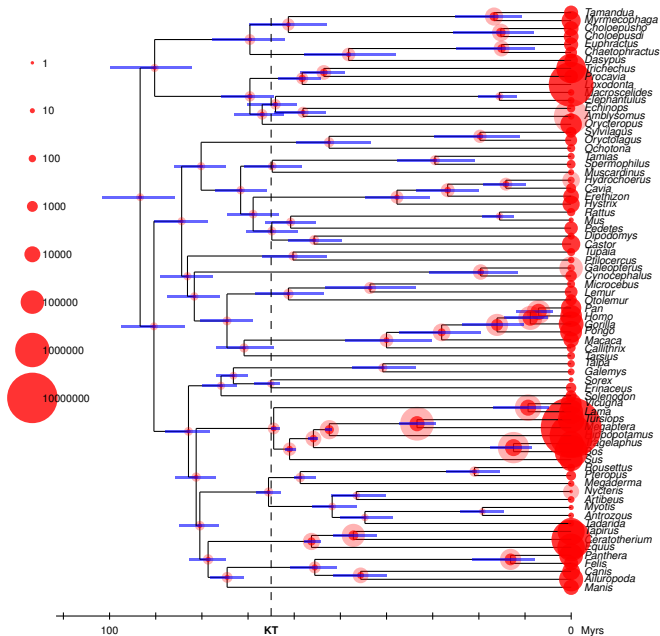
Estimated regression and Cope's trend parameters



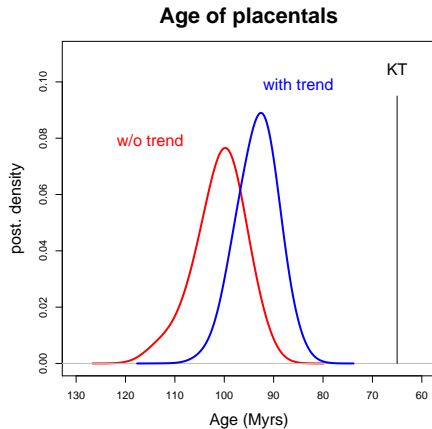
Fossil calibrations

- on dates (8 lower bounds, 5 upper bounds, Springer et al, 2003)
- on ancestral body sizes:
 - placental ancestor: $\ln m \sim N(4.5, 2)$ (5 g to 2 kg) (Alroy, 1996)
 - ancestor of carnivores (< 2kg), cetartios (< 2kg) (Kemp, 2006)
 - ancestor of primates (< 1kg), perissos (< 20kg) (Kemp, 2006)

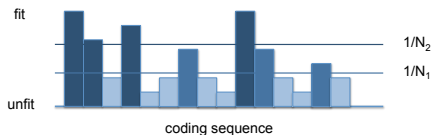
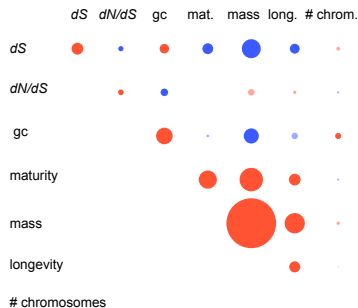
The evolution of body size (with trend)



Impact on divergence dates



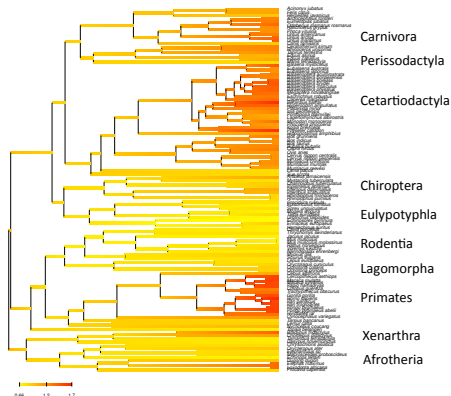
Absence of correlation between dN/dS and body-size



Possible cause

- interference between purifying selection and biased gene conversion
- biased gene conversion can promote fixation of deleterious alleles (Galtier et al 2009, Berglund et al 2009)

Alternative interpretations of K_r/K_c



- nearly neutral interpretation (population size effect)
- more adaptative substitutions in larger and long living animals
- adaptive and nearly-neutral substitutions differentially sensitive to mutation rate or to generation-time
- perspective: making correlations with mitochondrial polymorphism

The evolution of body size

