統計数理(2008)
第 56 巻 第 1 号 37–54

特集「分子進化と統計科学」
[研究詳解]

コドンモデルを用いた分岐年代のベイズ推定

徐 泰健[1] ・岸野 洋久[1] ・Jeffrey L. Thorne[2]

(受付 2007 年 9 月 18 日；改訂 2007 年 12 月 3 日)

要　　旨

進化研究において，種の分岐，環境による淘汰圧とそれに対する適応の推定は中心的な役割を担う．我々は 2004 年，分岐年代と同義・非同義置換速度の変動を同時推定する階層ベイズモデルを発表した．この方法は進化速度の一定性を仮定せず，これら 2 種類の分子進化速度に 2 変量幾何ブラウン運動の事前分布を導入する．本稿ではさらにこのモデルを拡張し，複数の遺伝子を解析する推定法を開発した．これにより，分岐年代の推定精度が向上するとともに，分子進化速度の変動の固有性と共通性，遺伝子間の相関を見ることで，背景にある因子に迫り，遺伝子間の共進化を検出する可能性が開かれた．ここでは，我々のベイズ推定法を説明し，哺乳類のミトコンドリアゲノムにコードされる 12 種類のタンパク質遺伝子を解析する．

キーワード： 分子時計，分岐年代，コドンモデル，適応進化，ベイズ法，階層モデル．

1. はじめに：分子時計と進化速度の変動

分子データに基づく分岐年代推定においては，一次近似として分子進化速度の一定性を仮定した分子時計(molecular clock)がしばしば用いられる(Kimura, 1983)．通常，進化速度は「単位時間当たり，座位当たりの塩基(またはアミノ酸)置換数」で表す．分子時計の理論的な根拠は「分子進化の中立説」(Kimura, 1983)に求められる．それは，「集団中に起きる突然変異の大部分は，適応度において集団中の他のものと同等である」とする．観測される分子進化は集団に定着した突然変異の履歴である．このため，大きさ N (2 倍体では遺伝子数 $2N$)の集団の分子進化速度 r は，遺伝子の突然変異率を ν，突然変異の集団への固定確率を f とすると，$r = 2N\nu f$ と表される．中立説の下では $f = \frac{1}{2N}$ となり，分子進化速度は集団の大きさに左右されず，遺伝子の突然変異率になる．従って，突然変異率が一定であるような状況では，分子時計が成立することになる．進化速度の一定性が認められる場合においては，化石データと一部の配列データから推定された進化速度を全系統樹に適用し，系統樹の全ての分岐点の年代を推定することができる．

しかし，しばしば進化速度は変動する．分子時計が成り立っているか否かは系統樹の形からある程度推測できる．図 1 は分子時計が成り立っている時の系統樹と，成り立っていない時の系統樹を模式的に対比させたものである．枝の長さは，その系統で生じた置換の数を表現している．進化速度が一定であれば，共通祖先からそれぞれの配列に至るまでの進化距離は等しく，系統樹の右端が揃う(図 1 (a))．これに対して，進化速度が変化する場合は，分岐点からの時間経過が同じでも，進化距離が異なり，共通祖先からそれぞれの配列に至るまでの進化距離に

[1] 東京大学 農学生命科学研究科：〒 113–8657 東京都文京区弥生 1–1–1
[2] Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27606, U.S.A.

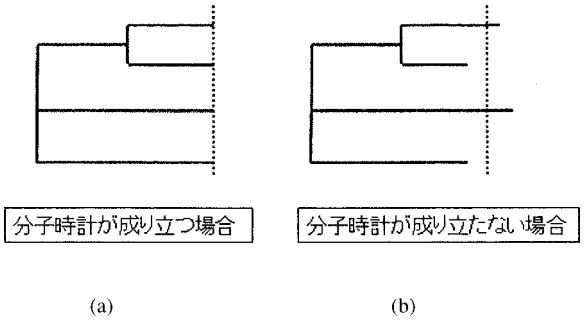


図 1. 分子時計が成立する場合と成立しない場合の系統樹の形(模式図).

違いが生じる(図 1 (b)). 分子進化は確率的な現象であるため, 解析する遺伝子配列の長さが有限である限り, 枝の長さも確率的に変動する. そのばらつきが分子時計を仮定した確率的な変動の範囲であるか, あるいはこれを超えて分子進化速度自体が変化しているかは, 尤度比検定により検証することができる(Felsenstein, 1981).

進化速度が変動する要因にはさまざまなものが考えられる. ここではタンパク質と直接関係するコード領域の分子進化に的を絞り, 同義・非同義 DNA 塩基置換速度の変動の背景にある要因を検討する. 同義置換(synonymous substitution)はアミノ酸を変えない DNA 塩基置換を, 非同義置換(nonsynonymous substitution)はアミノ酸を変える DNA 塩基置換を意味する. 例えば, フェニルアラニンをコードするコドンである「TTT」の 3 番目の T が C に変わって「TTC」になっても同じフェニルアラニンをコードするが(同義置換), 3 番目の T が A に変わって「TTA」になると, ロイシンをコードするようになる(非同義置換). 同義置換については, GC 含量の変化を通したアミノ酸翻訳効率の変化などは認められるが, アミノ酸を変えないため, 比較的自然淘汰の影響を受けないと信じられている. これに対して, 非同義置換はアミノ酸を変えるため, 自然淘汰の影響を受けやすい.

アミノ酸の変異の多くが適応度において優れたものである場合には, 非同義置換は同義置換より早く集団の中で定着するだろう. ウイルスのタンパク質における抗体との結合部位, 海中において雑多な種の精子と遭遇するアワビの卵子などでは, タンパク質の局所構造の変化が促される. 一方, こうした多様化圧を受けない場合には, アミノ酸の変異の多くは, タンパク質の構造の安定性や複合体の結合能を低下させるため, その突然変異を持つ個体は集団から取り除かれる傾向がある. そのため, 非同義置換速度は同義置換速度より遅くなる. 速度の違いは, タンパク質の性質や機能的な重要性に依存する. このように, 非同義置換の定着率は淘汰の影響を受けるため, 環境による淘汰圧や機能的な制約が変化すると, 非同義置換速度も変化する.

集団の大きさも非同義置換速度に影響を与える. 大きな集団では, 有利な突然変異は確実に定着し, 有害な突然変異は確実に排除される. これに対して, 集団の大きさが小さくなると, 突然変異の定着に不確実性が伴ってくる. 従って, 突然変異の多くが有害か弱有害である場合には, 分子進化速度は集団の大きさに対して負の相関を持つことになる. 非同義置換速度は世代の長さにも影響される. 仮に世代あたりの突然変異率に変化がなくても, 世代の長さが変化すると, 単位時間あたりの突然変異率は変化する.

同義置換・非同義置換の速度比を推定する方法はいくつか知られている(例えば, Goldman and Yang, 1994; Muse and Gaut, 1994; Miyata and Yasunaga, 1980; Nei and Gojobori, 1986 など). この方法は, 分岐年代の推定とは切り離して遺伝子配列にかかる多様化圧を測ることが

できる点で優れた指標であるが，他方，同義置換・非同義置換の速度比では汲み取れない重要な情報があることも事実である．速度比の変化には上に挙げたさまざまな要因が複合的に関っており，それらのうちどの要因が本質的に効いているか特定することは難しい．淘汰圧や集団の大きさの変化は，主として非同義置換速度の変化に関係し，同義置換速度にはあまり影響を与えない．これに対して，世代の長さの変化や世代あたりの突然変異率の変化は，同義置換速度と非同義置換速度の両方に影響する．さらに，ゲノム上のすべての遺伝子座に一様に影響する．Thorne et al. (1998) は分子進化速度の確率変動を事前分布に導入した階層ベイズモデルを開発し，分子進化速度の変化と分岐年代を同時推定する方法を提案した．これを同義置換・非同義置換 2 変量モデルに拡張することにより，こうした情報を漏れなく抽出することが可能となる (Seo et al., 2004)．以下に，我々のベイズ法の詳細と哺乳類ミトコンドリアデータ解析への適用例を紹介する．

2. コドンモデルと速度変化の階層ベイズモデル

2.1 コドンモデルと同義・非同義置換

コドンモデルは，コドン間の置換を Continuous time Markov モデルで表現したものである．3 つの塩基座位のそれぞれに 4 種類 DNA 塩基 (A，C，T，G) が可能であることから，コドンの総数は $4^3 = 64$ 種類になるが，標準コドンテーブル (Standard codon table) の場合，TAG，TAA，TGA の 3 種類のコドンはアミノ酸を指定しない終止コドンであるため，61 種類のコドン間の置換を考える．

我々の方法では，系統樹の枝ごとに非同義置換・同義置換の比が異なることを許す．系統樹の m 番目の枝で，コドン i からコドン j に変わる瞬間速度は次のように表現される．

$$q_{ij}^{(m)} = \begin{cases} 0 & \text{DNA 塩基置換が 2 回以上起きる} \\ u^{(m)}\pi_j & \text{同義置換かつトランスバージョン} \\ u^{(m)}\kappa\pi_j & \text{同義置換かつトランジション} \\ u^{(m)}\omega^{(m)}\pi_j & \text{非同義置換かつトランスバージョン} \\ u^{(m)}\omega^{(m)}\kappa\pi_j & \text{非同義置換かつトランジション} \end{cases} \tag{2.1}$$

ここで，π_j はコドン j の頻度を，κ はトランジション (transition; purine (A, G) 或いは pyrimidine (C，T) 同士の置換) のトランスバージョン (transversion; purine と pyrimidine の間の置換) に対する相対的な速度を，$\omega^{(m)}$ は m 番目の枝における非同義置換と同義置換の速度比を表す．$u^{(m)}$ は 1 単位時間の間起こるコドン置換が 1 になるように標準化するパラメーターであり，$\pi_j (1 \le j \le 61)$, κ, $\omega^{(m)}$ がデータから推定されると自動的に決まる．式 (2.1) で，i と j が等しい時は，$q_{ii}^{(m)} = -\sum_{j:j\neq i} q_{ij}^{(m)}$ にして，(61×61) のマルコフ転移行列 ($\boldsymbol{Q}^{(m)}$) を定義する．すると，進化距離 $b^{(m)}$ に対応する転移確率は $\boldsymbol{P}(b^{(m)}) = e^{b^{(m)}\boldsymbol{Q}^{(m)}}$ で計算できる．枝ごとに $\boldsymbol{P}(b^{(m)})$ を計算し，Felsenstein の pruning algorithm (Felsenstein, 1981) を用いて系統樹の尤度が計算できる．

簡単のため，A と B，C と D がそれぞれ姉妹群をつくる 4 本の配列 A，B，C，D について，尤度表現を確認しておく．A と B は共通祖先 E からそれぞれ進化距離 $b^{(A)}$，$b^{(B)}$ ほど離れており，配列 C と D は共通祖先 F からそれぞれ進化距離 $b^{(C)}$，$b^{(D)}$ ほど離れているとする．ここで，配列 E と F はともに未知であるが，それらの間の進化距離は $b^{(E)}$ であったとする．配列 A，B，C，D の第 h 番目の座位において，p，q，r，s コドンが観測された場合，この座位の尤度は

$$f\big(p,q,r,s|b^{(A)},b^{(B)},b^{(C)},b^{(D)},b^{(E)}\big)=\sum_{i=1}^{61}\pi_i p_{ip}(b^{(A)})p_{iq}(b^{(B)})\sum_{j=1}^{61}p_{ij}(b^{(E)})p_{jr}(b^{(C)})p_{js}(b^{(D)})$$

である．各座位の対数尤度を足し合わせることにより，配列全体の尤度を得る．

進化距離，$b^{(m)}$ はコドン座位当たり置換数であるが，

$$b_s^{(m)}=b^{(m)}\times\sum_i\pi_i\sum_{j:a_j=a_i}q_{ij}^{(m)}$$
$$b_n^{(m)}=b^{(m)}\times\sum_{i_j}\pi_i\sum_{j:a_i\neq a_j}q_{ij}^{(m)}$$

と，コドン座位当たりの同義置換数 ($b_s^{(m)}$) と非同義置換数 ($b_n^{(m)}$) に分離することができる．m 番目の枝の時間間隔，平均同義置換速度，平均非同義置換速度をそれぞれ，$t^{(m)}$，$\bar{r}_s^{(m)}$，$\bar{r}_n^{(m)}$ にすると，$b_t^{(m)}=\bar{r}_s^{(m)}\times t^{(m)}$，$b_n^{(m)}=\bar{r}_n^{(m)}\times t^{(m)}$ が成り立つ．分岐年代と同義・非同義置換速度の同時推定を可能にするために，$t^{(m)}$，$\bar{r}_s^{(m)}$，$\bar{r}_n^{(m)}$ に事前分布を導入する．

2.2 進化速度の 2 変量確率変動モデル

Thorne et al. の速度変動モデル (Thorne et al., 1998) を拡張し，同義置換速度・非同義置換速度の 2 変量モデルを考える．すなわち，j 番目のノードにおける非同義置換速度 (r_n^j) と同義置換速度 (r_s^j) に対して，親ノード $a(j)$ における状態で条件付けしたときの分布が

$$\begin{pmatrix}\log r_n^j\\ \log r_s^j\end{pmatrix}\Bigg|\begin{pmatrix}\log r_n^{a(j)}\\ \log r_s^{a(j)}\end{pmatrix}\sim N\left(\begin{pmatrix}\log r_n^{a(j)}\\ \log r_s^{a(j)}\end{pmatrix},\ \begin{pmatrix}\nu_n t^j & 0\\ 0 & \nu_s t^j\end{pmatrix}\right)\tag{2.2}$$

と，ブラウン運動するとする．時間 t^j は，ノード j とその親ノード $a(j)$ の間の時間間隔である．超パラメーター ν_n，ν_s は，非同義置換速度，同義置換速度の単位時間当たり変動係数を表している．

2.3 分岐年代・進化速度の事前分布と事後分布

分岐年代の事前分布 ($\mathrm{P}(\boldsymbol{T})$) として，ディリクレ分布を仮定する (Thorne et al., 1998)．化石データによる分岐年代の上限・下限に関する情報 ($\boldsymbol{C}$) も，$\mathrm{P}(\boldsymbol{T})$ のサポートを制限することにより容易に取り入れることができる ($\mathrm{P}(\boldsymbol{T}|\boldsymbol{C})$)．また，系統樹の根元での非同義置換速度の事前分布 ($\mathrm{P}(r_n^{(root)})$)，同義置換速度の事前分布 ($\mathrm{P}(r_s^{(root)})$) として，指数分布を仮定する．さらに，超パラメータ ν_n，ν_s には超事前分布として指数分布を導入する．$\mathrm{P}(\boldsymbol{T}|\boldsymbol{C})$，$\mathrm{P}(r_n^{(root)})$，$\mathrm{P}(r_s^{(root)})$ が決まると，式 (2.2) により，すべてのノードでの非同義置換・同義置換速度の事前分布が決まる．

ベイズの公式により，事前分布と尤度から事後分布が求められる．

$$\begin{aligned}\mathrm{P}(\boldsymbol{T},\boldsymbol{r}_n,\boldsymbol{r}_s,\nu_n,\nu_s|\boldsymbol{X},\boldsymbol{C})&=\frac{\mathrm{P}(\boldsymbol{X}|\boldsymbol{T},\boldsymbol{r}_n,\boldsymbol{r}_s,\nu_n,\nu_s,\boldsymbol{C})\mathrm{P}(\boldsymbol{T},\boldsymbol{r}_n,\boldsymbol{r}_s,\nu_n,\nu_s|\boldsymbol{C})}{\mathrm{P}(\boldsymbol{X}|\boldsymbol{C})}\\&=\frac{1}{\mathrm{P}(\boldsymbol{X})}\mathrm{P}(\boldsymbol{X}|\boldsymbol{T},\boldsymbol{r}_n,\boldsymbol{r}_s)\mathrm{P}(\boldsymbol{r}_n,\boldsymbol{r}_s|\boldsymbol{T},\nu_n,\nu_s)\mathrm{P}(\boldsymbol{T}|\boldsymbol{C})\mathrm{P}(\nu_n)\mathrm{P}(\nu_s)\mathrm{P}(r_n^{(root)})\mathrm{P}(r_s^{(root)})\end{aligned}\tag{2.3}$$

マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo, MCMC) により，事後分布を計算する (Thorne et al., 1998; Seo et al., 2004)．

2.4 複数の遺伝子の同時解析

式 (2.3) は単一遺伝子を用いたベイズ法の事後分布を表している．複数の遺伝子に対応する，尤度・速度の事前分布・速度変動係数の事前分布を導入することにより，これらの遺伝子を同時解析することが可能となる．

N 個の遺伝子が互いに独立であると仮定すると，尤度は各遺伝子の尤度の積で表現できる．非同義置換・同義置換の速度は一般に遺伝子ごとに異なることを許し，それらの事前分布には積型の分布 $\prod_{k=1}^{N} \mathrm{P}(\boldsymbol{r}_n^{(k)}, \boldsymbol{r}_s^{(k)} | \boldsymbol{T}, \nu_n^{(k)}, \nu_s^{(k)})$ を用意する．これに伴い，速度変動の超事前分布，および系統樹の根における速度の事前分布 $\mathrm{P}(\nu_n)\mathrm{P}(\nu_s)\mathrm{P}(r_n^{(root)})\mathrm{P}(r_s^{(root)})$ も遺伝子ごとのそれの積で表現する．分岐年代の事前分布 $\mathrm{P}(\boldsymbol{T}|\boldsymbol{C})$ は単一遺伝子の場合と異ならない．

2.5 同義置換速度・非同義置換速度変化の相関

世代の長さの変化や世代当たりの突然変異率の変化が進化速度の変動の主たる要因である場合には，同義置換速度と非同義置換速度の 2 種類の進化速度が同時に加速(或いは減速)すると予想される．これに対して，淘汰圧の変化が主たる要因である場合には，2 つの速度の変化は互いに関連しない．また密接に相互作用するタンパク質においては，これらにかかる淘汰圧の変化が協調することから，これらタンパク質をコードする遺伝子の非同義置換速度も共時的に変動するであろう．

同義置換速度と非同義置換速度の事後分布に基づき，2 種類の速度の変化の相関を調べることができる．同様に，異なる遺伝子の非同義置換速度変化(あるいは同義置換速度変化)の相関を調べることができる．ここでは単一遺伝子の中で，同義置換速度・非同義置換速度の変化の相関を調べる方法(Seo et al., 2004)を紹介する．この方法は，異なる遺伝子間の同義置換速度・非同義置換速度の変化の相関を調べるのにも容易に応用できる．

i 番目のノードで，MCMC 標本中 j 番目の非同義置換速度と同義置換速度のサンプルを各々 $r_n^{i,j}$，$r_s^{i,j}$ と表し，$\delta(i,j)$ 関数を次のように定義する．

$$\delta(i,j) := \begin{cases} 1 & \text{if } (r_n^{i,j} - r_n^{i(p),j})(r_s^{i,j} - r_s^{i(p),j}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

ここで，$i(p)$ はノード i の親ノードを意味する．すると，$i(p)$ 番目のノードと i 番目のノードの間の速度変化の相関は $\delta(i,j)$ の平均で推定することができる．

$$\hat{q}(i) := \frac{1}{N} \sum_{j=1}^{N} \delta(i,j)$$

ここで，N は MCMC 標本の大きさである．もし，進化速度の変化に相関がなければ，N 個のサンプルに対して $\delta(i,j)$ には 1 または 0 がランダムに割り当てられ，$\hat{q}(i)$ は 0.5 に近いだろう．もし，強い正の相関があれば $\hat{q}(i)$ は 1.0 に近いと予想される．そこで，系統樹全体の速度変化の相関を表す尺度 S を $S := \frac{1}{B} \sum_{i=1}^{B} \hat{q}(i)$ のように定義し，S が 0.5 から乖離した値をとる場合には，2 種類の進化速度の変動には相関があるとみなす．速度変化の方向を無作為化するシミュレーションにより，有意性を評価する．

3. 哺乳類ミトコンドリアゲノムの分子進化

哺乳類の起源と適応進化は，これまで多くの関心を引きつけてきた．分岐年代の推定に関する先行研究としては，Springer et al. (2003)，Hasegawa et al. (2003) がある．Hasegawa et al. (2003) は Thorne et al. (1998) の階層ベイズモデルをミトコンドリアゲノムにコードされた 12 のタンパク質コード領域に適用し，真獣類 42 種の分岐年代と DNA 塩基置換速度を推定した．さらに Murphy et al. (2001) が真獣類 64 種について解析した 12 の核遺伝子，2 つのミトコンドリアリボソーム RNA を再解析し，分岐年代を推定している．ここでは，ミトコンドリアゲノムにコードされた 12 のタンパク質を解析し，我々の方法による分岐年代推定値と先行研究による推定値を比較する．併せて，同義・非同義置換速度の変動パタンを推定する．

3.1 データ

Nikaido et al. (2003) は 69 種の哺乳類から得られたミトコンドリアゲノムデータを用いて，哺乳類の系統関係（トポロジー）を明らかにした．ここでは，このトポロジーを仮定する．哺乳類のミトコンドリアには 13 種類のたんぱく質遺伝子が存在する．Nikaido et al. のデータ解析と同様に，L 鎖に座乗する NADH6 を除く 12 種類のタンパク質遺伝子を解析対象とする．

3.2 分岐年代のベイズ推定

Kielan-Jaworowska の研究結果 (1992) に基づき，Hasegawa et al. (2003) は真獣類と有袋類の間の分岐年代に，上限：180MYA，下限：140MYA という制限を加えた．ここでは，上限・下限を設定して厳しく縛ることはせずに，ガンマ分布を仮定し，その平均と標準偏差をそれぞれ 160MYA (Million Years Ago) と 10MYA に設定した．Hasegawa et al. (2003) と Cao et al. (2000) の研究で使われた分岐年代の上限・下限の情報を，6 ヵ所の分岐点に加えた．図 2 は，分岐年代の事前分布の中央値と 95% 信頼区間 (credibility interval) を表している．

分岐年代は，12 種類の遺伝子において共通であるが，同義・非同義置換速度は違うことを許す．したがって，12 種類の遺伝子に対して，異なる $\mathrm{P}(r_n^{(root)})$ と $\mathrm{P}(r_s^{(root)})$ を考えた．$\mathrm{P}(r_n^{(root)})$ と $\mathrm{P}(r_s^{(root)})$ の平均は進化距離の情報を利用して決めた．即ち，各遺伝子において，共通祖先からそれぞれの配列にいたるまでの，同義・非同義距離の和を求めた．その和の中央値を年代の事前分布の平均である 160 で割り，その値が事前分布の中央値になるように $\mathrm{P}(r_n^{(root)})$ と $\mathrm{P}(r_s^{(root)})$ を決めた．$\mathrm{P}(\nu_n)$ と $\mathrm{P}(\nu_s)$ は 12 種類の遺伝子で共通であると仮定し，平均 0.005 の指数分布を仮定した．

図 3 は分岐年代の事後分布の中央値と 95% 信頼区間 (credibility interval) を表している．図 2 の事前分布と比べると信頼区間の幅が狭くなったことが分かる．配列データが持っている情報が事後分布に反映されることにより，分岐年代を特定することができ，その結果，信頼区間が狭くなったと言える．

表 1 で，我々の年代推定値と Springer et al. (2003) の推定値，Hasegawa et al. (2003) による推定値を比較した．Hasegawa et al. (2003) により再解析された Murphy et al. (2001) のデータと Springer et al. (2003) のデータは，一部ミトコンドリアデータも含んでいるが，主に核の遺伝子のデータである．また，Springer et al. (2003)，Hasegawa et al. (2003) はアミノ酸置換の情報から年代推定を行っている．

アフリカ獣上目 (Afrotheria) の起源の年代推定値と，真無盲腸類 (Eulipotyphla) の起源の年代推定値において結果に違いは見られるものの，他のノードにおいては，解析するデータ，モデルにより大きく結果が異ならないことが分かる．アフリカ獣上目と真無盲腸類の起源の年代推定値が先行研究の結果と違う主な原因として，配列データの違いと仮定する系統関係の違いが考えられる．Hasegawa らは，アフリカ獣上目からは，ツチブタ (aardvark) とゾウ (elephant) の 2 種からの配列を解析したが，我々はその後加わったデータも含め，8 種からの配列を解析した．また，Hasegawa らは，真無盲腸類 (Eulipotyphla) からはオオアシトガリネズミ (long clawed shrew) とヨーロッパモグラ (European mole) の 2 種からの配列を解析し，塩基組成の偏りなどの問題の多いハリネズミの仲間をはずしたのに対し，我々は 8 種を解析に入れた．ハリネズミの仲間の影響評価を充分に検討していないが，一般には，我々の解析ではこれら 2 グループではより多くのメンバーを解析しているため，その共通祖先は古くなる傾向がある．

アフリカ獣上目と真無盲腸類の起源の年代推定値が先行研究の結果と違うもう 1 つの原因として，モデルの違いが考えられる．他のノードはこれらの影響をあまり受けず，比較的ロバストな結果を見せるが，哺乳類の共通祖先に近い古いノードにおいては，これらの影響が大きいのではないかと思われる．

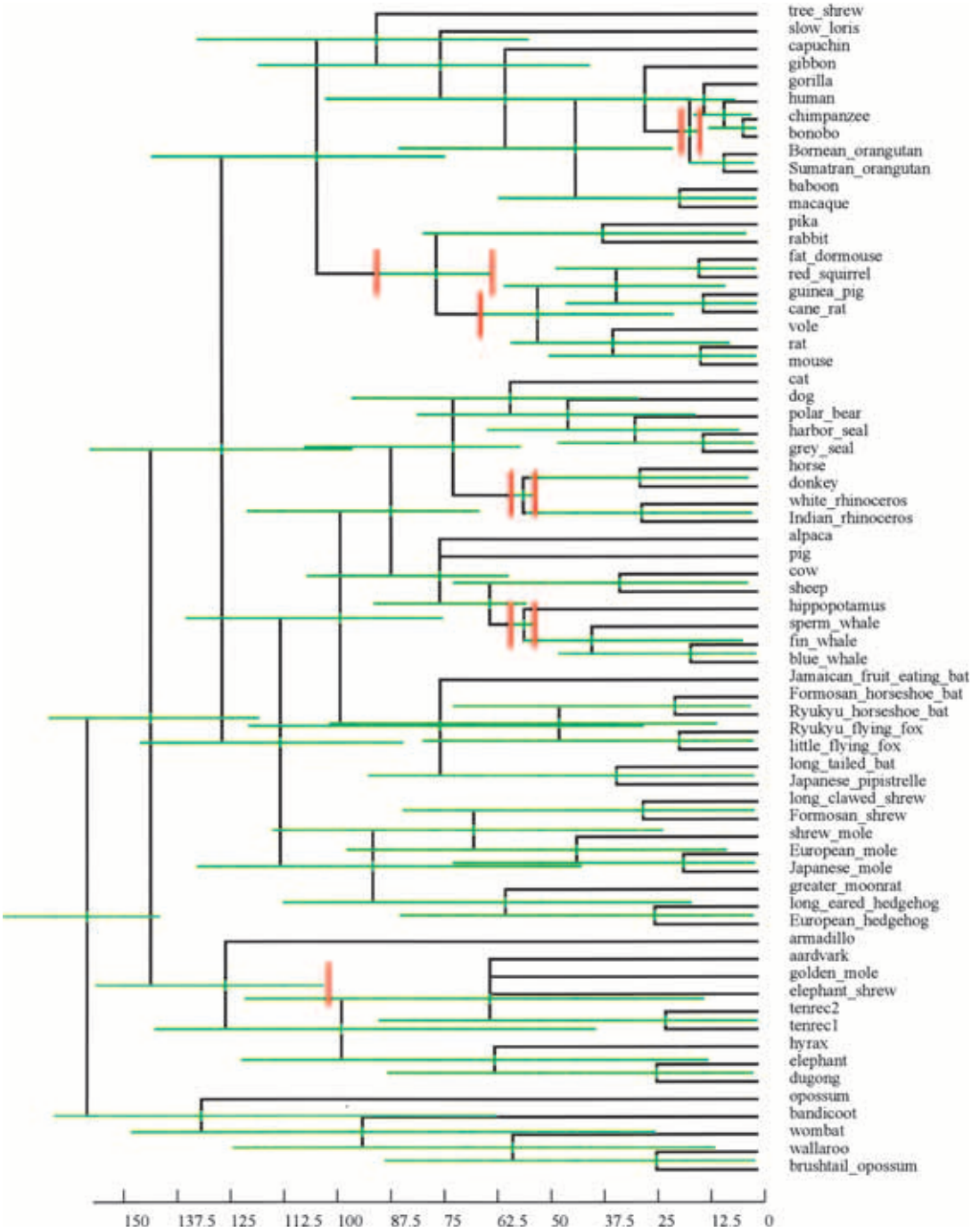


図 2. 哺乳類の進化：分岐年代の事前分布．緑色の線は事前分布の 95%信頼区間を表す．赤い色の区画は，化石データによる分岐年代の上限・下限を表す．

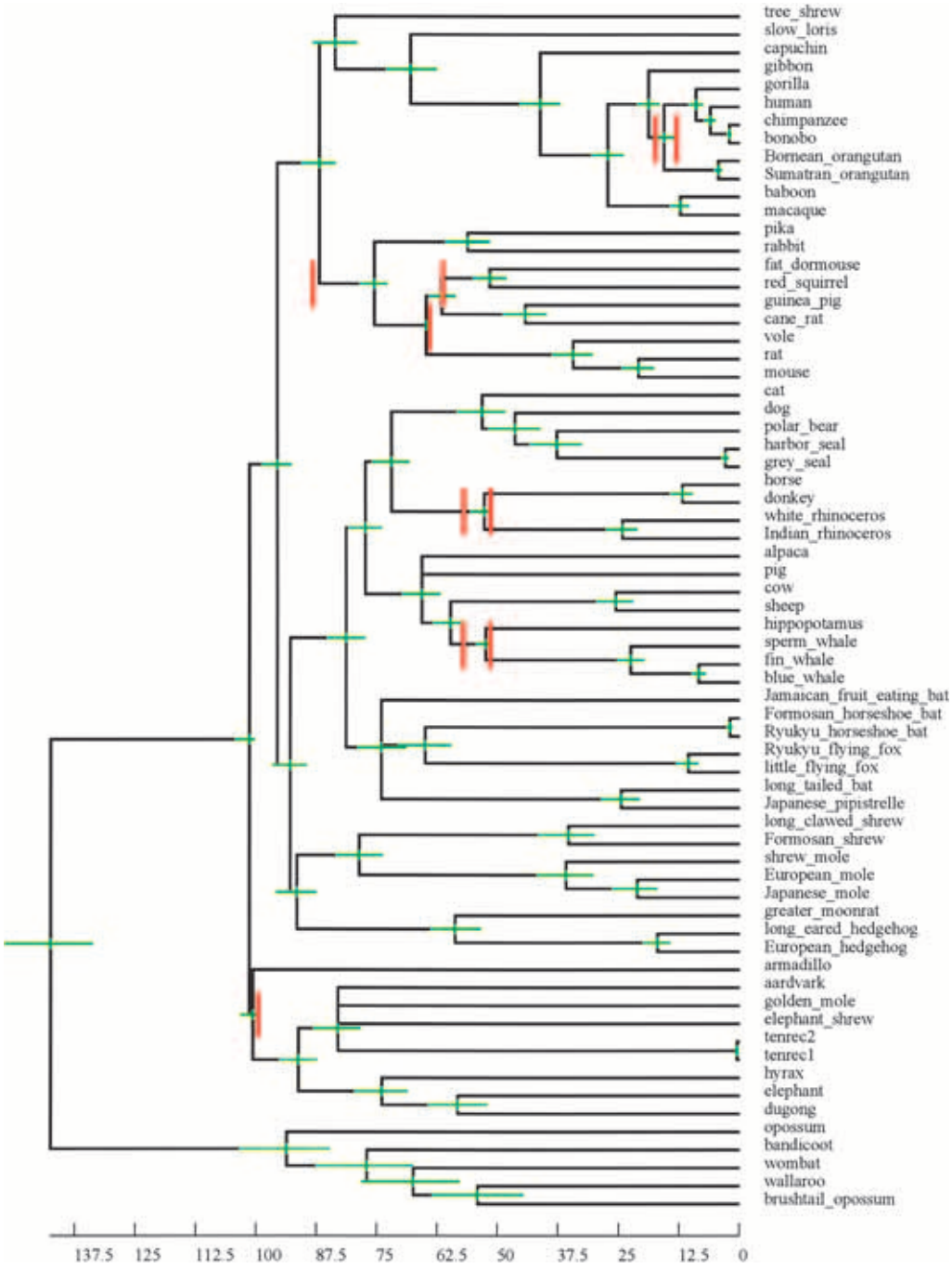


図 3. 哺乳類の進化：分岐年代の事後分布.

表 1. 哺乳類分岐年代の推定値.

Branching	Mt-Codon[a]	Mt-proteins[b]	12 Nuclear + 2 Mt[c]	19 Nuclear + 3 Mt[d]
Base of Afrotheria	91.3 ± 2.0	79.9 ± 2.9	82.8 ± 2.7	79.9 ± 3.0
Euarchontoglires/Laurasiatheria	95.7 ± 1.6	–	91.2 ± 1.8	94.0 ± 3.4
Euarchonta/Glire	87.0 ± 1.8	89.0 ± 1.9	81.6 ± 1.8	87.3 ± 3.2
Base of Euarchonta	83.7 ± 2.4	–	78.4 ± 2.2	86.0 ± 3.1
Base of Primates	68.2 ± 2.7	–	73.1 ± 2.7	77.1 ± 3.3
Patyrrhini/Catarrhini	41.3 ± 2.2	–	37.5 ± 3.1	–
Hominoidea/Cercopithecoidea	27.3 ± 1.7	34.6 ± 1.6	25.5 ±2.7	–
Human/gibbon	18.9 ± 1.3	21.7 ± 1.0	15.6 ± 2.1	–
Human/chimpanzee	6.13 ± 0.60	7.4 ± 0.7	–	–
Base of Lagomorpha (rabbit/pika)	56.3 ± 2.4	–	50.5 ± 3.2	50.9 ± 4.0
Mouse/rate	21.0 ± 1.8	16.2 ± 1.4	16.0 ± 1.9	16.3 ± 2.2
Base of Eulipotyphla	91.6 ± 2.2	61.0 ± 3.1	75.3 ± 3.1	75.9 ± 2.3
Base of Chiroptera	74.1 ± 2.6	65.2 ± 2.9	74.9 ± 3.0	65.3 ± 1.4
Base of Carnivora	53.3 ± 2.6	49.0 ± 2.7	56.8 ± 3.2	55.1 ± 2.5
Base of Cetartiodactyla	65.7 ± 2.1	64.1 ± 2.3	67.3 ± 2.7	63.8 ± 0.8
White rhinoceros/Indian rhinoceros	24.3 ± 1.7	26.1 ± 2.3	–	–
Opossum/wallaroo	93.8 ± 4.8	99.5 ± 5.4	–	–

a) 我々のベイズ法による推定値
b) Hasegawa et al. (2003)による推定値
c) Murphy et al. (2001)のデータを再解析した，Hasegawa et al. (2003)の推定値
d) Springer et al. (2003)による推定値

我々は，最近，アミノ酸置換モデルとコドン置換モデルを統計的に且つ定量的に比較する方法を提案した．これを用いた実証分析で，哺乳類のミトコンドリアゲノムの分子系統解析においては，同義置換が情報を保持しており，アミノ酸置換モデルを用いるよりもコドン置換モデルを用いる方が，信頼できる分子進化の推測ができることを示した(Seo and Kishino, 2007)．

同義置換速度の変化パタンと非同義置換速度の変化パタンが似ている場合は，コドンモデルから得られた系統樹とアミノ酸モデルから得られた系統樹はほぼ相似した形をしている．こうした場合には，コドンモデルとアミノ酸モデルいずれを用いても，分岐年代の推定には大きな違いはないことが期待される．しかし，一般には 2 種類の速度の変化パタンが異なる．こうした場合には，非同義置換のみを対象とするアミノ酸モデルよりも，同義置換と非同義置換の情報をともに取り入れるコドンモデルの方が，より信頼性の高い分岐年代の推定を可能にする．

3.3 同義置換の傾向的加速と非同義置換速度の不均質性の拡大

分岐年代の事後分布は 12 個の遺伝子の間で共通であるが，進化速度の事後分布は遺伝子ごとに異なり，ノードごとにも異なる．図 4 (a)，(b)，(c) は，COXI (Cytochrome oxidase subunit I) の例であり，各ノードにおける同義置換速度・非同義置換速度・ω の事後中央値の対数を表している．この図によって，進化の過程のどの時点，どの生物群で同義・非同義置換速度が加速・減速したか視覚的に分かる．Seo et al. (2004) は COXI を解析した．そこでは，2 種類のげっ歯類 (Rodent) を外群とし，19 種の霊長類の分岐年代と非同義速度・同義置換速度・ω の事後分布を求めた．その結果，類人猿 (anthropoid) のグループで非同義置換速度が増加したことが認められ，Wu et al. (2000) 先行研究と合致する結果を得た．今回，複数の遺伝子を同時解析したが，類人猿のグループ (図 4 の 3 番目の capuchin から 12 番目の macaque まで) において，COXI の非同義置換速度に加速傾向が認められた (図 4 (b))．

図 5 は，各ノードにおける進化速度の事後中央値をプロットし，時間変化を追ったものである (値は対数値)．12 遺伝子の内，nadh2， nadh4， nadh5， cytb の 4 遺伝子において，同義置換速度に傾向的な加速が見られた．ここではこれらについて，同義置換速度，非同義置換速度，ϖ の時間変化を表している．同義置換速度は次第に全体として増加する傾向が見てとれる．一方，非同義置換速度は明確なトレンドは見られず，時間とともに系統間の違いが拡大している．

3.4 シミュレーションによる不偏性の検証

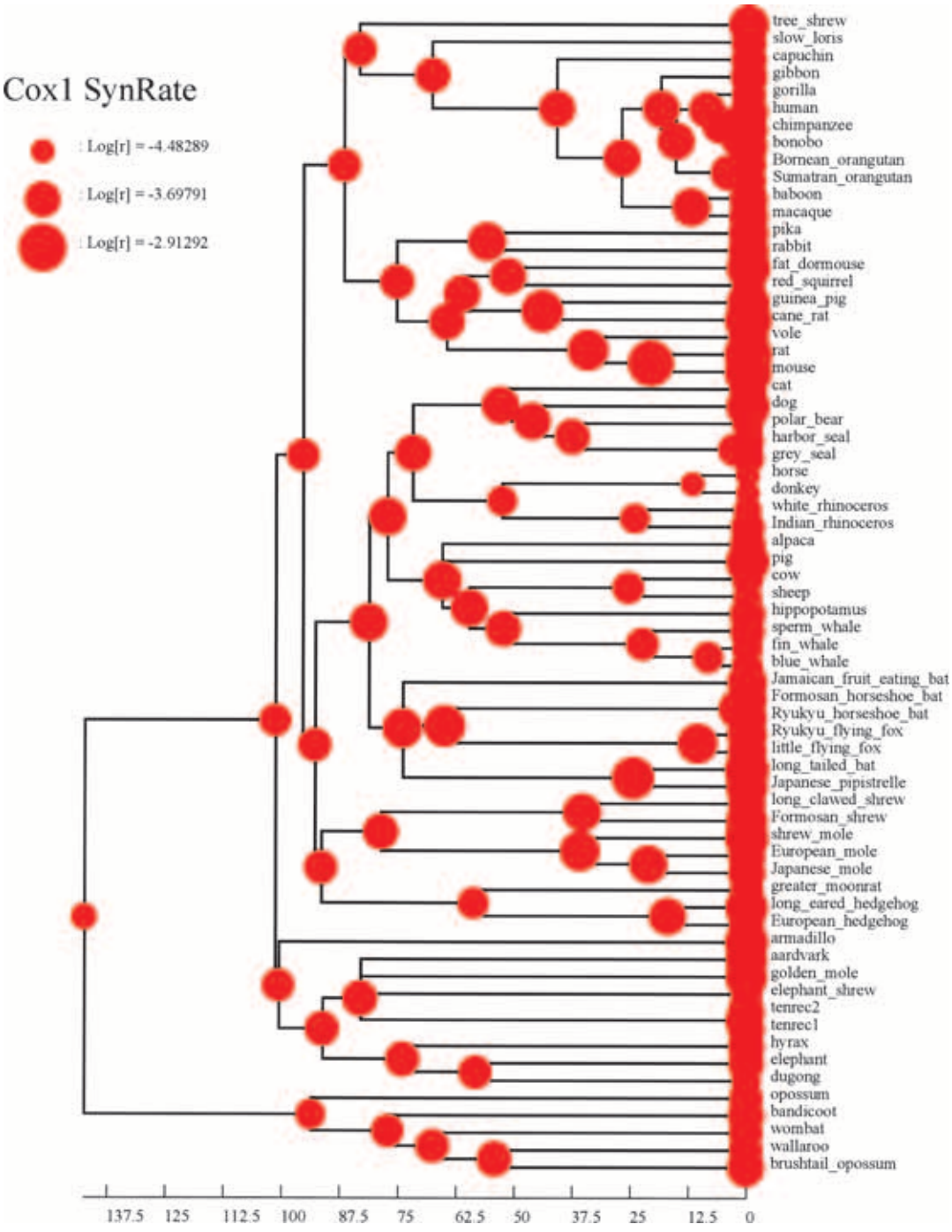
ここで注意しなければならないのは，多くの場合同義置換速度は非同義置換速度よりも大きく，過去に遡るにつれ同一のサイトで 2 回以上置換が起きる，いわゆる多重置換の現象が無視できなくなることである．このため，共通祖先に近い過去の同義置換量は過少評価され，結果として見かけ上同義置換速度が増加するように推定された，という可能性を疑う必要がある．そこで我々は，哺乳類のミトコンドリアの進化の蓄積量を反映させたシミュレーションを実施し，同義置換の飽和による進化距離の過少評価が起こるか，検証した．

同義置換速度と非同義置換速度ともに変化のない場合，同義置換のみ単調に加速する場合の 2 通りの進化シナリオについて，ランダムに配列データを生成した．配列の数・系統関係・長さ・分岐年代・共通祖先の速度などは nadh1 遺伝子から推定されたパラメーターを用いた．なお，同義置換が加速するシナリオにおいては，現在の同義置換速度が，哺乳類の共通祖先の速度よりも 10 倍速くなるように設定した．

図 6 の左側の列は真の進化シナリオを，右側の列は我々のベイズ法で推定された結果を表している．同義置換速度のトレンドがない場合，推定される速度にも明確なトレンドは観察されず，年代が遡るにつれ同義置換速度を過少推定する現象は認められない (図 6 (a))．さらに，同義置換の加速度も正しく検出された (図 6 (b))．このシミュレーションの結果から，ミトコンドリアの 4 種類の遺伝子の同義置換速度増加パタンは，実際の進化の様子を反映すると思われる．

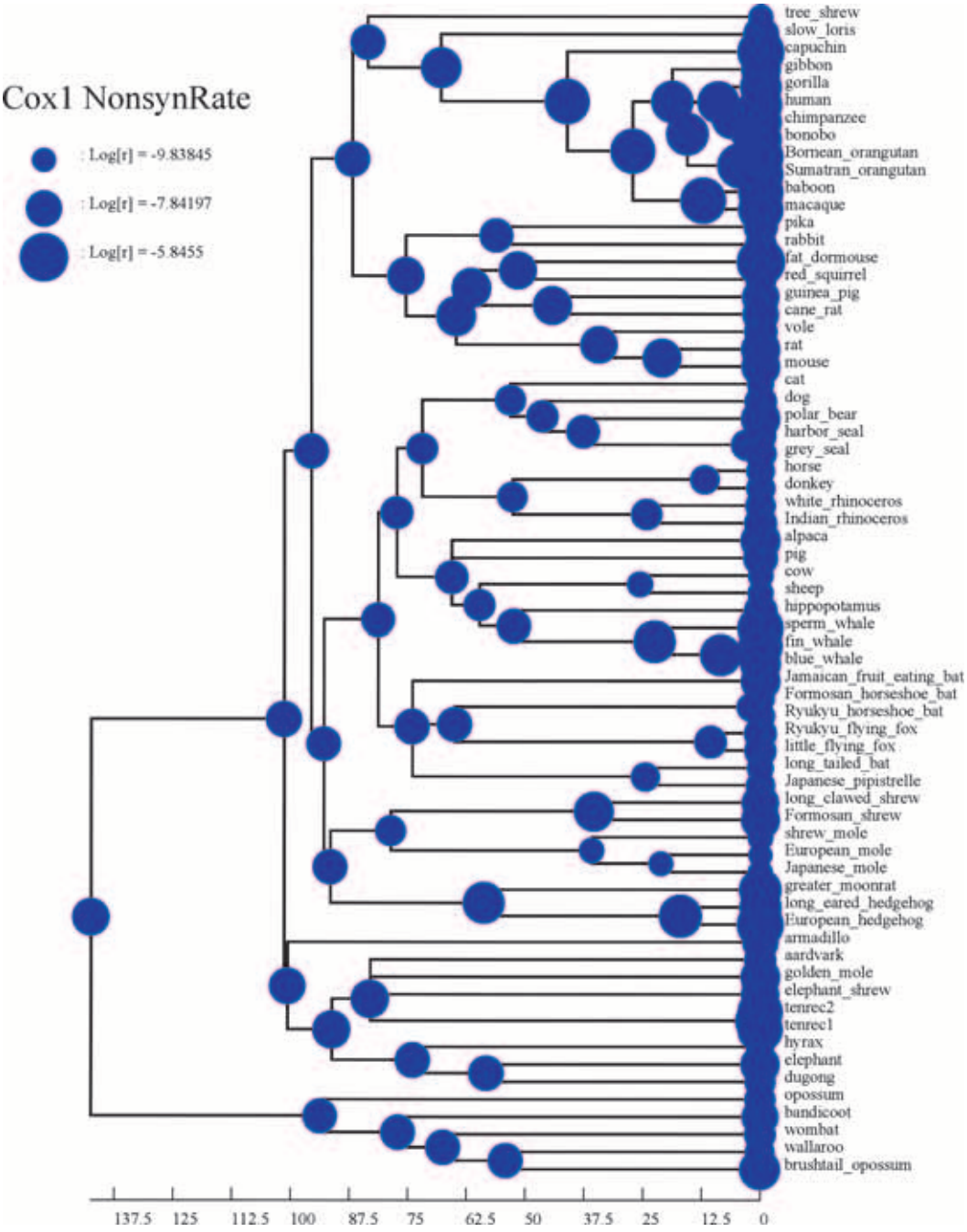
3.5 進化速度の変化と変化の相関

2.5 節で紹介した S 統計量を応用し，非同義置換速度の変動と同義置換速度の変動について，遺伝子間の相関を調べた．図 7 の右上の三角形部分は非同義置換速度変動の相関を表す S 統計量の P 値を，左下の三角形部分は同義置換速度変化の S 統計量の P 値を表す．すべての非同義置換速度変動の相関は有意である．これに対して，同義置換速度の変動については，有意な相関が認められた遺伝子ペアは限られていた．有意性が検出されなかったペアは，少なくとも一方の配列の長さが短い．実際，NADH3，NADH4L，ATP8 の長さは，ギャップを含めて，それぞれ 129, 98, 70 コドンである．データの持つ情報が少ないため相関を検出するのに失敗したと思われる．



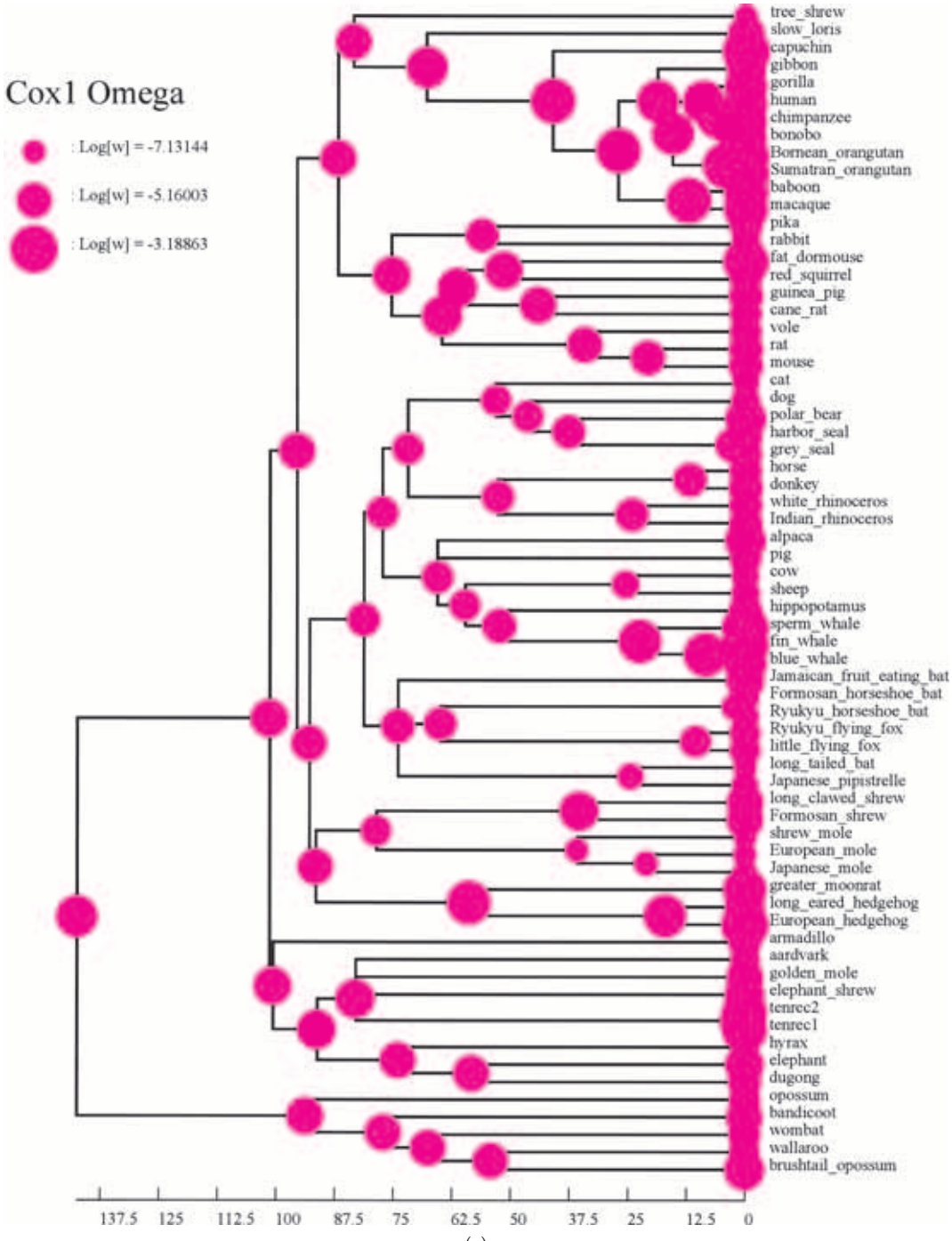
(a)

図 4. COX1 に見られる分子進化速度の変化. (a) 同義置換速度, (b) 非同義置換速度, (c) 同義置換・非同義置換の速度比 ω の各ノードにおける推定値 (事後分布中央値の対数).


Cox1 NonsynRate
: Log[r] = -9.83845
: Log[r] = -7.84197
: Log[r] = -5.8455
tree_shrew
slow_loris
capuchin
gibbon
gorilla
human
chimpanzee
bonobo
Bornean_orangutan
Sumatran_orangutan
baboon
macaque
pika
rabbit
fat_dormouse
red_squirrel
guinea_pig
cane_rat
vole
rat
mouse
cat
dog
polar_bear
harbor_seal
grey_seal
horse
donkey
white_rhinoceros
Indian_rhinoceros
alpaca
pig
cow
sheep
hippopotamus
sperm_whale
fin_whale
blue_whale
Jamaican_fruit_eating_bat
Formosan_horseshoe_bat
Ryukyu_horseshoe_bat
Ryukyu_flying_fox
little_flying_fox
long_tailed_bat
Japanese_pipistrelle
long_clawed_shrew
Formosan_shrew
shrew_mole
European_mole
Japanese_mole
greater_moonrat
long_eared_hedgehog
European_hedgehog
armadillo
aardvark
golden_mole
elephant_shrew
tenrec2
tenrec1
hyrax
elephant
dugong
opossum
bandicoot
wombat
wallaroo
brushtail_opossum
137.5
125
112.5
100
87.5
75
62.5
50
37.5
25
12.5
0

(b)

図 4. (つづき)


Cox1 Omega
: Log[w] = -7.13144
: Log[w] = -5.16003
: Log[w] = -3.18863
tree_shrew
slow_loris
capuchin
gibbon
gorilla
human
chimpanzee
bonobo
Bornean_orangutan
Sumatran_orangutan
baboon
macaque
pika
rabbit
fat_dormouse
red_squirrel
guinea_pig
cane_rat
vole
rat
mouse
cat
dog
polar_bear
harbor_seal
grey_seal
horse
donkey
white_rhinoceros
Indian_rhinoceros
alpaca
pig
cow
sheep
hippopotamus
sperm_whale
fin_whale
blue_whale
Jamaican_fruit_eating_bat
Formosan_horseshoe_bat
Ryukyu_horseshoe_bat
Ryukyu_flying_fox
little_flying_fox
long_tailed_bat
Japanese_pipistrelle
long_clawed_shrew
Formosan_shrew
shrew_mole
European_mole
Japanese_mole
greater_moonrat
long_eared_hedgehog
European_hedgehog
armadillo
aardvark
golden_mole
elephant_shrew
tenrec2
tenrec1
hyrax
elephant
dugong
opossum
bandicoot
wombat
wallaroo
brushtail_opossum
137.5
125
112.5
100
87.5
75
62.5
50
37.5
25
12.5
0

(c)

図 4. （つづき）

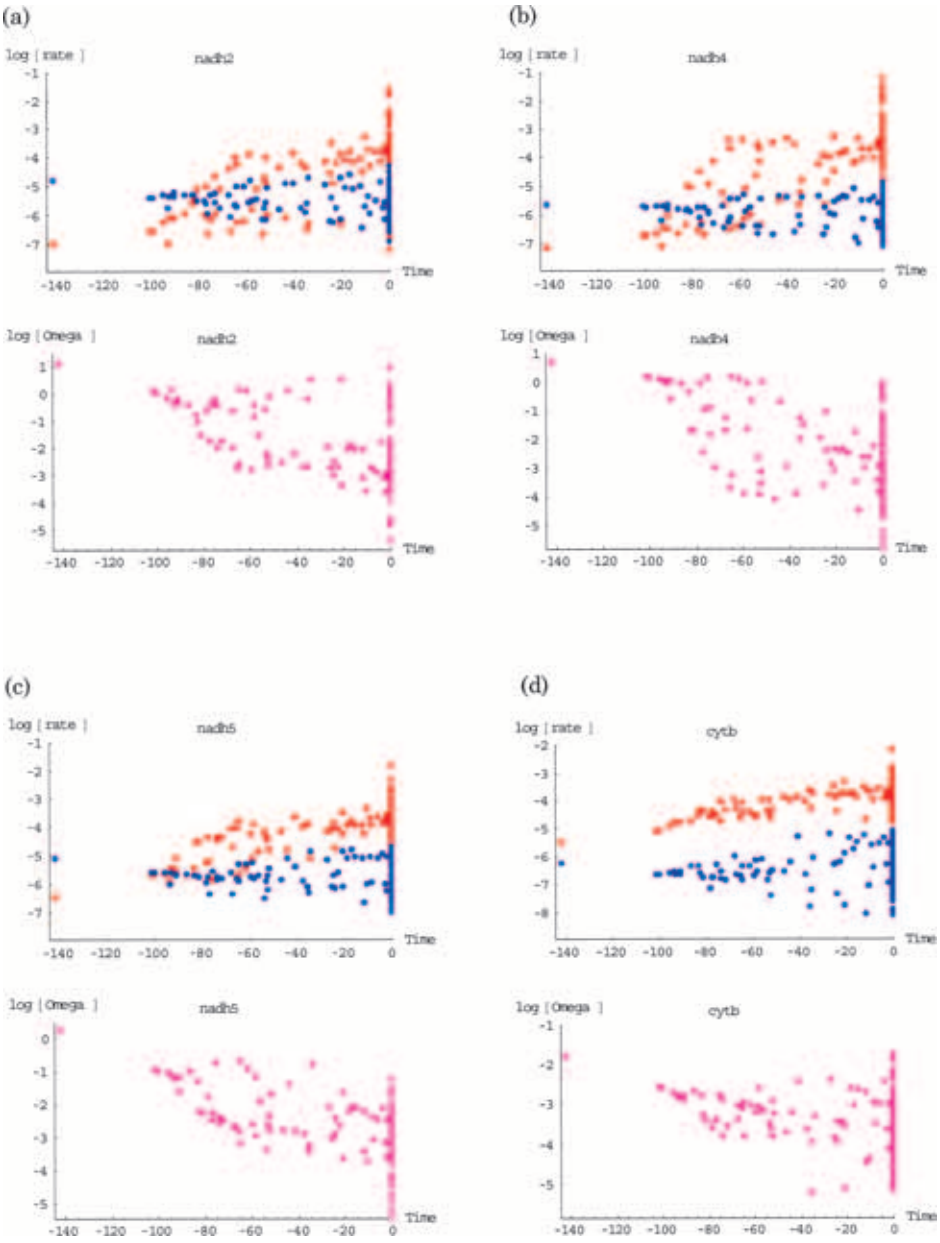


図 5. 同義置換速度・非同義置換速度・ω の変化．(a) nadh2, (b) nadh4, (c) nadh5, (d) cytb．それぞれにおいて上は同義置換速度 (赤) と非同義置換速度 (青) の時系列変化，下は ω の時系列変化をプロットしており，横軸に推定分岐年代をとり，各ノードにおける事後中央値の対数を示している．

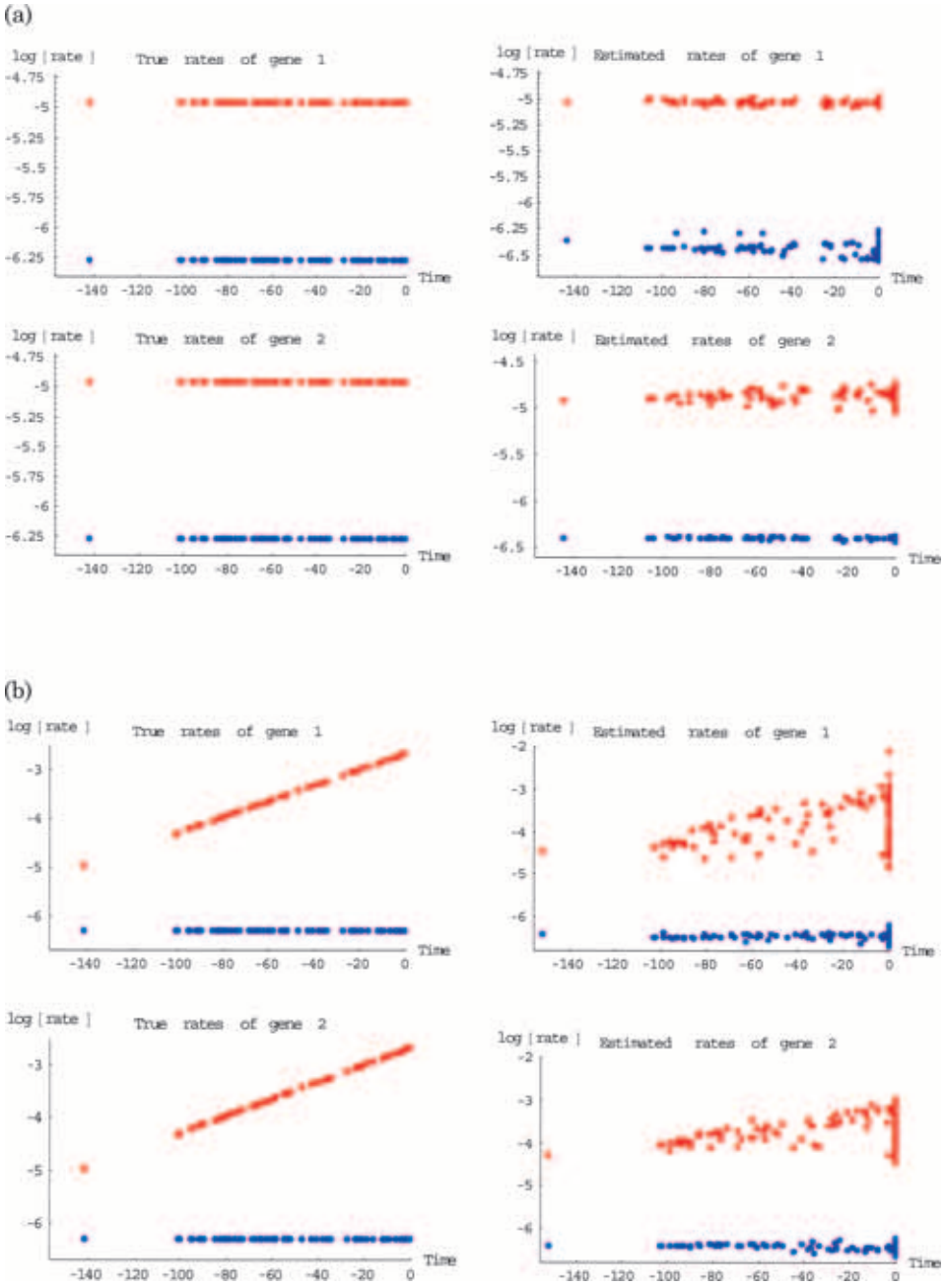


図 6. シミュレーションによる同義・非同義置換速度の時系列的変化の推定の検証. (a) 同義置換速度・非同義置換速度ともに一定の場合, (b) 同義置換速度のみ傾向的に加速する場合. 左側の列は真の進化速度のトレンドを, 右側の列は推定された進化速度のトレンドを表す.

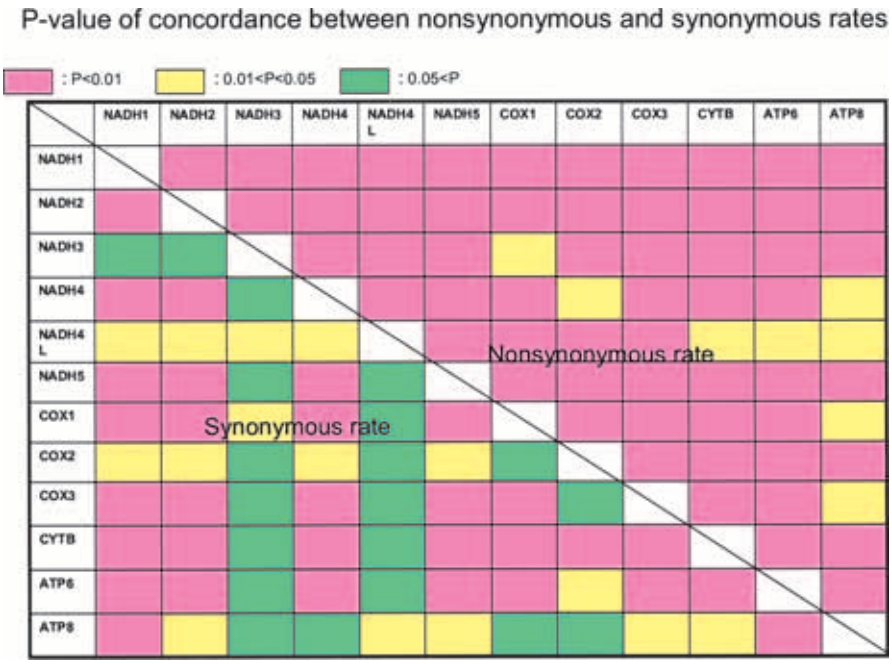


図 7. 哺乳類ミトコンドリア遺伝子の同義・非同義置換速度変化の相関．シミュレーションで得られた，S 統計量 (2.5 節) の P 値．右上・左下の三角形部分はそれぞれ，非同義置換速度変動の相関を表す S 統計量の P 値・同義置換速度変動の相関を表す S 統計量の P 値を表す．

4. おわりに

哺乳類の系統間でミトコンドリアゲノムの非同義置換速度の違いが開いていく現象は，哺乳類が適応放散した後にミトコンドリアの機能が多様化していることを示唆している．一方，同義置換速度が傾向的に加速する現象に関連して，哺乳類のミトコンドリアゲノムにおいて，T から C へ塩基の偏りが進行していることが指摘されている (Reyes et al., 1998; Gibson et al., 2005)．塩基レベルの非対称な突然変異の加速とこれに伴う組成の変化が，アミノ酸の変化を促し，哺乳類のミトコンドリアゲノムの多様化を可能にしたように見える．

ここでは我々が開発した同義・非同義置換速度変化モデル (Seo et al., 2004) を紹介し，これを複数の遺伝子を同時解析するプログラムに拡張した．複数の遺伝子を解析することにより，分岐年代の推定精度が向上するだけでなく，遺伝子間の同義・非同義置換速度変動の相関を調べることが可能となる．異なる遺伝子の間の同義・非同義置換速度変動の相関は，たんぱく質機能と相互作用を分子進化学的なアプローチから推定する有力な手段を提供することが期待される．

参 考 文 献

Cao, Y., Fujiwara, M., Nikaido, M., Okada, M. and Hasegawa, M. (2000). Interordinal relationships and time-scale of eutherian evolution as inferred from mitochondrial genome data, *Gene*, **259**, 149–158.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach, *Journal of Molecular Evolution*, **17**, 368–376.

Gibson, A., Gowri-Shankar, V., Higgs, P. G. and Rattray, M. (2005). A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods, *Molecular Biology and Evolution*, **22**, 251–264.

Goldman, N. and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences, *Molecular Biology and Evolution*, **11**, 725–736.

Hasegawa, M., Kishino, H. and Thorne, J. L. (2003). Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution, *Genes and Genetic System*, **78**, 267–283.

Kielan-Jaworowska, Z. (1992). Interrelationships of Mesozoic mammals, *Historical Biology*, **6**, 185–202.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.

Miyata, T. and Yasunaga, T. (1980). Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitution from homologous nucleotide sequences and its application, *Journal of Molecular Evolution*, **16**, 23–36.

Murphy, W. J., Eizirik, E., O'Brien, S. J., et al. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenies, *Science*, **294**, 2348–2351.

Muse, S. V. and Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome, *Molecular Biology and Evolution*, **11**, 715–724.

Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, *Molecular Biology and Evolution*, **3**, 418–426.

Nikaido, M., Cao, Y., Harada, M., Okada, N. and Hasegawa, M. (2003). Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla, *Molecular Phylogenetics and Evolution*, **28**, 276–284.

Reyes, A., Gissi, C., Pesole, G. and Saccone, C. (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of Mammals, *Molecular Biology and Evolution*, **15**, 957–966.

Seo, T.-K. and Kishino, H. (2007). Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins (to appear).

Seo, T.-K., Kishino, H. and Thorne, J. L. (2004). Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences, *Molecular Biology and Evolution*, **21**, 1201–1213.

Springer, M. S., Murphy, W. J., Eizirik, E. and O'Brien, S. J. (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary, *Proceedings of National Academy of Sciences, U.S.A.*, **100**, 1056–1061.

Thorne, J. L., Kishino, H. and Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution, *Molecular Biology and Evolution*, **15**, 1647–1657.

Wu, W., Schmidt, T. R., Goodman, M. and Grossman, L. I. (2000). Molecular evolution of cytochrome c oxidase subunit I in primates: Is there coevolution between mitochondrial and nuclear genomes?, *Molecular Phylogenetics and Evolution*, **17**, 294–304.

Bayesian Divergence Time Estimation Using Codon Model

Tae-Kun Seo[1], Hirohisa Kishino[2] and Jeffrey L. Thorne[3]

[1]Professional Programme for Agricultural Bioinformatics, University of Tokyo
[2]Laboratory of Biometry and Bioinformatics, University of Tokyo
[3]Bioinformatics Research Center, North Carolina State University

Because evolutionary rates of molecular data can change over time, it is unreasonable to assume a molecular clock to estimate divergence times. Changes in mutation rate, effective population size and selective pressure may cause changes in either or both of the rates of synonymous and nonsynonymous substitutions. Recently, we developed a new Bayesian method to estimate divergence times and absolute rates of synonymous and nonsynonymous substitutions. Instead of assuming a molecular clock, we assume that both rates change over time following a log-normal process. By adopting a Markov chain Monte Carlo procedure, we can estimate the posterior probabilities of divergence times, synonymous and nonsynonymous rates, and rate variation parameters. This paper discusses the extension of our method to the analysis of multilocus sequence data, and shows the analysis of mammalian mitochondrial protein-coding genes.

Key words: Codon model, synonymous substitution, nonsynonymous substitution, molecular clock, divergence time.