

ゲノム進化の階層モデル

東京大学・農学生命科学 岸野洋久

1 序

数多くあるゲノムの変異の中から、生物の適応進化に本質的に結びついたものを探し出す努力が多くの生物学者によって積み重ねられてきた。現在のところ、ゲノムレベルでの適応進化の爪痕を検出する際の第一の探索的アプローチとして系統間で分子進化速度を比較する方法が有力視されている。重複遺伝子の運命 (Zhang *et al* (1998); Lynch and Conery (2000))、菌類の共生と分子進化速度の加速化 (Lutzoni and Pagel (1997)) など、90年代半ばから興味深い研究が次々に出てきた。

筆者らはこれまで、分子進化速度が変動する様子を、事前にグループ分けするなど強い制約を置くことなく柔軟に推定する階層モデルを提案した (Thorne *et al* (1998))。そして、分子時計を仮定した解析との対比において、分岐年代推定の頑健性を調べた (Kishino *et al* (2001))。

生物の適応・多様化とゲノム進化のフィードバックのありさまを高感度に検出することを可能とする分析手法を開発できれば、これを重用遺伝子の探索のための有効な武器として用いることが可能となるであろう。それは種内変異に基き遺伝子を探索する連鎖解析に対置され、機能や形態の広範な種間変異の情報から関連遺伝子を探索する方法論を提供する。

2 進化速度の確率変動モデルと共進化の検出

マルコフ過程で記述される分子進化の推移速度行列をモデル化することにより、分子系統樹の尤度が表現される。相同な s 本の配列を比較して系統関係を推定する場合を考える。配列の長さを n とすると、系統樹 T の対数尤度は $l(\theta|\mathbf{X}) = \sum_{h=1}^n \log f(\mathbf{X}_h|\theta)$ と表わされる。ここで θ_i は進化のプロセスを規定するパラメータである。配列の変化の統計モデルとしてはマルコフ過程によるモデル化が妥当する。分岐後それぞれの種の配列は独立に進化すると仮定すると、 $f(\mathbf{X}|\theta)$ は

$$f(\mathbf{X}|\theta) = \sum_{Z_{i_0}} \pi_{Z_{i_0}} \prod_{j \in \text{node}(T) \setminus i_0} \sum_{Z_j} P_{Z_{anc(j)}, Z_j}(t_{anc(j), j}) \quad (1)$$

と簡単に表される。ここで、 $\text{node}(T)$ は系統樹 T の節を表し、 i_0 はその根である。無根系統樹の場合には、任意の節を指定する。 $\text{anc}(j)$ は j に隣接する祖先となる節である。 $P_{xy}(t)$ は時間 t を経た推移確率である。

進化速度が変動する背景要因として、選択圧の変化を中心とした環境変動、集団の大きさの変動、世代の長さの変動などが考えられる。これらはいずれも自己相関を持って変動することが予想される。そこで、事前分布として速度 $r(t)$ の対数をとったものが簡単な拡散過程に従うとする。すなわち、 $\tilde{r}(t) = \log r(t)$ は正規マルコフ過程で、任意の2時点 t, s ($t > s$) に対して

$$\begin{aligned} E[\tilde{r}(t)|\tilde{r}(s)] &= \tilde{r}(s) - \frac{\nu}{2}(t-s) \\ V[\tilde{r}(t)|\tilde{r}(s)] &= \nu(t-s) \end{aligned}$$

を仮定する。移流項は進化速度の期待値が傾向的に変動させないためにつけたものである。

ところで、遺伝子間で独立を仮定した事前分布に対する速度の事後平均を遺伝子間で比較することにより、共進化を検出することができる。ここで、速度の系統間の系列相関があることに注意が必要である。2遺伝子における各節での速度の事後平均 \bar{r}_{mj} ($m = 1, 2, j \in \text{node}(T)$) を求め、順位相関を取る。無相関の帰無仮説の下での分布は、速度変化の確率変動と推定誤差による不確実性を踏まえていなければならない。ここ

では、速度変化の方向を無作為化することにより分布を得る (Thorne and Kishino (2002))。すなわち、帰無仮説の下での各節における速度の事後平均 \bar{r}'_{mj} を

$$\begin{aligned}\bar{r}'_{m,i_0} &= \bar{r}_{m,i_0} \\ \bar{r}'_{m,j} &= \bar{r}'_{m,anc(j)} + W_{mj} (\bar{r}_{m,j} - \bar{r}_{m,anc(j)}) \quad (j \in node(T) \setminus i_0)\end{aligned}$$

により生成する。 W_{mj} は 1, -1 をそれぞれ確率 $\frac{1}{2}$ でとる互いに独立な確率変数である。

3 HIV-1 の宿主内変異とその分布

Shankarappa (1999) は 9 人のエイズ患者について *env* 遺伝子の遺伝的多様度をはかった。その結果、潜伏期間においては一定の速度で多様度が増加して行き、ある時点に達すると速度が低下することを見出した。この時、ウイルス集団が免疫系の網の目を潜り抜けて爆発的に増大した。それから程なくして症状が現れた。ウイルス集団の宿主内ダイナミクスとその宿主間変異を分子レベルで見ることにより、集団遺伝・分子進化と疫学を結びつけ、より感度の高い発病予測と対処方法を開発することが可能となるかも知れない。

遺伝子の種間変異は分子系統樹により記述される。これに対して、遺伝子の種内変異を記述するものは遺伝子の系図である。集団から取られた大きさ i の標本のうちのどれか 1 つのペアが共通祖先に行きつく時間の分布は、指数分布 $p(t_i|N_e) = \frac{i(i-1)}{2N_e} \exp\left\{-\frac{i(i-1)}{2N_e}t_i\right\}$ で表現される。 N_e は有効な集団の大きさである。RNA ウイルスは進化速度が速いため、異なる時点から得られた配列を比較分析することにより、進化速度の絶対値および分岐時点を精度良く推定することができる (Rambaut (2000))。打切りを含む実験が複数回ある状態を想定することにより、異なる時点から得られた配列の分岐時点の尤度が、たとえば 2 時点からの抽出の場合

$$L_1 = \left[\prod_{i=n_1}^{c+1} p(t_i|e(t_i), N_e) \right] \times p(t_c^*, s_{c+n_2}|e(t_c^*), k_2, N_e) \times \left[\prod_{i=c+n_2-1}^2 p(s_i|e(s_i), N_e) \right].$$

と求められる。

実際には t もデータから推定する必要があるため、擬似最尤法により有効な集団の大きさを求める。すなわち、分子系統樹の尤度関数を最大化させてこれを推定し、この推定値を用いて上の遺伝子の系図の尤度関数を最大化する。 t の推定精度を良くするためには、サンプリング時点の分散を大きく取ることが望ましい (Seo *et al* (2002))。Shankarappa の解析した 9 人の患者における HIV-1 *env* 遺伝子について、有効な集団の大きさと進化速度の関係を見ると、負の相関が認められた (Seo *et al* (2002))。

参考文献

- Kishino, H., Thorne, J. L., and Bruno, W. J. (2001) *Mol. Biol. and Evol.* **18**: 352–361.
Lutzoni, F. and Pagel, M. (1997) *Proc. Natl. Acad. Sci., USA.* **94**: 11422–11427.
Lynch, M. and Conery, J. S. (2000) *Science.* **290**: 1151–1155.
Rambaut, A. (2000) *Bioinformatics* **16**, 395–399.
Seo, T-K., Thorne, J.L., Hasegawa, M. and Kishino, H. *Bioinformatics.* in press.
Seo, T-K., Thorne, J.L., Hasegawa, M. and Kishino, H. *Genetics.* accepted.
Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., Huang, X.L. and Mullins, J.I. (1999) *J. Virol.* **73**, 10489–10502.
Thorne, J.L. and Kishino, H. *Syst. Biol.* accepted.
Thorne, J.L., Kishino, H. and Painter, I.S. (1998) *Mol. Biol. Evol.* **15**: 1647–1657
Zhang, J., Rosenberg, H. F. and Nei, M. (1998) *Proc. Natl. Acad. Sci., USA.* **95**: 3708–3713