

遺伝子型の分子進化と表現型の適応進化

渡部 輝明¹・徐 泰健²・岸野 洋久²

(受付 2007 年 8 月 29 日; 改訂 2007 年 10 月 10 日)

要 旨

本稿では、著者たちの最近の研究をベースとして、配列の分子進化から淘汰圧を受けた表現型の適応進化へのつながりを、定量的に説明する試みを紹介する。合体過程と分子進化の統計モデルを通して、宿主内の HIV 集団に配列進化と集団の大きさの間の負の相関が見られることを示す。この負の相関は、分子進化の多くが中立または弱有害の突然変異で駆動されているとする説と矛盾しない。続いて、タンパク質構造のデータベースの情報を利用して、配列の構造への適合度を経験尤度の形で定義し、HIV の V3 ループにおける微細構造の感染後の変異の大きさが分子進化速度と正に相関することを見る。さらに、タンパク質複合体の結合能を、結合状態と自由状態の尤度比で定義し、インフルエンザウイルス HA と 4 種類の抗体との結合能の長期変化を推定する。淘汰圧の強さとパターンが表現型のダイナミクスに影響を与えていることを見る。

キーワード：配列進化，構造進化，合体過程，分子進化。

1. RNA ウイルス

表現型の進化は、数多くの突然変異と配列の分子進化により支えられている。突然変異により生成された遺伝的多様性が、環境の淘汰圧による篩にかけられて、集団は柔軟に適応していく。しかしながら、配列進化と表現型の進化の間の関連を定量的に分析することは、必ずしも容易でない。高等生物では数多くの要因が個体の生存力に影響するため、突然変異の適応度を一義的に評価することは困難である。これに対して、RNA ウイルスはこうした研究に最適と言える。第一に、世代の長さが短いことと逆転写の際にエラーが混入しやすいことから突然変異率が高いことである。第二に、単純な生活環を持つことから、環境への適応度を数少ない因子で説明できることである。

ウイルスは、自分自身で世代交代を行うことはできないが、宿主の細胞に侵入し、自身のゲノムを宿主のゲノムに組み込む。そして、宿主細胞のゲノムが複製されるとともに、ウイルスのゲノムも複製され、やがて新たなウイルス粒子となって巣立っていく。それらの一部がほどなく新たな宿主細胞に侵入し、複製される。自身ではエネルギーを利用しないので生物というよりは物質であるが、宿主を利用して世代交代を行う。宿主細胞とそれを取り巻く環境に依存した生活環を持つことから、絶えず環境の強い淘汰圧に晒されている。さらに、特に DNA を持たず RNA ゲノムを持つ RNA ウイルスは、世代の長さが短い上に RNA から DNA へと逆転写の際にエラーが混入しやすいことから、突然変異率が非常に高い。そのため、ウイルス集

¹ 高知大学 医学部医学情報センター：〒 783-8505 高知県南国市岡豊町小蓮

² 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

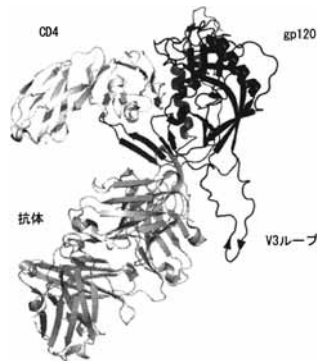


図 1. CD4 受容体に結合した HIV の gp120 とその V3 ループ. すぐ近くに抗体が結合し、補受容体に吸着できない (Huang et al., 2005, PDB コード: 2B4C).

団においては環境への適応進化を現在進行形の形で観察できる。また、単純な「生活環」を持つことから、遺伝子型の進化と表現型の進化、それに基づく環境への適応を定量的に分析することを可能にする。

図 1 は、抗体が HIV-1 (エイズウイルス) の V3 ループ近辺に結合し、細胞への侵入を阻止している様子を示している (Huang et al., 2005)。ウイルス粒子の膜表面にある突起(スパイク)を構成するタンパク質である gp120 は、CD4 という宿主細胞の受容体(レセプター)に結合し、これが軸足となって V3 ループが CCR5 という補受容体(コレセプター)に吸着し、細胞への侵入を開始する。V3 ループは宿主細胞と親和性があるため、細胞は自身のものとして細胞内にウイルスを取り込む。

ウイルスにとっては、抗体に結合されると宿主細胞の CD4 に結合できなくなる。したがって、抗体との結合部位においては、微細構造を変化させるよう強い淘汰圧が働く。突然変異によりコードするアミノ酸の種類が変わると、タンパク質の形状が変わる。これが結合部位の結合能を変化させる。たとえば、親水性の強いアミノ酸は疎水性のアミノ酸をはじく。疎水性のアミノ酸どうして結合していた部位が親水性のアミノ酸に置き換わると、結合能は低下する。一方、結合部位における微細構造の変化は、これに隣接する V3 ループにも影響を与え、これが不都合に歪むと、コレセプターへの結合がうまくいかなくなる。微妙なバランスを保ちながら、ウイルス集団は免疫系の攻撃をすり抜けていく。

微妙なバランスの難しさは、結合部位において多様度が高いことからうかがい知ることが出来る。突然変異はランダムにおきるが、その多くは有害あるいは弱有害である。したがって、平時においてはこれらの多くは集団から脱落する。ところが強い淘汰圧にさらされると、数多くの突然変異の中で、完璧とまで行かなくても、コストを払いながらも抗体との結合を下げる変異が集団に残る。

本稿ではまず、gp120 をコードする env 遺伝子配列を解析し、エイズ患者におけるウイルス集団の大きさと分子進化速度の間の負の相関関係を調べた研究を紹介し、集団遺伝学的な検討を行う。続いて、タンパク質立体構造のデータベースを解析して、溶媒への露出度などアミノ酸を取り巻く局所環境別に選好されるアミノ酸とアミノ酸対相互作用を経験分布の形で算出する。これを元に、遺伝子配列のタンパク質立体構造への適合度を尤度の形で表現し、宿主内で V3 ループの微細構造が感染後変化するスピードを配列進化のスピードと対比させる。さらに、タンパク質複合体の結合能を結合状態と自由状態の尤度比の形で表現する。1968 年大流行した

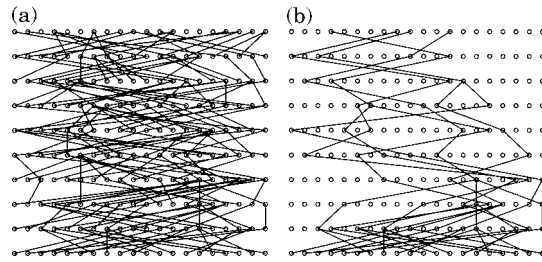


図 2. 合体過程. (a) 世代をまたいだ親から子への遺伝子の継承, (b) 現存する遺伝子の共通祖先.

香港風邪で、インフルエンザウイルスの赤血球凝集素 (HA) が宿主細胞に結合するのを阻止する抗体 4 種類について、立体構造が調べられていた。そこで、その後長年にわたりサンプリングされてきているインフルエンザ HA 配列データについて、これら抗体への結合能の変化を追い、環境への適応とそれに払うコストについて考察を加える。

2. 合体過程と配列の尤度

集団の大きさと遺伝的多様度の間には密接な関係がある。小さな集団からサンプリングされた個体は互いに近縁関係にある傾向があるのに対し、大きな集団からのサンプルは遺伝的に離れている。そこで、サンプリングされた個体の遺伝子配列を比較することにより、集団の大きさを間接的に推定することができる。遺伝子は世代をまたいで親から子へ継承されていく (図 2 (a): 上から下へ世代が進んでいる)。中には、途絶え、現在に受け継がれない遺伝子も数多くある。逆に、標本個体の遺伝子の祖先を手繰っていくと、やがてそれらは次々に共通祖先に合体していく (図 2 (b))。

合体するまでの世代数が集団の大きさと関係している。簡単のため、それぞれの遺伝子の祖先は、祖先集団から無作為に決まる無作為交配集団を仮定する。 N 個の遺伝子からなる母集団からの大きさ $n (\ll N)$ の標本を考える。標本に含まれる遺伝子のいずれかの対が共通祖先に合体するまでの世代数 T_n は

$$P(T_n > t) = \left(\prod_{k=1}^{n-1} \left(1 - \frac{k}{N} \right) \right)^t \sim \left(1 - \sum_{k=1}^{n-1} \frac{k}{N} \right)^t \sim \exp \left(-\frac{n(n-1)t}{2N} \right)$$

と指数分布で近似される。同様の計算を繰り返すことにより、標本が相次ぎ合体していき、単一の共通祖先に行き着くまでの相次ぐ合体の時間間隔 T_n, T_{n-1}, \dots, T_2 の尤度が

$$(2.1) \quad L(t_n, t_{n-1}, \dots, t_2) = \prod_{j=2}^n \frac{j(j-1)}{2N} \exp \left(-t_j \frac{j(j-1)}{2N} \right)$$

と求められる (Kingman, 1982; Kuhner et al., 1995)。

分子進化をマルコフ過程でモデル化し、サイト間の独立性を仮定すると、長さ n の s 本の配列データ

$$\mathbf{X} \equiv (\mathbf{X}_1, \dots, \mathbf{X}_n) = \begin{pmatrix} X_{11} & \cdots & X_{1h} & \cdots & X_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{p1} & \cdots & X_{ph} & \cdots & X_{pn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{s1} & \cdots & X_{sh} & \cdots & X_{sn} \end{pmatrix}$$

の対数尤度は

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{X}) &= \sum_{h=1}^n \log f(\mathbf{X}_h|\boldsymbol{\theta}) \\ (2.2) \quad f(\mathbf{X}_h|\boldsymbol{\theta}) &= \sum_{Z_{i_0h}=\text{T,C,A,G}} \pi_{Z_{i_0h}} \prod_{j \in \text{node}(T) \setminus i_0} \sum_{Z_{jh}=\text{T,C,A,G}} P_{Z_{anc(j)h}, Z_{jh}}(b_j) \\ &\equiv \sum_{Z_{i_0h}=\text{T,C,A,G}} \pi_{Z_{i_0h}} f_0(\mathbf{X}_h|\boldsymbol{\theta}, Z_{i_0h}) \end{aligned}$$

と、表される (Felsenstein, 1981). すなわち、各サイトの尤度は、各枝における推移確率の積和の形をしている。ここで $\boldsymbol{\theta}$ は、分子進化速度行列を規定するパラメータ $\boldsymbol{\theta}_0$ と系統樹を構成する枝の長さ (期待置換数) \mathbf{b} からなる。 $\text{node}(T)$ 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。 $f_0(\mathbf{X}_h|\boldsymbol{\theta}, Z)$ は祖先における状態を所与としたときの条件付尤度である。尤度を最大化することにより、系統樹の形や枝の長さを最尤推定する。

3. 集団の大きさと分子進化速度

エイズ予防と治療のために、広く同性愛の男性に協力してもらい、半年に一度、定期検診をしている活動がある。 Shankarappa et al. (1999) は、エイズの兆候を示す抗体が確認された9人の患者について、血液サンプルの分析を行うことの許可を得、ウイルス配列を比較した。図3は、ある患者Aにおけるウイルスの env 遺伝子の分子系統樹で、別の患者Bのウイルス遺伝子配列を外群にとっている。枝の長さは、サイトあたり塩基置換数を表している。

先に述べたように、RNAウイルスは突然変異率がきわめて高く、感染してから発症にいたるまでの数年間でも数多くの突然変異が集団に蓄積する。このため、進化速度が不変とする分子時計を仮定して(2.2)式を最大化させることにより、分子進化速度と合体時間を同時推定することができる (Rambaut, 2000)。

また、合体時間の尤度((2.1)式)も、継時的なサンプリングを考慮に入れて、若干の修正を行う。いまから $k_0=0, k_1, \dots, k_s$ 世代前に、それぞれ n_0, n_1, \dots, n_s 個の遺伝子をサンプリングするとする。現在から過去に遡る合体過程全体として見ると、それぞれのサンプリング時点ごとに、それより後にサンプリングされた遺伝子の祖先にその時点でサンプリングされた遺伝子が加わる形をしている。

いま、 $k_{s'}$ 世代前において、それより後にサンプリングされた遺伝子の共通祖先が $c_{s'}$ 個存在していたとする。これにサンプリングされた遺伝子を加えた $\tilde{n}_{s'} = n_{s'} + c_{s'}$ が、この時点における標本サイズである。一つ前の $s'+1$ 回目のサンプリングの時点では、共通祖先の数は $c_{s'+1}$ になっているので、二つのサンプリングの間に $\tilde{n}_{s'} - c_{s'+1}$ 回の合体が起きたことになる。相次ぐ合体の時間間隔を $t_{\tilde{n}_{s'}}, t_{\tilde{n}_{s'}-1}, \dots, t_{c_{s'+1}+1}$ とし、 $t_{c_{s'+1}}^* = k_{s'+1} - k_{s'} - \sum_{j_{s'}=c_{s'+1}+1}^{\tilde{n}_{s'}} t_{j_{s'}}^{s'}$ とお

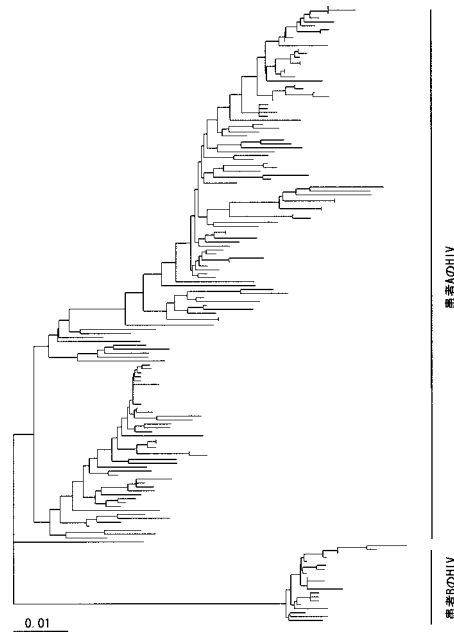


図3. エイズ患者から継時にサンプリングされた env 遺伝子配列の分子系統樹 (患者 B から部分サンプルを外群とした). 配列の継承番号: AF137629–AF138163, AF138166–AF138263, AF138305–AF138703 (Shankarappa et al., 1999).

くと, $\mathbf{t}^{s'} = (t_{\tilde{n}_{s'}}^{s'}, t_{\tilde{n}_{s'}-1}^{s'}, \dots, t_{c_{s'+1}+1}^{s'}, t_{c_{s'+1}}^{s'*})$ ($s' = 0, \dots, s$) の尤度が

$$(3.1) \quad L(N|\mathbf{t}^0, \mathbf{t}^1, \dots, \mathbf{t}^s) = \prod_{s'=0}^s \exp\left(-\frac{c_{s'+1}(c_{s'+1}-1)}{2N} t_{c_{s'+1}}^{s'*}\right) \prod_{j_{s'}=c_{s'+1}+1}^{\tilde{n}_{s'}} \frac{j_{s'}(j_{s'}-1)}{2N} \exp\left(-\frac{j_{s'}(j_{s'}-1)}{2N} t_{j_{s'}}^{s'}\right)$$

として表現される.

宿主に感染したウイルス集団は, 免疫系の攻撃を受ける. 均衡状態 (潜伏期間) が続いた後, 免疫系に打ち勝った集団は膨張し, 宿主は発症する. ここでは, この潜伏期間におけるウイルス集団について, 平均的な集団サイズと分子進化速度を解析する. 先行研究に基づき, ウイルスの世代の長さを仮に 1.47 日として, 9 人の患者に感染したウイルスについて集団の大きさを求めた (Seo et al., 2002).

遺伝子配列が得られると, 分子系統樹の対数尤度 ((2.2) 式) と合体過程の対数尤度 ((3.1) 式の対数をとったもの) の和として, 配列の尤度が書き下される. 集団の大きさの推定においては, 合体時間は局外母数である. ここでは簡単のため, プロファイル尤度のアプローチをとる. すなわち, まず (2.2) 式を最大化させて合体時間間隔の推定値を求め, これを合体過程の対数尤度に代入することにより, 近似プロファイル尤度を得る. これを最大化させることにより, 集団の大きさの推定値を得る. 推定量の分散は, 合体過程の確率誤差と分子系統樹の推定誤差からなる. ここでの状況を反映させたシミュレーションでは, 長さ 600 塩基の配列を 4 時点にわたり, 計 100 本サンプリングした場合, 集団の大きさは 5% 程度過大推定されることが確認された. この程度の偏りは, 次に見る解析結果の解釈を質的に変えるものではないが, 現在では局

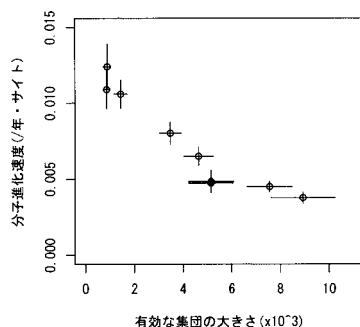


図 4. 9 人のエイズ患者における HIV の有効な集団の大きさと env 遺伝子分子進化速度. 線分は標準誤差を表す.

外母数に関する不確実性を取り込んだ階層ベイズモデルが提案されている (Drummond et al., 2002). 性能比較を将来の課題にする.

図 4 を見ると, 横軸のスケールは数千から 1 万程度である. しかし実際には, 宿主体内に存在するウイルス粒子の数は恐らくこれよりも桁桁も大きい. ウイルス粒子は雨あられと宿主細胞に衝突するが, そのうち首尾よく細胞に結合し, 侵入するものはほんの一握りである. この一握りのウイルスが, 数多くの子を残す. 私たちは, (2.1), (3.1) 式で表される尤度を最大化させることにより, 集団の大きさを推定した. そこでは, 各遺伝子について親は互いに独立に決まっていた. ところが, ウイルスの場合には, 多くの遺伝子が親を共有するのである. このため, 合体時間が予想外に短くなるのである. したがって, 遺伝子数 N がウイルス数に比してずっと小さくなったのである. 図 4 から, 有効な集団の大きさは分子進化速度と負の相関があることがはっきり読みとれる. 宿主内で勢力を持ち, 大きな集団では分子進化速度が小さいのに対し, 宿主の免疫力が強く, 細ったウイルス集団はより多くの変異を受け, 多様化している. こうして, 免疫系の網の目を潜り抜ける適応株の出現するチャンスを広げている.

4. 集団遺伝学の理論からの考察

集団の大きさに依存して分子進化速度を変化させ, 集団を頑健に適応させていくのは, ウイルスが獲得した巧みな生存戦略と映る. しかし, これは実は, 突然変異が確率的に起きていること, そしてその多くが有害であることから自然に導かれた帰結なのである.

先にも書いたように, ゲノムの情報は世代をまたいで受け継がれていく. その間にエラーが混入し, 変異を受ける. この突然変異が集団に多様性を導入する. 変異は多くの場合有害であるか, 適応度では集団中の他のものと差異は見られない. 後者は中立な突然変異という. まれに有利な突然変異が出現する. 有利な変異は他を押しつけて集団に固定されやすいが, 中立であるか有害な突然変異は, 集団から消滅する傾向がある. 従って, 遺伝的多様度は突然変異の起こりやすさとこれにかかる淘汰圧のバランスの上に成り立っている. 固定確率, 消滅確率は集団の大きさに依存する.

突然変異により集団内に変異体が生まれた場合を想定し, その後のダイナミクスを調べてみよう. 無作為交配する 2 倍体の集団で議論を進めるが, 無性生殖集団は, 遺伝子数を個体数に等しくすればよい. それまで集団を占めていた対立遺伝子 (a アレル) に対して, 突然変異体の A アレルの適応度は $1+s$ であったとする. 大きさ N の集団中の A アレルの割合 $X(t)$ の

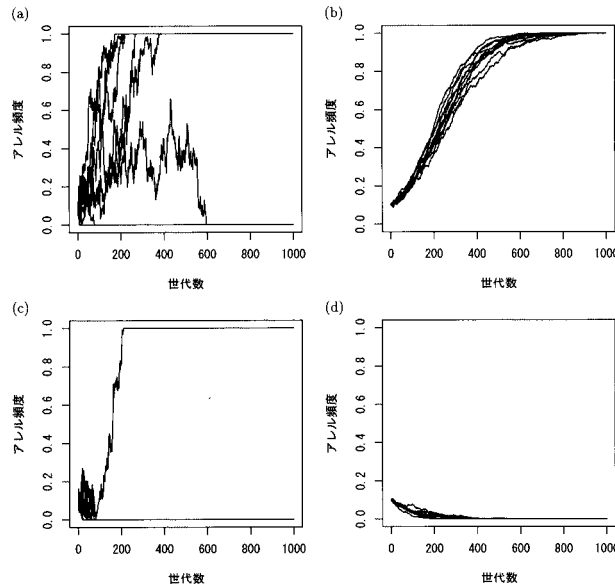


図 5. 適応度、集団の大きさと対立遺伝子頻度のダイナミクス. 初期頻度を 0.1 とし、1000 世代にわたる標本過程をそれぞれのケースで 10 本示している: (a) $s=0.01, N=100$, (b) $s=0.01, N=10000$, (c) $s=-0.01, N=100$, (d) $s=-0.01, N=10000$.

Δt における微小変化は、

$$E[X(t + \Delta t)|X(t)] = X(t) + sX(t)(1 - X(t))\Delta t + o(\Delta t)$$

$$V[X(t + \Delta t)|X(t)] = \frac{X(t)(1 - X(t))}{2N}\Delta t$$

で近似される. これより、確率微分方程式

$$(4.1) \quad dX(t) = sX(t)(1 - X(t))dt + \sqrt{\frac{X(t)(1 - X(t))}{2N}}dB_t$$

$$\equiv m(X(t))dt + \sigma(X(t))dB_t$$

を得る (Wright, 1931; Kimura, 1962). 第二項は二項分布のばらつきを表現しており、 B_t はブラウン運動である. 図 5 は、適応度が正の場合、負の場合おのおのについて、標本過程を 10 本ずつ描いたものである. 集団が大きくなると不確実性は減少し、適応度において勝るアレルは確実に集団に定着し、劣るアレルは集団から排除される. 小さな集団においては、勝る変異も集団から排除されたり、逆に劣る変異も集団に固定される可能性が出てくる. どちらのアレルも固定せず、集団が多型性を保つ時間も長い.

このことは、Kimura (1962) が導出した理論式から確認することができる. すなわち、集団中の確率 (アレル頻度) が x であったアレルが t 時間以内に集団に固定する確率 $u(t, x)$ は、(4.1) 式より偏微分方程式

$$\frac{\partial u(t, x)}{\partial t} = \frac{\sigma(x)^2}{2} \frac{\partial^2 u(t, x)}{\partial x^2} + m(x) \frac{\partial u(t, x)}{\partial x}$$

を満たす. したがって、いずれ集団に固定する確率 $u(x) = u(\infty, x)$ は、常微分方程式

$$\frac{\sigma(x)^2}{2} \frac{d^2 u(x)}{dx^2} + m(x) \frac{du(x)}{dx} = 0$$

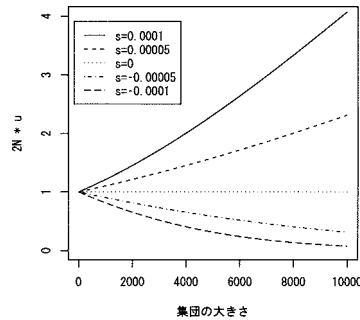


図 6. 突然変異の適応度と集団あたりの変異の定着率.

$$u(0) = 0$$

$$u(1) = 1$$

を満たす. この方程式の解は

$$u(x) = \frac{1 - e^{-4N_s x}}{1 - e^{-4N_s}}$$

と陽に書き下すことが出来る. 突然変異により生じたアレルの頻度は当初 $\frac{1}{2N}$ であるから, このアレルが集団に定着する確率は

$$u = u\left(\frac{1}{2N}\right) = \frac{1 - e^{-2s}}{1 - e^{-4N_s}} \sim \frac{2s}{1 - e^{-4N_s}} \quad (|s| \ll 1)$$

である. 突然変異率を μ と書くと, 分子進化速度は単位時間当たり期待突然変異数 $2N\mu$ に固定確率 u を乗じ, $r = 2N\mu u$ として求められる. 図 6 は, $2Nu$ を集団の大きさの関数として見たものである. 分子進化の多くが有利な突然変異により駆動されている場合には, 集団の大きさが大きいほど進化速度が速い. 他方, 突然変異の多くがまわりよりも劣る有害な突然変異である場合には, 集団の大きさが小さいほど分子進化速度が速い. 中立の場合 (Kimura, 1968) は, 分子進化速度は集団の大きさによらない. こうしてみると, 図 4 に見られた分子進化速度と集団の大きさの間の負の相関は, この後者のシナリオと矛盾しないことになる (Ohta, 1972). 変遷する環境に適応することの本質は, 無数のノンストラテジックな変異が作り出す多様性にあったのである.

5. 配列進化と構造進化

コード領域におけるコドンの変異は, アミノ酸の置換を引き起こす. アミノ酸置換が起きると, 一般にタンパク質の構造が影響を受ける. しかし, 立体構造を不安定化する置換は, 集団から排除される. このため, 配列進化に比べ構造進化のスピードは遅い. ただし, 今問題にしている V3 ループは, 構造上柔らかく縛りもないことから, 変異を受けやすい. したがって, 感染後配列進化のスピードと微細構造の変化の大きさは, 正に相関している可能性がある. このことを, 先の Shankarappa らのデータで確認してみる.

タンパク質の立体構造は, 結晶化させて X 線を照射し, その回折の様相を分析したり, 磁場と原子核の間の共鳴現象を分析したりすることにより, 精密に測定される. しかし, これには多大な労力と時間を要するため, 集団からの標本を数多く測定することができない. そこで,

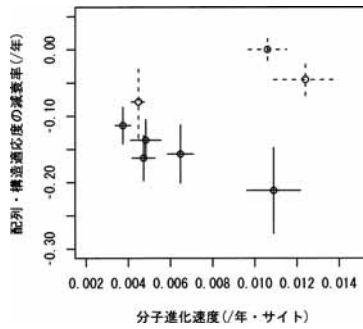


図 7. 8 人のエイズ患者における env 遺伝子分子進化速度と V3 ループの配列・構造適合度の減衰率. 線分は標準誤差を表す: 配列・構造適合度が有意に低下した 5 人の患者のウイルス集団は実線で, 他の 3 人の患者のウイルス集団は破線で示している.

タンパク質の構造データベースに見られる特徴を利用して, V3 ループの微細構造の変化の割合を間接推定する.

構造データベースには, タンパク質を形作るアミノ酸の位置情報が登録されている. そこで, これを下に, アミノ酸残基の埋もれ度合などの詳細な局所環境別に選好されるアミノ酸を, 頻度の形で求めることができる. 主鎖のみの位置関係の安定性を大まかに掴むのであれば, 配列を構成する各アミノ酸 $A = (a_1, \dots, a_n)$ について, その α 炭素の周辺 10 オングストローム以内にある α 炭素の数で局所環境 $E = (e_1, \dots, e_n)$ を定義する. 立体構造に照らし併せた配列の適合度 SSF (sequence-structure fitness) を, 構造環境下における配列の尤度

$$(5.1) \quad P(A|E) \sim \left[\prod_{i=1}^n p_1(a_i|e_i) \right] \left[\prod_{j < k} \frac{p_2(a_j, a_k | d_{jk}, e_j, e_k)}{p_1(a_j|e_j)p_1(a_k|e_k)} \right]$$

で指標化する. 第 1 項は局所環境におかれたアミノ酸の選好度, 第 2 項は近接アミノ酸対の選好度を表現しており, $p_1(a|e)$, $p_2(a, a' | e, e')$ はそれぞれ構造データベースから得られる局所環境下のアミノ酸頻度とアミノ酸対の同時分布, d_{ij} は 2 つのアミノ酸の間の距離である (Simons et al., 1999; Watabe et al., 2006).

構造データベースには, V3 ループの構造が, 不完全データも含め 5 つ収められていた. そこで, 患者から集められた配列とこれらの構造との平均的な適合度をはかり, 潜伏期間における時系列変化を調べた. 9 人の患者のうち 1 人における配列は, V3 ループにアミノ酸の挿入が検出され, 今回の解析の枠組みでは構造との関連付けに曖昧さを伴うため, 解析からはずした. そして, 残る 8 人の患者について, V3 ループの配列・構造適合度を感染後の年数で対数線形回帰させ, その減衰率を求めた. 図 7 は, これを env 遺伝子配列の分子進化速度と対比させたものである. 配列レベルでは大きく多様化しつつも V3 ループの構造は変化しないものと, 遺伝子配列の変化とともに構造も変化していることが想定されるものとの大きく 2 つの集団に分かれていることが読み取れる. 後者については, 分子進化速度が速いものほど, 配列・構造適合度の減衰率が大きく, 配列進化が素地となって構造変化の可能性を高めていることが想像される.

さらに直接的に環境の淘汰圧を受けるのは, 抗体との結合能, 宿主細胞受容体との結合能である. 通常, 抗体との結合部位は宿主細胞受容体との結合部位に隣接するか, あるいは重なり合っている. したがって, 抗体との結合能を弱めるような突然変異の多くは宿主細胞受容体と

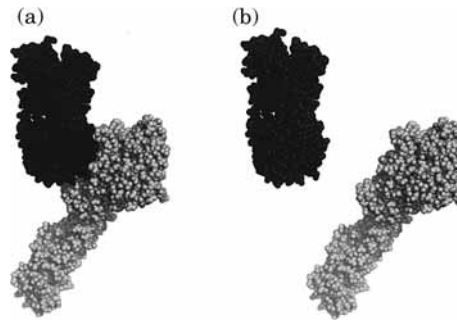


図8. インフルエンザ HA (明灰色) とそれに対する抗体 (BH151, 暗灰色). (a) 結合状態, (b) 遊離・自由状態.

の結合能をも低下させる可能性がある. 後者を弱めることなく前者を低下させることが, ウイルスの適応戦略と言える.

互いに相互作用する二種類のタンパク質 A および B 結合能は, 平衡状態における結合状態にあるタンパク質複合体の濃度 $[AB]$ と互いに遊離して自由状態にある両タンパク質の濃度の積 $[A][B]$ の比で定義される. ところで, 統計物理における確率分布による分子の速度分布のモデリングからの類推によれば, タンパク質の濃度は配列の構造に対する適合度を表現する尤度 ((5.1) 式) としてモデル化される. すると, 結合能は, 結合状態と自由状態の尤度比で表現できよう (Watabe et al., 2007).

図8 (a) は, 1968 年香港風邪をもたらしたインフルエンザウイルスの赤血球凝集素 (HA) とこれに結合する抗体 (BH151) の立体構造である. これに対して, 同図 (b) では HA と抗体が自由状態にある. 自由状態においては, HA と抗体の結合部位がそれぞれ表面に露出する. したがってたとえば, この部位において, 表面に露出した状態が自然なアミノ酸への置換が起きれば, 結合能は下がることになる.

このインフルエンザ分離株について, HA と 4 つの抗体 Fab との複合体の構造が得られている. 4 つの Fab のうち 2 つは宿主のレセプターとの結合領域にかぶった場所と結合し, 直接的に宿主細胞への侵入を阻止する. 他の二つの Fab はレセプターとの結合部位とは離れたところと結合し, 間接的にレセプターとの結合を阻止する. これら 4 つの Fab との結合部位について, 抗原・抗体結合能の長期変化を解析した. その結果, 間接的に感染を阻止する抗体については, 結合が緩やかに低下するのに対し, 直接レセプター結合部位に結合する Fab については, HA は一旦結合能を失効させるが, その後ある期間を経て再び結合を許し再度許した結合を失効させていく様子が読みとれた. ウィルス集団が, 構造変化に伴うコストと免疫系からの回避による利得を巧みにバランスさせながら適応していく様子が示唆された.

6. まとめと考察

エイズウイルス, インフルエンザウイルスを材料として, 遺伝子型の進化と表現型の進化を結びつける統計手法について紹介した. RNA ウィルスは突然変異率が高く, 分子進化と適応進化を現在進行形の形でモニタリングすることができる. また, 生存力にはさまざまな要因が複雑に関係する高等生物と異なり, 単純な生活環を持つことから, 比較的少数の表現型が環境の淘汰圧への適応を規定している. 本稿では, その中の重要な形質のひとつである, 宿主細胞受容体あるいはコレセプターとの結合部位の微細構造に注目し, 集団中の配列進化が表現型の

多様性を高めることを見た。さらに、淘汰圧のターゲットとしての抗体との結合能をモデル化し、淘汰圧の強さが結合能のダイナミクスに影響を与えることを見た。確率的に起きる突然変異のそれぞれに対して、今後適応度を評価することが可能となってくるであろう。これにより、たとえば、ウイルス集団の集団遺伝情報をモニタリングしながら、それが淘汰圧の種類と強さに応じてどのように適応していくか、予測することが可能になってくるであろう。そこでは、突然変異体の適応度を統計的にモデリングし、精度良く推定することが鍵となるであろう。

参 考 文 献

- Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data, *Genetics*, **161**, 1307–1320.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach, *Journal of Molecular Evolution*, **17**, 368–376.
- Huang, C.-C., Tang, M., Zhang, M.-Y., Majeed, S., Montabana, E., Stanfield, R. L., Dimitrov, D. S., Korber, B., Sodroski, J., Wilson, I. A., Wyatt, R. and Kwong, P. D. (2005). Structure of a V3-containing HIV-1 gp120 core, *Science*, **310**, 1025–1028.
- Kimura, M. (1962). On the probability of fixation of mutant genes in a population, *Genetics*, **47**, 713–719.
- Kimura, M. (1968). Evolutionary rate at molecular level, *Nature*, **217**, 624–626.
- Kingman, J. F. C. (1982). On the genealogy of large populations, *Journal of Applied Probability*, **19A**, 27–43.
- Kuhner, M. K., Yamato, J. and Felsenstein, J. (1995). Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling, *Genetics*, **140**, 1421–1430.
- Ohta, T. (1972). Population size and rate of evolution, *Journal of Molecular Evolution*, **1**, 307–314.
- Rambaut, A. (2000). Estimating the rate of molecular evolution: Incorporating non-contemporaneous sequences into maximum likelihood phylogenetics, *Bioinformatics*, **16**, 395–399.
- Seo, T.-K., Thorne, J. L., Hasegawa, M. and Kishino, H. (2002). Estimation of effective population size of HIV-1 within a host: A pseudo-maximum likelihood approach, *Genetics*, **160**, 1283–1293.
- 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). Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection, *Journal of Virology*, **73**, 10489–10502.
- Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C. and Baker, D. (1999). Improved recognition of native-like protein structures using combination of sequence-dependent and sequence-independent features of proteins, *Proteins*, **34**, 82–95.
- Watabe, T., Kishino, H., Okuhara, Y. and Kitazoe, Y. (2006). Fold recognition of the HIV-1 V3 loop and flexibility of its crown structure during the course of adaptation to a host, *Genetics*, **172**, 1379–1383.
- Watabe, T., Kishino, H., Martins, L. O. and Kitazoe, Y. (2007). A likelihood-based index of protein-protein binding affinities with application to influenza HA escape from antibodies, *Molecular Biology and Evolution*, **24**, 1627–1638.
- Wright, S. (1931). Evolution in Mendelian populations, *Genetics*, **16**, 97–159.

Modeling Evolution of Genotypes and Phenotypes

Teruaki Watabe¹, Tae-Kun Seo² and Hirohisa Kishino²

¹Center of Medical Information Science, Kochi University

²Graduate School of Agricultural and Life Sciences, University of Tokyo

Molecular evolution drives evolution of phenotypes. Genetic diversity of a population depends on the balance between mutations and selections. This paper introduces our recent studies to make a bridge between molecular evolution and adaptive evolution of phenotypes. First, we analyze the molecular evolution of the HIV env sequences within hosts. With the model of the coalescent process of molecular evolution, we show the negative correlation between the effective population size and the evolutionary rate. This negative correlation is consistent with nearly a neutral model of molecular evolution. With the model of the sequence-structure fitness (SSF), we show that the rate of sequence evolution is correlated with the change in micro-structure of the V3 loop after infection. Finally, we develop a model to predict the binding ability of the protein complex in terms of the ratio of the SSF of the complex protein to that of the free-state protein. The dynamics of the ability of binding between the influenza HA sequences and four types of antibodies implied the cost for the adaptation.