

平成22年3月31日現在

研究種目： 基盤研究 (B)
 研究期間： 2007 ~ 2009
 課題番号： 19300094
 研究課題名 (和文) 抵抗性株の出現と動態の統計遺伝学的モデリングと階層ベイズ法の開発
 研究課題名 (英文) Development of statistical genetic models and hierarchical Bayes procedures to predict emergence and dynamics of resistant alleles
 研究代表者
 岸野 洋久 (KISHINO, HIROHISA)
 (東京大学・大学院農学生命科学研究科・教授)
 研究者番号：00141987

研究成果の概要 (和文) :

本研究はウイルスを主たる研究対象とし、寄生種の遺伝的多様性を評価し、抵抗性株の出現を予測する新規の方法を開発した。ゲノムの部位間で系統関係を対比し、ベイズ法で組換えを検出する方法を開発した。尤度の枠組みで遺伝子配列、立体構造とデータベースの情報を組み合わせ、ウイルスと受容体/抗体との結合能を見積もる方法を開発した。結合能の変化を個体群動態モデルの係数の変化に結び付け、宿主集団の間に固定する確率の高い突然変異のタイプと出現のタイミングを計算する方法を開発した。

研究成果の概要 (英文) :

Setting a virus as a main target of the research, we developed new statistical models and procedures to estimate its genetic diversity and to predict the emergence of resistant alleles. Our hierarchical Bayes model detects inconsistency of phylogenetic relation between the segments of the genomes and estimates the recombination history. By our knowledge-based measure of the binding ability of a protein complex, it became possible to predict the change in the values of the coefficients that specify a population dynamic model. As a result, it became possible to predict the types and timing of new resistant alleles that are likely to be fixed among host population.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	5,700,000	1,710,000	7,410,000
2008年度	4,600,000	1,380,000	5,980,000
2009年度	3,800,000	1,140,000	4,940,000
年度			
年度			
総計	14,100,000	4,230,000	18,330,000

研究分野：統計科学、統計遺伝学

科研費の分科・細目：情報学・統計科学

キーワード：MCMC, 立体構造データベースと尤度, 配列進化と構造進化, 分子系統樹, ベイズ階層モデル, 統計遺伝学, 適応度ランドスケープ, ゲノム組換え

1. 研究開始当初の背景

ヒトが周囲の生態系へ及ぼす影響が観察できるときは、これをモニタリングし、適正に管理することができる。急激な変化には警鐘を鳴らすことも可能である。ところが、生態系は目に見えないところで、ヒトの淘汰圧に適応し、時として人間社会の脅威となることがある。抗生物質に対するバクテリアの抵抗性、殺虫剤や農薬への昆虫の抵抗性、薬剤・抗体へのウィルスの抵抗株の出現などである。また、飼料滅菌の抗生物質が、回りまわって人の腸内細菌の多様度を下げていることも懸念されている。したがって、彼らの個体群動態を精確にモニタリングし、抵抗性株の出現を予測しつつ、適切な処方を検討することが喫緊の課題となっている。

2. 研究の目的

ヒトの淘汰圧を受けた寄生者の個体群動態は、簡単な数理モデルを用いて書き下すことができる。ところが実際には、個体群動態はパラメータに大きく影響される。淘汰圧の強さは実験的に評価することが可能であるが、感受性株から抵抗性株への突然変異率は本質的な因子でありながら、評価が困難である。無作為に起こる突然変異の中から抵抗性株が出現するハザードレイトを見積もることは、容易でない。このため現状では、個々のケースに対応して実質的な戦略を提示できていない。

そこで本研究では、統計遺伝学によるアプローチにより困難を克服することを目的とした。抵抗性株が出現し、他と競合しながら集団に定着し、成長するプロセスをモデリングし、シミュレートするためには、分子進化学と集団遺伝学による洞察が不可欠である。さらに、構造生物学による知見を尤度表現して統計モデルに取り込むことにより、ランダムな突然変異の中からタンパク質の微細構造の変化を通じて適応株が出現する確率とタイミングを予測するモデルを提案することを目的とした。

3. 研究の方法

(1) ゲノム組換えのベイズ階層モデル

ゲノムの組換えは、異なるゲノムをモザイク上に組み合わせることにより、時として大規模な多様化を達成し、適応進化の可能性を高める。ウィルスゲノムにおける隣接領域間の系統関係(トポロジー)の食い違いとして、

組換えの歴史を推測することができる。

本研究ではトポロジーの間の食い違いを組換え回数で説明する組換え距離を近似するアルゴリズムを開発した。トポロジー T が決まると、長さ n の配列 $\mathbf{X}=(\mathbf{X}_1, \dots, \mathbf{X}_n)$ の各サイトの尤度は、

$$f(\mathbf{X}_i | \boldsymbol{\theta}, T) = \sum_{Z_0} \pi_{Z_0} \prod_{j \in \text{node}(T) \setminus \sqrt{0}} \sum_{Z_j} P_{Z_{\text{anc}(j)}, Z_j} (t_{\text{anc}(j), j})$$

と、分子進化のマルコフ過程モデルを用いて、推移確率の積として表現される。サイトごとにトポロジーを割り当てる。組換え距離 $\mathbf{d}=(d_1, \dots, d_{n-1})$ に

$$P(\mathbf{d} | \lambda) = \prod_{i=1}^{n-1} e^{-\lambda} \frac{\lambda^{d_i}}{d_i!}$$

と、ポアソン分布の事前分布を導入することにより、隣接サイト間のトポロジーの食い違いに対するペナルティを課す。これにより、超多母数問題における推定の不確実性を抑えつつ、稀な組換えのみを抽出する縮小推定量を得る(図1c)。MCMCにより、組換え位置と組換え回数、配列の各サイトのトポロジーの事後分布を求める。

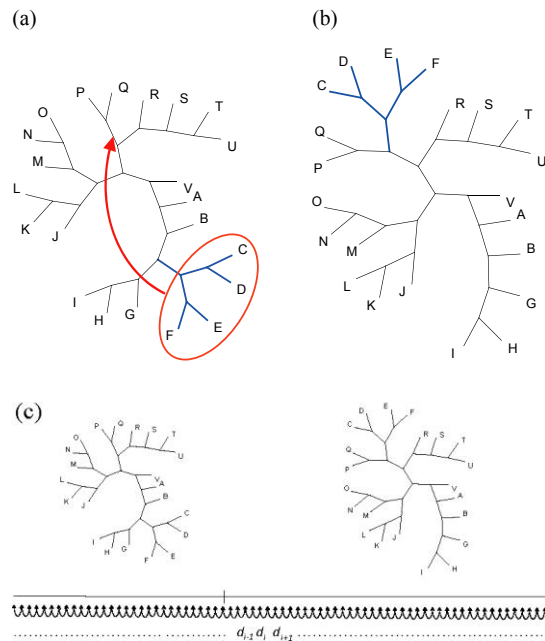


図1 ゲノム組換えとトポロジーの食い違い。系統樹(b)は系統樹(a)の部分系統樹を移す作業を1回行うことにより得られる。(c) 隣接サイト間の組換え回数に対するペナルティを組換え距離にポアソン分布を導入するベイズ階層モデルにより実現。

(2) ウイルスの宿主内動態のモデリングと突然変異の固定確率

図2はSARSウイルスが抗体および宿主受容体と結合しているスナップショットを重ね合わせたものである。実際にはウイルスの結合相手は抗体か宿主受容体のいずれかである。見方を変えると、この両者がウイルスをめぐって席取り競争を行っている。ウイルスV、抗体W、正常細胞X、感染細胞Yの宿主内における個体群動態は、数理モデル

$$\begin{aligned}\frac{dX}{dt} &= \lambda - dX - \beta XV \\ \frac{dY}{dt} &= \beta XV - aY \\ \frac{dV}{dt} &= kY - uV - \beta XV - qVW \\ \frac{dW}{dt} &= rV - hW - qVW\end{aligned}$$

により記述される。突然変異により宿主受容体、あるいは抗体との結合部位における形状や疎水度、電荷が変化すると、結合能 β, q が β', q' へと変化する。突然変異の適応度は、変異株、野生株の宿主内サイズの比から

$$s = \frac{\bar{V}(\beta', q')}{\bar{V}(\beta, q)} - 1$$

と定義され、変異株の固定確率は集団遺伝の拡散方程式を解いて

$$u(s; p) = \frac{1 - e^{-(2N-1)\log(1+sp)}}{1 - e^{-(2N-1)\log(1+s)}}$$

と求められる。ここで p は初期頻度である。ところで、変異株はまず宿主内で野生株との競争にさらされる。そこで、野生株と変異株の宿主内動態を数理モデルで記述し、変異株が1次感染で受け継がれる宿主個体の数の期待値を求める。これを初期頻度として、固定確率を計算する。

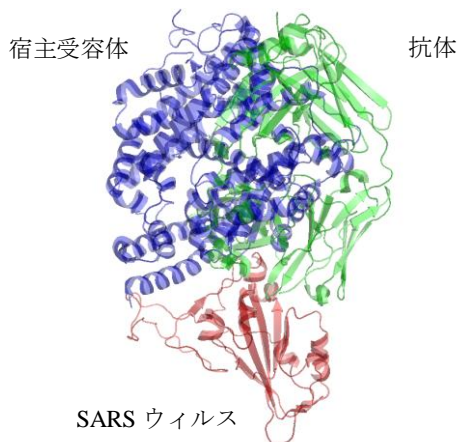


図2 SARSウイルスをめぐり抗体と宿主受容体の席取り競争

(3) 構造情報による経験尤度にもとづくタンパク質複合体の結合能の評価

タンパク質立体構造データベースPDBの情報から、周辺に何個のアミノ酸があるか、といった局所環境別にアミノ酸の頻度、アミノ酸対の頻度を計測する。これを下に、立体構造を所与とした配列の経験尤度を求める。図3は上述のSARSウイルスで、自由状態にある場合と抗体に結合した状態を対比している。結合部位のアミノ酸は結合状態では内部に埋もれているのに対し、自由状態では表面に露出している。2状態における配列の対数尤度比により複合体の結合能を得る。これにより、突然変異により数理モデルの係数がどのように変化するか、定量的に予測することが可能となる。

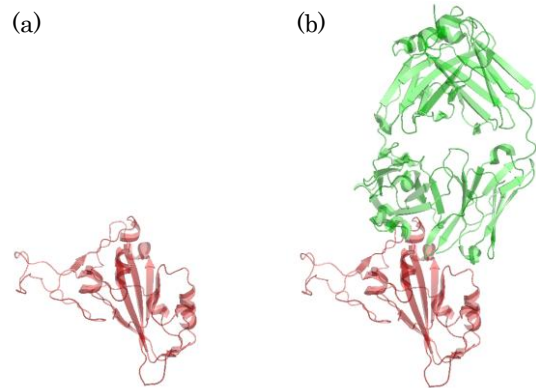


図3 SARSウイルスと抗体。自由状態(a)と、抗体と結合した状態(b)の配列の経験尤度を比較することにより、結合能を評価する。

4. 研究成果

(1) ゲノム組換えのベイズ推定：シミュレーションによる有効性の検討とHIV-1集団の解析

シミュレーションにより、新規に開発した組換え距離の計算アルゴリズムが、トポロジー間の組換え回数を偏りなく推定することを確認した。また、遺伝子配列内の3か所で組換えのあるシミュレーションにより、組換え位置が偏りなく推定されることが示された。さらに、隣接領域間のトポロジーの近さに関する制約を組換え距離の事前分布の形で導入したことにより、配列の各領域における系統関係の推定精度は既存の方法を凌駕することが示された。そこで、このモデルを南アメリカのHIV-1集団に適用し、同一サブタイプ内でも頻りに組換えが起きていることを見出すとともに、ゲノム上高頻度で組換えの起きている位置を新たに検出した。

(2) 経験尤度に基づく複合体結合能の指標:妥当性評価とインフルエンザウィルスHAの解析

TEM-1 β -Lactamase はペニシリンを含む β -ラクタム系抗生物質を加水分解する。これに対する阻害タンパク(BLIP)についてアラニン置換を行い、結合能の変化を測定して阻害に寄与する部位を調べた実験がある。そこでこの測定結果と上記経験尤度に基づく方法により得られた値を比較し、提案する指標の妥当性を確認した。また、PDBに登録されている746の複合体について指標値を計算し、結合領域と結合能の間の関数関係を測定した。

さらに、インフルエンザHAの配列・構造データに本手法を適用し、有効性を検証した。1968年に大流行を引き起こした香港風邪のインフルエンザHAについて、これと結合する4つの抗体の複合体立体構造が測定されている。このうち2つは宿主細胞の受容体との結合部位に隣接した領域に結合することにより、間接的にウィルスの宿主細胞への侵入を阻止する。他の2つは、受容体との結合部位に重なる形で結合し、直接的に侵入を阻止する。

大流行につながる突然変異の傾向を調べることが目的として、HAの配列の継時データが得られている。配列情報は比較的容易に測定できるのに対し、構造情報の測定は時間と労力を要する。このため、その時系列変化を直接観察することはできない。そこで、上記指標により結合能の長期変動を推定した。その結果、受容体との結合部位と隣接した領域に結合する抗体との結合能はあまり変化しなかったのに対し、受容体との結合部位にかぶる形で結合する抗体との結合能は、1968年以降大きく低下し、その後揺り戻しを持つ、ダイナミックな長期変動をした。受容体との結合能のある程度の低下というコストを伴いながら抗体の攻撃へ適応をしていることが窺われた。

(3) 突然変異の結合能の評価と集団への固定確率の予測

SARS ウィルスのスパイクタンパクについて、2つの抗体との複合体、および宿主受容体との複合体の構造が測定されている。そこでこの構造情報を下に、上述の方法で野生株の結合能 β, q と変異株の結合能 β', q' を推定した。これらを係数値にもつ数理モデルを解くことにより野生株に対する変異株の適応度を評価し、固定確率を計算した。その結果、突然変異のうちいくつかは、ワクチン接種を受けた宿主集団において高い確率で固定されることが予測された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計23件)

- ① Watabe, T. and Kishino, H. (2010). Structural considerations in the fitness landscape of a virus. *Molecular Biology and Evolution*. in press. (査読有)
- ② Miyashita, S. and Kishino, H. (2010). Estimation of the size of genetic bottlenecks in cell-to-cell movement of Soil-borne wheat mosaic virus (SBWMV) and the possible role of the bottlenecks in speeding up selection on variation in trans-acting genes or elements. *Journal of Virology*. **84**: 1828–1837. (査読有)
- ③ de Oliveira Martins, L. and Kishino, H. (2009). Distribution of distances between topologies and its effect on detection of phylogenetic recombination. *Annals of the Institute of Statistical Mathematics*. **62**: 145–159. (査読有)
- ④ Seo, T.-K. and Kishino, H. (2009). Statistical comparison of nucleotide-, amino-acid-, and codon-substitution models for the evolutionary analysis of protein-coding sequences. *Systematic Biology*. **58**: 199–210. (査読有)
- ⑤ de Oliveira Martins, L., Leal, É. and Kishino, H. (2008). Phylogenetic detection of recombination with a Bayesian prior on the distance between trees. *PLoS ONE*. **3** (7): e2651. (査読有)
- ⑥ Seo, T.-K. and Kishino, H. (2008). Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Systematic Biology*. **57**: 367–377. (査読有)
- ⑦ Kitada, S., Kitakado, T., and Kishino, H. (2007). Empirical Bayes inference of pairwise F_{ST} and its distribution in the genome. *Genetics*. **177**: 861–873. (査読有)
- ⑧ Tanabe, K., Escalante, A., Sakihama, N., Honda, M., Arisue, N., Horii, T., Culleton, R., Hayakawa, T., Hashimoto, T., Longacre, S., Pathirana, S., Handunnetti, S., and Kishino, H. (2007). Recent independent evolution of *msp1* polymorphism in *Plasmodium vivax* and related simian malaria parasites. *Molecular & Biochemical Parasitology*. **156**: 74–79. (査読有)
- ⑨ Thorne, J. L., Choi, S. C., Yu, J., Higgs, P. G., and Kishino, H. (2007). Population genetics without intraspecific data. *Molecular Biology and Evolution*. **24**: 1667–1677. (査読有)

- ⑩ Watabe, T., Kishino, H., de Oliveira Martins, L., 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. (査読有)

[学会発表] (計 26 件)

- ① Kishino, H. (代表) Co-evolution of proteins and covariation of evolutionary rate. 生物物理学会. 2009年10月30日. 徳島文理大学
- ② 岸野洋久(代表). 感染症研究における統計遺伝学的アプローチの可能性. 統計関連学会. 2009年9月7日. 同志社大学
- ③ Seo, T.-K. (代表) and Kishino, H. Statistical Comparison of nucleotide-, amino-acid-, and codon-substitution models for evolutionary analysis of protein-coding sequences. Evolution 2009, the joint annual meeting of the Society for the Study of Evolution (SSE), the Society of Systematic Biologists (SSB), and the American Society of Naturalists (ASN). 2009年6月16日. University of Idaho
- ④ Watabe, T. (代表) and Kishino, H. Structural consideration on the fitness-landscape of virus. Society for Molecular Biology and Evolution. 2009年6月7日. University of Iowa
- ⑤ Kishino, H. (代表) Measuring diversity and evolution through statistical modeling. Workshop on Measuring Evolution. 2008年10月3日. ノルウェイ科学アカデミー, Oslo, Norway
- ⑥ 岸野洋久. (代表) ゲノム進化と変異の統計解析. 応用統計学会. 2008年6月7日. 筑波大学
- ⑦ Kishino, H. (代表) Using a database and a prior for the estimation of structural evolution and recombination. Phylogenetic workshop at the 15th HIV Evolution and Dynamics meeting. 2008年4月27日. Los Alamos国立研究所, Santa Fe, U. S. A.
- ⑧ de Oliveira Martins, L. (代表), Leal, E. S., 岸野洋久. 系統樹間の距離とゲノム組換えのベイズ推定. 統計関連学会. 2007年9月7日. 神戸大学

[図書] (計 1 件)

- ① 岸野洋久 (2008). 分子進化の統計科学. (21 世紀の統計科学 II : 自然・生物・健康の統計科学 (小西貞則・国友直人編)). 東京大学出版会

6. 研究組織

(1) 研究代表者

岸野 洋久 (東京大学・農学生命・教授)
研究者番号 : 00141987

(2) 研究分担者

渡部 輝明 (高知大学・医・講師)
研究者番号 : 90325415
北添 康弘 (高知大学・医・名誉教授)
研究者番号 : 90112010

(3) 研究協力者

Jeffrey L. Thorne (ノースカロライナ州立大学・バイオインフォマティクス研究センター・教授)
Leonardo de Oliveira Martins (ヴィゴ大学・生化遺伝・ポスドク)
Élcio de Souza Leal (パラ連合大学・バイオテクノロジー・教授)