氏　　名　鈴木善幸

学位（専攻分野）　博士(理学)

学位記番号　総研大乙第78号

学位授与の日付　平成１２年３月２４日

学位授与の要件　学位規則第４条第２項該当

学位論文題目　Molecular evolution of pathogenic viruses

論文審査委員　主　査　教　授　池村　淑道
教　授　舘野　義男
教　授　石濱　明
助教授　溝上　雅史（名古屋市立大学）

**Abstract**

The molecular evolutionary analyses have been conducted to clarify the evolutionary mode and history of pathogenic viruses. The evolutionary mode and history include (1) the phylogenetic relationships, (2) the rates of nucleotide substitutions, (3) the divergence times, (4) the patterns of nucleotide substitutions, and (5) the natural selection.

In Chapter I, the significances of analyzing the above subjects are summarized. (1) The investigation of the phylogenetic relationships among virus strains is known as the molecular epidemiology. Once the phylogenetic relationships among virus strains are established, it is possible to identify the transmission routes of the virus within human population. The identification of the transmission route is then useful to infer the possible transmission mode of viruses. The investigation of the phylogenetic relationships among virus strains is also useful to clarify the geographical origin of viruses. Moreover, the comparison of the phylogenetic relationships among virus strains obtained from various host species with the phylogenetic relationships among the host species may indicate the possible occurrence of interspecies transmissions.

(2) The studies of the rate of nucleotide substitutions for various viruses clarified that the RNA viruses can be divided into two categories, according to their rates of nucleotide substitutions. The first category consists of the rapidly evolving RNA viruses with the rate of nucleotide substitutions of the order of $10^{-3}$ to $10^{-4}$ per site per year. The second category includes the slowly evolving RNA viruses with the rate of nucleotide substitutions of the order of $10^{-6}$ to $10^{-7}$. It implies that the evolutionary theories so far proposed can be tested experimentally using rapidly

evolving RNA viruses. The evolutionary rate of viruses is also useful to predict the possibility of developing effective vaccines against viruses.

(3) Applying the rate of nucleotide substitutions to the phylogenetic tree reconstructed for virus strains, the divergence times among virus strains can be estimated. The comparison of the divergence times among virus strains with the divergence times among their host species indicates the possible interspecies transmission of viruses.

(4) In general, the exact knowledge of the pattern of nucleotide substitutions for a particular organism is important to choose appropriate nucleotide substitution models in the molecular evolutionary analyses for that organism. The study for the pattern of nucleotide substitutions for viruses is also useful for developing new drugs, particularly nucleotide analogues, against virus infections.

(5) The factors determining the mode of molecular evolution include the mutation rate, the random genetic drift, and the natural selection. The mutation rates for the rapidly evolving RNA viruses seem to be more than million times faster than the mutation rate for humans, as is the case for the rate of nucleotide substitutions. According to the neutral theory of molecular evolution, the great majority of evolutionary changes at the molecular level are caused not by positive selection but by random drift of selectively neutral or nearly neutral mutants. However, positive selection operating at the amino acid sequence level has been detected on many protein coding genes of viruses.

In Chapter II, studies on human T-cell lymphotropic virus types I (HTLV-I) and II (HTLV-II) are briefly reviewed from the viewpoint of molecular evolution, with special reference to the evolutionary rate and evolutionary relationships among different isolates of these viruses. In particular, it appears that, in contrast

to the low level of variability of HTLV-I among different isolates, individual isolates form quasispecies structures. Elucidating the underlying mechanisms of these two phenomena will be one of the future problems in the study of the molecular evolution of HTLV-I and HTLV-II.

In Chapter III, with the aim of elucidating evolutionary features of GB virus C/hepatitis G virus (GBV-C/HGV), molecular evolutionary analyses were conducted using the entire coding region of this virus. In particular, the rate of nucleotide substitutions for this virus was estimated to be less than $9.0 \times 10^{-6}$ per site per year, which was much slower than those for other RNA viruses. The phylogenetic tree reconstructed for GBV-C/HGV, by using GB virus A (GBV-A) as outgroup, indicated that there were three major clusters (the HG, GB, and Asian types) in GBV-C/HGV, and the divergence between the ancestor of GB and Asian type strains and that of HG type strains first took place more than 7,000-10,000 years ago. The slow evolutionary rate for GBV-C/HGV suggested that this virus cannot escape from the immune response of the host by means of producing escape mutants, implying that it may have evolved other systems for persistent infection.

In Chapter IV, molecular evolutionary analyses for Ebola and Marburg viruses were conducted with the aim of elucidating evolutionary features of these viruses. In particular, the rate of nonsynonymous substitutions for the glycoprotein (GP) gene of Ebola virus was estimated to be, on the average, $3.6 \times 10^{-5}$ per site per year. Marburg virus was also suggested to be evolving at a similar rate. Those rates were a hundred times slower than those of retroviruses and human influenza A virus, but were of the same order of magnitude as that of hepatitis B virus. When these rates were applied to the degree of sequence divergence, the divergence time between Ebola and Marburg viruses was estimated to be more than

several thousand years ago. Moreover, most of the nucleotide substitutions were transitional and synonymous for Marburg virus. This observation suggests that purifying selection has operated on Marburg virus during evolution.

In Chapter V, a method was developed for detecting the selective force at single amino acid sites, given a multiple alignment of protein coding sequences. The phylogenetic tree was reconstructed using the number of synonymous substitutions. Then, the neutrality was tested for each codon site using the numbers of synonymous and nonsynonymous changes throughout the phylogenetic tree. Computer simulation showed that this method estimated accurately the numbers of synonymous and nonsynonymous substitutions per site, as long as the substitution number on each branch was relatively small. The false positive rate for detecting the selective force was generally low. On the other hand, the true positive rate for detecting the selective force depended upon the parameter values. Within the range of parameter values used in the simulation, the true positive rate increased as the strength of the selective force and the total branch length, namely the total number of synonymous substitutions per site, in the phylogenetic tree increased. In particular, most of the positively selected codon sites, with the relative rate of nonsynonymous substitution to synonymous substitution being 5.0, were correctly detected when the total branch length in the phylogenetic tree was 2.5 or more. When this method was applied to the *human leukocyte antigen* (*HLA*) gene, which included antigen recognition sites (ARSs), positive selection was detected mainly on ARSs. This finding confirmed the effectiveness of the present method with actual data. Moreover, two amino acid sites were newly identified as positively selected in non-ARSs. Three-dimensional structure of the HLA molecule indicated that these sites might be involved in

antigen recognition. Positively selected amino acid sites were also identified in the envelope protein of human immunodeficiency virus and the influenza virus hemagglutinin protein. This method is helpful for predicting functions of amino acid sites in proteins, especially in the present situation that sequence data is accumulating at an enormous speed.

論文の審査結果の要旨

鈴木善幸氏の博士論文は、病原性ウイルスの進化機構および歴史を明らかにすることを目的に、各種ウイルスについての分子進化学的解析を行っている。特筆すべき成果の第一として、GBウイルスC/G型肝炎ウイルス（GBV-C/HGV）の解析により、このウイルスのゲノム全体における塩基置換速度が座位当り年当り$9.0\times10^{-6}$未満と推定し、RNAウイルスとしては非常に遅い進化速度を持つことを示した。アフリカ型、アメリカ・ヨーロッパ型、およびアジア型の３つの主要な型に分類し、アメリカ・ヨーロッパ型が7,000-10,000年以前に最初に分岐したと推定している。また、GBV-C/HGVは垂直感染をし、持続感染し、進化速度が遅いことから、ヒトの移動の歴史を解明するために有用であること、ならびにGBV-C/HGVに対する効果的なワクチンの作製が可能との予測を提唱している。審査委員会は、ヒトに持続感染する機構を解明する上での基礎知見を提供している価値の高い成果と判断した。様々なRNAウイルスにおける進化速度を比較し、RNAウイルスは座位当り年当り$10^{-3}$-$10^{-4}$という速い進化速度を示す高速進化型RNAウイルスと、座位当り年当り$10^{-6}$-$10^{-7}$という遅い進化速度を示す低速進化型RNAウイルスに分類できることを示した。審査委員会は、GBV-C/HGVを含む遅い進化速度を示すRNAウイルスの存在の指摘は、ウイルスの進化機構に新しい重要な課題を提供していると評価し、その分子機構に関するモデルの提唱やさらなる議論を期待するとの判断をも加えた。

特筆すべき成果の第二の点は、塩基配列の多重整列を用いて、各アミノ酸座位における自然選択を検出する方法を開発したことである。系統樹を作成後に、各コドン座位について系統樹全体における同義置換数と非同義置換数を推定、比較することにより中立性を検定している。コンピュータ・シミュレーションにより、方法の妥当性を検証した後に、ヒト白血球抗原（HLA）遺伝子に適用し、正の自然選択を主に抗原認識部位において検出しており、遺伝子の配列解析に有用であることを明らかにした。さらに抗原認識部位として報告されていなかった２つのアミノ酸座位においても正の自然選択を検出しており、HLA分子の立体構造からこれらの座位も抗原認識に関与している可能性を指摘している。ヒト免疫不全ウイルスの外被糖蛋白質やヒトインフルエンザA型ウイルスのヘマグルチニン蛋白質についても正の自然選択が働いているアミノ酸座位を同定し、主にエピトープに存在していることを明らかにした。配列データが飛躍的に増大している状況を考えた場合、正に時宜にかなった有力な方法論の開発と審査委員会は判断した。蛋白質中のアミノ酸座位の機能推定などに有用になっていくとも考えられる。

これらの研究成果を含めて、病原性ウイルスの進化機構に関して１０編の原著論文を国際誌に発表しており、この分野への貢献が多大と評価し、審査委員一同は博士論文としての条件を充分に満たすと判断した。