氏名	伊　奈　康　夫
学位（専攻分野）	博士（理学）
学位記番号	総研大甲第８４号
学位授与の日付	平成　６　年　３　月２４日
学位授与の要件	生命科学研究科　遺伝学専攻 学位規則第４条第１項該当
学位論文題目	Estimation of the numbers of synonymous and nonsynonymous substitutions with special reference to viral evolution
論文審査委員	主　査　教　授　原　田　朋　子
	教　授　池　村　清　久
	教　授　高　畑　尚　之
	助教授　舘　野　義　男
	教　授　堀　　寛（名古屋大学）

論文内容の要旨

In studies of molecular evolution, it is of great importance to estimate the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per site separately. In particular, accurate estimation of d_S and d_N is essential for a statistical test for the neutral theory of molecular evolution because $d_S \geq d_N$ is expected under the neutral mutation hypothesis. Various methods for estimating d_S and d_N have ever been proposed. Among them, Miyata and Yasunaga's (MY) methods, Li, Wu, and Luo's (LWL) method, and Nei and Gojobori's (NG) method have been widely used. Recently, Pamilo and Bianchi and Li proposed another (PBL) method for estimating d_S and d_N. However, properties of these methods are poorly understood.

In this study, first, I proposed a method of computer simulations to evaluate the accuracies of methods for estimating d_S and d_N. Moreover, I developed new methods for estimating d_S and d_N. The new methods take into account a transition/transversion-bias. By the simulation method, I evaluated the accuracies of the MY, LWL, NG, and PBL methods and the new methods. In addition, by the NG and PBL methods and the new methods, I analyzed statistically nucleotide sequences of the hemagglutinin 1 (HA1) gene of human influenza A viruses (H1 subtype). In the analysis, the nucleotide sequences were divided into antigenic and nonantigenic sites, and then the nucleotide diversities for antigenic and nonantigenic sites of the HA1 gene were computed at synonymous and nonsynonymous sites separately. The main purpose of the analysis is to clarify whether or not positive selection operates on antigenic sites of the HA1 gene. I summarize the results obtained in this study as follows.

(1) The MY, LWL, and NG methods give overestimates of d_S and underestimates of d_N. This result calls for reexaminations of some genes. This is because evolutionary pictures of genes have often been discussed on the basis of results obtained by the MY, LWL, and NG methods, which are favorable for the neutral theory of molecular evolution. The major cause for the biased estimation of d_S and d_N is that these three methods underestimate the number (S) of synonymous sites and overestimate the number (N) of nonsynonymous sites. The biased estimation of S and N result from the assumption of equal mutation

rates of the four nucleotides. As the transition/transversion ratio at the mutation level increases, the extents of the biased estimation of d_S, d_N, S, and N become larger.

(2) The PBL method gives better estimates of d_S and d_N than those obtained by the MY, LWL, and NG methods. Although the PBL method originally does not estimate S and N, we can estimate S and N if we assume that nucleotide mutations follow Kimura's 2-parameter model. Under such an assmption, I derived formulae for estimating S and N by the PBL method. Estimates of S and N obtained by the formulae are better than those obtained by the MY, LWL, and NG methods.

(3) The new methods also give better estimates of d_S and d_N than those obtained by the MY, LWL, and NG methods. In addition, estimates of S and N obtained by the new methods are reasonably accurate. Moreover, the number of inapplicable cases is much smaller for the new methods than for the PBL method. This indicates that sampling variances of estimates of d_S and d_N are smaller for the new methods than for the PBL method.

(4) In some cases, the new methods and the PBL method give biased estimates of substitution numbers. However, from the number (d_3) of nucleotide substitutions at the 3rd position of codons, we can examine whether estimates obtained by the new methods are good or not. This is because estimates of d_S obtained by the new methods are strongly correlated with estimates of d_3 obtained by Kimura's 2-parameter method; the new methods give good estimates of d_S when estimates of d_3 obtained by Kimura's 2-parameter method are essentially the same for those obtained by methods which consider a nucleotide-frequency-bias (*e.g.*, Tajima and Nei's method, Gojobori, Ishii, and Nei's method, Tamura and Nei's method). On the other hand, we cannnot make an examination of estimates obtained by the PBL method. This is because no clear-cut relationships were observed between estimates of d_S obtained by the PBL method and estimates of d_3 obtained by Kimura's 2-parameter method, mutation patterns, or the degree of functional constraints.

(5) When there are strong transition/transversion- and nucleotide-frequency-biases like mitochondrial genes, all of the above methods give biased estimates of substitution numbers. In such cases, Kondo *et al.*'s method is

recommended to be used for estimating d_S, although their method cannot estimate d_N and is time-consuming.

(6) For nonantigenic sites of the HA1 gene of human influenza A viruses (H1 subtype), the nucleotide diversities were larger at synonymous sites than at nonsynonymous sites. This is consistent with the neutral theory of molecular evolution. For antigenic sites, however, the nucleotide diversities at nonsynonymous sites were larger than those at synonymous sites. Thee results suggest that positive selection operates on antigenic sites of the HA1 gene of human influenza A viruses (H1 subtyupe). For antigenic sites, the $\bar{d}_N/\bar{d}_S$ ratios within and between (sub) groups of human influenza A viruses were not the same, where $\bar{d}_S$ and $\bar{d}_N$ are the nucleotide diversities at synonymous and nonsynonymous sites, respectively. This result can be explained by the following possibilities: (i) Patterns of evolution of the HA1 gene varied from virus to virus. $N_e s$ may have changed with time, where N_e is the effective population size of human influenza A viruses and s is the selection coefficient. (ii) Even in antigenic sites of the HA1 gene, the fraction of nondeleterious mutations at nonsynonymous sites is not so large and nonsynonymous substitutions occur at restricted sites. If this is the case, the number of nonsynonymous substitutions between distantly related sequences is underestimated by methods which assume the uniform substitution rate among nonsynonymous sites. Although it is now unclear which of the above possibilities is correct, it is likely that positive selection operates on antigenic sites of the HA1 gene of human influenza A viruses (H1 subtype). Probably, some of amino acid changes in antigenic sites of the HA1 gene product are advantageous to escape of human influenza A viruses from host immune systems. If this is the case, the present study is the first report that clearly showed advantageous evolution of viruses at the molecular level by refined statistical methods of molecular evolutionary analyses.

論文の審査結果の要旨

遺伝子塩基配列の比較分析において、アミノ酸に変化をもたらす非同義置換とアミノ酸の変化をもたらさない同義置換とを分けて推定することは、自然淘汰の効果を検出する上できわめて有益な方法である。今までいくつかの方法が開発されていたが、いづれも単純化された塩基置換のモデルに基づいたものである。たとえばNei-Gojoboriの方法は突然変異がA、T、G、C、の間でランダムに起るという仮定に基づいている。実際の遺伝子進化ではしばしばtransitionがtransversionより起りやすく、上の仮定は成り立たない。そのため推定値はバイアスした値となる。出願者　伊奈康夫君は膨大な量のシミュレーションを行って既存の方法の検定を行うとともに、新しい方法を開発した。この方法では、実際のデータからtransitionとtransversionの起りやすさの違いを推定し、その上で同義置換と非同義置換の数を推定する。シミュレーションの結果既存のどの方法よりも新しく開発した方法の方がすぐれていることが分かった。

伊奈君はこの方法を用いて、ヒトのインフルエンザAウイルスのヘマグルチニン1の遺伝子のデータ解析を行った。この遺伝子には抗原部位が存在することが知られていて、この部位での進化が自然淘汰によるのか中立かについて明らかでない。伊奈君は抗原部位と他の領域にわけて、同義置換および非同義置換のパターンを調べた。その結果、抗原部位では非同義置換が同義置換を上まわり、自然淘汰が働いたことを推定できた。これらの結果はProc. Natl. Acad. Sci. およびJ. Mol. Evol. に投稿中である。以上の内容は学位論文として充分であると判断する。