

氏 名 間野 修平

学位（専攻分野） 博士（理学）

学位記番号 総研大乙第189号

学位授与の日付 平成20年9月30日

学位授与の要件 学位規則第6条第2項該当

学位論文題目 STOCHASTIC MODELS IN POPULATION GENETICS

論文審査委員	主査	特任教授	長谷川 政美
		准教授	印南 秀樹
		准教授	高野 敏行
		教授	五條堀 孝

論文内容の要旨

Stochastic models have played important roles in population genetics. They have given theoretical understanding on evolutionary mechanism of maintaining genetic diversity within and between species. Following in a line of Fisher (1930) and Wright (1945), in 1950-1980's Kimura and his coworkers had given foundations of evolutionary theories by developing stochastic models based on the diffusion process. By applying their theoretical predictions to emerging molecular data at that time, many important aspects of molecular evolution have been revealed so far. The most significant prediction is probably the neutral hypothesis of molecular evolution, which was advocated by Kimura (1968). In early 1980's, a stochastic model, which is now called the coalescent model, is introduced (Kingman, 1982; Tajima, 1983; Hudson, 1983). The coalescent process is a stochastic process of ancestors of a sample, which are taken from a population evolving under the diffusion model. The coalescent model has given a useful framework of statistical analysis of a sample taken from a population; tests based on the model and efficient simulation schemes to generate samples evolving under arbitrary hypothesis have been developed so far. In this dissertation, the author will present several analytical results on stochastic models in population genetics, obtained by the author and coworkers. These models cover various aspects, but with special reference to multi-locus diffusion models (Chapter 2, 3) and the relationship of the diffusion model and the coalescent model under selection (Chapter 4, 5). They are central issues under development in the current population genetics theory.

In chapter 2, effects of random genetic drift upon linkage disequilibrium are investigated in terms of a two-locus diffusion models. An analytic expression of conditional expectation of transient gamete frequency, given that one of the two loci remains polymorphic, is obtained by calculating the moments of the distribution. Using this expression, a model where linkage disequilibrium is introduced by a single mutant is investigated. The random genetic drift should have large important on the model, since the mutant is prone to disappear from a population. The conditional expectation of the gamete frequency given that the locus with the mutant allele remains polymorphic is presented. The behavior is significantly different from the monotonic decrease observed in the deterministic model without random genetic drift. Then, evolution of linkage disequilibrium of the founders in exponentially growing populations is investigated in terms of a time-inhomogeneous stochastic model, which is an extension of the diffusion approximation of the Wright-Fisher model. As a measure of linkage disequilibrium, the squared standard linkage deviation, which is defined by a ratio of the moments, is investigated. A system of ordinary differential equations that these moments obey is provided. In addition, by a perturbative series expansion in a growth parameter, an asymptotic formula for the squared standard linkage deviation after a large number of generations is obtained. According to the formula, the squared standard linkage deviation tends to be $1/(4Nc)$, where N is the current size of the population and c is the recombination rate between two loci. It depends on neither of the initial effective size of the population, the growth rate, nor the mutation rate. In exponentially growing populations, linkage disequilibrium will be asymptotically the same as that in a constant size population, the effective size of which is the current size.

In Chapter 3, evolutionary rates of duplicated genes under concerted evolution by gene conversion are investigated. Effects of directional natural selection and bias in conversion rate on fixation of a single mutant in a locus, where the mutant spreads in a multigene family by gene conversion, are investigated. For the directional selection, a model in which selection operates on the number of the mutant in a diploid is assumed. Because of gene conversion between loci, either the mutant or wildtype allele will eventually fix in all the loci. An analytic expression of the fixation probability is obtained in terms of a two-locus diffusion model. For the genic selection, the formula is given by the well known formula for the single-locus problem, replacing the effective population size by the twice and the initial allele frequency by arithmetic mean of the initial frequencies in the two loci. The expression depends on neither of the initial linkage disequilibrium, the recombination rate, nor the conversion rate. According to simulations, the simple correspondence between the formula for a single locus and for two loci holds when number of the locus is larger than two. For the biased gene conversion, an analytic expression of the fixation probability is obtained in terms of a n -locus diffusion model. With these formula of the fixation probabilities, effects of gene conversion on the rate of molecular evolution in a multigene family under concerted evolution are discussed. It is shown that selection and bias in conversion rate operate more efficiently in a large multigene family.

In Chapter 4, the ancestral selection graph, which is an analogue to the coalescent genealogy, is investigated. The number of ancestral particles, backward in time, of a sample of genes is an ancestral process, which is a birth and death process with quadratic death and linear birth rate. An explicit form of the number of ancestral particle is obtained, by using the density of the allele frequency in a diffusion model obtained by Kimura (1955). It is shown that fixation is convergence of the ancestral process to the stationary measure. The time to fixation of an allele is studied in terms of the ancestral process.

In Chapter 5, an approximate sampling formula for the infinite allele model at the end of a selective sweep is obtained, in terms of a weighted binomial mixture of the Ewens sampling formula. The approximate sampling formula is based on the hitchhiking model proposed by Maynard Smith and Haigh (1974). The formula will give a simple and useful framework for theoretical understanding of allele frequency distribution at the end of a sweep. By using the approximate sampling formula for the infinite allele model at the end of a sweep, a new likelihood based test to detect recent selective sweep is presented. Although the test seems slightly less powerful than the test based on the frequency of the most common allele when the mutation rate (size of the neutral region) is low, however, the test gives estimates of the selection coefficient and the position of the target of the selection.

論文の審査結果の要旨

本論文は、集団遺伝学の基礎理論を大きく発展させるものである。とくに、複数の遺伝子座が存在し、その間に自然選択などの相互作用があるという、非常に複雑なモデルにおけるいくつかの難問に対して、主に拡散理論を用いて解を与えている。集団遺伝学では、拡散理論の重要性は80年以上も昔から認識されてきたが、その難解さ故に最近の発展は乏しい。実情としては、現在の集団遺伝学ではコンピュータによる計算が主流になっている。そのような状況のもと、本研究の学術的価値は非常に高いものである。数式解は、あらゆるパラメータに対して、瞬時に解を与えるというコンピューターシミュレーションにはまねのできない利点と魅力があるからである。

論文ではまず、2遺伝子座モデルを用いて、連鎖不平衡が時間とともにどのように変化していくかという過程を調べている。これは、これまでの連鎖不平衡の理解を大きく飛躍させるものである。次に、重複遺伝子の進化において、遺伝子変換の役割を理論的に研究している。モデルは、単純な2遺伝子座モデルを用いて解析解を与えているが、そこから複数遺伝子座モデル（3遺伝子座以上）への発展も行っている。遺伝子変換の方向性のバイアスを加味したモデルも興味深い。さらには、自然選択を組み込んだ難解モデルに対しても、いくつかの解を与えている。ひとつは、Ancestral selection graph という、自然選択の可能性を考慮に入れた系図構築理論のもと、過去のどの時点でどのような系図構造になっているかを調べている。さらには、Selective sweep という、有用な突然変異が正の自然選択によって種内に固定する過程で起こる DNA 多型のパターンの変化を理論的に研究している。これは、DNA 多型データを理解する上で非常に重要な情報を提供するものである。

最近の DNA 多型データの蓄積にともない、集団遺伝学の重要性は、とくに医学、農学、生態学の分野で注目されている。ポストゲノム時代と呼ばれる現在の問題点は、有り余るデータの量に対して、解析が追いつかないということにある。このような現状のなか、集団遺伝学の基礎理論を大きく発展させた本研究の意義は計り知れないものがあり、世界的にも大きなインパクトを与えるものである。それを的確に記述した学位論文を厳密に審査した結果、審査委員全員一致で博士（理学）の学位にふさわしいと判断した。