

論文の内容の要旨

生産・環境生物学専攻

平成 27 年度博士課程入学

呉佳齊 (Jiaqi Wu)

指導教官：岸野洋久 教授

Statistical modeling of species and population level genomic evolution

(種と集団レベルのゲノム進化の統計モデリング)

Description and understanding the differences of the living organisms surrounding us is one of the central activities of human through histories. From era of Charles Darwin, especially owing to his publication of *On the origin of species* (1859), evolution became a basis that explains the diversity of living organisms. Darwin generated his theory based on large amount of observations of phenotypic variations, and emphasized nature selection is the main power for species evolution. In 1968, Motoo Kimura calculated the rate of genomic evolution and found it is much faster than Haldane's limit on the speed of beneficial evolution. This finding gave birth to neutral theory of molecular evolution. Neutral theory that declares mutations in molecular level are largely neutral became central idea of molecular evolution and population genetics. In genomic era, it is possible to develop a new approach to link the molecular evolution with the phenotypic variations of living organisms.

A variant of a gene or an allele can be considered as a state of a trait of a genome.

Summary statistics with careful statistical modeling can be applied to study the genomic variations. In this thesis, I summarize the current inference methods of evolutionary biology (Chapter 2.1-2.4) and propose a new model, which extends the neutral theory of molecular evolution to describe genome evolution (Chapter 2.5). Also I report the result of population genomic analysis based on the model of the joint allele frequency spectrum (Chapter 5). The former investigates species level genomic evolution and the latter investigates population level genomic evolution. Both studies describe models of the summary statistics to infer genomic evolution. The new model of species level genomic evolution enables robust estimation of the divergence times and links the gene variations and the life history traits evolution of mammals.

In Chapter 3 and 4, I formulate the models in more detail by analysing the evolutionary rates of 1,185 genes on a phylogeny of 89 mammals. In Chapter 3, I show the branch effect-based divergence time estimation approach provides robust estimate of divergence times. Remarkably, DNA, codon or protein-level analyses give the same result. Also this measures firstly the variation of genomic rate among species. In Chapter 4, I show the efficacy of the new rate-based ancestral state reconstruction approach. By using this new approach, I reconstructed the history of 10 discrete traits related to activity, diet and social behaviours. The results indicate the ancestor of placental mammals was solitary, seasonally breeding, insectivorous and likely nocturnal. The predictor genes of the traits can be automatically selected from genomic data, without relying on the pre-knowledge of gene annotations. This approach has a potential to link the Darwin's natural selection theory of phenotypic variations and Kimura's neutral theory of molecular evolution. The method with its case study introduced in this thesis will hopefully inspire the field of genomic study.