

氏 名 山下 永香

学位(専攻分野) 博士(理学)

学位記番号 総研大甲第 2370 号

学位授与の日付 2022 年 9 月 28 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6条第1項該当

学位論文題目 Population genomics of weak evolutionary forces: Base
composition evolution in coding and non-coding gene regions
in *Drosophila*

論文審査委員 主 査 北野 潤
遺伝学専攻 教授
齋藤 都暁
遺伝学専攻 教授
池尾 一穂
遺伝学専攻 准教授
藤戸 尚子
遺伝学専攻 助教
長田 直樹
北海道大学 大学院情報科学研究院 准教授

(Form 3)

Summary of Doctoral Thesis

Name in full Yamashita, Haruka

Title

Population genomics of weak evolutionary forces:

Base composition evolution in coding and non-coding gene regions in *Drosophila*

One of the central questions in evolutionary genetics is whether natural selection is a primary force acting on new mutations. However, our understanding is limited, especially when natural selection is weak, because of technical challenges. Detection of natural selection requires large sample sizes (numbers of DNA changes) and reliable ancestral inference methods. This study aims to address these limitations in order to understand the role of weak natural selection in molecular evolution.

Codon usage bias, non-random usage of synonymous codons, can be described in population genetic models that include fitness differences among codons. We can test model predictions given ancestral and derived states for observed changes. Because the model applies roughly across codon positions, synonymous changes from different genes can be pooled to provide statistical power in the analysis.

I assembled a data set of genome sequences from species in the *Drosophila melanogaster* subgroup from the public databases. The species are an excellent system for our analyses in several aspects. Since these species are closely related, the degree of sequence differences reflects the number of mutations in a lineage. This feature allows us to reduce errors in ancestral inference compared to lineages with long genetic distances. Genome sequences are available from natural population samples and outgroup

species. I employed DNA sequences for strains from Rwanda and Madagascar populations of *Drosophila melanogaster* and *D. simulans*, respectively, and reference sequences of *D. yakuba* and *D. erecta* as outgroups.

I focused on distinguishing among directional forces, natural selection and GC-biased gene conversion (gBGC), acting on synonymous mutations and intronic mutations. Since the two forces are predicted to give overlapping signatures in polymorphisms and divergence, I refer to them as “fixation bias”. To distinguish fixation biases from mutation biases, I analyzed site frequency spectrum (SFS), distributions of allele frequencies among variants segregating in a population. Mutations that have positive fitness effects are expected to be segregating at higher frequencies in a population than mutations that have negative fitness effects. The expectation can be tested by comparing SFS between two classes of mutations that are physically interspersed within DNA (*e.g.*, T→C vs. C→T). To construct SFS for each mutation class, I employed an ancestral inference method that our lab has developed previously. Our lab has confirmed the method’s reliability in studying the base composition evolution in the *Drosophila* species. The combination of large-scale data and a reliable method is one of the advances from previous studies.

In *D. simulans*, synonymous and intronic mutations show strong support for fixation biases consistent with the direction to increase GC content. I employed the categorization of synonymous families, groups of synonymous codons that differ by a single nucleotide. I estimated fixation biases for mutations for each synonymous family using SFS data. I found that the absolute values of the estimates are heterogeneous among synonymous families and between synonymous and intronic mutations. Because gBGC

should not distinguish functional classes (*e.g.*, coding vs. non-coding regions), it cannot explain the heterogeneity in fixation biases. Fixation bias estimates are greater at X-linked loci than at autosomal loci. This difference may reflect the differences in factors, such as heterogamety in males, dosage compensation, gene content, and population-scale recombination rates. These factors have been predicted to differentiate the magnitudes of fixation biases between X-linked and autosomal genes but this work does not address which mechanisms contribute to the autosome vs X chromosome difference in fixation biases.

In *D. melanogaster*, fixation bias is strongly supported for synonymous mutations, but is acting in the opposite direction from that in *D. simulans*; AT-increasing changes are segregating at higher frequencies compared to GC-increasing changes for synonymous mutations only within NAY codons. This pattern is not found for other mutations but is consistently found for NAY codons at autosomal and X-linked loci. Moreover, the absolute values of fixation bias estimates are greater at X-linked loci than at autosomal loci. The results cannot be explained by gBGC. The results provide strong evidence that synonymous mutations in *D. melanogaster* are under ongoing adaptive evolution toward fitness optimum different from the closely related species, *D. simulans*. NAY codons are the only ones translated by tRNA that can undergo post-transcriptional replacement of guanosine with its modified form at wobble positions, called queuosine modification. The levels of queuosine modification are known to depend on substrate abundance in diets. One possibility is that diet changes may underly change in directions of selection between *D. simulans* and *D. melanogaster*. This study shows that population genomics approaches can contribute to the discovery of functional

effects of mutations that were not detected in experimental or ecological approaches.

博士論文審査結果

Name in Full
氏名 山下 永香

Title
論文題目 Population genomics of weak evolutionary forces: Base composition evolution in coding and non-coding gene regions in *Drosophila*

自然選択が突然変異に作用する主要な力であるかどうかは進化遺伝学の中心的な課題の一つである。しかし、弱い自然選択を検出するのは技術的に困難であり、多くの DNA 変異を解析に用いる必要があるため、極めて挑戦的な課題となる。山下さんは、キイロショウジョウバエ (*Drosophila melanogaster*) とその近縁種オナジショウジョウバエ (*D. simulans*) の大規模な塩基多型のデータセットを公開データベースから得ることでこの課題に取り組んだ。キイロショウジョウバエの近縁種は、互いに近縁であり祖先推定における誤差を減らすことができる。また、集団内の多型情報に加えて、アウトグループ種について、ゲノムスケールの配列が利用可能である。具体的には、*D. melanogaster* のルワンダ集団、*D. simulans* のマダガスカル集団の DNA 配列に加えて、アウトグループとして *D. yakuba* と *D. erecta* の参照配列を用いた。

まず、固定バイアス（自然選択と遺伝子変換バイアス）と突然変異バイアスを区別するために、サイト頻度スペクトル (Site Frequency Spectrum) を利用した集団ゲノミクスアプローチを採用した。さらに、大規模な塩基多型のデータを用いることで、同義置換とイントロン、常染色体と X 染色体の比較などを行って自然選択と遺伝子変換バイアスの効果を分けることを試みた。これら一連の解析では祖先多型と派生多型の正確な情報が必要となるが、その目的のために、出願者の所属する研究室ですでに開発されていた精度の高い祖先推定法を活用した。

その結果、まず、*D. simulans* では、コドン第三塩基において GC が AT よりも固定バイアスが高いことを見出した。この傾向はコドンによって異なっていたことから、単純に遺伝子変換バイアスでは説明できないと解釈された。なぜなら、遺伝子変換バイアスは、変異が遺伝子機能をどのように変えるかには関係なく生じると推定されるからである。ついで、*D. melanogaster* では、*D. simulans* とは逆の方向、すなわち、コドン第三塩基において AT への固定バイアスが高いという傾向を見出した。しかし、特定のコドン (NAY) に関してのみ固定バイアスの証拠が見つかったことから、遺伝子変換バイアスではなく自然選択の作用であると推定された。なぜなら、遺伝子変換バイアスは、コドンの種類に関係なく作用すると推定されるからである。さらに、これらの固定バイアス推定値は常染色体よりも X 染色体で大きかった。突然変異バイアスや遺伝子変換でこの現象を説明するには、かなり複雑な前提条件を必要とするため、自然選択の効果によるものと推定された。以上の結果を総合すると、*D. simulans* と *D. melanogaster* では自然淘汰の作用が異なっており、自然淘汰の作用が進化的に短い時間スケールで変動しうることを示唆している。

以上の結果は、研究が困難な弱い自然選択の作用についての新しい知見を与えるものであり、審査委員一同で、博士論文にふさわしいと判断した。