

氏 名 Mishra, Neha

学位(専攻分野) 博士(理学)

学位記番号 総研大甲第 1864 号

学位授与の日付 平成28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 Inferring evolutionary forces from regional and temporal base
composition variation in *Drosophila*

論文審査委員 主 査 教授 中村 保一
教授 有田 正規
准教授 池尾 一穂
助教 石川 麻乃
助教 野澤 昌文 首都大学東京

論文内容の要旨
Summary of thesis contents

My research goal is to understand the variation in global forces of evolution at the genomic level taking *Drosophila* as a model system. Although the global forces of genome evolution have largely been established, the variations in their strength within and between genomes are less understood. Patterns of base composition can reveal variation in evolutionary forces, such as selection for codon usage, mutation bias and biased gene conversion. The base composition of synonymous sites is known to evolve under the affect of selection for translational efficiency and accuracy, mutation and drift. Most introns, on the other hand, evolve under much lesser selective constraint. Hence, variations in various evolutionary parameters within and between genomes can be studied using base composition of synonymous sites and introns. Base composition comparison among different nucleotide classes can also help in distinguishing the causes of base composition heterogeneity. I studied the variation in base composition (GC content) within and across genes in the *D. melanogaster* genome and across different lineages in the *D. melanogaster* subgroup.

I found that base composition at synonymous sites and introns varies at a within-gene as well as genome-wide level. Within genes, GC content of small introns decreases from the 5' end to the 3' end. The GC gradient near the 5' end is sensitive to the transcriptional level of the genes with highly transcribed genes having a steeper gradient. The intron divergence exhibits a positive gradient near the 5' end, which is also steeper for genes with higher transcriptional level. The 5' GC gradient in small introns is also observed in genes expressed at low levels in the germline, suggesting it to be under selection. On the other hand, GC content at synonymous sites shows a sharp increase at the 5' end and then declines towards the 3' end. The base composition at the synonymous sites near the 5' end seems to be under strong selection for low GC content, which might be functionally important for translation of mRNA since it is not observed in introns.

At the genome-wide level, base composition is heterogeneous within as well as between chromosome arms. GC content at synonymous sites and introns shorter than 100bp is significantly higher on the X chromosome compared to the autosomes. GC content at the synonymous sites is the most heterogeneous among all GC classes, suggesting that selection intensity might vary at a shorter scale than mutation in the *D. melanogaster* genome, since base composition of synonymous sites is thought to be evolving predominantly under the selection for codon usage whereas other nucleotide classes are mainly affected by mutation bias and biased gene conversion.

(別紙様式 2)

(Separate Form 2)

To study the base composition variation across different genomes, I examined the changes in the synonymous base composition patterns in >5000 genes from the 7 *Drosophila melanogaster* subgroup species. I used the existing genome data from five species and also added data of two more species in the *D. melanogaster* subgroup. Next-Generation RNA sequencing and Genome sequencing on the *D. tessieri* and *D. orena* transcriptomes and genomes was performed in the lab. I conducted a rigorous analysis of the RNA-seq and Genome-seq data and developed protocols for *de novo* gene and intron annotation. For this task, I used the available data from the sequenced *Drosophila* species. I developed several methods and also used some of the softwares that are already available for NGS data analysis. The substitutions occurred in each of the lineages were inferred using maximum likelihood approach. I found that all of the lineages showed strong departures from the equilibrium states and in some lineages both effective population size and mutation bias seemed to have fluctuated. These findings suggest that magnitude of forces governing base composition at synonymous sites may have varied frequently in a lineage-specific manner.

Summary of the results of the doctoral thesis screening

ゲノム進化の要因についてはこれまでに多くの研究があるが、ゲノム内部あるいは異なるゲノム間でのそれらの影響力の差についての知見は少ない。出願者は、コドン使用・変異のバイアス・GCバイアス遺伝子変換などの結果としてのゲノム上の塩基組成パターンを詳細に解析することで、進化を推進する要因の多様性を明らかにできると考え、研究を行った。

タンパク質コード遺伝子の同義置換サイトの塩基組成は、翻訳効率と精度、変異と浮動の選択の結果である。一方、多くのイントロン、とくにイントロン内に制御領域が存在しないと考えられる全長100塩基以下の小型イントロンには、そのような制約がない。そのため、ゲノム内・ゲノム間の進化要因の多様性を同義置換サイトとイントロンの塩基組成の比較解析により明らかにすることができると出願者は考えた。この博士論文研究では、*Drosophila melanogaster*とその近縁種において、遺伝子構造における位置的な塩基組成の変動を検討し、ゲノム全体、遺伝子内の同義サイトとイントロンでの塩基組成分布の多様性を明らかにした。

出願者はまず、遺伝子内の塩基組成の変動を解析した。小型イントロンではGC含量は塩基配列の5' から3' 側にしたがって減少する傾向を示した。遺伝子コード領域ならびにイントロンの5' 末では転写量が多い遺伝子ほどGC含量は高く、3' への減少は強いことが観察された。イントロンのGC含量の勾配は生殖細胞系で転写される遺伝子では低い。そうした違いを生成する選択圧の存在が示唆された。一方、同義置換サイトのGC含量は5' 末側で鋭いピークを示したのち、3' 側へと減少していた。このようなコード領域の同義置換サイトの5' 末端の低GC含量への選択圧はイントロンでは観察されない。GC含量の違いは、mRNAの転写調節に関わる意義をもつものと考えられた。

次に、ゲノム全体ならびに染色体長短腕間の比較においても塩基組成の多様性を検討した。X染色体は常染色体に比べ小型イントロンのGC含量が著しく高く、同義置換サイトでのGC含量はきわめて不均一であった。これは選択圧が短い時間で大きく影響していることを示唆した。同義置換サイトにおいてはコドン使用頻度が主として進化を推進する要因ではあるが、その外の部位では主として変異と遺伝子変換が進化の要因となっているものと考えられた。

最後に、ゲノム間の塩基組成の多様性を観察するために、*D. melanogaster* の7近縁種由来の5,000以上の遺伝子の同義置換サイトの塩基組成パターンを網羅的に解析した。この解析のために出願者は既存の5種に加え、NGSを用いて新規に *D. tessieri* と *D. oreana* のゲノム塩基配列決定並びに RNA-seq 解析を実施した。得られた塩基配列を元に独自の配列解析パイプラインを構築し、新規2種のゲノム遺伝子構造注釈とイントロン位置の正確な特定を行った後に、それらの配列を用いて塩基組成パターン解析を行った。その結果、いずれの種に於いても平衡状態からの乖離を示しており、有効集団サイズと変異バイアスの双方で変動が観察された。これらの発見は、同義置換サイトの塩基組成を支配する力は種特異的に大きく異なることを示した。

(別紙様式 3)

(Separate Form 3)

Nehaさんは進化を推進する要因の多様性を理解するために、進化のモデルシステムとしてのショウジョウバエを活用し、ゲノム情報解析系も独自に構築し研究を推進した。本研究はゲノム進化を推進する要因とその関係について理解する上で重要な知見を与えるもので、学位論文審査に値する研究であると判断した。