Natural selection in a population of Drosophila melanogaster explained by changes in gene expression caused by sequence variations in core promoter regions

論文内容要旨

（博士論文題名）Natural Selection in a Population of *Drosophila melanogaster* Explained by Changes in Gene Expression Caused by Sequence Variations in Core Promoter Regions

（ショウジョウバエ野外集団におけるコアプロモーター配列変異によって生じる発現量変化への自然選択）

東北大学大学院生命科学研究科

生態システム生命科学専攻

佐藤光彦

Understanding the evolutionary forces that influence variations in gene regulatory regions within natural populations is an important issue for evolutionary biology. Recent studies have emphasized that the evolution of gene regulatory sequences is important for adaptive evolution and thus, regulatory regions may play a major role in adaptation. Variations in gene expression are considered to affect phenotypic consequences in morphology, physiology, behavior, and disease susceptibility. Transcriptomic technologies, such as microarray and high-throughput RNA sequencing (RNA-seq), make it possible to observe variations in gene expression within natural populations of some species including humans, fish, *Drosophila*, and yeast. However, these studies could not clarify the relationship between sequence variation in the detected regions and change in gene expression levels. For *Drosophila melanogaster*, genome and transcriptome data from a natural population in North Carolina, USA, are available. Here we focused on sequence variations in core promoter regions, which are critical regions for gene regulation in higher eukaryotes. Core promoter regions are generally defined as DNA regions that direct the accurate initiation of transcription by RNA polymerase II and contain various sequence motifs that interact with basal transcriptional factors. We identified the core promoter sequence variations associating with differences in gene expression levels that been subject to natural selection.

We identified genes whose expression levels were changed significantly by polymorphisms in core promoter regions using linear model. To find potential regions evolving under directional or balancing selection in the natural population, we calculated Tajima's *D* value for the core promoter regions. Outlier regions from the observed Tajima's *D* influenced by demographic events in North Carolina population were identified by coalescent simulation. We performed 100,000 simulations and defined that CPRs with observed Tajima's *D* more extreme than 99% confidential intervals (CIs) had been subjected to natural selection. Furthermore, for candidate core promoter regions influenced by natural selection, change of binding sites of basal transcriptional factors were estimated using sequence motif analysis.

We found two genes under directional selection, two genes under purify selection and nine genes under balancing selection. One of the genes under directional selection was *CHKov1*, which confers resistance to the sigma virus and related insecticides (Figure 1). For this gene,

nucleotide changes in core promoter regions caused the loss of two basal transcriptional factor binding sites and acquirement of one transcriptional factor binding sites, resulting in decreased gene expression levels. Consequently, sequence variants with these nucleotide changes increased in frequency through selective sweep. In several core promoter regions, such as N-methyl-D-aspartate receptor-associated protein (*Nmda1*), the brain tumor gene (*brat*), Cytochrome P450-4d1 (*Cyp4d1*), and the six unknown genes (*CG15743*, *CG9044*, *CG6950*, *CG10463*, *CG14253*, and *CG33506*), the variant sequences might have different binding sites of basal transcriptional factors and expression levels that could be maintained by balancing selection. Our results revealed the evolutionary process through natural selection for differences in gene expression levels caused by sequence variations of core promoter region in a natural population. The present study indicated that the sequences of core promoter regions were diverse even within a population, which might provide a source for natural selection.

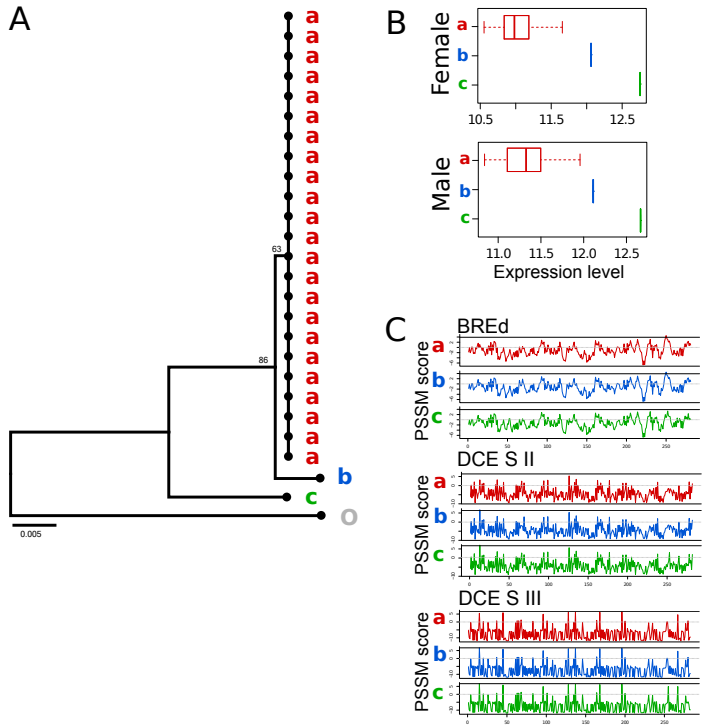

Fig. 1 Results of CPR for *CHKov1*

(A) Phylogeny by neighbor-joining tree for different alleles (a-c) of the CPR sequences found in the natural population is drown. *D. simulans* was used as an outgroup (o). Bootstrap values are shown for nodes with greater than 60% support. (B) Expression level of each allele of the CPR were from the database using microarrays in female and male flies (Ayroles et al. 2009) found in natural population. (C) PSSM scores (log odds finding motifs) for the binding sites of BREd, DCE S II, and DCE S III at all the positions along CPR sequences for the strand were shown. Gray dotted horizontal lines indicate a PSSM score of zero, and black horizontal lines indicate PSSM scores at threshold values above which each transcription factor is likely to bind.