論文の内容の要旨

生産・環境生物学専攻
平成28年度博士課程入学
氏名　趙新威
指導教員 岸野洋久

論文題目：Variation of transcription factor profiles and their roles in mammalian evolution

Recently unexpectedly large variations were found in regulatory proteins. These variations may significantly affect regulatory network evolution by changing the expressions, molecular interactions, and post-translational modifications of the regulator. These events, however, have been assumed to be occasional or rare in the context of cis-regulatory research for a long time. Trans-species conserved nature was confirmed by comparing orthologous TFs from different species. On the contrary, macro evolution of TF families and its potential effect on species specificity are less studied. It may be acquisition of new TFs and loss of existing TFs that mainly make species unique. In this thesis, I had attempted to answer to this question by analysing the evolution of the mammalian TF families.

Multiple isolated transcription factors act as switches and contribute to species uniqueness

In some mammals, although there are good genome sequencing data, there are few studies on the identification of transcription factors in these species. By using the standard hidden Markov model (HMM), I constructed a database of TFs from the entire genomes of 96 species of mammals. We further annotate our database with orthologous groups information by OrthoDB. Now, there exist several animal TF databases, such as humanTFs, animalTFDB3, Riken mouse TFdb,FlyTF, TFCat, TFCONES, and ITFP. My database and the others are based on the similar pipeline by DNA binding domain (DBD) and HMMER. AnimalTFDB3 contains 125,135 TFs from 97 genomes, which ranges from Caenorhabditis elegans to mammal species like human. My database contains 140,821 TFs, focusing on 96 mammal species. This may provide a better solution in mammal's TF evolution history. Numbers of TFs varied largely between species, with minimum of 1,113 in platypus and maximum of 1,905 in chimpanzee.

To interpret the membership variation of TF families in evolutionary context, I constructed the phylogenetic trees of TF families and estimated the events of gene duplication and gene loss

by reconciling the gene trees with the species tree. It was found that, in mammalian evolution, the TF families had increased their members by 37.8% and lost their members or their DBDs by 15.0%. As a result, each species has its unique set of TFs.

Largely because existing TF databases had insufficient coverage, previously constructed gene regulatory networks (GRNs) for mammals were mostly limited to TF-to-TF relationships and cover only a small subset of orthologues. They revealed the conservation between orthologues, with all other TFs ignored. To account for the contribution of birth and death of TFs, I constructed more comprehensive GRNs (1,200+ TFs and 8000+ nodes and containing TF-to-TF and TF-to-TG edges) for mouse and rat generated from whole species gene interaction networks using STRING. Because mouse and rat have the same common ancestor not long ago, any differences between their GRNs can be assumed to be caused by recent changes.

By comparing the human TF-to-TF interaction network with that of mouse, I confirmed that two-thirds of TFs in the largest connected component were conserved. On the other hand, the other about 500 isolated TFs, which had been mostly treated as out of target in the preceding studies, were variable and prone to be lost from the genome. By comparing the number of isolated and primary connected TFs, I found that a large amount of human isolated TFs were enriched in the Cys2His2-like (zf-C2H2) TF family. The difference in the members of C2H2 zinc finger proteins, the largest TF family, was associated with lower expression of their interaction genes in human compared with mouse. Knock-out mice that lacked these interacting genes had abnormal phenotypes, such as a short tail or hairless skin, which are observed in humans.

Pruning of transcription factors and the macroevolution of mammals

The most recent common ancestor of extant mammals dates back to approximately 170 Mya. After splitting from Marsupialia, Placentalia diverged into Afrotheria, Xenarthra and Boreoeutheria about 100 Mya. The rate of TF amplification before the divergence of Placentalia and Marsupialia was higher than the rate of loss (58.1/My vs. 8.9/My). Most TFs arose before the common ancestor of placental animals. TFs were rapidly lost, however, during the evolution of placental mammals, whereas only a few new TFs were generated. Early Euarchontoglires, Laurasiatheria and marsupials appeared between 100 and 95 Mya and underwent initial diversification. During this time, the average TF loss rate reached its first

peak of 192.3/My. The second peak of TF loss occurred near the Cretaceous–Paleogene (K–Pg) boundary, at 66 Mya. The TF loss rate was as high as 68.2/My within a 2-Mya window of the K–Pg boundary. Major mammalian orders appeared during this time. In terms of loss events on branches, we detected 1,967 TF loss events occurring at a rate of 284.7/My along the ancestral branch of Euarchontoglires. In addition, 2,879 TF loss events at a rate of 345.0/My were inferred along the ancestral branch of the Laurasiatherian lineage, while 3,819 taking place at a rate of 209.3/My were uncovered on the ancestral branch of marsupials. TF loss also occurred rapidly on subsequent branches, especially where common ancestors of different mammalian orders started to differentiate, consistent with the above results. The TF pool became simplified during mammalian species diversification, thus indicating that such an event may have contributed to the early diversification of mammals. In contrast, gain events did not follow any specific pattern; therefore, the amplification of TFs may have contributed little to mammalian species diversification.

The effect of TF loss

To quantify the effect of TF loss on the target genes (TGs), I examined TG evolutionary rates and found the deceleration effect of TF loss on TGs over the long term. Because the rate of molecular evolution of a gene is negatively correlated with the strength of a functional constraint, the loss of a TF is expected to have a role in the adaptive evolution of the regulatory system. The molecular evolutionary rate of a gene is negatively correlated with its expression level. Consistently, in human and mice, TGs lacking TFs had higher expression levels. Functional annotation of TGs revealed functions mainly related to the cell cycle, cell migration, signal transduction, and inflammation. By comparing GRNs of mouse and rat, I found that loss of TF genes lost all its edges and DBD changes resulted in an average loss of 41.1% of the edges, much larger than 22.6% edge loss due to all other factors besides DBD. So, TF and its DBD are the main factors affecting its subnetworks in GRN.

I used gene ontology (GO), a standard functional annotation method for comparative research among species, to reveal relationships between TFs and their functions. I found that 92.7% of TFs (1430 out of 1543 TFs) own unique set of GO items when comparing with any other TFs in mouse. Considering the low resolution of some GO items, this means each TFs can be taken as unique and have their own functions. In the list of GO items that are governed by single TFs, the regulating TFs vary among human, mouse and rat. Through the enrichment analysis of this

small GO item list, I found that majority of these GO items are phenotype relate functions. These functions were found to be regulated by different TFs among species, suggesting changes in the regulatory pattern.

To examine the effect of TF loss on life history traits, we regressed four binary traits—sociality, diurnality, reproductive seasonality, and insectivory—on the states (presence/absence) of 1-to-1 orthologous TFs. In total, 24 TF genes were significantly associated, mostly positively, with diurnality. Functions of TFs associated with diurnality shed light on their adaptation. Overexpressed ZNF667 (MIPU1) decreases oxLDL-induced cholesterol accumulation. ZNF648 is associated with high-density lipoprotein cholesterol levels. The phenotype of the knockout ZFP92 gene increases the amount of total body fat. GBX1-knockout mice exhibit reduced thermosensory functions. A higher hunger and cold tolerance is likely to promote adaptation to a nocturnal situation.

Among genes associated with sociality, half were related to immunity-associated phenotypes. Knockout mice of SEBOX, PHOX2A and KLF8 genes, that are associated with insectivory, increases levels of circulating total protein or improves glucose tolerance. Knockout of NSL1 and ZNF648 leads to increased circulating cholesterol levels. Among genes associated with sociality, half were related to immunity-associated phenotypes. Knockout mice of SEBOX, PHOX2A and KLF8 genes, that are associated with insectivory, increases levels of circulating total protein or improves glucose tolerance. Knockout of NSL1 and ZNF648 leads to increased circulating cholesterol levels. TFs positively correlated with the reproductive-seasonality trait were lost among year-round breeding animals. More than half of the reproductive seasonality-associated TFs were ZNFs. KRAB-ZNFs play a major role in the recognition or transcriptional silencing of transposable elements. With regard to the phenotype associated with these genes, 13 out of 23 are related to fetal viability or formation.

In the early stage of mammalian species formation, the number of TFs increased greatly, but later with the production of macroevolution, the loss of TFs accounted for the main part. The increase and decrease in the number of these TFs mainly exist in isolated TFs. The isolated TFs were lost with mild fitness effect. Faced with the serious environment changes, such as the K-Pg boundary, the loss of such TFs rewired the regulatory network, set the functional constrain on the target gene additionally, and facilitated the survival of the organism.