

氏 名 松波 雅俊

学位（専攻分野） 博士（理学）

学位記番号 総研大甲第 1522 号

学位授与の日付 平成 24 年 3 月 23 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 The evolutionary analysis of the vertebrate two-round
whole genome duplications

論文審査委員 主 査 教授 城石 俊彦
教授 明石 裕
准教授 高野 敏行
准教授 木村 暁
教授 佐藤 矩行 沖縄科学技術大学院大学

Two rounds of whole genome duplications (2R WGDs) occurred in the vertebrate ancestors, and they generated large numbers of duplicated protein-coding genes and their regulatory elements. These events could contribute to the emergence of vertebrate-specific features. However, the evolutionary history and influences of the 2R WGDs are still unclear. To address this issue, I conducted comprehensive studies on both protein-coding and non-coding sequences found in the conserved synteny blocks generated by the 2R WGDs. Such conserved synteny blocks are expected to retain duplicated protein-coding and gene regulatory sequences. Consequently, evolutionary changes or some constraints relating to these blocks would have played important roles in the evolution and diversification of vertebrates. On the basis of this view, I focused on evolution of both protein-coding and non-coding sequences of the vertebrate genomes, especially Hox clusters.

Because a part of gene regulatory elements are expected to be conserved according to their functional importance, evolutionarily conserved non-coding sequences (CNSs) might be good candidates of gene regulatory elements. In addition, portion of the paralogous protein-coding genes retained after the 2R WGDs show overlapping expression pattern. Therefore, paralogous genes might share gene expression regulatory mechanisms. Paralogous CNSs have a possibility to control overlapping expression patterns of those paralogs. Thus, detecting paralogous CNSs and inferring the relation between paralogous gene and CNSs is important to understand evolution after the 2R WGDs.

Four or more paralogous Hox clusters exist in vertebrate genomes because of the 2R WGDs. The paralogous genes in the Hox clusters show similar expression patterns, implying shared regulatory mechanisms for expression of these genes. Previous studies partly revealed the expression mechanisms of Hox genes. However, *cis*-regulatory elements that control these paralogous gene expression are still poorly understood. Toward solving this problem, I searched CNSs within vertebrate Hox clusters. I compared orthologous Hox clusters of 19 vertebrate species, and found 208 intergenic conserved regions. I then searched for CNSs that were conserved not only between orthologous clusters but also among the four paralogous Hox clusters. I found three regions that are conserved among the all four clusters and eight regions that are conserved between intergenic regions of two paralogous Hox clusters. In total, 28 CNSs were identified in the paralogous Hox clusters, and nine of them were newly found in this study. One of these novel regions bears a RARE motif. These CNSs are candidates for gene expression regulatory regions among paralogous Hox clusters. I also compared vertebrate CNSs with amphioxus CNSs within the Hox cluster, and found that two CNSs in the HoxA and HoxB clusters retain homology with amphioxus CNSs through the 2R WGDs.

The duplication histories of vertebrate Hox clusters are controversial. Under the assumption of the 2R WGDs, phylogenies of Hox gene should show a symmetrical topology. However, some previous studies did not support this symmetrical topology. I thus carried out exhaustive phylogenetic analysis of deuterostome Hox genes. First, to identify outgroup genes of each vertebrate Hox paralog group, I inferred the correct ortholog/paralog relationships among deuterostome posterior Hox genes by comparing available Hox genes. Amphioxus Hox9-11 were generated by amphioxus specific tandem duplications. Because vertebrate Hox10-12 and Hox14-15 genes have no counter parts in amphioxus Hox genes, they were probably lost in the amphioxus lineage. Secondly, the duplication histories of vertebrate Hox genes were inferred by constructing phylogenetic trees and phylogenetic networks. My analysis suggested that the ((A,B), (C,D)) topology is the most suitable explanation of Hox cluster duplications.

I then carried out genome-wide identification of paralogous CNSs. A sensitive BLAST search of each synteny block revealed 7,924 orthologous CNSs and 309 paralogous CNSs conserved among 8 high quality vertebrate genomes. I newly detected 194 paralogous CNSs. Their locations are biased nearby the transcription factors coding regions shown expression in brain and neural systems. The existence of these paralogous CNSs is difficult to explain by previous duplication models. Because these sequences have same transcription factor binding motifs, they might be backup of paralogous gene expression and/or contribute to the interaction between paralogs.

The 2R WGDs occurred after the split of the urochordate ancestors but before the diversification of extant gnathostomes (jawed vertebrates). However, there is no clear evidence whether the timings of the 2R WGDs are before or after the split of agnathans (jawless vertebrates including lamprey) and gnathostomes. To clarify this problem is highly important for study of vertebrate evolution and development. The lamprey gene data are also useful for molecular function and developmental studies. Thus, I analyzed the mRNA sequences of Japanese brook lamprey (*Lethenteron reissneri*) and estimated the relative timing of the 2R WGD by combining newly obtained sequence data from Japanese brook lamprey and sea lamprey (*Petromyzon marinus*) data in the database.

The Japanese brook lamprey cDNAs were synthesized from the mRNAs of ammocoetes larva and were sequenced by using Roche 454 GS FLX titanium system. After the assembly of 426,476 sequence reads, I obtained 7,708 contigs with 336 bp length on average. Additionally, I also analyzed the sea lamprey mRNA sequence data in the SRA database. Including 119,412,170 reads, they were assembled to 78,947 contigs. Based on these lamprey data, I analyzed putative orthologous and paralogous gnathostome sequences corresponding to the lamprey contigs to estimate the relative timing of the 2R WGDs. From the homologous gene clustering, phylogenetic trees of 358 gene families are reconstructed. However, if I restrict trees which contain two

duplication events and have high statistical supports, only 55 trees were left. The majority (49) of them showed the pattern that two genome duplications both occurred before the lamprey divergence.

Recently, the sea lamprey (*Petromyzon marinus*) genome sequences appeared in the public database including 11,429 genes. I also investigated the possibility that gene losses caused misunderstanding of true ortholog/paralog relationships by using these newly released sea lamprey data, as well as the genome data of 13 gnathostomes and 6 nonvertebrate species. I reconstructed phylogenetic trees of 545 gene families, and there were 127 trees with one agnathan (A) and two gnathostomes (G) clusters. Although 69 trees showed topology ((A,G),G) suggesting two duplications before the agnathans/gnathostomes divergence, the remaining 58 trees had topology ((G,G),A). I compared the branch lengths connecting the gnathan common ancestor and the agnathan/gnathostomes common ancestor, and found that ((G,G),A)-topology trees had the significantly longer branch than ((A,G),G)-topology trees. This suggests that agnathan genes were lost in the lamprey lineage in ((G,G),A)-topology trees, and the occurrence of duplications erroneously looked like after the agnathans/gnathostomes divergence. I thus conclude that 2R WGDs both occurred before agnathans/gnathostomes divergence.

Two-rounds of whole genome duplication (2R WGD) occurred in the vertebrate ancestors, which gave rise to large numbers of duplicated protein-coding genes and their regulatory *cis*-elements. These events likely contributed to emergence of the vertebrate features. However, the process and mode of the 2R WGD are still elusive. To address these issues, Mr. Matsunami conducted four studies including *de novo* mRNA sequencing and computational analyses of both protein-coding and non-coding sequences.

First, Mr. Matsunami searched paralogous conserved non-coding sequences (CNSs) within the vertebrate *Hox* clusters. Four paralogous *Hox* clusters are thought to emerge from 2R WGD. He identified 28 CNSs in the paralogous *Hox* clusters, in addition to those between the orthologous clusters. These paralogous CNSs are good candidates as the *cis*-regulatory elements, which allow similar expression pattern observed among the paralogous *Hox* genes.

Second, duplication history of the vertebrate *Hox* clusters, A, B, C and D, has been contradictory in previous studies. He conducted exhaustive analyses of the phylogenetic tree and phylogenetic network and proposed that the symmetrical topology ((A,B), (C,D)) is the best supported history of the *Hox* cluster duplications.

Third, genome-wide identification of the paralogous CNSs has been incomplete thus far. He conducted sensitive BLAST search of synteny blocks among eight high quality vertebrate genomes and newly identified 194 paralogous CNSs in addition to previously reported 115 CNSs. Their locations may be biased toward regions close to the transcription factors (TFs) expressed in brain and neural system. Because these sequences have same TF binding motifs, they might be backup of paralogous gene expression and/or contribute to the interaction between the paralogous genes.

Lastly, the timing of the 2R WGD in the course of the agnathan-gnathostome divergence is still controversial. He conducted *de novo* mRNA sequencing of Far East brook lamprey (*Lethenteron reissneri*), and carried out comprehensive phylogenetic analyses with the new sequencing data and a publicly-available genome sequence of sea lamprey (*Petromyzon marinus*). The result suggested that both rounds of WGD occurred before the agnathans-gnathostomes divergence.

The results of the above four studies in this thesis contributed to better understanding of the process and mode of the 2R WGD, and they would provide insights into its impact on the vertebrate evolution.