

氏 名 Kim Hyung-cheol

学位（専攻分野） 博士（理学）

学位記番号 総研大甲第 1004 号

学位授与の日付 平成 18 年 9 月 29 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 EVOLUTIONARY ANALYSIS OF THE MAMMALIAN
DLX GENE CLUSTERS

論文審査委員	主 査 教授	相賀 裕美子
	教授	大久保 公策
	助教授	池尾 一穂
	助教授	川上 浩一
	教授	岡田 典弘（東京工業大学）

論文内容の要旨

Every organism is product of a long evolutionary history, and it is very important to elucidate key genetic changes that caused significant phenotypic transformation. Evolutionary studies at the molecular level have been mainly focused on protein evolution. This is partly because of our lack of knowledge on the evolutionary patterns of non-coding regions, except for promoter regions and some enhancers. Here, “non-coding” regions include intergenic regions, introns, short repeat sequences such as SINE and VNTR, and nontranslated region of exons. This old framework of molecular evolutionary studies gradually shifted to current genome-wide approach from 1980’s, because not only protein coding regions but also non-coding regions were started to be sequenced. We, at the beginning of the 21st Century, are in the position of utilizing vast amount of genomic sequences. It is now the age of “comparative genomics” or more appropriately, “evolutionary genomics”. Some of these nucleotide sequence comparisons were connected with developmental studies, and the role of “cis-regulatory” elements on certain morphological features have been elucidated. We now have possibility of delineating important genomic changes through comparison of the vast amount of genome sequence data. The purpose of this study is thus to gain a better understanding of the evolutionary significance of non-coding regions in mammalian genomes.

As the model system, I chose the *Dlx* gene clusters. *Dlx* genes are involved in the development of the vertebrate forebrain, branchial arch, sensory organ, and limbs. The vertebrate *Dlx* genes are homologous to *Drosophila Distal-less (Dll)* gene, that is responsible for limb outgrowth, and probably arisen as a result of a tandem gene duplication followed by a number of large genomic scale duplications, according to phylogenetic analyses including my own. There are six *Dlx* genes in mammalian genomes, located in three chromosomes as three pairs of duplicated genes, namely, *Dlx1-2*, *Dlx5-6*, and *Dlx3-7*. These paired genes are convergently transcribed, and the intergenic regions of each pair contain enhancer elements. Expression patterns of these known enhancers are often conserved, even between distantly related organisms such as between mouse and zebrafish. However, evolution has two faces; conservation among different lineages of organisms and diversification at each lineage of organism. Therefore, it is possible that an evolutionarily conserved region in the common ancestral genome diversified as speciation created morphologically different lineages. A certain non-coding sequence changes may be responsible for some morphological diversity of placental mammals.

I first analyzed *Dlx* gene cluster sequence data of the following 9 species; human, chimpanzee, baboon (macaque for *Dlx5-6*), mouse, rat, cow, dog, elephant, and whale. All sequences except for whale were retrieved from the DDBJ/EMBL/GenBank International Nucleotide Sequence Database. Because I was interested in the variation of this gene cluster within mammals, I newly determined *Dlx1-2*, *Dlx5-6*, and *Dlx3-7* genomic sequences (a total of 63.9 kb) for whale (*Antarctic minke whale; Balaenoptera bonaerensis*), whose morphology is quite different from other mammals. Whale sequences always clustered with cow in the phylogenetic trees for three clusters, as expected, and the branch length for the whale lineage was slightly shorter than that for cow, suggesting a lower evolutionary rate in whale than cow. This difference may be caused by longer generation times in whale than cow. The evolutionary rate for the *Dlx3-7* cluster was highest among the three clusters for all the 15 branches of the phylogenetic tree except for two. SINEs were observed in much higher numbers in the *Dlx3-7* cluster than the other two clusters, and its density was much higher than the genomic average for human and mouse genomes. This is probably partly because of high GC content (50-60%) compared to a typical mammalian genome (ca. 40%). In contrast, LINES, that tend to be

found in low GC content regions, were rare in neither the *Dlx3-7* cluster nor the other two clusters.

I also noticed that three highly conserved regions in *Dlx1-2* and *Dlx5-6* clusters contain “ultra-conserved elements” (100% identity for more than 200bp among human, mouse, and rat genomes), and they coincided with functionally important enhancers, I12a, I12b, and I56i. There is thus a possibility that other ultra-conserved elements are also responsible for some unknown important evolutionary changes of placental mammals. I thus sequenced 105 whale and elephant sequences that are homologous to ultraconserved elements. Although many sets of ultraconserved element sequences exhibited extremely high levels of conservation between human and whale and between human and elephant, there were some changes, and the rate of nucleotide substitution for non-exonic elements was slower than that for exonic ones. Ultraconserved elements are also highly conserved in non-mammalian vertebrates but the substitution rate progressively decreased from the common ancestor of vertebrates to that of mammals; the rate outside amniotes is in the order of 10^{-9} /site/year, while rates for the bird and the mammalian lineages are in the order of 10^{-11} /site/year, with 2-3 times higher rate for the bird lineage than the mammalian one. Among mammalian species, dog and whale lineages showed slower rate (2.7 and 3.1×10^{-11} /site/year, respectively) than that (4.7×10^{-11} /site/year) for elephant. This result suggests that even the ultra-conserved regions experienced lineage specific changes and some of these changes may be responsible for lineage specific phenotypes.

As the *Dlx3-7* cluster had more heterogeneity than the other two *Dlx* clusters, there is a higher chance to find species or lineage specific changes in this cluster. I thus determined nucleotide sequences of the *Dlx3-7* cluster for morphologically quite diverse five mammalian species; dolphin, elephant, armadillo, otter, and sea lion. Dolphin (Dall's porpoise; *Phocoenoides dalli*) belongs to Order Cetartiodactyla, elephant (Asiatic elephant; *Elephas maximus*) and armadillo (*Orycteropus afer*) belongs to Order Afrotheria, and otter (Eurasian river otter; *Lutra lutra*), and sea lion (California sea lion; *Zalophus californianus*) belong to Order Carnivora. I first determined their mitochondrial ribosomal RNA sequences, and confirmed species identity by comparing them with already available sequences deposited in the DDBJ/EMBL/GenBank International Nucleotide Sequence Database.

Four long PCR amplifications were conducted for each species to cover the entire *Dlx3-7* cluster region. I designed PCR primers from highly conserved regions residing within and nearby the cluster. All long PCR amplifications were successful, and amplicons were subjected to shotgun sequencing. In total, I determined about 137.2 kb (average = 27.4kb), and 84.5kb (average = 16.9kb) were in the intergenic region. Although majority of the determined sequences are non-coding, there are many regions with high conservation, and the multiple alignment of the entire region was possible using ClustalW. The phylogenetic trees constructed for the entire region and the intergenic region were both concordant with the known mammalian phylogeny, indicating orthologous relationship of these newly determined six species including whale and already available 12 species (human, chimpanzee, baboon, lemur, rabbit, mouse, rat, dog, bat, cow, pig, and armadillo, retrieved from the DDBJ/EMBL/GenBank International Nucleotide Sequence Database).

I compared those 18 genomic sequences of the *Dlx3-7* cluster in order to investigate species specific changes occurred on the evolutionarily conserved regions among the mammalian species using mVISTA and MultiPipMaker. A total of nine regions of multispecies conserved non-coding sequences were found. These conserved nine regions were almost identical with those previously found (Sumiyama et al., 2002; Sumiyama, personal communication), indicating their highly conserved nature throughout mammalian evolution. I then multiply aligned these nine regions, and extracted species or lineage specific changes based on the species phylogeny. These species-specific changes included both

substitutions and insertions/deletions. There was no clear characteristic in substitutions. Among three species of Cetartiodactyla, whale and dolphin experienced more deletions than insertions compared to cow. Deletions were also more frequent in three species (dog, otter, and sea lion) of Carnivora.

I found many putative transcription factor binding sites from top five conserved regions using TRANSFEC. A certain numbers of whale and dolphin lineage specific changes were found from these multispecies conserved sequences. In particular, one conserved region had 9bp (AGTGCCTGG) deletion existing in only whale and dolphin. I therefore examined enhancer activity of this whale sequence using transgenic mice. Although this region contained several substitutions and a 9bp deletion, mice embryos with whale transgenes showed interdigit and AER expression in limb, more or less similar expression pattern to that of mouse lines. This result suggests that this region of the whale and mouse genomes share essential *cis*-elements for limb expression.

In this study, I analyzed non-coding regions of the three *Dlx* clusters, including newly determining the three *Dlx* cluster sequences for whale and the *Dlx3-7* cluster region for five mammalian species, as well as analysis of ultraconserved non-coding regions for various mammalian species. Some of these non-coding regions have been highly conserved during mammalian radiation, yet I discovered many species and lineage specific changes. These changes may be connected to species-specific phenotypic features and await experimental verifications in the near future.

従来、遺伝子の進化に関する研究は、タンパク質をコードしている領域に大きく偏っていた。しかし、近年大規模な塩基配列決定が可能になるにつれて、いわゆる非コード領域の塩基配列の進化も研究対象として重要視されている。特に進化的に保存されている非コード領域には、エンハンサーなどの機能があり、遺伝子発現の調節に重要な働きを有する場合がある。

Kim 君はこのような非コード領域に着目し、またモデルシステムとして *Dlx* 遺伝子クラスターの解析を行なった。*Dlx* 遺伝子はショウジョウバエ *Dll* のホモログとして哺乳類で発見された遺伝子で、2 個の直列重複した遺伝子のクラスターとして哺乳類ゲノムには 3 種類 (*Dlx1-2*, *Dlx5-6*, *Dlx3-7*) 存在している。どのクラスターにおいても、中央部の非コード領域中に遺伝子発現を調節する、進化的に保存された部位があることが知られている。Kim 君はまず彼が新たに決定したクジラの配列を含む 10 種の哺乳類の 3 クラスターの塩基配列の分子進化的解析を行った。その結果、3 種類の *Dlx* クラスターの中の *Dlx3-7* 遺伝子クラスターの進化速度 (塩基置換速度) が最も速いことを見出した。またこの *Dlx* 遺伝子クラスターを詳細に分析する過程で、Kim 君は *Dlx1-2* と *Dlx5-6* クラスター中にいわゆる「超保存配列」(ultraconserved element) が存在することを発見しこの中に既知の転写調節領域が含まれることを明らかにした。そこで Kim 君は、超保存配列の進化的意義に興味をもち、ヒト、マウス、ラット、クジラ (新規に配列を決定)、ゾウ (新規に配列を決定)、イヌ、ニワトリ、アフリカツメガエル、ゼブラフィッシュの他の遺伝子に見られる 105 個の超保存配列の分子進化的解析を行なった。その結果、「超保存配列」にも種特異的、系統特異的な変化が蓄積していることを明らかにした。塩基置換速度を推定したところ、ゾウ進化速度 (およそ 5×10^{-11} /site/year) はイヌとクジラよりも若干高く、また鳥類 (ニワトリ) の系統は哺乳類の数倍高い速度であり、さらに魚類の系統は哺乳類の 100 倍の速度と推定された。このように、当初 1 億年以上のあいだまったく変化が観察されないという定義であった超保存配列が系統によって大きく異なる進化速度を持っていることは、興味深い発見である。

一方、*Dlx* 遺伝子クラスターの中で最も進化速度の高かった *Dlx3-7* 遺伝子クラスターに関して、進化的な形態変化と転写調節領域の変化の関連を調べるため、彼が新たに配列を決定した 6 種を含めた 18 種の哺乳類において保存性の高い転写調節領域の比較を行い、クジラとイルカで共有されている 9 塩基の欠失を見出した。そこで、クジラのゲノム DNA を含むベクター DNA をマウスの受精卵に入れて、トランスジェニックマウスを作成し、いろいろな発生段階における発現パターンを観察したが、マウス本来の発現パターンと本質的な差はなかった。

本研究において、Kim 君は *Dlx* 遺伝子クラスターをモデルとして、哺乳類における非コード領域の進化について、データベースに存在しないいくつかの種の塩基配列を全体で約 250kb 決定し、それらと既知の配列を詳細に分析した。その結果、いわゆる「超保存配列」にも種により、あるいは領域により、進化的な保存度の違いがあることを明らかにした。種特異的な変化のうち、特にクジラで見いだされた 9 塩基欠失のある保存領域に焦点をあ

て、トランスジェニックマウスでの発現を確かめた。クジラとマウスで大きな変化はなかったものの、大規模塩基配列決定と発生の実験を組み合わせる点で、分子進化学の研究として重要であると評価できる。