

氏名	Hettiarachchi Nadeeka Nilmini
学位(専攻分野)	博士(理学)
学位記番号	総研大甲第 1842 号
学位授与の日付	平成28年3月24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	Heterogeneous characteristics of Conserved Noncoding Sequences (CNSs) in Eukaryotes
論文審査委員	主査 教授 角谷 徹仁 教授 倉田 のり 教授 藤山 秋佐夫 准教授 池尾 一穂 教授 颯田 葉子

論文内容の要旨
Summary of thesis contents

Comparative genomics approach has made it feasible to determine potential functional elements in the eukaryote genome via computational analyses. Studies on Conserved Noncoding Sequences (CNSs) have reached its climax where the computationally discovered functional elements in genomes have also been experimentally verified where they were found to function as regulatory elements governing gene expression. CNSs have been verified to function as enhancers, and as well repressors. Earlier all noncoding region of genomes was referred to as “junk DNA” but now it became clear that some noncoding regions which are evolutionarily conserved, namely CNSs, are as important as the coding regions.

I first determined lineage specific CNSs of eudicots, grasses, monocots, angiosperms, and land plants and plants. Here I identified 27, 6536, 204, 19, 2 eudicot, grass, monocot, angiosperm, vascular plant specific CNSs respectively. Average percentage identity for CNSs in all groups was more than 80% and the average length varied across groups.

Since the number of CNSs varied considerably across lineages, I tried to determine if this could be due to evolutionary rate, number of species in each group or relatively short divergence time for grass lineage. Even with pairs of species from eudicots and grasses with same divergence time, eudicots had less number of CNSs than monocots. Next I showed that the number of species in each group cannot be a reason for the difference by considering same number of species from eudicots and monocots. In this case also eudicots had less number of CNSs (69) than monocots (204). With respect to the evolutionary rate I found that eudicots have a saturated Ds (2.44) compared to grasses and monocots. If evolutionary rate played a key role in number of CNSs, lineage specific genes should follow the same pattern in abundance as CNSs. Surprisingly Eudicots had more lineage specific genes (2439) than CNSs whereas monocots and grasses (113, 444) had less lineage specific genes. I also found that UTR (untranslated region) CNSs were overrepresented with respect to other regions (introns and intergenic regions). The GO (gene ontology) analysis for the likely target genes of CNSs showed that these genes are related to transcription regulation and development. Another discovery was that CNSs are flanked by sequences showing an increase of GC content and CNSs were also GC rich. Next I tested if these high GC regions are recombination hot spots, as recombination hot spots are known to be related with high GC content in the human genome. None of the CNSs overlapped with recombination hot spots, while high GC CNSs have a higher propensity to form nucleosomes.

(別紙様式 2)
(Separate Form 2)

The lineage specific plant CNSs I identified were all GC rich, but mammalian CNSs have been reported to have low GC content. CNSs of diverse lineages follow different patterns in abundance, sequence composition and location. I therefore conducted a thorough analysis of CNSs in diverse groups of Eukaryotes with respect to GC content heterogeneity. I examined a total of 55 Eukaryote genomes (24 fungi, 19 invertebrates, and 12 non-mammalian vertebrates) as to find lineage specific features of CNSs. Fungi and invertebrate CNSs are predominantly GC rich, whereas non-mammalian vertebrate CNSs are GC poor, similar to mammalian CNSs. This result suggests that the CNS GC content transition occurred from the ancestral GC rich state of invertebrates to GC poor in the vertebrate lineage probably due to the enrollment of GC poor binding sites that are lineage specific. To test how the transition of GC content could have occurred, I determined the GC contents of transcription factor (TF) binding sites and their level of conservation in outgroup lineages. GC content of CNSs also showed a correlation with the location in the genome; GC poor CNSs showed a higher probability to be located in open chromatin regions and GC rich CNSs showed a tendency to locate in heterochromatin regions. The histone modification signal analysis showed that CNSs overlapped with more H3K27Ac and H3K4Me1 compared to the random expectation. These histone marks are signatures of active enhancer regions. The predicted target genes of CNSs also agreed with the previous analysis where the most overrepresented GO term was related to transcription and development. Vertebrate ubiquitous TF binding sites are significantly GC rich compared to tissue specific ones. In contrast, plants have GC rich tissue specific binding sites compared to ubiquitous ones. This heterogeneity in GC content therefore must be attributable to TF binding sites and I found that vertebrate tissue specific TFs are more lineage specific than ubiquitous ones, whereas plant tissue specific and ubiquitous binding sites showed no significant difference in conservation implying the underrepresentation of tissue specific TFs among conserved TFs is a specific feature in vertebrate lineage.

(別紙様式 3)
(Separate Form 3)

博士論文の審査結果の要旨

Summary of the results of the doctoral thesis screening

ゲノム中の遺伝子をコードしない領域にも種間で保存された配列があり、これらを Conserved Noncoding Sequence (CNS) と総称する。配列情報から同定された CNS のいくつかは遺伝子発現制御に重要であることが、動物で実験的に示され、配列情報解析から出発するこのアプローチはゲノム機能の探索法として有力と考えられる。一方、植物での CNS 解析はあまり行われてこなかった。出願者は植物の CNS 探索から出発して、以下の二つの研究を行った。

(1) まず、15 種類の植物種のゲノム情報をもとに、双子葉植物、イネ科植物、単子葉植物、被子植物、維管束植物のそれぞれに特異的な CNS を探索した。不思議なことに、単子葉植物特異的な CNS は双子葉植物特異的な CNS よりもずっと多かった。このことは、調べた種の数や分岐時間の違いでは説明できなかった。単子葉植物に特異的な体制や発生過程との関連が検討されている。一般に CNS は遺伝子内非翻訳領域に濃縮されており、転写制御や発生に関わる遺伝子の近くに多かった。植物の CNS は一般に GC-rich であることがわかり、これは、哺乳類の CNS が GC-poor なのと対照的である。また、この GC-rich CNS はヌクレオソームを形成する傾向があった。

(2) 植物の CNS は GC-rich で哺乳類の CNS は GC-poor である。この傾向の違いを理解するため、さらに広い範囲の種の CNS を調べた。菌類や無脊椎動物の CNS は植物と同様に GC-rich で、哺乳類以外の脊椎動物の CNS は GC-poor だった。GC-poor な CNS 形成は、脊椎動物に特異的な現象が関与すると示唆された。GC-poor な CNS は開いたクロマチンにあることが多く、GC-rich な CNS はヘテロクロマチンにあることが多かった。また、このような CNS の傾向と同様に、脊椎動物の組織特異的な転写因子の結合部位は、ユビキタスに発現するものより GC-poor であり、植物では、この関係は逆転する。上記の CNS の塩基組成の違いは系統特異的な転写因子結合部位の違いとして理解できる。

これらの研究は、植物、脊椎動物をはじめ多くの生物の CNS の進化と遺伝子制御における役割について、これまでの知見を大きく前進させるものである。審査員全員で審査した結果、本大学院における学位授与の水準を十分に満たす論文であると判断した。

なお、本研究のうち (1) の部分は、すでに出願者が筆頭著者の論文として国際誌に出版されている。また、(2) の部分も、結果はほぼまとめ、現在投稿論文を執筆中である。