

氏 名 太 田 聡 史

学位（専攻分野） 博士(理学)

学 位 記 番 号 総研大甲第353号

学位授与の日付 平成10年9月30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4条第1項該当

学 位 論 文 題 目 Development of an integrated system for molecular
evolutionary study and its application

論 文 審 査 委 員 主 査 教 授 館野 義男
教 授 池村 淑道
教 授 五條堀 孝
教 授 長谷川 政美
教 授 西川 建

Development of an Integrated System for Molecular Evolutionary Study and Its Application

1. "DeepForest": A molecular evolutionary analysis system

The DeepForest is a molecular phylogenetic analysis system implemented in a parallel logic programming language; KL1/KLIC. This is not a set of application programs, but a set of libraries. The major objectives of the DeepForest are (1) reconstruction of phylogenetic trees, (2) sequence data operation, and (3) tree superimposition. One of the most prominent features in the DeepForest is parallel execution. I implemented an application program on a shared memory parallel machine; CRAY CS-6400, to examine its efficiency, and found that powerful symbol operation and flexible parallel computation in the DeepForest could provide efficient sequence analyses.

2. "ThreeTree": A new method to reconstruct phylogenetic trees

Although various methods have been proposed to reconstruct molecular phylogenetic trees, no method has been shown to be perfect. Different methods sometimes produce different trees from the same data. Conversely, if different methods produced the identical result, the obtained tree could be reliable. Therefore, it is meaningful to invent a new method based on a different criterion from existing ones.

At present, many tree-making methods are in general based on minimum evolution criterion. I developed a method, the ThreeTree method, based not on that criterion, but on examining branch lengths for all the possible combinations of three OTUs at a time. This method is guaranteed to lead to the true tree, when pure distance additivity holds. When distances are not purely additive and the guarantee does not hold, the reliability can still be improved through the maximum likelihood estimation of branch lengths.

I developed a program of this method, and performed computer simulations to examine its reliability. The result shows that the new method is quite reliable and robust, even if the evolutionary rate is set to be different from branch to branch. However, its ability to infer the correct tree is lower than the neighbor joining method, when diversity of among sequences used is large. On the other hand, when amino acid sequence data are used, the new method complements the neighbor joining method in the case that the latter performs poorly. When the new method makes a wrong bifurcating node at a step owing to a short internal branch involved, it has a tendency to produce a trifurcating node instead of a worse bifurcating node in the next step. This means that this method is more resistant to adverse influences originated in the preceding steps

than other stepwise methods.

3. Phylogenetic relationship of muscle tissues deduced from superimposition of gene trees

There has been no systematic study on the phylogenetic relationships of muscle tissues. The tissues can be divided into six classes such as smooth, fast skeletal, slow skeletal and cardiac muscle tissues, striated and smooth muscle. I reconstructed a phylogenetic tree of each of five protein genes that are expressed in the six muscle tissues, and inferred the phylogeny the muscle tissues by superimposition of those gene trees. To carry out the superimposition, I developed a new program by use of libraries of the DeepForest.

It is noted that actin gene, one of the five genes, is known to be duplicated before vertebrates and arthropods diverged. The gene consequently produced a tree which consisted of two subtrees; one from one duplicated gene and the other from another. In other words, actin gene is said to be paralogous. To construct the superimposed tree mentioned, therefore, I actually used six gene trees in total instead of five.

With respect to the muscle evolution of arthropods and chordates, the result shows that while arthropod striated, urochordate smooth, and vertebrate skeletal-cardiac muscles share a common ancestor, arthropod nonmuscle and vertebrate smooth-non muscle share another common ancestor. The superimposed tree also indicates that the emergence of skeletal-cardiac muscle type tissues preceded the vertebrate and arthropod divergence (ca. 700 million years ago). Somewhat surprisingly, vertebrate smooth muscles seem to have evolved independently from other muscles. Furthermore, while skeletal muscles are not monophyletic, cardiac and slow skeletal muscles make a clade. The result finally implies that an ancestral form of skeletal muscle emerged probably as early as vertebrate nonmuscle type tissue emerged.

論文の審査結果の要旨

太田君の学位論文は4章から成り、第1章は序論、第2章は分子進化解析ツールの開発、第3章は系統樹作成法の開発、そして第4章は筋組織系統進化の解析という内容になっている。各章は緩やかな関係を保ちながら、全体としては分子進化学の解析方法の開発とその生物学的応用という2点を狙った構成になっている。序論を除けば、方法の開発が3分の2、生物学的応用が3分1という量的構成になっているが、主眼は生物学的応用に置かれていると思われる。

第2章では、分子進化解析ツールを、コンピュータの並列処理言語からなるモジュールの組み合わせで用途別に作っていくという独自の考えを展開している。このため、太田君は、70もの並列言語モジュールを開発し、これを用いて20の応用プログラムを作成した。いわば、モジュールが言語という考えである。これらの開発の中で特に独創的と思われるのは、2つ以上の系統樹を重ね合わせて、その評価を行うという点である。2つの系統樹間の樹形の違いを量的に計る方法は既に存在しているが、彼はこの方法を応用して、重ね合わさった複数の系統樹のスコアを求め、その評価を定量的に行う方法を考案した。

第4章では、筋組織の分子進化学的に解析するため、まず筋組織を6つのクラスに分類し、それぞれの組織に存在する遺伝子データを多くの生物種についてデータベースから獲得した。この過程では、データベース検索だけでなく多く原著論文を読んで目的に適合するミオシン、アクチン、トロポニン遺伝子など6種を得ている。太田君は、「筋組織の進化は存在する遺伝子の進化に反映される」という考えのもとに遺伝子の系統解析を行った。この解析において、第2章で考案した系統樹の重ね合わせとその評価法を応用している。つまり6つに分類された筋組織のおのおのについて、該当する種別の遺伝子の配列情報から系統樹を求め、それらを重ね合わせて複合系統樹を作成しその評価を行った。重ね合わせには多くの組み合わせがあるが、その中で最も良い重ね合わせスコアとなった複合系統樹を選び出すのである。この結果、骨格-心筋型組織の出現は脊椎動物と節足動物の分岐に先行しているという新たな考察を導いている。また、脊椎平滑筋が他の筋から独立に進化していることや、骨格筋の進化は同一祖先型ではないが、心筋の進化は同一祖先型であるという興味ある議論も展開している。

第3章は第4章の仕事を遂行する過程で生じた彼の考えをまとめたものである。その考えとは、系統樹作成で1つの原理となっている最小進化の概念を用いずに、3つの枝をもつ樹を単位として、その枝の長さを評価しながら系統樹を作成するというものである。彼はこの新しい方法をコンピュータシミュレーションで比較評価し、最尤法や近隣結合法など既存の方法が犯す間違いを補正する性質があることを見い出している。審査員一同は、太田君の学位論文は、以上に述べたように、問題設定の有意性、そしてその解決における遂行能力と独創性をよく表しており、専攻の基準を満たしていると判断した。