

Title	Analysis of invariant sequences among 266 complete genomes
Author(s)	後藤, 直久
Citation	大阪大学, 2007, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/48769
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏 名	後 藤 直 久
博士の専攻分野の名称	博 士 (理 学)
学 位 記 番 号	第 2 1 6 3 5 号
学 位 授 与 年 月 日	平成 19 年 12 月 20 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 理学研究科生物科学専攻
学 位 論 文 名	Analysis of invariant sequences among 266 complete genomes (266 種のゲノム配列間で保存された不変配列の解析)
論 文 審 査 委 員	(主査) 教 授 安 永 照 雄 (副査) 教 授 岡 田 雅 人 教 授 中 村 春 木

論 文 内 容 の 要 旨

バクテリアから動植物・ヒトまでにわたる全ての生物のゲノムに保存されている配列はいったい何であろうか。それも、挿入・欠失なく完全に保存されている配列（不変配列）は存在するだろうか。全生物に保存されている不変配列は、全生物の共通祖先のゲノム配列から現在まで変わらず存在し続け、生物の生命活動に必須の重要な役割を担うと考えられる。現在ゲノム全配列が入手可能な 250 種以上の生物すべてに保存された不変配列を求めることにより、この疑問に対する答えに迫ることが可能である。

しかし、ゲノム全配列は非常に長くデータ量が多いため、既存のソフトウェアではこの解析を行うことは困難であった。このため、複数のゲノム全配列に保存されている不変配列を高速に検出するソフトウェア“CONSERV”を開発した。CONSERV は Generalized suffix tree と Ukkonen のアルゴリズムを使用し、完全一致しか検出できないが、対象ゲノム配列の合計長に比例した時間・メモリ消費量で高速に保存配列を検出できる。塩基配列だけでなくアミノ酸配列にも対応し、真核生物など複数の染色体を持つ生物にも対応している。

2005 年 7 月までにゲノム全配列が決定した 266 種（真正細菌 217 種、古細菌 22 種、真核生物 27 種）を対象として不変配列を求めた。まず、代表として 143 種（真正細菌 125 種、古細菌 16 種、真核生物 2 種）について、CONSERV を使用し、143 種全てのゲノム配列に保存されている長さ 11 塩基以上の不変配列を求めた。そして、残りの 125 種のゲノムに 143 種での不変配列が存在するかを確認し、266 種における不変配列を求めた。その結果、3 配列が不変配列として検出された。最長の不変配列は AAGTCGTAACAAGGT (15 塩基) であった。この配列は 16S/18S rRNA の 1492 塩基め (大腸菌 *Escherichia coli* K12 の値) に存在した。この配列の位置する領域は、従来から 16S/18S rRNA の中でも特に高く保存されていると言われていたが、本研究は、この配列がゲノムにおいてもっとも長く連続して保存された配列であることを明らかとした。

同様の解析を 266 種の全遺伝子のアミノ酸配列についても行った。その結果、長さ 6aa 以上の不変配列は 6 配列検出された。8aa の配列 GHVDHGKT と DTPGHVDF が最長の不変配列であった。GHVDHGKT は EF-Tu および IF-2、DTPGHVDF は EF-G に存在する配列であった。この 2 配列は、真核生物ではミトコンドリア性のタンパク質に存在したが、その遺伝子は核の染色体に存在した。真核生物のこれらのホモログは不変配列と 1-2 残基異なる配列を持っていた。この結果はミトコンドリア細胞内共生説と一致する。

一方、不変配列の一部を持たない例外的生物の存在も明らかとなった。特に、超好熱性古細菌 *Nanoarchaeum equitans* には、塩基配列 2 配列、アミノ酸配列 3 配列の不変配列が存在しなかった。理由として、高温高压の環境や、他生物との共生の影響が考えられる。

本研究で明らかとした不変配列は、全生物の共通祖先のゲノムに存在し、進化の過程で変わらず保存され続けた、進化的・機能的に非常に重要な配列であると示唆される。また、最長の不変配列が存在する 16S/18S rRNA、EF-Tu、IF-2、EF-G がリボソームにおけるタンパク質合成に関与する分子であったことは、タンパク質合成が生命の基本的で重要な要素であることを従来とは別の面から強調する。

論文審査の結果の要旨

後藤直久君は、共通祖先から現存生物に至る進化の過程でゲノム中に不変に保たれた塩基配列およびアミノ酸配列が存在するのか、存在するとすればそれはどのような配列かという問題に興味を持ち研究を行った。この問題を解析するために、情報科学分野で研究が進んでいた suffix tree と呼ばれるデータ構造を用いて、ゲノム配列間に保存された配列を高速に見出すことが出来る新たな手法を開発した。この手法を用いてバクテリアからヒトまで 266 種のゲノム配列に保存されている長さ 11 塩基以上の不変配列が 3 配列存在することを見出した。最長の不変配列は AAGTCGTAACAAGGT (15 塩基) で、16S/18S rRNA の 1492 塩基め (大腸菌 *E. coli* K12) に存在した。この領域は、従来から特に強く保存されていることが知られていたが、本研究によりこの配列が全生物のゲノム中で最も長く連続して保存された配列であることが明らかとなった。また同様の解析を蛋白質アミノ酸配列でも行い、最長の不変配列が長さ 8 アミノ酸で、IF2/EF-Tu 中の GHVDHGKT、および EF-G 中の DTPGHVDF であることを示した。真核生物ではこれらの配列は、ミトコンドリアで働く蛋白質に存在した。このことはミトコンドリアの細胞内共生説を新たな視点から支持するものである。本研究により明らかになった不変配列は、機能的に重要な配列であり、現存生物の共通祖先ゲノムに存在し進化の過程で保存され続けたと考えられる。これらの結果はゲノム進化の研究において新たな知見を与えるものである。よって本論文は博士 (理学) の学位論文として十分価値あるものと認める。