

論文審査の結果の要旨

論文提出者氏名

内古閑 伸之

本論文は1編、5章から成り、第1章では序論、第2章では解析方法と配列、第3章では結果、第4章では考察、第5章では結論が述べられている。

第1章の序論では、本学位論文で行われた研究の背景と目的について詳しく述べられている。ゲノムDNAはタンパク質をコードしている領域とコードしていない領域に分けられる。それらの領域における塩基配列は決してランダムではなく、連続する k 塩基長配列である k -タプルの使用頻度には偏りが存在する。論文提出者は、最初に、 k -タプル解析を利用した研究により明らかにされているタンパク質コード領域と非コード領域での塩基配列構造の特徴について概説している。そして、それらの研究の中で、特にタンパク質コード領域と非コード領域で2、3-タプルの使用頻度に相関があることを示唆したManiの報告(J. Theor. Biol. 158, 429-445, 1992)に注目し、それが本論文の研究を行う動機のひとつになったと述べている。機能がまったく異なるコード領域と非コード領域の塩基配列構造に類似性があるとすれば、コード領域と非コード領域を含むゲノムDNAの進化について重要な知見を得ることができると考えられる。また、その知見はゲノムの塩基配列からタンパク質をコードしている領域を予測するというゲノムインフォマティクスにおける最も基本的な解析の精度を向上させるために利用できるかもしれない。Maniの解析には非コード領域のデータにコード領域が含まれている等の問題点があったため、論文提出者は塩基配列構造の類似性を k -タプルの使用頻度の偏りから明らかにする新たな方法を考案するとともに、最近になって決定された多数のゲノムDNAの全塩基配列を利用して、タンパク質コード領域と非コード領域の塩基配列構造の類似性を明らかにする研究を行った。

第2章の解析方法と配列では、本論文の解析において中心的役割を果たしている高頻度タプル(SPT: significantly preferred tuple)と全ゲノム高頻度タプル(WGSPT: whole genome SPT)の定義、解析に使用したゲノムの塩基配列とコンピュータの環境などについて述べられている。SPTは各領域に対して定義されるタプルで、その領域と同じ塩基組成をもつランダム配列での出現頻度に対して実際の出現頻度が統計的に有意に多い長さが k 塩基長のタプルである。一方、WGSPTは各ゲノムに対して定義されるタプルで、コード領域と非コード領域に共通して高い割合で見られる長さが k 塩基長のSPTである。解析には25種の単細胞生物ゲノムの全塩基配列が使用されているが、それらは古細菌が6種、真正細菌が18種、真核生物が1種(16染色体)である。

第3章では、第2章の方法に従って解析を行った結果について述べられている。最初に、各領域に対して定義されたSPTが領域内で一様に分布し、領域内の一部に局在している繰り返し配列などに起因するものではないことが示されている。次に、タンパク質コード領域と非コード領域におけるタプルの使用頻度の偏りを比較するために、各SPTが見られるコード領域の割合と非コード領域の割合との相関が調べられている。多くのゲノムで高い正の相関があることが5塩基長以下のSPTについて示され、コード領域と非コード領域でタプルの使用頻度の偏りが統計的に有意に類似していることが明らかにされている。

見出された正の相関は、タンパク質コード領域と非コード領域の両方で統計的に有意に多用されているタプルであるWGSPTが存在することを意味している。論文提出者はこのWGSPTを利用して、コード領域と非コード領域の塩基配列構造の類似性について詳しい解析を行っている。そして、各ゲノムの

WGSPT は互いに相補的塩基配列をもつものが多いこと、WGSPT が見られる領域がゲノム全体に一樣に分布していることを示している。また、3 塩基長 WGSPT は、コード領域ではリーディングフレームに最も高い頻度で出現してコドンに関連していたが、非コード領域では仮想的に設けたフレームとまったく関連がないことを示した。偽遺伝子を含む非コード領域は偽遺伝子を含まない非コード領域に比べ高い割合で WGSPT をもち、WGSPT が偽遺伝子と深く関係していることも明らかにした。それに対して、RNA 遺伝子や繰り返し配列を含むタンパク質非コード領域ではあまり顕著な特徴は見られなかったと述べている。さらに、ゲノム間で WGSPT の比較を行ない、進化的距離が近いと考えられる同じ属のゲノム同士は WGSPT を高い割合で共有していたが、進化的距離の遠い異なる属のゲノム同士では WGSPT の共通性が見られないことを示した。この傾向は特に 4 塩基長以上の WGSPT で顕著であった。

第 4 章では第 3 章で得られた結果について考察を行っている。論文提出者はまず得られた結果と他の研究で得られた結果との比較検討を行っている。Mani の研究により示唆されたタンパク質コード領域と非コード領域の塩基配列構造の類似性が明確に証明できたこと、Karlin や Nakashima らのタプル解析の結果からもこの塩基配列構造の類似性が読みとれることについて述べている。次に、コード領域と非コード領域でのタプルの使われ方が類似している原因について議論している。そして、第 3 章で述べられている結果は、ゲノムがその一部分が重複と逆挿入を繰り返して進化してきたことの証拠であると結論づけている。

以上のように、本論文は 25 種類の単細胞生物ゲノムにおいて、タンパク質をコードしている領域とコードしていない領域で塩基配列の構造が統計的に有意に似ていることをはじめて明確に示した研究である。この結果は生物のゲノム DNA が重複と逆挿入を繰り返しながら進化してきたことを示す重要な証拠を与えるとともに、ゲノムの塩基配列からタンパク質をコードする領域を予測する方法に対して重要な知見を与えるものであると考えられる。

なお、本論文は陶山明との共同研究であるが、論文提出者が主体となって研究全体を行ったもので、論文提出者の寄与が十分であると判断する。したがって、博士(学術)の学位を授与できると認める。