

論文内容の要旨

申請者氏名 福島 敦史

A deep understanding of the structural characteristics of living organisms requires an exhaustive analysis of the DNA sequence that constitutes the blueprint of their organic architecture. One of these characteristics of the genomic DNA is the latent periodicity in base pairs, the elucidation of which may play an important role in unveiling basic genome architecture of organisms and thus a way of discriminating among species based on their sequence along. This thesis performed analysis of the periodicities in nucleotide sequence of prokaryotes genomes, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Drosophila melanogaster*, *Anopheles gambiae*, and *Homo sapiens* using a methodology based on power spectrum and to characterize them. Power spectrum analysis is a widely used methodology in physics, especially in signal recognition theory to measure the periodicity of sound and light signals. The application of this methodology to the analysis of the DNA sequences, also attempted in previous studies, has shown a high degree of effectiveness in molecular biology, leading to the finding of long-range correlations in genomic DNA. To characterize these periodical regions in genomic DNA, I propose periodic nucleotide distribution represented by a parameter F_k , computed as the frequency of periodic pairs separated by a number of k nucleotide. For the *C. elegans* genome, a periodicity of 68 bp was found in chromosome I, a 59 bp periodicity in chromosome II, and a 94 bp periodicity in chromosome III. In *A. thaliana*, I obtained three periodicities (248, 167, and 126 bp) in chromosome 3, three (174, 88, and 59 bp) in chromosome 4, and four (356, 174, 88, and 59 bp) in chromosome 5. These findings are in good agreement with ORFs consisting of Gly-rich amino acid sequences. Additionally, in the human genome, periodicity was detected in chromosomes 21 and 22 with a factor of 167 or 84. The periodicity of 167 bp is identical to the DNA that forms two complete helical turns in nucleosome. For *D. melanogaster* and *A. gambiae*, it was found that the G and C content contains flat regions in middle frequency, which are associated with randomness of base sequence composition. This unique property has not been observed in *Saccharomyces cerevisiae*, *C. elegans*, *A. thaliana*, and *H. sapiens*. The present study goes beyond the simple determination of this type of periodical correlations and attempts to determine their biological and physiological implications in the genomic structure of organisms such as repetitive sequence, long-range correlation, and mainly genome evolution.

論文審査結果の要旨

申請者氏名 福島 敦史

生命の起源や進化の問題を自己組織化あるいは複雑系という立場から理解しようとする試みは、カウフマン、スタンレイなどの複雑系物理学者によりはじまった。その一つにゲノム配列内に潜む組織構造を数理解析により解明する試みがある。本論文では(1)塩基配列上に潜む周期構造をバクテリアからヒトにおけるゲノムを対象に解析し、生物種固有のゲノム周期構造を検出した。検出された配列の特徴を分子生物学に基づく知見により考察した。また、(2)パワースペクトルを用いることによりゲノム配列に潜む周期構造をフラクタル性により説明することが可能となる。そこで、ゲノム配列に潜むフラクタル性をゲノムに含まれる繰り返し構造ならびに遺伝子の配置との関連から考察した。本論文では、以下の成果を得ている。(1)線虫(*C. elegans*)において 68 塩基周期(第 1 染色体)、59 塩基周期(第 2 染色体)、94 塩基周期(第 3 染色体)の配列が数十から数百回繰り返されている領域がゲノムの中央領域に検出された。線虫ゲノムはホロセントリック染色体をもち、分散したキネトコア構造をもつことから、本研究で検出した繰り返し構造がセントロメアの機能と関連する可能性を示した。解析した多くのゲノムにおいて GGN を核とする周期性が見出された。GGN の繰り返し配列はゲノムの折りたたみなどの高次構造をつくるために重要な配列であることから、GGN の繰り返し構造がゲノム中に一定の周期でゲノム中に散在すると結論付けられる。ヒトゲノムにおいて、167 塩基および 84 塩基における周期性がゲノム全体で検出することができた。これらの周期構造の核となる配列もまた GGN である。また、167 塩基周期はヌクレオソームにおける DNA の二巻構造と対応する。これらのことから、GGN を核とする配列がヌクレオソーム構造の形成に関与することが示された。(2)塩基における周期の逆数(1/bp)を周波数(f)とよぶ。周波数の対数($\log(f)$)に対するパワースペクトル強度の対数($\log S(f)$)の傾き β ($S(f) \propto f^\beta$) はゲノムの自己組織構造を検出するための指標となる。ヒトにおけるそれぞれの染色体について 10 キロ塩基付近における β 値とゲノム GC% には有意な負の相関($r=-0.84$)があり、第 19 染色体において最も -1 に近い β 値となった。10 キロ塩基のサイズはおおまかには遺伝子の大きさと対応する。また、ヒトゲノムにおいては GC% が高いほど遺伝子密度が高くなる傾向がある。すなわち、フラクタル性の高い染色体ほど遺伝子密度が高いことから、ヒトゲノムにおけるフラクタル性は主に遺伝子密度により説明することができる。このことはバクテリアを含む多くの生物種に対する解析とも一貫性を有する。さらにこのようなフラクタル性を生み出す原因を数理モデルを用いて考察した。

以上のように、本論文はゲノムの構造を分子生物学と情報科学の両面から検討したもので、学術上、応用上貢献するところが少なくない。よって審査委員一同は、本論文が博士(バイオサイエンス)の学位論文として価値あるものと認めた。