

DNA 塩基配列におけるコード・非コード領域の周期性と自己組織化

理工学研究科 物質地球科学専攻

発表者：坂本 博之

指導教官：清野 光弘

生物の基本情報であるゲノムはヌクレオチドの重合体である塩基配列で構成される。このゲノムには、蛋白質を生成するためのアミノ酸配列を指定するコード領域と、その発現の制御に関わる調節領域や偽遺伝子などを含む非コード領域が含まれている。また、非コード領域は蛋白質の発現調節子や偽遺伝子の他にも蛋白質合成に関与しない mRNA の情報なども含まれており、この領域は他の未知の作用が含まれている可能性もある。

塩基配列における周期性の研究は 1990 年代より H. E. Stanley や R. F. Voss、W. Li によって行われ、 $1/f$ ノイズや 3 周期性が存在することが示されてきた。本研究においては、*Eschericia Coliphage* 及び *Caenorhabditis Elegans* の塩基配列を用いて、周期性が配列中のコード領域・非コード領域でどのように観測されるかを観察し、その領域の周期性の特性を調べた。

周期成分は塩基配列のパワースペクトルにより求められた。塩基配列 $\mathbf{x}$ とそれに対応する数値配列 $\mathbf{X}$ は、塩基配列のあるサイトから n までの部分配列を $\mathbf{x}_t = \{x_t, x_{t+1}, x_{t+2}, \dots, x_{t+n-1}\}$ 、長さ n の塩基パターンを $\mathbf{p}_k = \{p_{k1}, p_{k2}, \dots, p_{kn}\}$ 、数値配列の t 番目の数値 $X(\mathbf{x}_t)$ として $X(\mathbf{x}_t) = a_k$ ($\mathbf{x}_t = \mathbf{p}_k$)、 $X(\mathbf{x}_t) = 0$ (その他) の規則により数値化し、そのパワースペクトルを求める。塩基パターンは A、T、C、G の組み合わせにより構成され、その塩基パターンの数と同数のパワースペクトルが観測される。

各試料は数千のコード・非コード領域を持ち、その各領域に対してパワースペクトルが計算された。この結果、*E. Coli*、*C. Elegans* のコード領域はともに 3 周期性を持つことがわかった (図 1)。また、*C. Elegans* の非コード領域は主に A、T の塩基が 10 周期性を持つことがわかり、*E. Coli* と *C. Elegans* の非コード領域に違いがあることが示唆された (図 1)。

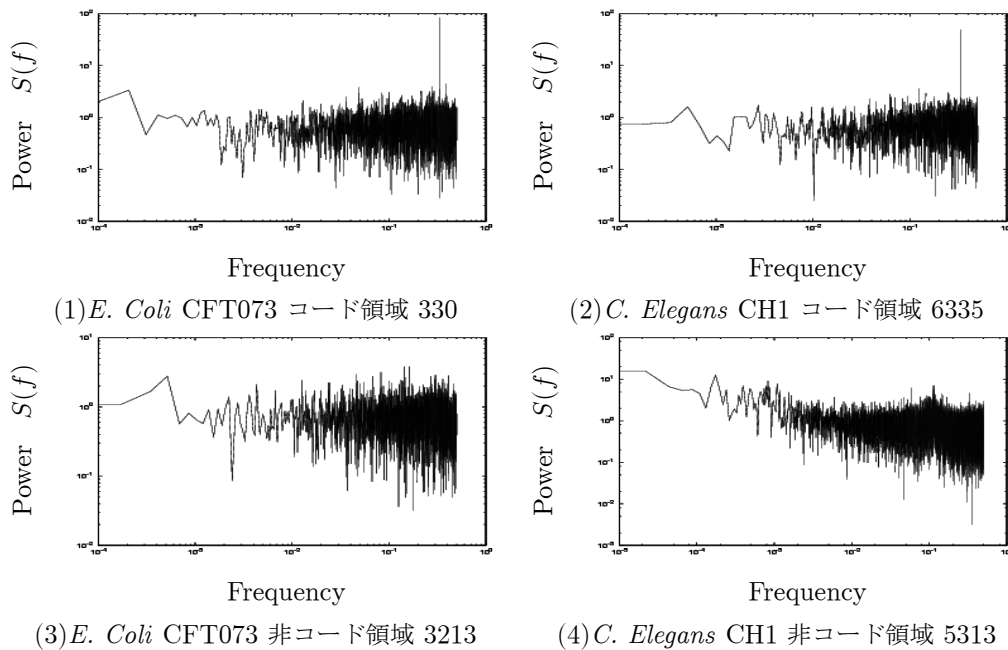


図 1 コード領域・非コード領域における平均パワースペクトル

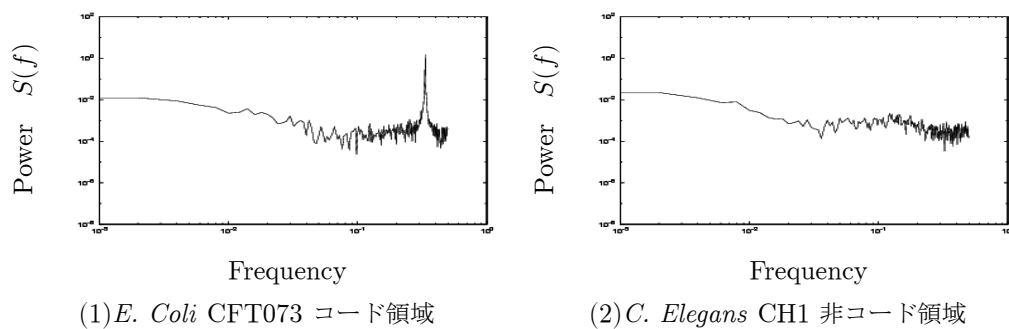
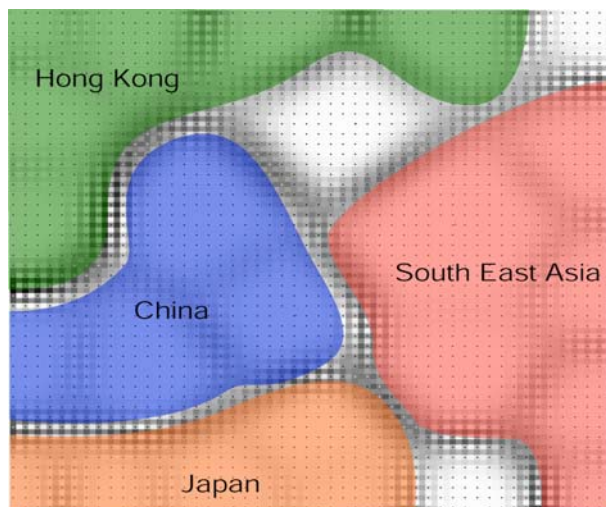


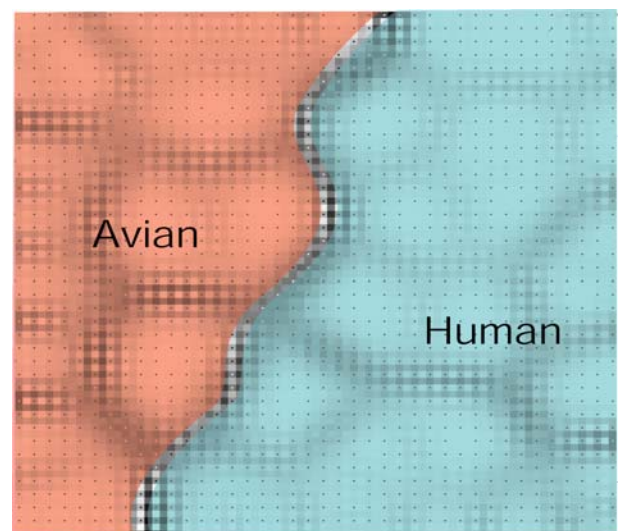
図 2 出現頻度に対する平均パワースペクトル

E. Coli、*C. Elegans* のコード・非コード領域の塩基出現頻度も計算され、出現頻度のパワースペクトルから 3 及び 10 周期性がゲノムの広範囲にわたる現象であることが観測された (図 2)。

これにより、ゲノムの塩基配列はコード領域において 3 周期性をもつことが判明し、非コード領域が活用されている生物では 10 周期性の塩基配列の特性をもつことがわかった。特に、10 周期性は *E. Coli* の非コード領域から観測されなかったことから、*C. Elegans* をはじめとする高等生物では、原生生物には無い機構が非コード領域に含まれていることが示唆された。



(1) HA(H5N1、アジア地域)



(2) M1(H2N2、H3N2、H5N1、H9N2 等、日本)

図 3 HA 及び M1 蛋白に対する SOM

自己組織化は *Influenza A Virus* の変異と地域性や宿主の違いに対して行われた。トリインフルエンザとして知られる H5N1 型のウィルスのヘマグルチニン (HA) は図 3-(1) のような相似性を持つことがわかった。この結果より、現在までのヒト感染能力を持つ H5N1 型ウィルスは、特定の地域で発生したものが伝染しているわけではなく、各地域のトリ感染性の H5N1 型ウィルスが変異することにより発生していることがわかった。一方、今までに流行したヒト及びトリの各型のインフルエンザウィルスは構造蛋白 (M) と非構造蛋白質 (NS) に違いがあり、ウィルスがそれぞれの型ごとに変異を繰り返しているものと考えられる (図 3-(2))。