

論文の内容の要旨

論文題目

単細胞生物における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

内古閑 伸之

本研究の目的はゲノム塩基配列に存在する k -タプル(連続する k 塩基長配列)の使用頻度の偏りからゲノム塩基配列構造を明らかにすることである。塩基配列の k -タプル解析はコード領域、非コード領域ともになされている。コード領域には 3-タプルであるコドンが存在し、コドンはひとつのアミノ酸に対して縮退している。同じアミノ酸をコードする複数のコドンを同義語コドンという。同義語コドンの使用頻度は一様ではなく、遺伝子やゲノムにより異なる偏りが存在している。非コード領域ではおもに 2-タプル解析がなされており、それらの使用頻度も一様ではない。また、コード領域と非コード領域の 2,3-タプルの使用頻度に相関があることが示唆されているが、そこで扱われた非コード領域のデータはコード領域を含んでおりコード領域と非コード領域の相関を正確に示した結果ではない。

そこで、本研究では 25 種の単細胞生物ゲノム(古細菌 6 種、真正細菌 18 種、真核生物 1 種)をコード領域と非コード領域に分割し、それぞれの領域で統計的有意に多く存在する k -タプル(Significantly Preferred Tuple: SPT)を定義した。SPT は実際の出現頻度と、調べる領域と同じ塩基組成を持つランダム配列での出現頻度との差を計算しその差が標準偏差を単位にして、ある距離以上のタプルとして定義される。したがって、SPT は調べる領域の塩基組成の影響を差し引いて得られた高頻度に存在する k -タプルである。ゲノム内で SPT_w が見られるコード領域の割合と非コード領域の割合をそれぞれ $FP(w)$ 、 $FN(w)$ とする。そして、各ゲノムのすべての SPT について FP 、 FN の相関を見ると正の相関が見られた(例:図 1a)。6-タプルではあまり高い相関が見られないが、5 塩基長以下のタプルでは多くのゲノムで高い正の相関が見られる(図 1b)。したがって、5-タプル以下についてコード領域と非コード領域のタプルの出現頻度に相関があることを

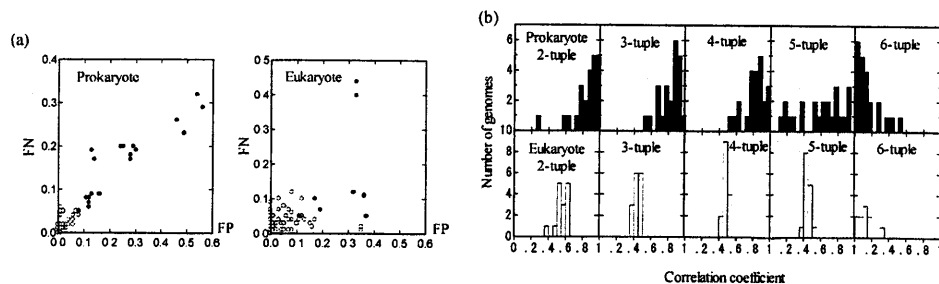


図:コード領域と非コード領域のk-タプル出現頻度の相関関係
 a: *A. pernix*(Prokaryote), *S. cerevisiae*(Eukaryote)の例。黒い点はWGSPT。
 b: 各ゲノムの相関係数の分布

明確に示すことができた。このことは5-タプル以下についてコード領域と非コード領域でk-タプル出現頻度の偏りに類似性があることを意味する。そして各ゲノムについて、コード領域と非コード領域に共通して高い割合(上位20)で見られるSPTを全ゲノムSPT(Whole Genome SPT: WGSPT)として定義した。5-タプル以下についてProkaryoteよりもEukaryoteの相関係数は小さく(図1b)、EukaryoteのWGSPTの数はProkaryoteよりも少なかった(例:図1aの黒い点の数)。

WGSPTをもちいて、ゲノム塩基配列の特徴を調べた。その結果、WGSPTが見られる領域がゲノム内に一様に分布していることと、各ゲノムのWGSPTは互いに相補的なものが多いことが示された。この結果から、ゲノム塩基配列構造は領域の重複や逆挿入により生成されたと考えられる。

コード領域、非コード領域の各領域についてWGSPTが見られる要因について調査した。コード領域には同義語コドンの使用頻度の偏りが存在するため、WGSPTはコドンに影響されていると考えられる。実際、3塩基長WGSPTの多くはリーディングフレームに位置しており、WGSPTはコドンに関連していることがわかった。一方、非コード領域についてはゲノム内に多く存在する偽遺伝子、RNA遺伝子、繰り返し配列に注目し、非コード領域に見られるWGSPTはどの領域に影響されているかを調べた。その結果、RNA遺伝子や繰り返し配列ではあまり顕著な特徴は見られなかったが、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が見られた。偽遺伝子はコード領域が変異などにより機能しなくなった領域で、同義語コドンの使用頻度の偏りが十分残っていると考えられる。コード領域と非コード領域の類似性はコード領域についてはコドン、非コード領域については偽遺伝子の配列の特徴による影響が強いと考えられる。

WGSPTの進化的考察をするためゲノム間で比較を行なった。3塩基長WGSPTについて進化的距離が近いと考えられる同じ属のゲノム同士は高い割合で共通していたが、例外的に少数の進化的距離の遠い属(異なる属)同士でも共通性が見られた。4塩基長WGSPTでは同じ属のゲノム同士でのみ共通性が見られた。このことから、4塩基長のWGSPTは3塩基長のWGSPTよりもゲノムの進化を考えるのに適当な長さであると考えられる。個々のゲノムでWGSPTが見られるが、進化的距離に応じてWGSPTの共通性が異なるのは変異した領域がゲノム全体に重複して起こることが原因と考えられる。ゲノム内で、変異をもった領域がゲノム全体で一様に重複すれば、異なるがWGSPTがみられる。したがって、それぞれのゲノムではWGSPTが見られるが、ゲノム間では進化的距離に応じてWGSPTが異なると考えられる。本研究の結果から、ゲノム塩基配列構造は個々のゲノムについてだけでなく、進化的考察についても領域の重複により説明ができる。