

原始コドン仮説に基づいた遺伝符号の進化に関する研究

三上 達¹、森 太郎²、水田 智史¹

¹ 弘前大学大学院理工学研究科、² 弘前大学工学部電子情報工学科

1 イントロダクション

1.1 コドンとタンパク質合成

遺伝子上では、20種のそれぞれのアミノ酸は3個の塩基の組(コドン)によって符号化されており、現存する生物におけるDNAからタンパク質への合成は



というプロセスによって行われる。

転写はDNAの二重らせん構造の片側を鋳型とし、DNAの塩基と対になる塩基(A-U, C-G)を一つずつ結合させるプロセスである。

また、翻訳によってmRNAのコドンとtRNAが持つmRNAのコドンと対応する塩基配列(アンチコドン)が結合し、tRNAが運んできたアミノ酸がペプチド結合され、その結果タンパク質が合成される。よって、コドンとアミノ酸とを具体的に対応付けているのはtRNAということになる。

tRNAはDNAの塩基配列のtRNA遺伝子が転写されたもので、mRNAのコドンと対応するアミノ酸を運ぶ役割を担っている。基本的に5'末端側から34番目から36番目の塩基がmRNAのコドンと対応するアンチコドンを持っており、これがmRNAのコドンと結合することでアミノ酸が選択される。

1.2 原始コドン仮説

以上のプロセスは進化の過程において得られたものであり、進化が単純な形態からより複雑な形態へと進むという方向性を考慮すると、過去においては塩基1個または2個の組(以下「原始コドン」と称する)でアミノ酸を符号化していた段階があったと考えられる。

原始コドンによってアミノ酸を符号化する場合、1個の原始コドンに複数のアミノ酸が対応することになり、アミノ酸の決定論的な選択性が犠牲になる。しかし、自然淘汰の観点からすれば、確率的に選択されたアミノ酸から作られる配列がタンパク質として生命体

の生存に適した機能を持った場合にだけ生き残り、やがてその組み合わせが確実なものになるように現在のような3個組のコドンのパターンが形成されてきたと考えられる。

そこで、本研究ではゲノムのDNA配列を調べることにより、どのような過程を経て現在のような符号化パターンができてきたのかを明らかにすることを目的とする。

2 方法

本研究で用いたDNA配列は、GenBank[1]からダウンロードした。

2.1 tRNA 遺伝子からのアプローチ

まず、tRNA遺伝子をアミノ酸およびアンチコドン別にそれぞれ分類したもの、そしてtRNA遺伝子をランダムに1000本選択したものについてそれぞれペアワイズアライメントを行いそのスコアの分布を調べた(図1,2)。その結果、アンチコドン別に分類したものの平均スコアが最も高かったので、アンチコドン別に分類したもからコンセンサス配列を求め、さらにアンチコドン部を1個または2個の塩基に置換した配列(以下「原始tRNA遺伝子配列」と称する)を作成し、相同性検索技術等を用いて真核生物の遺伝子間領域を主な検索対象として類似配列を探し出す。

ここで、コンセンサス配列は、tRNA遺伝子のアンチコドン部を堺に3'末端側と5'末端側に分け、それぞれをマルチプルアライメントすることにより作成する。

2.2 タンパク質遺伝子からのアプローチ

原始コドンによってタンパク質を符号化した遺伝子の痕跡が残っている可能性を考え、タンパク質遺伝子の各コドンの第3塩基を削除した配列(以下「原始タ

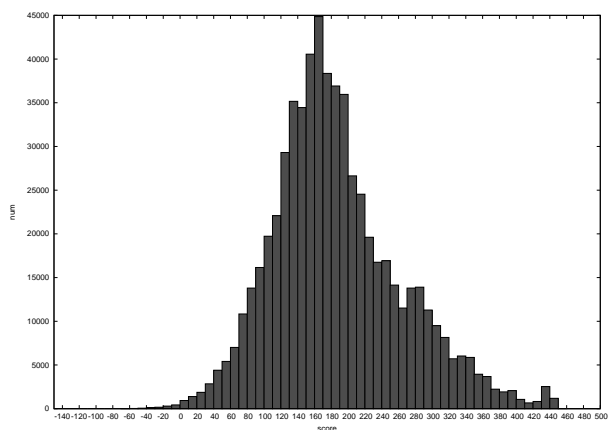


図 1: アンチコドン taa で分類した tRNA 遺伝子間のペアワイズアライメントのスコア分布 (平均 186.0、標準偏差 76.5)

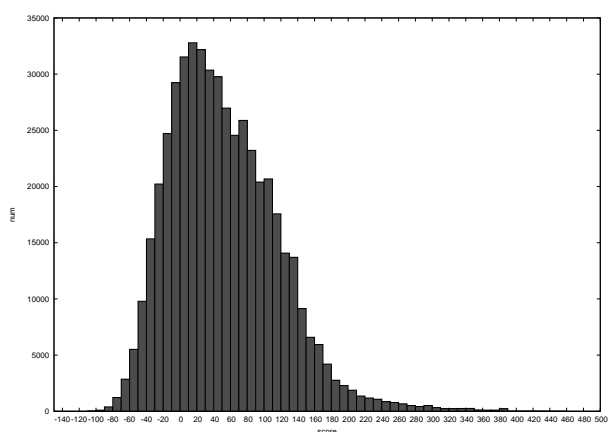


図 2: ランダムに選択した tRNA 遺伝子間のペアワイズアライメントのスコア分布 (平均 51.7、標準偏差 64.5)

ンパク質遺伝子配列」と称する)を作成し、ゲノム配列中を検索して類似配列を探し出す。

対象とする遺伝子としては、生命維持にとって基本的な役割をもったものが適当であると考えられること、および原始タンパク質遺伝子配列を作成する際にイントロンを含んでいると取扱いが難しくなるということから、バクテリアのゲノムから取り出すこととした。

3 結果と考察

tRNA 遺伝子 51,525 本をアンチコドン別に分類し、合計で 1,220 本のコンセンサス配列が得られた。次の配列は、アンチコドン taa に関して得られたコンセンサス配列である。

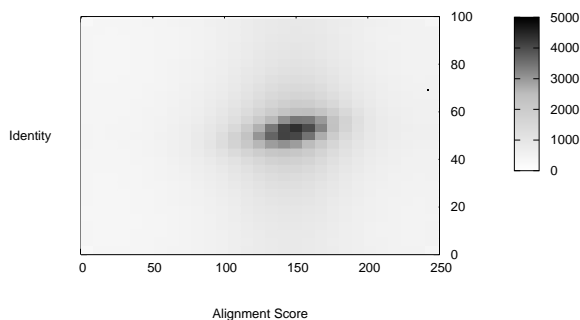


図 3: ランダム配列の出現頻度分布 (実際の原始タンパク質遺伝子配列の最良結果:アライメントスコア 241、一致度:69.3%)

```
TGCCCGCGGTGGCGGAATTGGTAGACGCAAGGGACT-A-AA
ATCCCTCGGTGGTTAAACCGCGACCGTGCCGGTTCGA
```

これらコンセンサス配列をさまざまな生物種のゲノム配列を対象に検索し、原始 tRNA 遺伝子の痕跡を探す。

図 3 は S50 ribosomal protein L36 のタンパク質遺伝子から作られた原始タンパク質遺伝子配列を基にランダム配列を 6 本作成し、シロイヌナズナのゲノムを対象に検索した結果である。アライメントスコアと一致度によってその出現頻度の合計を示している。

実際の原始タンパク質遺伝子配列を検索した最良の結果は遺伝子内のイントロン領域に見付き、そのアライメントスコア、および一致度はそれぞれ 241、69.3%である。この値はランダム配列ではほとんど得ることのできない値であることから、原始タンパク質遺伝子配列の有力な候補であると考えられることができる。

今後、対象とする遺伝子の種類を増やすとともに、より詳細に検索結果を調べることにより、遺伝符号の進化に関する手掛かりを求めていく。

参考文献

- [1] GenBank: <http://www.ncbi.nlm.nih.gov/>