

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 6 月 4 日現在

機関番号：11101
 研究種目：挑戦的萌芽研究
 研究期間：2012～2012
 課題番号：24650152
 研究課題名（和文）原始遺伝符号の痕跡の探索
 研究課題名（英文）Searching for the relics of primitive coding system

研究代表者
 水田 智史（MIZUTA SATOSHI）
 弘前大学・大学院理工学研究科・准教授
 研究者番号：90250589

研究成果の概要（和文）：遺伝符号（暗号）とは、遺伝子に記述されているタンパク質の設計図の書式のことを言います。それは、大変、合理的に組み立てられているのですが、生命の初期段階では書式が異なっていたかもしれません。本研究では、ゲノム配列やタンパク質のアミノ酸配列の中から、いくつかの方法を用いてその痕跡を探索しました。その結果、確証を得るためにはさらに研究を進める必要がありますが、いくつかの候補を得ることができました。

研究成果の概要（英文）：The genetic code is the format of the blueprint of proteins that is stored in genes. The format is assembled very reasonably, but it may have been different in the early stage of life. In this study, we search for the relics of the past genetic code in genomic sequences and amino acid sequences of proteins using several methods. As a result, we got some candidates for the relics, although we have to study further in order to obtain conclusive evidences.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	700,000	210,000	910,000

研究分野：バイオインフォマティクス

科研費の分科・細目：情報学・生体生命情報学

キーワード：ゲノム、コドン、遺伝符号、進化、アミノ酸配列

1. 研究開始当初の背景

生命体を構成するタンパク質は、ゲノム配列に含まれる遺伝子の塩基配列から転写、翻訳のプロセスを経てアミノ酸の1次元配列として合成される。4種の塩基に対してタンパク質を構成するアミノ酸は20種あるため、決定論的にアミノ酸を符号化するためには3個の塩基の組(コドン)によって1個のアミノ酸を符号化する必要があり、現存する生物では実際にそのような符号化が行われている。しかし、太古に生命が誕生し、その後、転写、翻訳のプロセスが形成された当初から、このような合理的な符号化パターンが存在していたとは考えにくい。進化が単純な形態からより複雑な形態へと進むということ

考えた場合、決定性は犠牲になるものの、塩基1個ないし2個の組(以下、原始コドンと記す)によって複数のアミノ酸が符号化され、その内のいずれかのアミノ酸が確率的に選ばれていたという段階があり、長年にわたる進化の過程を経て現在のような決定論的な符号化パターンが形成されてきたと考えるのが自然であろう。しかし、現在のところそのプロセスは明らかになっていない。

2. 研究の目的

過去に原始コドンによりアミノ酸が符号化されていたとすれば、当時のタンパク質遺伝子の痕跡がゲノム配列中に残されている可能性がある。また、原始コドンからなる遺伝子を基に、確率的に選択されたアミノ酸が

1st	2nd			
	U	C	A	G
U	Phe/Leu	Ser	Tyr/STOP	Cys/Trp/STOP
C	Leu	Pro	His/Gln	Arg
A	Ile/Met	Thr	Asn/Lys	Ser/Arg
G	Val	Ala	Asp/Glu	Gly

図 1. 原始コドン表

ら合成される異なる配列のいくつかが生命にとって有用な機能をもったとすれば、ある種の相同タンパク質として現在まで残っていることも考えられる。そこで本研究では、パターンマッチング、および相同性検索等の技術を利用し、様々な生物種のゲノム配列およびタンパク質のアミノ酸配列を検索し、上で述べた痕跡を探し出すことが目的である。

3. 研究の方法

本研究では、現在の 3 個組のコドンにおいて、いくつかのアミノ酸が第 1 および第 2 塩基のみで特定されるという点に着目し、コドンの第 3 塩基を削除した 2 個組の塩基を原始コドンであるという作業仮説を立て、公開されているゲノム配列およびタンパク質のアミノ酸配列データベースを用いて研究を進めた。図 1. はこのようにして作成した原始コドン表である。

また、図 2. は本研究において想定しているコドン進化のシナリオである。図 1 に示されるように、原始コドンはアミノ酸と多対多の対応をもつため、1 個の原始遺伝子から一般に複数のアミノ酸配列が確率論的に合成

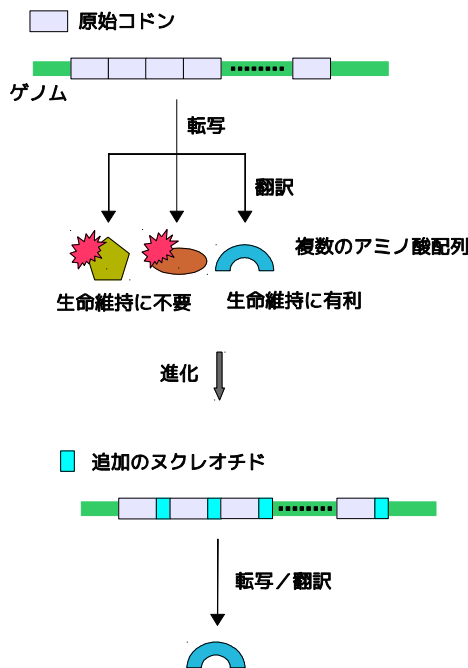


図 2. コドン進化のシナリオ

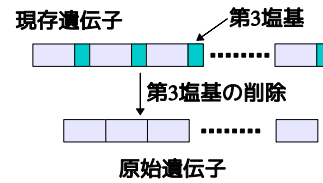


図 3. 原始遺伝子

される。やがてその中で、生命の維持にとって重要な機能をもち合わせたアミノ酸配列が決定論的に合成されるように、原始コドンに対して 1 個の塩基が付加され 3 個組に進化してきたものとする。

このシナリオから、原始コドンの痕跡として次の 2 つの形態が考えられる。

- (a) 原始コドンから構成されたタンパク質遺伝子の痕跡
- (b) 原始コドンから構成された遺伝子を基に合成された相同タンパク質

(a) の痕跡を探索するために、既存の遺伝子の第 3 塩基を削除することにより、原始遺伝子配列を作成し(図 3)、ゲノム配列を対象にして相同性検索を行った。

このための遺伝子としては、できるだけ生命にとって基本的な機能をもったタンパク質が適切であると考え、マイコプラズマ由来の遺伝子を用いた(合計 483 配列)。また、(a) の痕跡は存在するとすればゲノムの非符号化領域であることが想定されるため(図 4)、検索対象には、大きな非符号化領域をもつ真核生物からシロイヌナズナを選択した。

(b) は共通の原始遺伝子から合成された複数のタンパク質が現在まで残されていることを想定している(図 4)。我々はこれを準相同タンパク質と呼ぶことにする。準相同タンパク質の関係にあるタンパク質の間においては、塩基の 2 個組と 1 対 1 の関係にある 5 個のアミノ酸(図 1 の原始コドン表において色の付いたセルにあるアミノ酸)は、挿入・欠損が無い限り、配列の同じ位置に出現することがわかる(図 5)。このことを手掛かりに、上で述べた 5 個のアミノ酸の出現位置の一致数が大きいアミノ酸配列ペアを探索した。

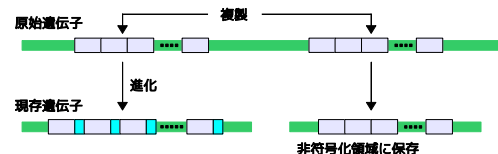


図 4. 原始遺伝子の痕跡

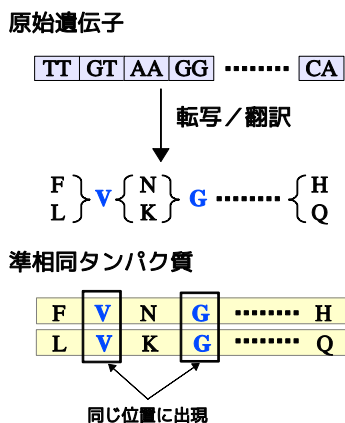


図 5. 準相同タンパク質

ここでも、(a)の場合と同じ理由で、検索要求としてはマイコプラズマ由来のタンパク質配列を選択した。ただし、検索対象は特に限定する必要がないため、SwissProt に登録されている全タンパク質配列を対象とした(合計 531473 配列)。

(b)については挿入・欠損についても考慮し、アミノ酸の出現位置の一致数だけでなく、本研究用に新たに作成したアミノ酸の置換行列を用いた相溶性検索も行った。図 6 は検索に用いた置換行列で、BLOSUM30 を基に、コドンの第 1-2 塩基を共通にもつアミノ酸の間の置換確率は大きいことが期待されることなどを考慮して、網掛けの部分を実験用に修正したものである。検索は、FASTA3 パッケージの glsearch を用いて行った。

なお、準相同タンパク質はお互いに大きく異なる機能をもっているものと仮定する。なぜなら、それらが類似した機能をもっていた場合は、いずれか 1 つの配列を残してその他は淘汰されてしまっている可能性が大きいからである。

4. 研究成果

まず (a) の検索の結果、50S ribosomal protein L32 と 50S ribosomal protein L36

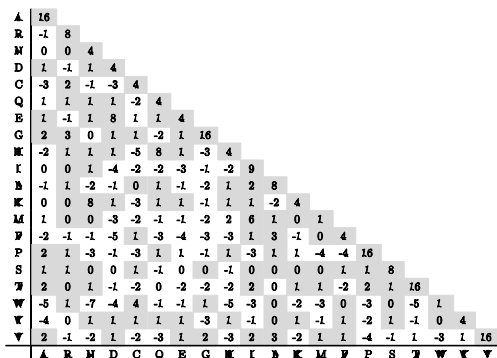


図 6. 置換行列

を元にした 2 個の原始遺伝子候補が見つかった。検索要求の塩基をランダムに入れ替えて作成した配列を用いて評価した p-value の値はそれぞれ 9.9e-03 と 2.6e-03 である。

次に (b) の 5 個のアミノ酸の一致数に基づいた検索の結果、検索要求 50S ribosomal protein L32 に対して UDB-N-acetylmuramate dehydrogenase (*Magnetococcus marinus*)、Collagen alpha-2(I) chain (Human)、Protein smf (*Escherichia coli* K-12) の 3 つの候補が見つかった(カッコ内は由来の生物種)。いずれもアミノ酸の一致数は 5 個であり、ランダム配列を用いた p-value の評価値はそれぞれ 3e-07、1e-07、6e-07 である。

また、(b) の置換行列を用いた検索の結果、次の 3 組の候補が見つかった。いずれも上段が検索要求で、下段が検索結果である。検索結果の e-value は glsearch によって求められ、それぞれ 1.5e-32、3.8e-08、1.8e-07 である。

Dihydrofolate reductase
Segregation and condensation protein (<i>M. genitalium</i>)
Mgp-operon protein 1
50S ribosomal protein L9 (<i>Onion yellows phytoplasma</i>)
P32 adhesin
Protein transport protein SEC31 (<i>D. discoideum</i>)

以上が現時点における検索結果である。いずれも、原始遺伝子符号の痕跡として有望な候補であると考えられる。特に、50S ribosomal protein L32 は (a)、(b) 双方の候補として取り出されており、原始遺伝子符号の痕跡を表すものとして重要であると言える。しかし、現時点においては、検索要求と検索の結果得られた配列の関係が原始遺伝子符号以外のものであるという可能性が排除されないため、それが原始遺伝子符号の痕跡であるという確証を得るためには、今後、さらに研究を進めていく必要がある。具体的には、検索の結果得られた配列の機能を中心に、より詳細な解析を進めているところである。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

- (1) Satoshi Mizuta and Taro Mori, Searching for the Relics of Primitive Codons, International Journal of Advanced Computer Science, Vol.3, 2013 (to be published), 査読有。

[学会発表] (計 3 件)

- (1) 水田智史、森太郎、タンパク質配列にお

ける原始遺伝符号の痕跡の探索、情報処理学会第 75 回全国大会、2013 年 3 月 8 日、東北大学.

- (2) Satoshi Mizuta and Taro Mori, Searching for the Relics of Primitive Codons, The 2012 ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB 2012), October 7-10, 2012, Orlando, Florida, U. S. A.
- (3) 森 太郎、水田 智史、原始遺伝符号の痕跡の探索、平成 24 年 電気学会 電子・情報・システム部門大会、2012 年 9 月 7 日、弘前大学.

6. 研究組織

(1) 研究代表者

水田 智史 (MIZUTA SATOSHI)
弘前大学・大学院理工学研究科・准教授
研究者番号：90250589