

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を

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

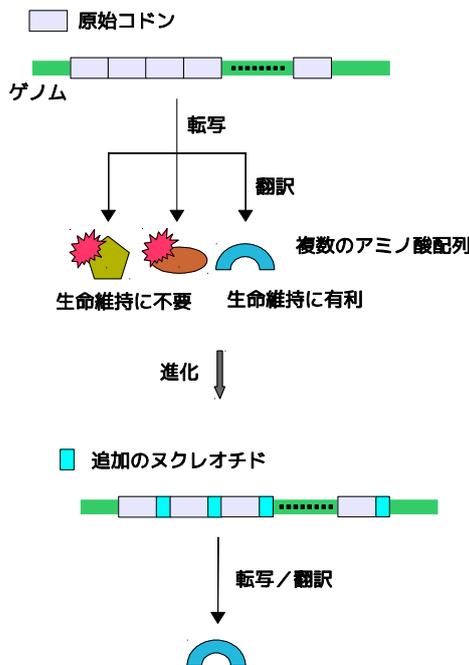


図 2. コドン進化のシナリオ

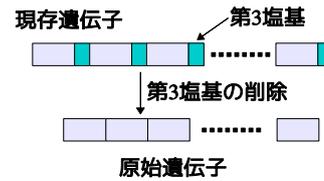


図 3. 原始遺伝子

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。

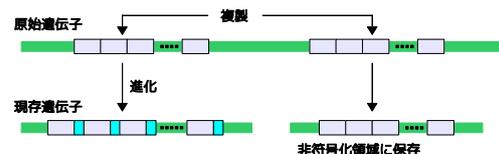


図 4. 原始遺伝子の痕跡

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