

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 17 日現在

機関番号：38005

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24770070

研究課題名(和文) 核ゲノムデータを用いた条鰭類の系統解析に向けて：相同遺伝子選定パイプラインの構築

研究課題名(英文) Resolving ray-finned fish phylogeny with genomic data: Developing an analysis pipeline for identification of homologous genes

研究代表者

井上 潤 (Inoue, Jun)

沖縄科学技術大学院大学・Mathematical Biology・スタッフサイエンティスト

研究者番号：10596779

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、条鰭魚類の系統関係を推定するのに適した遺伝子配列を、膨大なゲノムデータから選び出す解析パイプラインを構築した。パイプラインの作成は、洗練された分子系統解析の手法を組み合わせ、脊椎動物 15 種の全タンパク質遺伝子配列を解析することで達成した。種間でゲノム構造を比較したところ、条鰭魚類の中心をなす真骨魚類のゲノムは、進化の初期に生じた全ゲノム重複の直後に共通した構造をすでに形成していたと推察された。

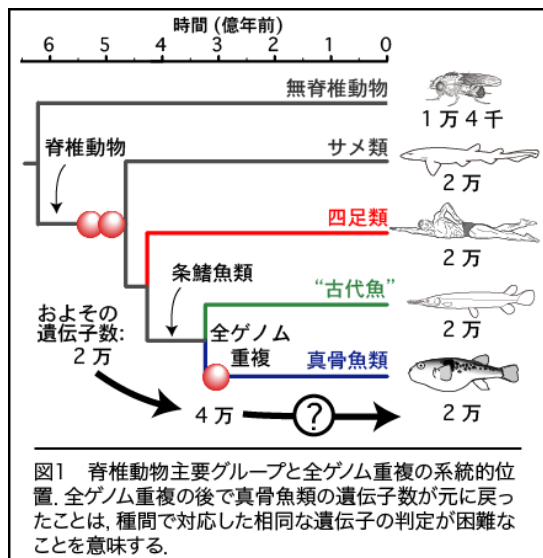
研究成果の概要(英文)：This study constructed an analysis pipeline to identify marker genes from genomic data for phylogenetic studies of ray-finned fishes. The pipeline was developed by implementing cutting-edge methods of molecular phylogenetics and by analyzing all protein-coding gene sequences of 15 vertebrates. Genome structural comparisons among vertebrates suggest that common structure of teleost genomes were established just after whole genome duplication occurred in the ancestor of all extant teleosts.

研究分野：生物学

キーワード：全ゲノム重複 真骨魚類 分子系統

## 1. 研究開始当初の背景

核ゲノムデータを解析すれば、条鰭魚類(図1)の解明困難な分岐順序が明確になると期待されている。しかし、哺乳類や植物など研究が盛んなグループと同様、条鰭魚類の研究でも、数十遺伝子からなる中規模データから、核ゲノムという大規模データの解析への転換に技術が追いついていない。



### (1) 中規模データ解析に基づく魚類の大系統分析の展開と課題

条鰭魚類に含まれる主要な系統の分岐順序は、中規模データの分析で解明されてきた。研究代表者が所属する研究グループは、核ゲノムとは別にミトコンドリア内部に保存されたゲノムの全遺伝子配列データに注目することで、低迷していた魚類の系統研究を飛躍的に前進させた (e.g. Miya et al. 2003, 2013; Inoue et al., 2003, 2010; Lavoue et al., 2010; Saitoh et al. 2006, 2010)。系統解析に際して、核ゲノムという大規模データに比べてミトコンドリアゲノムは、1) 37 遺伝子と決まった数の遺伝子を持ちコピー遺伝子が存在しないため、相同性の疑いがない。2) 細胞あたり 1000 あまりのコピーがあるため、配列解読が容易である、と言った優れた特徴を持つ。しかし、あまりにも急激に問題を解決してきた我々の研究グループは、概して、「核ゲノムデータより進化速度が速いミトコンドリアゲノムデータは高次系統解析に不適切」という理由で強い批判を受け (e.g., Zheng et al. 2011), 論文の出版すら難しい状態に陥った。

### (2) 大規模データ解析手法の開発

次世代シーケンサーの登場による急速な核ゲノムデータの蓄積は、盲目的に大量データを解析する風潮を生み出した。膨大な核ゲノムデータには多数のコピー遺伝子が含まれるため、種間でゲノム比較を行うには遺伝子の対応関係を慎重に吟味する必要がある。しかし、ほとんどのゲノム研究はデータ量で結

果の価値を主張するのみでデータの選択を軽視している。特に本研究が対象とする真骨魚類は祖先で独自に全ゲノム重複を経験しており(図1)、その独特な遺伝子構成と進化パターンは安易な解析を寄せ付けない。例えば真骨魚類は、真骨魚類全ゲノム重複を経験していない四足類とほぼ同じ遺伝子数を持つ。このことは、四足類と真骨魚類の間で遺伝構成が異なり、遺伝子の対応関係が複雑なことを意味する。しかし、もし遺伝子の対応関係を明確にできれば、核ゲノムという大量データを分析困難であった系統関係の解析に適用できる。

こうした研究活動の岐路に立ち、研究代表者は、実際の系統解析に適用する前に、核ゲノムデータから系統解析に有用な遺伝子を抽出するのを感じた。本研究課題を終え論文を作成する段階に至り、核ゲノムデータを用いた系統推定、という最終目標は、脊椎動物ゲノムの進化過程を推定すると言う、より根源的な問題と双壁をなすことがわかってきた。すべての脊椎動物はその進化の初期で 2 回の全ゲノム重複を経験している(図1)、より最近のイベントである真骨魚類ゲノムに焦点を当てた本研究は、脊椎動物のゲノム進化を解明する第一歩である。

## 2. 研究の目的

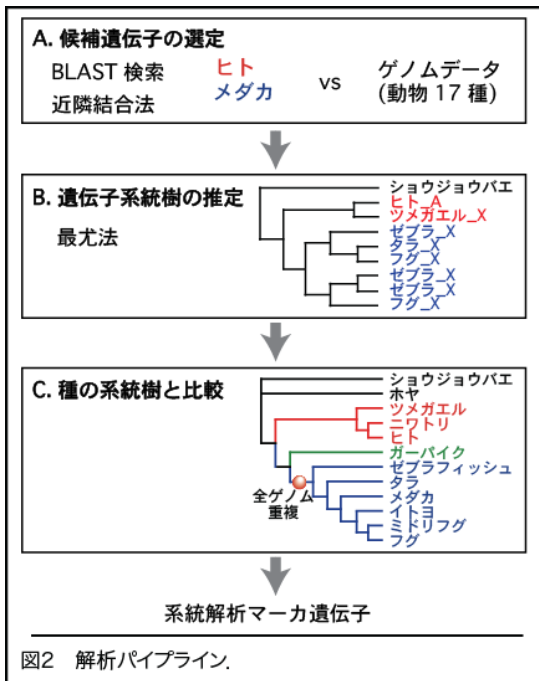
大きな目標を目前にして、目的を二つに分けた: 1) 核ゲノムデータから系統解析に有用な遺伝子を選定する解析パイプラインを作成する。2) 作成したパイプラインで選定した遺伝子配列を用いて、下位条鰭魚類の時間軸付き系統樹を推定する。

## 3. 研究の方法

本研究課題は上記 1) の、解析パイプラインの作成、にほとんどの時間を費やした。その手法は、ゲノムに含まれる全遺伝子の系統樹を推定する、という数年前では考えられない規模のものであった。

研究代表者が専門とする分子系統解析の分野では、プログラミングを専門とする研究者が様々な解析プログラムをウェブ上で公開している。その手法は多様かつ難解で、目的に沿ったプログラムの選定と解析には、専門的な知識が必要である。研究代表者は本研究課題を開始する以前、5 年間にわたり海外の研究室で最先端の分析方法を学んだため、解析にはその知識を生かすことができた。

脊椎動物は約 2 万の遺伝子を持つため、すべての解析ステップを自動化する必要がある。本研究ではプログラミングによって解析の自動化を行った。解析には、沖縄科学技術大学院大学およびヒトゲノム解析センターが提供するスーパーコンピュータを使用した。



#### 4. 研究成果

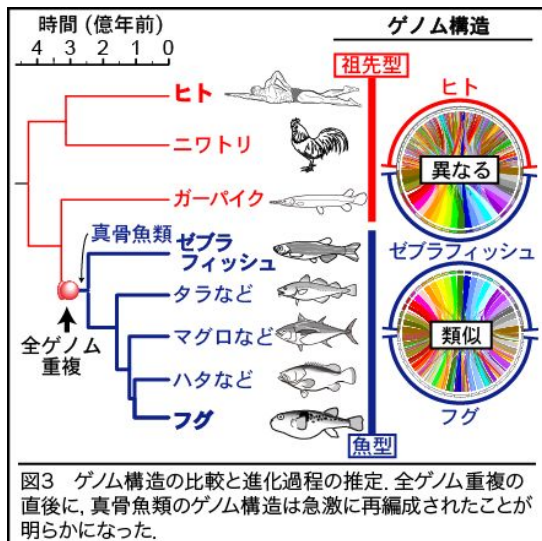
研究成果は Inoue et al. (2015) にまとめた.

##### (1) 解析パイプラインの構築

分子系統解析のプログラムを組み合わせることにより、A) 候補遺伝子の選定、B) 遺伝子系統樹の推定、C) 種の系統樹と比較、という 3 ステップからなる解析パイプラインを作成した (図 2). およそ 2 万からなる遺伝子の系統樹は様々なパターンが含まれるため試行錯誤を繰り返し、パイプラインの完成には 3 年以上の年月が必要であった.

##### (2) 系統解析マーカの選定

作成した解析パイプラインを用いて、約 1000 個の系統解析マーカ遺伝子を選定した. 解析パイプラインを完成させれば即座に得られる結果ではあるが、分子系統解析が核遺伝子データを必要としてから数十年にわたる問題に、はじめて科学的な根拠に基づく答えを出したことになる. 遺伝子重複ではなく種分化で別れた遺伝子を選ぶため、以下 2



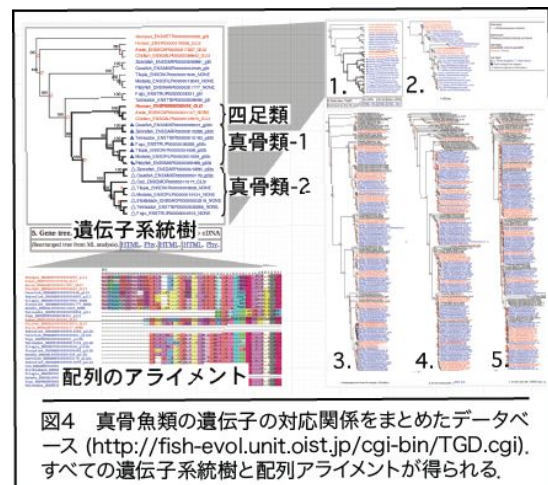
点をマーカ遺伝子選択の基準とした: 1) 真骨魚類全ゲノム重複の直後にペアとなった片方の遺伝子を失った. 2) 解析に用いたすべての種に存在するが、種内部でコピー遺伝子を持たない. 当初は得られたマーカ遺伝子を野生個体からも収集し、条鰭魚類の時間軸付き系統樹を推定する計画であった. しかし、配列解読の技術的な進歩が予想以上に遅かったため、この目的は次の研究展開として保留した.

##### (3) ゲノム構造の比較と進化過程の推定

解析マーカ遺伝子を選定する作業は、当初考えていた本研究課題の目的と同等かそれ以上の重要なテーマを浮かび上がらせた. それは 100 時間を超える研究協力者とのミーティングの末に得られた. 本課題が目的とした、系統解析に必要な解析マーカ遺伝子の選定には、すべての遺伝子の由来を明らかにする必要があった. その結果、遺伝子の集合体であるゲノムの構造の比較と進化過程の詳細な推定が可能になった (図 3). さらに、研究グループは、ミトコンドリアゲノムデータ解析によって魚類の時間軸付き系統樹を推定している (e.g., Azuma et al. 2008). 近代的なゲノム比較解析の準備が整っていた.

解析パイプラインから得られた全遺伝子の対応関係をもとにゲノム構造を比較したところ (図 3), 真骨魚類の内部ではゲノム構造が極めて類似していた. しかし、真骨魚類のゲノム構造はヒトやニワトリなど四足類のものとは大きく異なっていた. このことは、祖先型のゲノムが重複した直後に、急激に真骨魚類で共有される魚型のゲノム構造が形成されたことを意味する. さらに、得られた結果を真骨魚類の種の歴史と照らし合わせたところ、ゲノム進化と種の繁栄の関連が示されていた. 脊椎動物最大のグループである真骨魚類の多様化には、染色体同士が関わる大規模なゲノム構造変化はほとんど寄与せず、染色体内部で生じた遺伝子配置の変化や配列の変異が主に関わるのが推察される.

数理的な解析を行うことにより、全ゲノム



重複後に真骨魚類の遺伝子が 2 万まで減少した過程を推定した(図 1 の“?”). 数理モデルを用いた解析は, 遺伝子の七割以上が欠失したこの急激な減少が, 遺伝子が染色体規模のブロックとして欠失することによって引き起こされたことを示唆した. これらの分析には魚類の時間軸付き系統樹(図 3)や高度な数理解析などが必須である. 本研究課題の以前に研究グループが展開してきた研究と分野を超えた協力体制があつてはじめて, ゲノム進化研究の新展開を示す結果が得られた.

#### (4) データの公開

本研究では, 解析で得られたすべての結果をウェブサイトから公開している(図 4). これまでゲノム解析を対象とした論文の多くは, 結果を表として発表するのが主であり, 実際の解析結果まで公開していなかった. ウェブサイトでは, 得られた結果として, 真骨魚類を中心とした脊椎動物の遺伝子系統樹と配列アライメントのすべてを公開している.

本研究課題は, 研究の総括として作成した論文をトップジャーナルに掲載することで終了した. しかし, 得られた研究成果を確固たるものにし, 競争の激しいゲノム進化研究の分野で主導権を握るには, 密度の高い研究を次々と展開する必要がある. 実際, 論文が受理されるには, 他の研究グループとの思いもよらぬ駆け引きが必要であった. 国際 DNA バンクで公開されていたガーパイクのゲノムデータを用いた解析結果は, ゲノムを解読したグループが主要な論文を公開していないという理由で, 編集部の要請に従い論文から除外しなくてはならなかった. 当初は誰も予想していなかったが, 真骨魚類全ゲノム重複直前に分岐したガーパイクのゲノム構造は, ニワトリのものと酷似していた(図 3). このことは, 両系統が四足類+条鰭魚類の祖先種が持っていたゲノム構造をいまだに保持していることを意味する. 論文投稿の後に生じたこれらのやり取りは, 混乱したゲノム比較研究に対し, 進むべき正しい道を本研究課題が照らし出した証拠である.

#### 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)

Inoue, J., Sato, Y., Sinclair, R., Tsukamoto, K., Nishida, M. 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proceedings of the National Academy of Sciences of the United States of America* 112(48):14918-14923. doi: 10.1073/pnas.1507669112, 査読有り.

[学会発表](計 5 件)

- 2015 年 11 月 6~7 日. Inoue J., Sato, Y., Sinclair, R., Nishida, M. Identification of phylogenetic marker gene considering whole genome duplication and species tree. *Introduction to the Analysis of Molecular Evolution in the Genomic Era. Shanghai (Invited)*.
- 2015 年 8 月 23 日. 井上 潤, 佐藤 行人, シンクレア ロバート, 塚本 勝巳, 西田 睦. 全ゲノム重複と種の系統樹を考慮した系統解析マーカー遺伝子の選定. 日本進化学会年会「夏の学校」. 中央大学, 東京 (招待講演).
- 2014 年 3 月 10~12 日. Inoue J., Sato, Y., Sinclair, R., Nishida, M. Selecting molecular markers from transcriptome data for fish phylogenetics. *Workshop on Methods for Biodiversity Research, Fudan University, Shanghai (Invited)*.
- 2013 年 8 月 28~31 日. 井上 潤, 佐藤 行人, 西田 睦. 脊椎動物 3 億年のゲノム進化: ゲノム重複に由来する遺伝子の保持と進化に寄与した自然選択項の検出. 日本進化学会年会. 筑波大学, つくば.
- 2012 年 8 月. 井上 潤. 真骨魚類ゲノムの進化特性: トランスクリプトーム相同遺伝子データベースの構築. 日本進化学会年会. 首都大学東京, 東京.

[その他]

ホームページ等

FishOrthoDB

<http://fish-evol.unit.oist.jp/cgi-bin/TGD.cgi>

#### 6. 研究組織

(1) 研究代表者

井上 潤 (INOUE, Jun)

沖縄科学技術大学院大学・Mathematical Biology・研究員

研究者番号: 10596779

(2) 研究協力者

西田 睦 (NISHIDA, Mutsumi)

琉球大学・理事

研究者番号: 90136896

塚本 勝巳 (TSUKAMOTO, Katsumi)

日本大学・海洋生物資源科学科・教授

研究者番号: 10090474

シンクレア ロバート (SINCLAIR, Robert)

沖縄科学技術大学院大学・Mathematical Biology・准教授

研究者番号: 50423744

佐藤 行人 (SATO, Yukuto)  
東北大学・東北メディカルメガバンク機構・  
助教  
研究者番号：20566418