

令和 6 年 6 月 19 日現在

機関番号：82609

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K06132

研究課題名（和文）祖先転写制御の実験的復元に基づく遺伝子の創成・消失過程の解明

研究課題名（英文）Elucidation of de novo gene birth and pseudogenization processes via reconstruction of ancestral transcriptional regulation

研究代表者

原 雄一郎（HARA, Yuichiro）

公益財団法人東京都医学総合研究所・ゲノム医学研究センター・主席研究員

研究者番号：70709708

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：遺伝子が生まれる、あるいは消失するときに、「タンパク質コード領域」「転写調節領域」のどちらが主となり進化を駆動してきたかを解明すべく、類人猿のヒトに至る系統で生じた「遺伝子の新規創成」「偽遺伝子化」の過程を祖先配列の推定により復元した。計算機上での配列解析では推定が困難であった祖先転写制御領域の活性を、祖先塩基配列を用いた超並列レポーターアッセイにより実験的に計測した。本研究は、これまで主にタンパク質コード配列のみで考えられてきた遺伝子の生成/欠失において、転写制御の共役的進化という新たな観点を与える。

研究成果の学術的意義や社会的意義

これまで困難であった祖先遺伝子の機能を実験的かつ網羅的に復元するアプローチは、進化学ひいてはゲノム科学の根源的な問いでもある転写制御と遺伝子の共役的な進化過程を解明する基盤として、先駆性を持って確立されると期待できる。進化生物学においては、重複遺伝子の多様化やがんゲノム進化など転写制御活性の変化とも共役し得る遺伝子進化を紐解くにあたり、進化時間に沿って転写制御活性を計測するという方法論を確立した。

研究成果の概要（英文）：We elucidated whether protein coding regions or transcription regulatory sites have mainly driven gene evolution in the process of gene birth and loss. To achieve this, we reconstructed the ancestral gene regulation activities of the de novo genes and pseudogenes that occurred in the primate lineage leading to humans. We measured the transcription regulatory activities of the ancestral promoters and enhancers of these genes, which had been difficult to infer in silico, by employing the massively parallel reporter assay with synthesized nucleotide sequences. This study provides a new perspective on the collaborative evolution of transcriptional regulation with the open reading frames in gene birth and death that have been considered primarily only in terms of protein coding regions.

研究分野：進化生物学

キーワード：de novo遺伝子 偽遺伝子 超並列レポーターアッセイ プロモーター エンハンサー 祖先配列

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

生物の進化では、ゲノムにコードされた遺伝子のレパートリーが種ごとに変化し、表現型の多様性が作られてきた。遺伝子は、遺伝子重複、*de novo gene* (遺伝子の新規創成)などによって新たにゲノムに加わり、ゲノム領域の欠失、偽遺伝子化などによって失われる。現在では多くの生物種のゲノムが解読され、遺伝子進化の高解像度な解析が可能となりつつある。研究代表者は、ヤモリ、サメ、クマムシなどのゲノム解読を通して、遺伝子レパートリー進化を引き起こすメカニズムおよび表現型進化との関連を明らかにしてきた(Hara et al., Nat. Evol. Ecol. 2018; Hara et al., BMC Biol, 2018; 他)。それでも、遺伝子レパートリー進化の理解は未だ断片的で、転写制御配列の進化を含めた統合的なモデルの研究は進んでいない。

遺伝子レパートリー進化のうち、*de novo gene* はゲノム配列という混沌の中に秩序を生み出し、偽遺伝子化は秩序を再び混沌に戻す。遺伝子における秩序とは、アミノ酸配列をコードする読み取り枠(Open Reading Frame; ORF)であり、調節された転写機構である。では、ORF、転写調節配列のうちどちらが *de novo gene*/偽遺伝子への進化をいざなうであろうか。遺伝子進化の根源を成すこの問いはこれまでも“ORF-first vs. transcription-first 仮説”として提起されているが、現在までに決定的な答えを得られていない。配列として定義される ORF は分子系統解析により祖先状態を推定できるが、祖先遺伝子もつ転写活性を *in silico* で推定することは技術的に困難である。研究代表者が研究開発当初に所属していた東京都医学総合研究所 ゲノム医学研究センターでは、遺伝子転写制御の複雑性を解明する手段として合成 DNA を用いた大規模なレポーターアッセイを導入しており、この方法を活用して転写調節領域の祖先配列における転写活性の計測を実現しこれまでの技術的困難を解決することを考案した。

### 2. 研究の目的

本研究は、遺伝子の新規創成/消失が、「タンパク質コード領域」「転写調節領域」いずれの創成/消失によって駆動されるかを解き明かすことを目的とする。そのために、(1)類人猿のヒトに至る系統で生じた *de novo gene* および偽遺伝子において、遺伝子調節領域を含めた進化過程の復元、(2)ヒト集団で多型的に存在する偽遺伝子に起きている遺伝子調節の変化、を明らかにする。この目的を達成するために、超並列レポーターアッセイ(Massively Parallel Reporter Assay; MPRA)を用いて祖先配列における転写調節機能を網羅的に復元する。超並列レポーターアッセイは、1万種類を超えるプロモーター/エンハンサー配列をベクターのレポーター遺伝子の近傍に導入し、培養細胞でのレポーター遺伝子の発現に基づき各配列の転写活性を同時に計測する技術である。このアッセイでは実存する細胞のゲノム DNA だけではなく *in vitro* で合成した DNA もターゲットにできるため、現代には存在し得ない推定祖先配列の検証も可能である。本課題は、進化学で長らく未解決の問題に最新の計測技術で取り組み、遺伝子の創成/消失における「タンパク質コード領域」「転写調節領域」の創成/消失の役割を統合的に理解する研究として位置付けられる。

### 3. 研究の方法

本研究では、ヒトに至る類人猿の系統およびヒト集団で生じた *de novo gene*、偽遺伝子の祖先配列を推定し、祖先の転写調節領域の活性を計測することにより、遺伝子新規創成、偽遺伝子化の進化過程の順序を推定した。以下の3つを行程として研究を進めた。研究代表者の原は、研究を統括し、分子進化学的解析、超並列レポーターアッセイの配列設計およびデータ解析を行った。研究分担者の吉沢博士は超並列レポーターアッセイの実験プロトコルを最適化し、研究協力者

の実験補助員とともに実験を遂行した。研究協力者の川路博士より、アカゲザル、マーモセットの in-house 発現解析データを得た。加えて 2022, 2023 年度に先進ゲノム支援より HiChIP および Hi-C のサポートを受けた。

(1) ヒトに至る類人猿の系統で生じた *de novo* gene および偽遺伝子を同定する

ヒト、類人猿、サルゲノム・トランスクリプトームデータより、ヒトに至る類人猿の系統で ORF を獲得した *de novo* gene を同定した。これらのうち、FATOM プロジェクトの CAGE データおよび公共データベースの Ribo-seq データを解析し、ヒトゲノムにおいてプロモーター活性と翻訳のエビデンスをもつ *de novo* gene を抽出した。

同様に、サルでは完全な ORF をもち機能しているがヒトに至る類人猿の系統で ORF が壊れた遺伝子をヒト系統特異的偽遺伝子として同定した。研究協力員の川路博士が整備した D3G データベースのアカゲザル CAGE データより、サル遺伝子におけるプロモーター領域を調べた。また、ヒトの集団で common かつ loss of function の確度が高いナンセンス変異をもつ遺伝子を「隠れ偽遺伝子」と定義してヒト多型レポジトリ GTE<sub>x</sub> より取得し、CAGE データよりプロモーター領域を同定した。

(2) *de novo* gene・偽遺伝子の転写調節領域の祖先配列を推定し、転写調節に関する活性を計測する

(1)で推定したプロモーター領域において、ヒト、類人猿、霊長類の相同配列を用いた分子系統解析から、類人猿系統の祖先ノードで存在し得た塩基配列を推定した。また、隠れ偽遺伝子においてナンセンス変異と連鎖する転写調節領域のハプロタイプを推定する。

次に、ターゲットとする遺伝子のエンハンサー領域を探索した。近位エンハンサーはプロモーターの近傍 2kb 領域、遠位エンハンサーはプロモーター領域とコンタクトするゲノム領域を対象とした。このコンタクトを同定するために、先進ゲノム支援において、H3K4me<sub>3</sub> をターゲットとする HiChIP を行った。HiChIP はヒト 4 細胞種(HepG2, K562, SK-N-SH, WTC11) およびカニクイザル 1 細胞種に対して行なった。対象とする近位、遠位領域において、双方向性の CAGE ピークをもつ領域をエンハンサーとして同定した。

エンハンサー、プロモーターそれぞれにおいて、ヒトおよび類人猿の相同配列、祖先配列をまとめてオリゴプールとして合成した。エンハンサー、プロモーターそれぞれのオリゴプールを用いて、ヒト 4 細胞種(HepG2, K562, SK-N-SH, WTC11)に対し超並列レポーターアッセイを行い、各領域の転写活性を次世代シーケンサーにより同時に計測した。

(3) 遺伝子新規創成、偽遺伝子化の進化過程の順序を推定する

祖先転写領域配列における活性および ORF 領域の変化を時系列に並べ、遺伝子新規創成、偽遺伝子化が、タンパク質コード領域、転写調節領域どちらの創成/消失によって駆動されたか推定した。遺伝子によりいずれでも駆動されうる場合には、タンパク質コード領域/転写調節領域による駆動それぞれが成り立つときにおけるゲノム配列的および機能的な特徴を抽出し、それぞれにおける遺伝子の新規創成/消失の進化パターンを解明する。加えて「隠れ偽遺伝子」で計測したプロモーター活性により、多型的な偽遺伝子に生じた転写調節の変化を、隠れ偽遺伝子をもつ個体の表現型を遺伝性疾患データベース等を用いて調べた。

#### 4. 研究成果

2021 年度には、ヒトゲノムアセンブリならびに RNA-seq データに基づくヒト転写産物カタログから ORF を網羅的に抽出し、既存のタンパク質コード遺伝子にはオーバーラップしない新規

ORF を同定した。これらから、Ribo-seq データを用いて翻訳を確認できた ORF を *de novo* gene として選抜し、プロモーター配列の候補となる転写開始点を CAGE データを用いて探索した。また、類人猿と旧世界ザルのタンパク質コード遺伝子オーソログデータベースを構築し、ヒトに至る系統で CDS アノテーションを消失した遺伝子を偽遺伝子として同定した。これらの偽遺伝子のカニクイザルオーソログがもつプロモーター配列を CAGE データに基づき同定した。さらに、ヒト遺伝子発現データベース GTEx から、ホモでもつと発現が null になると推定される変異アリルを探索し、変異アリルとオーバーラップする CAGE のピークを対象遺伝子の「偽遺伝子化プロモーター」として同定した。同定したプロモーター配列と類人猿・旧世界ザルのオーソログ配列をもとに、分子系統学的手法を用いてプロモーターの祖先配列を推定した。*de novo* gene、偽遺伝子がもつプロモーター活性の進化的変遷を同定するために、変異アリルを含む現生生物のプロモーター配列と推定された祖先配列を用いて超並列レポーターアッセイを行うためのオリゴヌクレオチドを設計した。

2022 年度には、超並列レポーターアッセイを用いてプロモーター祖先配列の活性を計測した。対象とする遺伝子から 400 箇所のプロモーターを選び、その祖先配列をもとにして超並列レポーターアッセイのプロトコルに適合するように約 18,000 配列のオリゴヌクレオチドプールを設計、合成した。プールされたオリゴヌクレオチドにランダムバーコードを付加し、プロモーターを対象とする超並列レポーターアッセイ用に設計したバックボーンベクターに挿入してクローニングするとともに、ベクターをシーケンシングしてバーコードとオリゴヌクレオチドの対応関係を計測した。作製したプラスミドベクターをヒトの 4 種類の培養細胞に導入し、DNA, RNA 分子のバーコードをシーケンシングしてプロモーター祖先配列の活性を計測した。加えて、先進ゲノム支援のサポートを受けて、アクティブなプロモーターのヒストンマークである H3K4me3 をターゲットとする HiChIP 解析を行った。HiChIP には上記のヒト培養細胞の他にカニクイザルの細胞も用いている。先進ゲノム支援に HiChIP のライブラリ調整とシーケンシングを依頼し、納品された配列データを解析した。その結果、対象とするプロモーターの約半数でコンタクトするゲノム領域が見つかった。これらの領域において双方向性をもつ CAGE ピークを同定しエンハンサーの候補とした。

2023 年度は 2022 年度に同定したエンハンサーの祖先配列を推定し、超並列レポーターアッセイによりその活性を測定した。2022 年度に同定した約 600 箇所の近位・遠位エンハンサーにおいて祖先配列を推定し、それら配列をオリゴヌクレオチドプールとして合成した。オリゴヌクレオチドをランダムバーコードとともにエンハンサーを対象とする超並列レポーターアッセイ用のバックボーンベクターに挿入してヒトの 4 種類の培養細胞に導入し、超並列レポーターアッセイを行い、エンハンサー領域・祖先配列の転写制御活性を計測した。並行して、前年度に行ったプロモーター領域の超並列レポーターアッセイの解析を行った。研究の対象としたプロモーター領域では、新規に生じた遺伝子ではヒトに至る系統でプロモーター活性を獲得・増強し、失われた遺伝子ではヒトに至る系統でプロモーター活性を減弱・喪失していることが明らかになった。加えて、先進ゲノム支援のサポートを受けて、該当遺伝子を含むゲノム領域のゲノム構造を調べるべく HiC を行った。エンハンサー領域の超並列レポーターアッセイのデータを引き続き解析しており、ORF 領域の獲得・消失過程、プロモーター活性の獲得・消失過程、ならびに遺伝子が存在するゲノム領域の高次構造データと統合して遺伝子が生じる・失われる過程における転写制御の共役的進化について統合的な知見をまとめているところである。

## 5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Hara Yuichiro, Kuraku Shigehiro	4. 巻 12
2. 論文標題 The impact of local genomic properties on the evolutionary fate of genes	5. 発行年 2023年
3. 雑誌名 eLife	6. 最初と最後の頁 e82290-e82290
掲載論文のDOI（デジタルオブジェクト識別子） 10.7554/eLife.82290	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Hara Yuichiro, Shibahara Reira, Kondo Koyuki, Abe Wataru, Kunieda Takekazu	4. 巻 11
2. 論文標題 Parallel evolution of trehalose production machinery in anhydrobiotic animals via recurrent gene loss and horizontal transfer	5. 発行年 2021年
3. 雑誌名 Open Biology	6. 最初と最後の頁 200413 ~ 200413
掲載論文のDOI（デジタルオブジェクト識別子） 10.1098/rsob.200413	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Linard Benjamin, Ebersberger Ingo, McGlynn Shawn E, Glover Natasha, Mochizuki Tomohiro, Patricio Mateus, Lecompte Odile, Nevers Yannis, Thomas Paul D, Gabaldon Toni, Sonnhammer Erik, Dessimoz Christophe, Uchiyama Ikuo, QFO Consortium (incl. Hara Yuichiro)	4. 巻 38
2. 論文標題 Ten Years of Collaborative Progress in the Quest for Orthologs	5. 発行年 2021年
3. 雑誌名 Molecular Biology and Evolution	6. 最初と最後の頁 3033 ~ 3045
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/molbev/msab098	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計9件（うち招待講演 2件/うち国際学会 2件）

1. 発表者名 原 雄一郎
2. 発表標題 祖先転写制御の実験的復元に基づく遺伝子の創成・消失過程の解明
3. 学会等名 先進ゲノム支援 2023年度拡大班会議
4. 発表年 2023年

1. 発表者名 原 雄一郎、吉沢 直子、豊田 敦、川路 英哉
2. 発表標題 祖先転写制御の実験的復元から遺伝子が生まれる・消える過程を探る
3. 学会等名 第46回日本分子生物学会年会
4. 発表年 2023年

1. 発表者名 Yuichiro Hara, Shigehiro Kuraku
2. 発表標題 Gene fate spectrum as a reflection of local genomic properties
3. 学会等名 SMBE2023 (国際学会)
4. 発表年 2023年

1. 発表者名 原 雄一郎
2. 発表標題 祖先転写制御の実験的復元に基づく遺伝子の創成・消失過程の解明
3. 学会等名 先進ゲノム支援 2022年度拡大班会議
4. 発表年 2023年

1. 発表者名 原 雄一郎、齊藤紗希、和田涼子、川路英哉
2. 発表標題 FLAT-seq: Full-Length cDNA sequencing with Adaptive sampling of TSS
3. 学会等名 第45回日本分子生物学会年会
4. 発表年 2022年

1. 発表者名 原 雄一郎
2. 発表標題 不均一なゲノムの場合での突然変異と遺伝子進化のモード
3. 学会等名 日本進化学会年大会 第24回沼津大会 (招待講演)
4. 発表年 2022年

1. 発表者名 原 雄一郎
2. 発表標題 遺伝子の運命を左右するゲノムの場合
3. 学会等名 国立遺伝学研究所 研究会 「生命科学を支える分子系統学」 (招待講演)
4. 発表年 2022年

1. 発表者名 Hara Yuichiro, Shibahara Reira, Kondo Koyuki, Abe Wataru, Kunieda Takekazu
2. 発表標題 Parallel evolution of trehalose biosynthesis machinery in metazoans.
3. 学会等名 The 2nd AsiaEvo Conference (国際学会)
4. 発表年 2021年

1. 発表者名 原 雄一郎、工樂樹洋
2. 発表標題 種間比較から明らかにされる遺伝子欠失を受容するゲノム領域
3. 学会等名 第44回日本分子生物学会年会
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 原 雄一郎; (編) 井ノ上 逸朗, 今西 規, 河村 正二, 斎藤 成也, 颯田 葉子, 田嶋 敦	4. 発行年 2021年
2. 出版社 一色出版	5. 総ページ数 448
3. 書名 ヒトゲノム事典 12章 「12.3 遺伝子変換」, 「12.6 逆位」	

〔産業財産権〕

〔その他〕

研究室(東京都医学総合研究所 ゲノム医学研究センター) webページ <a href="https://www.igakuken.or.jp/genome-center">https://www.igakuken.or.jp/genome-center</a>
---

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	吉沢 直子 (須賀田直子)  (YOSHIZAWA-SUGATA Naoko)  (30344071)	公益財団法人東京都医学総合研究所・ゲノム医学研究センター・主席研究員   (82609)	

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	川路 英哉  (KAWAJI Hideya)  (20525406)	公益財団法人東京都医学総合研究所・ゲノム医学研究センター・副参事研究員   (82609)	
研究協力者	和田 涼子  (WADA Ryoko)	公益財団法人東京都医学総合研究所・ゲノム医学研究センター・技術補佐員   (82609)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------