

称号及び氏名	博士(応用生命科学)	北山 瑠璃
学位授与の日付	2024年5月31日	
論文名	Analysis of the atypical gene evolution of various organisms (様々な生物種における型破りな遺伝子進化の解析)	
論文審査委員	主査	青木 考
	副査	東條 元昭
	副査	佐々木 伸大
	副査	尾形 善之

## 論文要旨

生命進化を理解するにあたって、遺伝子進化が種分化とどのように相互作用し、進化の歴史においてどのような役割を果たしているのかを理解することが試みられている。遺伝子配列の解析技術や培養に頼らない解析技術が発展したことにより、利用可能な遺伝子情報は拡大した。また、情報科学的に遺伝子情報を解析し、二次的なデータベースも構築されている。

第1章の序論で解析技術の進化により判明した型破りな遺伝子進化と本研究の背景について説明する。遺伝子進化は分子系統樹の形で描写され、種分化の系統樹と比較されてきた。種分化の際に親の遺伝子と共通な遺伝子はオルソログと呼ばれている。一方で、遺伝子のコピーの際に重複した遺伝子が生じ、時に機能を変えて進化する。このような遺伝子配列が類似の遺伝子群はパラログと呼ばれる。これらのオルソログやパラログとは異なり、バクテリア等の原核生物では他の細胞の遺伝子を細胞内に取り込んで自身の遺伝子に組み込む形質転換や、ウイルスによって細胞内に遺伝子を運び込まれる形質導入によって他の種の遺伝子を取り込むことがある。このような世代交代を経ずに遺伝子が授受される現象は水平伝播と呼ばれている。ウイルスでは、他のウイルスと相同組換えにより大部分に変異が導入される現象がある。型破りな生物進化が行われる場合、分子系統樹を用いて遺伝子進化を正確に描写することは困難である

と考えられる。当研究室ではこれまで植物と真菌において遺伝子の相同性解析を行い、折れ線グラフやネットワークなどを用いた独自の 방법으로遺伝子の進化を描写したデータベースを構築してきた。そこで、本研究では当研究室で得られてきた方法や結果を踏まえて、様々な生物種（ウイルスを含む）において型破りな遺伝子進化の発見を試みた。SARS-CoV-2 ウイルスでは変異株の成立に不可欠な変異を辿り、逐次的にならない変異の導入に注目した。バクテリアでは相同性を有する遺伝子をもつ科間の相関ネットワークを描写し、系統分類と合致しないモジュールについて着目した。真核生物ではタンパク質クラスターに所属する系統を解析し、系統をこえて共有される遺伝子や系統を超えて遺伝子を共有する生物群に着目した。

第 2 章及び第 3 章では SARS-CoV-2 ウイルスの遺伝子進化と株の成立に関する研究について述べる。SARS-CoV-2 ウイルスは、2020 年初頭に登場して以来、遺伝子変異を繰り返しながらパンデミックを引き起こしてきた。COVID-19 に関して多数の研究報告やゲノム情報の開示が行われてきました。アメリカ国立生物情報工学情報センター（NCBI）の GenBank ウイルスデータベースや Nextstrain チームが運営する Nextstrain などの公開データベースには、100 万を超えるウイルスのゲノムと変異の解析結果が提供されており、毎日何万もの新しいゲノムがアップロードされている。当該データを用いてウイルスの進化と伝播の様子を地球規模および地域規模で綿密に追跡できるようになった。GenBank、Nextstrain、SARS-CoV-2 MAT などの公開データベースでは、ウイルス変異の分析とウイルス標本のメタデータに基づいて、系統全体または局所系統の規模でウイルスの系統樹を提供している。

ネットワークや樹形図は、複雑なシステムを視覚化するために使用されている。相関ネットワーク分析にも適用され、ノードが互いに密接に接続されているローカルコミュニティまたはネットワークモジュールを検出した。相関ネットワーク解析は、遺伝子共発現や遺伝子相同性ネットワークなどの生物学的ネットワークを説明するために使用されてきた。サンプリングされたウイルスゲノムが共有する変異間の関係を表すには、相関ネットワークアプローチが役立つ。そこで、変異を共有する検体の比率に着目し、相関ネットワーク解析によりウイルスの S タンパク質の変異型の違いを可視化した。ネットワークでは、Alpha 株と Delta 株で見つかった変異が、ノードが変異を表し、互いに密接に接続されたネットワークモジュールを形成した。ただし、Mu 株の変異は、複数のネットワークモジュールに分散されていた。得られた S タンパク質変異の相関ネットワークにより、Mu 株等は相同組換えなどに基づく複数の変異体由来する変異を持つと予想され、これらの変異体の進化的特徴について議論した。

第 4 章ではバクテリアの効率的な遺伝子解析と特徴的な生物群に関する研究について述べる。バクテリアは、温度や塩濃度などの極限環境や抗生物質に対する適応力を有している。耐性の獲得には遺伝子進化や水平伝播が寄与していると考えられる。そこで、バクテリアの遺伝子を解析し、系統によらない遺伝子進化の検出を試みた。バクテリアの冗長のない遺伝子は 2 億近く存在し、従来の総当たりの遺伝子解析では数百倍の時間が必要である。そこで効率的に解析を行うために、まずバクテリアのデータ

に対する配列相同性解析にかかる時間を推察した。その結果、バクテリアの科ごとの代表種を用いて解析することが十分可能であるとの見通しを立てた。バクテリアの同一ファミリー内の遺伝子の解析とファミリーの代表種の遺伝子を用いたファミリー間の遺伝子の解析を分けて行うことにした。ファミリー間の解析に用いる遺伝子は遺伝子配列のデータは欧州分子生物学研究所（EMBL）が運営するタンパク質配列に関するデータベースである UniProt の FTP サイトよりバクテリアの Reference proteome をダウンロードした。代表種の選別には、UniProt のタンパク質クラスターに関するデータベースである UniRef のクラスターに所属する遺伝子と代表配列の情報を用いてファミリー内の多くの種と相同性のある遺伝子をもつ種を選んだ。遺伝子の相同性解析には NCBI が提供するソフトウェアである BLAST+ を利用した。相同な遺伝子を用いて科間の相関性を評価し、ネットワーク解析を行った。その結果、 $\beta$ プロテオバクテリアと  $\gamma$ プロテオバクテリアが共有している遺伝子の構成において、型破りな遺伝子進化の特徴を捉えた。この特徴について議論する。

第 5 章では真核生物の型破りな遺伝子の効率的な検出方法の研究について述べる。真核生物は植物や動物など多様な系統が存在するが必ずしも系統によるものではない。例えば、バラ科やマメ科には木本植物と草本植物が含まれている。系統を跨ぐ形質と関与する遺伝子の検出には多様な系統間の解析が必要であり、総当たりによる相同性解析は困難である生物種間での解析を目指し、上記 UniProt から得られた UniRef50 のタンパク質クラスターに注目し、遺伝子を共有している生物群が同一の系統であるか調べた。真核生物についても、水平伝播などの型破りな遺伝子進化が知られているが、真核生物内の全生物規模、全ゲノム規模で検出した解析例は、おそらく初めてだと考えている。得られた結果として、真核生物全般にわたって、型破りな遺伝子と生物種を検出できた。例えば、鳥類では同科内での型破りな生物種の検出は少なかったが、魚類では多くの型破りな生物種を検出することができた。得られた遺伝子から生物種の特徴を議論する。また、得られた結果のデータベース公開を進めている。

## 審査結果の要旨

生命進化を理解するにあたって、遺伝子進化が種分化とどのように相互作用し、進化の歴史においてどのような役割を果たしているのかを理解することが試みられている。遺伝子配列の解析技術が発展したことにより、利用可能な遺伝子情報は拡大した。また、情報科学的に遺伝子情報を解析し、二次的なデータベースも構築されている。こうした二次的なデータベースにおいて、遺伝子進化は分子系統樹の形で描写され、種分化の系統樹と比較されてきた。種分化の際に祖先種の遺伝子と共通な遺伝子はオルソログと呼ばれている。一方で、遺伝子のコピーの際に重複した遺伝子が生じ、時に

機能を変えて進化する。このような遺伝子配列が類似の遺伝子群はパラログと呼ばれる。これらのオルソログやパラログとは異なり、バクテリア等の原核生物では他の細胞の遺伝子を細胞内に取り込んで自身の遺伝子に組み込む形質転換や、ウイルスによって細胞内に遺伝子を運び込まれる形質導入によって他の種の遺伝子を取り込むことがある。このような世代交代を経ずに遺伝子が授受される現象は水平伝播と呼ばれている。ウイルスでは、他のウイルスと相同組換えにより大部分に変異が導入される現象がある。型破りな生物進化が行われる場合、分子系統樹を用いて遺伝子進化を正確に描写することは困難であると考えられる。当研究室ではこれまで植物と真菌において遺伝子の相同性解析を行い、折れ線グラフやネットワークなどを用いた独自の方法で遺伝子の進化を描写したデータベースを構築してきた。そこで、本研究では当研究室で得られてきた方法や結果を踏まえて、ウイルスならびに様々な生物種において型破りな遺伝子進化の発見を試みた。

第 1 章で解析技術の発展により判明した型破りな遺伝子進化について説明する。SARS-COV-2 ウイルスでは変異株の成立に不可欠な変異を辿り、逐次的にならない変異の導入に注目した。バクテリアでは相同性を有する遺伝子をもつ科間の相関ネットワークを描写し、系統分類と合致しないモジュールについて着目した。真核生物ではタンパク質クラスターに所属する系統を解析し、系統をこえて共有される遺伝子をもつ生物群に着目した。

第 2 章では SARS-CoV-2 ウイルスの遺伝子進化と変異株の成立に着目した。本ウイルスは、2020 年初頭に登場して以来、遺伝子変異を繰り返しながらパンデミックを引き起こしてきた。COVID-19 に関して多数の研究報告やゲノム情報の開示が行われ、公共データベースには数百万を超えるウイルスのゲノムが提供されてきた。当該データを用いてウイルスの進化と伝播の様子を地球規模で綿密に追跡し、相関ネットワーク解析によりウイルスの S タンパク質の変異型の違いを可視化した。その中で、Mu 株の変異は、相同組換えなどに基づく複数の変異体に由来する変異を持つと予想された。

第 3 章ではバクテリアの効率的な遺伝子解析と特徴的な生物群に着目した。バクテリアは、温度や塩濃度などの極限環境や抗生物質に対する適応力を有している。耐性の獲得には遺伝子進化や水平伝播が寄与していると考えられる。そこで、バクテリアの遺伝子を解析し、系統によらない遺伝子進化の検出を試みた。バクテリア遺伝子は 2 億近く登録され植物等での総当たりの遺伝子解析の数百倍の時間が見込まれた。そこで効率的に解析を行うために、まずバクテリアのデータに対する配列相同性解析にかかる時間を推察し、科ごとの代表種を用いて解析することが十分可能であるとの見通しを立てて、バクテリア遺伝子の配列相同性解析を実施した。その結果、 $\beta$ プロテオバクテリアと  $\gamma$ プロテオバクテリアが共有している遺伝子の構成において、型破りな遺伝子進化の特徴を捉えた。

第 4 章では真核生物の型破りな遺伝子の効率的な検出に着目した。真核生物は植物や動物など多様な系統が存在するがこれらの多様性は必ずしも系統によるものではない。例えば、バラ科やマメ科には木本植物と草本植物が含まれている。系統を跨ぐ形質と関与する遺伝子の検出には多様な系統間の解析が必要であり、総当たりによる相同性解析は困難である生物種間での解析を目指した。UniProt から得られた UniRef50

のタンパク質クラスターに注目し、遺伝子を共有している生物群が同一の系統であるか調べた。その結果、真核生物全般にわたって、型破りな遺伝子と生物種を検出できた。得られた結果のデータベース公開を実現した。

本研究は、種分化と一致しない進化を示す遺伝子や生物種を、様々な分類群を対象にゲノム規模で検出することが可能であることを示した。理想的には全生物・全遺伝子間で総当たりでの配列相同性解析の実施が望まれるが、膨大な解析時間の短縮や解析作業の削減という困難を克服しなければならない。今回検討してきた効率的な遺伝子進化の追跡およびその方法論の開発はこの困難を克服し、ゲノム進化学分野の進歩に大いに貢献する成果である。したがって本論文の審査ならびに最終試験の結果と併せて、博士（応用生命科学）の学位を授与することを適当と認める。