膨大な量のゲノムデータを高精度に解析し、多様な微生物の全体像や絶滅動物の系統関係に迫る!

准教授　森（もり） 宙史（ひろし）　情報研究系　ゲノム多様性研究室

信州大学理学部生物科学科卒。奈良先端科学技術大学院大学博士前期課程修了。東京工業大学大学院生命理工学研究科博士後期課程修了。博士（理学）。2016年より 国立遺伝学研究所 情報研究系ゲノム進化研究室助教。2021年より、ゲノム多様性研究室准教授。

バイオインフォマティクスを駆使し、細菌群集のメタゲノム解析や絶滅動物の Ancient DNA 解析を進めてきた、森宙史准教授。並行して、情報解析初心者でも手軽に系統組成の推定ができる解析ツールや、メタゲノムの高速な配列相同性検索システムも開発。根底には、「生き物の生態を知りたい」という幼少期からの好奇心がある。

動物行動生態学から バイオインフォマティクス領域へ

子供の頃に昆虫や魚を捕まえて観察するのが好きだったこともあり、大学の学部では行動生態学を専攻しました。移動速度が遅いカタツムリは、大きな川などの地形によっても隔離がおきて別種になるなど、多様です。カタツムリの特徴として雌雄同体があげられますが、私は、同じ場所に生息する近縁種間の生殖行動の違いに着目して学部の卒業研究を行いました。当時（2000～2003年）ヒトゲノム解読計画の成果が華々しく報道されており、ゲノム研究にも興味を抱き、卒業研究の合間にはゲノム関連の書籍や論文も読んでいました。

カタツムリの生殖行動についての卒業研究は論文化[*1]も叶い、修士過程に進む段階で、研究をゲノム解析にシフトしようと考えました。当時は、生物のゲノムを解読するのは一大プロジェクトでした。一方で、ゲノム情報をコンピュータで扱うバイオインフォマティクスは、公共データベースのゲノム配列データを使えば個人でも研究できました。この点に魅力を感じ、奈良先端科学技術大学院大学に進学することにしました。遺伝研（総合研究大学院大学）も進学候補にしており、大学院見学会にも参加していたのですが、迷った末に奈良に決めました。修士課程では、現在、同じく遺伝研におられる黒川顕先生の元で、細菌ゲノムの比較解析や、当時データが出始めた細菌群集のゲノムを丸ごと解読するメタゲノム解析の研究を行いました。

先生が東京工業大学に移られたのにあわせて、私も博士課程から東工大に進学しました。修士時代には特定の細菌のゲノムを旧式のSanger型DNAシーケンサーで解読していたのですが、2006年ごろに電気泳動せずに高度な並列化を行える、いわゆる新型シーケンサーが登場し、急速に普及し始めました。新型シーケンサーのおかげで、細菌群集の短いゲノム断片を低コストで大量に読めるようになり、メタゲノム解析が一気に広まりました。新型シーケンサーによるメタゲノム解析では「非常に短い配

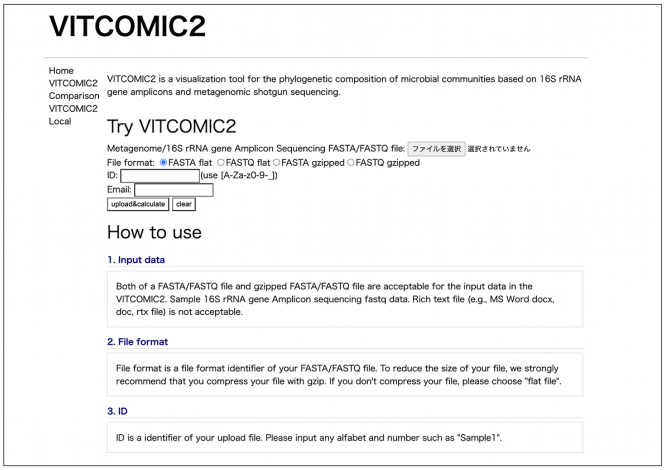
列」が大量に得られますが、この新たなデータから「細菌群集のメンバー組成（系統組成）や遺伝子機能組成」を推定するには、新たなバイオインフォマティクス解析手法やソフトウェアの開発が必須です。そこで私は、こうしたニーズを満たす解析手法やWebアプリケーションを開発して論文化し、東工大で博士号を取得しました。その後は、そのまま助教として大学に残り、研究者としてのキャリアをスタートしました。

そのころ（2010年ごろ）には、新型シーケンサーが普及してメタゲノム解析の有用性の理解も進み、世界中でメタゲノムデータが大量に産出されるようになりました。ただし、メタゲノム研究で得られたデータは「同じ群集を再現することが困難」という理由で国際的に整理されておらず、あちこちに散在している状態でした。そこで、メタゲノム解析データの情報解析手法研究と並行して、得られたメタゲノムデータを収集・整理し、環境ごとの微生物の解析や検索が容易にできる統合データベース作りをすることにしました。2016年に、黒川先生が遺伝研に移られることになり、私も助教として遺伝研に異動しました。遺伝研に来てからは、メタゲノム解析データの研究は継続しつつ、数万年前の生物の化石由来のDNA（Ancient DNA）から系統関係などを推定する「Ancient DNA研究」を本格的に始めました。2021年には准教授として自身の研究室を立ち上げ、今は、学生さんを指導しつつ自分自身でも複数の研究プロジェクトを同時に走らせる、といった忙しい日々を送っています。

手軽に系統組成を解析できるツール、VITCOMIC と VITCOMIC2 を開発

新型シーケンサーから得られる大量のメタゲノム配列データから高速かつ高精度に細菌群集の系統組成を推定するツール、「VITCOMIC(Visualization tool for taxonomic Compositions of microbial community)」[*2]、および、その改良版のVITCOMIC2を開発しました[*3]。VITCOMIC2は、Web上で公開しているWebアプリケーションになります。具体的には、系統名が既知のリファレンスDNA配列データに対して高速な配列類似性検索を行うことで、各配列の系統推定を行い、全ての配列における系統の組成の集計結果を視覚化することができます。

VITCOMIC2の「売り」は、「初心者でも使える手軽さ」です。細菌群集のメタゲノム配列データ（数十万から数千万本の断片配列）をアップロードすると、数分から十数分で自動的に解析を行い、系統組成が得られます。全ての操作がWebブラウザ上で完結しますので、バイオインフォマティクス解析が不得手なユーザーにも全く問題なく使っていただけます。計算には遺伝研のサーバーを使っており、1日平均十数件の解析をこなしている状況です。もちろん、私も自分の研究の解析に使っています。メタゲノム解析が普及した現在は、VITCOMIC2の他にも多数良い解析ツールが存在します。系統組成データはメタゲノム解析の基盤データですので、研究者は「系統組成のデータを用いてどのような解析をしたいか」によって、ツールを使い分けている現状といえますね。



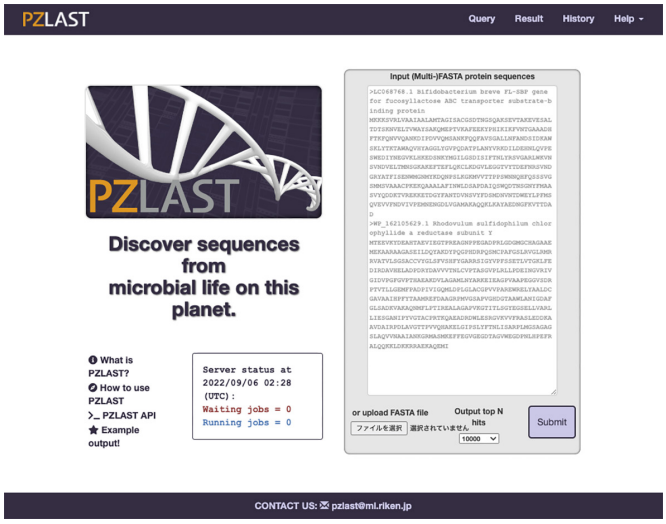
続いて、相同性検索を可能にする PZLAST やメタゲノムの統合データベースも

メタゲノム解析では、数百サンプルのデータを集めただけでテラバイトを超える膨大なデータ量になります。一方で、公共の塩基配列データベースでは、既に数十万サンプル以上のメタゲノム配列データが公開されています。テラバイトクラスのリファレンス配列データに対する相同性検索をしたいと思っても、BLASTなどの既存のよく使われているツールでは対応不能です。そこで私たちが開発したのが、「PZLAST」というWebアプリケーションです。既存のメタゲノムサンプル由来の約3.6テラバイトものアミノ酸配列データを対象に、配列相同性を検索するというツールです[*4]。3.6テラバイトの内訳は、予測遺伝子636億本、アミノ酸残基約2.4兆残基となっています。このような、大量のメタゲノム配列データに対する配列相同性検索ツールは、PZLASTの他にはほぼ存在しません。

日々、産出され、蓄積されるメタゲノムデータは、細菌が持つ多様な遺伝子の宝庫です。そこには産業応用可能な有用な遺伝子が多く秘められているはずですが、あまりにデータ量が大きいために、「埋もれた宝」を探せるツールがありませんでした。Web公開の PZLASTは、機能としては「メタゲノムへの配列相同性検索」と極めてシンプルですが、「特定の物質を分解する酵素と類

似した配列を持った酵素が、どのような環境に存在するのか」、「既存の酵素と同じ機能をもち、活性温度、活性pHなどが異なる酵素」といったように、様々な切り口での探索に活用できます。

このようにツール開発を行ってきましたが、今後は、「様々なツールの基盤となるデータベースの開発と整備」を重点的に進めたいと考えています。すでに今、「Microbiome Datahub」という「メタゲノムデータおよびメタゲノムから再構築した細菌ゲノムの統合データベース」を開発しています。背景には、この数年で、特定の環境中に生息する細菌の個別ゲノムを、不完全ながらもメタゲノムデータから再構築できるバイオインフォマティクス技術が発展し、多種多様な細菌ゲノム配列がメタゲノムデータ由来で得られるようになった状況があります。ただし、メタゲノム由来ゲノムデータは、論文ごとに作成方法やデータの公開の有無がバラバラで、全く整理されていません。開発中のMicrobiome Datahubではこの問題を解決し、将来的には「メタゲノムデータ」と「個別菌のゲノムデータ」を統合し、地球上に存在する細菌の多様性のほぼ全てを網羅したゲノム・メタゲノムの統合データベースとして発展させたいと考えています。



絶滅動物の遺骸由来 DNA を解析し、系統関係を推定する

さらに、絶滅動物の遺骸由来のDNAを解析するAncient DNA解析も進めています。Ancient DNA解析はメタゲノム同様、得られるDNAは多数の生物種由来の短い断片で、解析手法もメタゲノム解析と似ています。Ancient DNA解析の対象は、おおよそ「数十年よりも前に死んだ生物個体由来のDNA」ですが、Ancientか否かの明確な線引きはありません。化石由来のDNA解析は1980年代から行われていますが、新型シーケンサーの登場以降は、多数のサンプルを対象にしたゲノムレベルのAncient DNA解析が、世界各地で進むようになりました。Ancient DNA解析では、「絶滅生物の、現生の生物との系統関係、異なる年代における種の遺伝的多様性」などを調べることができます。絶滅した動物には、化石による形態情報のみで系統が定義されているものが多いため、Ancient DNAによる分子系統の情報は貴重といえます。私たちはこれまでに、マダガスカルのエピオルニス(*5)、本州のヒグマ(*6)とニホンオオカミ(*7)の、計3種の動物のAncient DNA解析を行いました。

エピオルニスは「象鳥」ともよばれる巨大な飛べない鳥で、人類がマダガスカルに来た約2000年前以降に、絶滅したと考えられています。最大のエピオルニス（エピオルニス・マキシマス）は体高3メートル以上もあり、鳥類としては最大級です。ダチョウやエミューなどの飛べない巨大な鳥や南米の飛べる鳥（シギダチョウ）は「古顎類」というグループに属しますが、古顎類の系統内でエピオルニスがどこに位置するかを、Ancient DNA解析で明らかにしたいと考えました。

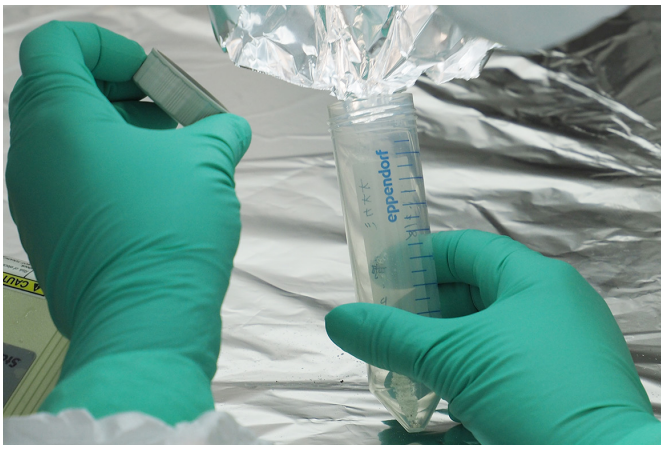
エピオルニスのDNAは、マダガスカルの大学の共同研究者が保存していた骨から抽出し、遺伝研等でシーケンスを行いました。得られたDNAの99%はカビや細菌由来でしたが、これらを排除した上でバイオインフォマティクスによる解析を行ったところ、核遺伝子の一部の配列と、ミトコンドリアの全長配列を得ることができました。これらを既知の古顎類の配列とくらべたところ、最も近い現生の種はニュージーランドに生息する小型の飛べない鳥キウイだとわかりました。太古にはマダガスカルとニュージーランドは大陸でつながっていたので、そこにいた共通祖先が2つの島に分かれ、それぞれに進化していったと考えられます。この知見は「Ancient DNA解析だからこそ得られた」と自負しています。

▮群馬県上野村の洞窟にてAncient DNA解析用のサンプリングを実施

より古い年代のAncient DNA解析が可能かを検証したいと考えて行ったのが、ヒグマの解析です。現生のヒグマは北海道のみに生息していますが、数万年前までは本州にも生息していました。ただし、本州から出土するヒグマの化石は現生の北海道の個体よりもサイズが大きく、本州のヒグマがどのような系統に位置し、いつ、どこからきたのかが不明のままでした。そこで、本州由来の3万2500年前のヒグマ化石を対象に、エピオルニスでの解析と同様の作業を進めました。その結果、本州のヒグマは現生では北海道や他のどの地域のヒグマとも異なる未知の系統であることがわかりました。また、現存の北海道南部のヒグマグループが姉妹系統に当たること、両者が約16万年前に分岐したこともわかりました。

これらの知見と、既知の化石記録を合わせると、ヒグマはユーラシア大陸から本州に少なくとも2回渡来したことが示唆されました。

こうして古い年代のAncient DNA解析にも成功したので、次にニホンオオカミの起源に迫ろうと考えました。日本固有とされる「極めて小型のニホンオオカミ」は、かつては本州、四国、九州に広く分布していましたが、約100年前に絶滅しました。これとは別に、化石からは、2万年前よりも以前の本州には「世界最大級の巨大なオオカミ」が生息していたことがわかっています。ところが、小型のニホンオオカミの起源や、巨大オオカミとの系統関係は、わかっていませんでした。私たちは、3万5000年前の巨大オオカミの遺骸、5000年前の小型のニホンオオカミの化石の二つを対象に、核DNAとミトコンドリアDNAを用いたAncient DNA解析を行いました。その結果、小型のニホンオオカミは、「古い年代（5万7000年～3万5000年前）に渡来した巨大オオカミ」と、「その後（3万7000年～1万4000年前）に渡来したオオカミ」の両系統が交雑して誕生したことがわかりました。



▌オオカミの骨のDNA解析用のサンプリングの様子

今は、数万年前まで生息していたナウマンゾウのAncient DNA解析を進めているところです。ナウマンゾウは各地で大量に出土しており、博物館などにも化石標本が展示されていますが、ゲノム解析の成功例はありません。すでに系統関係についての興味深い知見が得られ、論文を書いているところです。

厳しい状況に置かれた「日本のバイオインフォマティクス研究」

人材、予算の両面において、日本のバイオインフォマティクス研究は、欧米や中国等よりも厳しい状況にあります。昨今のAIブームもあり、多くの学生は、バイオインフォマティクス分野で学位を取得しても、民間企業への就職を選んでいます。学術界と民間企業では、大規模データの解析面や機械学習が得意な人材に対する給与などの待遇面に大きな差があること、バイオインフォマティクス研究者として学術界で長期間安定した職を得るのが難しい、などの問題があるからです。研究予算の削減も影響し、バイオインフォマティクスを含む、日本の生命科学研究全体の今後が危ぶまれています。私自身も人材育成のための情報解析講習会を開催したり、バイオインフォマティクス解析についての本を分担執筆したりと様々な努力を重ねていますが、なかなか改善しない現状にあります。

お話ししてきたように幅広く色々なことをやっていますが、メタゲノム解析を通じて、「微生物の多様性の全体像」を明らかにするのが究極の目標です。微生物の大半は、名前もない未知の存在のままです。どのような環境に、どのようなゲノム・生態の微生物が生息しているのかを、今後も研究していきたいと考えています。もちろん、一人でやれることは限られているので、積極的に共同研究も進めています。

まだまだ若手研究者の域を出ませんが、じっくり時間をかけて研究できるのが若手の特権だと思いますので、集中して取り組んでいます。ただ、自分の世界に籠り過ぎることのないよう、積極的に外部の人たちとコンタクトし、新しい情報と刺激を得るようにしています。遺伝研は学生よりも教員の数が多く、学ぶ環境として恵まれていると思います。バイオインフォマティクスを専門にしたい方等、大歓迎ですので、気軽に門戸を叩いてみてください。

聞　き　手：サイエンスライター　西村 尚子
写 真 撮 影：遺伝研 ORD　来栖 光彦

2022年8月

▌引用論文等

*1----------
▌Asymmetric reproductive isolation during simultaneous reciprocal mating in pulmonates
Biology Letters. 2009 Apr 23; 5(2): 240-3.

*2----------
▌VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences
BMC Bioinformatics. 2010 Jun 18; 11:332.

*3----------

▌VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing

BMC Systems Biology. 2018 Mar 19; 12(Suppl 2):30.

*4----------

▌PZLAST: an ultra-fast amino acid sequence similarity search server against public metagenomes

Bioinformatics. 2021 Jul 8; 37(21): 3944–3946.

*5----------

▌Phylogenomics and Morphology of Extinct Paleognaths Reveal the Origin and Evolution of the Ratites

Current Biology. 2017 Jan 9; 27(1): 68-77.

*6----------

▌Ancient DNA reveals multiple origins and migration waves of extinct Japanese brown bear lineages

Royal Society Open Science. 2021 Aug 4; 8 (8):210518.

*7----------

▌Paleogenomics reveals independent and hybrid origins of two morphologically distinct wolf lineages endemic to Japan

Current Biology. 2022 Jun 6; 32(11) 2494-2504.e5