



丁寧で高精度なゲノム解析と ゲノムバンク事業の国際連携を進める!

なかむら やすかず

教授 **中村 保一** 情報研究系 大量遺伝情報研究室

信州大学農学部卒、京都大学大学院農学研究科修士修了後、山梨県立総合農業試験場技師を経て、総合研究大学院大学生命科学研究科 遺伝学専攻(遺伝研)博士課程中退、1996年かずさDNA研究所研究員に着任、2006年研究室長昇任。2009年1月、遺伝研生命情報研究センター(当時)大量遺伝情報研究室教授として着任、現在同情報研究系教授・研究主幹。生命情報・DDBJセンター国際連携部門長を兼任。総合研究大学院大学博士(理学)

「4塩基からなる遺伝暗号文は、ゼニゴケでもヒトでも共通」--- 生物種を問わずに行えるゲノム解析に魅了され、多様な生き物の全ゲノムを丁寧に解析してきた中村保一教授。産官学で様々なレベルの研究を進める一方、遺伝研の代表として、塩基配列データバンク事業の拡充を図るための国際連携を担当している。

バイオインフォマティクスを専門とするまでの 紆余曲折

ベンチワークのバイオインフォマティシャンとして研究開発を進めていますが、大学院までは実験を手がける研究者でした。京都大学農学部の大山莞爾先生の研究室に所属し、植物として初めてゼニゴケのミトコンドリアの全ゲノム配列を決定する、といった研究に加わっていました。「大腸菌を使って、ゼニゴケのゲノムライブラリーを作り、ラジオアイソトープで配列を読む」、「バラバラに読んだゲノム断片を手作業でつなげる」といった、今では考えられない原始的なやり方で解析していました。途中で、手作業がバカらしくなり、研究室にあった16ビットの計算機を使い始めました。自分でプログラムを組み、自動で断片をつなげられるようにしたのです。振り返ると、このことがバイオインフォマティクスに向かうきっかけだったのかもしれない。

といっても、当時はまだ「バイオインフォマティクス」という概念がありませんでした。まわりを見渡すと、非常に優秀な先輩や同期ばかりで、そのまま博士後期課程に進んでも学問の世界で生き残るのは難しいと感じました。そこで、修士号取得後は、実家のある山梨県立総合農業試験場に就職することにしました。イネや野菜の栽培をしつつ、土壌の研究開発などを担当し、悪くない生活でしたが、年月とともに「これを一生続けるのか?」という疑問も浮かんできました。その頃に結婚したのですが、妻が経済的に自立していたこともあり、興味をもっていたコンピュータによる解析と生物学の両方を生かせる領域で博士号を取得しようと目論みました。

幸運なことに、遺伝研の池村淑道先生が学生を探していることを知りました。創立間もない総研大の「博士後期課程の募集要項」を目にしたのです。そこには「計算機生物学」という文字があり、「これだっ!」と思いました。1992年のことです。当時はヒト

ゲノム計画が全盛期を迎え、シーケンサーの登場によって塩基配列が高速に読めるようになってきていました。さっそく、池村先生に相談に行ったところ、「私のところでもヒトゲノムの大規模構造を解析するよ」と言われ、受験することにしたのです。

3期生として総研大に入学し、ヒトゲノムの解析が始まりました。対象がゼニゴケからヒトへとドラスティックに変わりましたが、どのような生物種であれ、遺伝情報は4種の塩基で書かれたものです。その意味で、バイオインフォマティクスは非常に自由度の高い領域といえます。池村先生は、実験と解析の両方で研究を進めていました。染色体を染めると濃いバンドと薄いバンドができますが、濃い部分にはGC含有量が多いこと(GCリッチ)、GCリッチ領域には遺伝子が多く存在することなどが知られていました。そのバンドの境界になにかシグナルがあるのではないかと考えて解析しておられました。

研究室には当時の金額で3000万円ほどの高性能なコンピュータがありました。私が使ったのはラッキーでした。私は、分担されたヒトゲノム解析をこなしつつ、「生物種ごとに少しずつ異なるアミノ酸のコードン」を自動で変換できるプログラムなどを作り、コードンデータベースを構築し始めました。コードンデータベースは池村先生が作り始めたものですが、先生はヒトゲノム解析で手一杯になっていたの、私が勝手に引き継いだかたちです。

データベースが完成すると、自力でウェブサーバーを立ち上げ、インターネットを介して公開しました。当時はインターネットが本格的に普及し始めた頃で、日本にはウェブサーバーがおそらく100~200ほどしかなく、その一つが私の立ち上げたものとなりました。

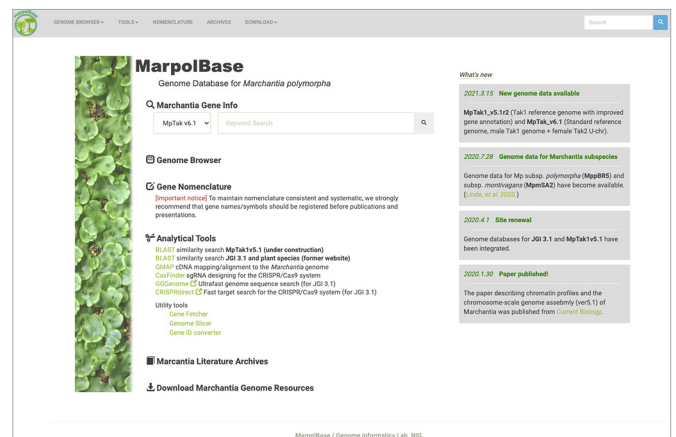
1996年に博士号を取得し、千葉県にあるかずさDNA研究所にバイオインフォマティクス研究者として就職しました。シアノバクテリアのゲノム解読を成功させたグループに入り、同じ年に配列が完成しました。光合成生物としては世界初の快挙で、バクテリアとしても世界で4番目の全ゲノム解読でした。この頃から、情報解析の責任者を務めるようになり、併任で遺伝研のデータバンク事業(DDBJ; DNA Data Bank of Japan)の運営委員にもなりました。2001年には高等植物のモデル生物として汎用されるシロイヌナズナのゲノム解析を完成させましたが、2008年に遺伝研から「教授として来ないか」との声がけをいただき、2009年に古巣に帰ることにしました。赴任当初はDDBJがメインの仕事でしたが、その後、国際連携も担当するようになり、現在は米NCBIや欧EBI、中国のゲノムバンクなどとの連携を進めています。もちろん、様々な生物種を対象にゲノム解析も続けています。これまでの主な解析対象には、ゼニゴケ、かんぎつ類、イネコなどがあります。

陸上植物の基本となるゼニゴケ

私にとって膨大な時間と労力を費やしたゲノム解析が、ゼニゴ

ケです。ゼニゴケは陸上植物としてもっとも原始的なしくみを残しており、陸上植物の形態や進化について研究する際の理想的な材料となります。モデル植物としての歴史は古く、植物の葉緑体やミトコンドリアの全ゲノムはゼニゴケで初めて解読されました。全ゲノムの解析も進められ、2007年には染色体レベルの概要成果が報告されたのですが、染色体丸ごとの詳細な解析は、なかなか進みませんでした。

このような状況で、遺伝研、近畿大学、神戸大学、基礎生物学研究所、東北大学、米エネルギー省、豪モナシュ大学など、国内外の39の大学・研究機関がコンソーシアムを作って、ゼニゴケの全ゲノム構造を解明するという大プロジェクトが立ち上がりました。DNAの塩基配列を比較的安価で、高速かつ高精度に読める次世代シーケンサーが本格的に導入されたのと、1万塩基以上の長いDNA配列を一続きに読める「ロングリード法」の開発が後押ししたのだと思います。私の研究室のグループは、遺伝研のスーパーコンピュータを使って、断片化されたゲノム配列を正確につなげたり、遺伝子を予測してデータベースを構築する、といった役割を担いました。



解析の結果、「約2万の遺伝子がある」、「遺伝子のスイッチとしてはたらく転写因子の数は他の陸上植物種よりも少ないが、陸上生物の基本となる遺伝子セットがある」といったことが明らかになり、陸上植物の進化、遺伝子構造、形態、情報伝達、代謝系などを研究する際にゼニゴケを多様な陸上植物の基本型として使うための情報が揃いました。論文は2017のCell誌に掲載されました(*1)。

ゼニゴケというと、一般の人にとっては「湿っぽい日陰に生えるコケ」といった認識しかないとと思いますが、植物研究者にとって、全ゲノムデータベースの完成は、待ちわびた価値ある成果でした。

温州みかんなど、多品種のカンキツ類ゲノムをカタログ化

遺伝研のある静岡県は、温暖な気候でみかんの名産地としても知られています。そのため、県内には農業・食品産業技術総合研究機構(農研機構)の果樹茶業研究部門が置かれています。カ

ンキツ類のゲノム解析は、同部門のカンキツゲノムユニット清水徳朗先生に声をかけていただいたことがきっかけとなりました。ゼニゴケの解析のように全ゲノムを一続きにして詳細に解析するのではなく、なるべく多くのカンキツ類ゲノムを大雑把に解析し、多品種のゲノムをカタログ化するのが目的でした。世界の市場では150以上の品種が出回っていますが、遺伝的多様性があまりにも大きく、親子関係については不明のものも多くありました。系統関係がわかれば、育種を行う上で非常に役立つと思われます。

まず、清水先生たちが、自身の研究所内に植えられている品種など269点を対象に、核とミトコンドリアのDNAを抽出し、各断片を解読しました。その配列データを私たちが受け取り、遺伝子の構造や配列の多型などについて詳細に比較解析しました。その結果、温州みかんをはじめとする22品種で「両親にあたる品種」が特定でき、種子親と花粉親の食み合わせも明らかになりました。また、45の在来品種について、片親や起源、親子関係といった系統情報を得ることができました。論文は2017年に出しました(*2)。

食味や耐病性などは、複数の遺伝子が関与するので、あらゆる遺伝子をカタログ的に概観できるデータベースは、現存品種を掛け合わせて「より美味しく、しかも病気に強い」といった品種を作出するのに役立つと期待できます。こちらも生データと解析データをDDBJに登録済みです。

内外で人気のアメリカンショートヘアの全ゲノムを解読

イエネコのゲノム解析をしたのは、私が無類のネコ好きであること、イエネコには「品種ごとの、発症しやすい遺伝子疾患」が知られていながら、ゲノム情報が蓄積されていないために予防や治療法の研究が進まない現状があったからです。イエネコとは、ヤマネコをもとに人類が家畜化し、コンパニオンアニマルとして育種されてきたものを指します。一部でヤマネコとの交雑もみられます。唯一、2007年に米国でアビシニアンという品種のゲノム解析がなされていますが十分なデータとはいえませんでした。例えば、アビシニアンは近親交配が進んでいる品種であるために、「他の品種と比較する際の遺伝的な目印(マーカー遺伝子)が少ない」という問題がありました。また、全ゲノムといっても、約4500のゲノム断片をバラバラに読んでつないだに過ぎませんでした。さらに、染色体レベルの配列構造にも未解明の部分が多くありました。

幸運なことに、今回の私たちのイエネコゲノム解析は、大手のペット保険会社として知られるアニコム損保のグループ会社(アニコム先進医療研究所)、かずさDNA研究所、香港中文大学(香港)というそれぞれの分野で研究を牽引してきた3組織との共同研究で進めることができました。解析する品種は、アメリカンショートヘア(メス)にしました。アメリカンショートヘアは、世界中で人気があり、スコテッシュフォールドなどの他品種との交

配にも使用されています。つまり、遺伝的に近縁なネコが多く、得られたゲノム情報を様々な応用でき、獣医療に関する研究にも寄与すると考えられます。また、アメリカンショートヘアはアメリカ大陸に初めて上陸したネコを祖先としており、イエネコの進化を考える上でも重要だと思われました。

DNAをシーケンスしたのはかずさDNA研究所で、私たちは遺伝研のスーパーコンピュータを駆使してシーケンスデータをもとにした遺伝子予測や遺伝子構造の解析を進めました。成功の鍵は、ゼニゴケの解析でも用いたロングリード法を駆使し、ネコにある19本の各染色体をほぼ全長で高精度に解析できたことにあります。解析の結果、アメリカンショートヘアのゲノムは全長で24億9314万1643塩基対に及び、2万3119個の遺伝子があるとわかりました(*3)。遺伝子予測には「遺伝子としては大き過ぎないか、分断されて予測されていないか」など、私たちのグループがもつこれまでの経験と勘をフル活用しましたが、こうした校正に今後はAIも活用できると良いと思います。

アメリカンショートヘアとアビシニアンの配列と比較することで、「両者は、染色体の一部が異なるゲノム構造になっている」といったこともわかりました。今後は、2つの品種の遺伝子をさらに詳しく比較することで、アメリカンショートヘアに多い肥大型心筋症などに関連した遺伝子についての知見が得られると期待できます。

民間企業との共同研究では、得られた成果を一般公開できないことが多いのですが、アニコムは快く認めてくださり、去年の5月からDDBJ上で公開しています。現状は配列データのみですが、獣医の皆さんに使っていただきやすいよう、医療情報や病因遺伝子との連携などの情報蓄積と整理を進めていきたいと考えています。

Cats-I: Cats' genome Informatics
Genome database for *Felis catus*
This web site contains genomic data obtained from the study for "AnAmS1.0: A high-quality chromosome-scale assembly of a domestic cat *Felis catus* of American Shorthair breed" (Iacob, Matsumoto, Chung, Sakamoto, et al. *BioRxiv*, 2020, [DOI](#)).

Search:

Keywords:

Our cat

Name	Genus
Breed	American Shorthair
Sex	Female
Genome sizes	20 sequences (19 chromosomes + 1 unplaced) 2.49 Gbp in total

Assembly and annotation version
The current version of the genome assembly is v1.0 (AnAmS1.0).
The genome sequences are also available from the INSDC under the accession number [BioProject:PL0283878](#).
Note that the current gene annotation is still in beta-version and is subject to change in future revisions.

Citation
The data provided in this web site is freely available for academic purposes. Please cite the preprint paper on [BioRxiv](#) if you use the data obtained from the web site.
DOI: [10.1101/2020.05.19.103708](#)

Inquiries and feedback
Genome Informatics Laboratory, National Institute of Genetics, [yn@nig.ac.jp](#)

Logos: [anicom](#), [国立遺伝学研究所](#), [香港中文大学](#)

Cats-I: Cats' genome Informatics / Genome Informatics Lab, NIG.

産官学を問わず、共同研究を進める理由

大学、研究所、企業と、産官学を問わず、多様な生物種、規模、レベルのゲノム解析を共同で進めています。その背景には、国による科学研究予算の不足や、規模の大きなゲノム解析を行える

インフォマティシャンがそう多くない、といった状況があります。「配列情報を重んじ、生物種を問わない」という私の研究姿勢も影響しているかもしれません。

産官学との連携は、引き続き進めていくつもりです。新型コロナウイルスの変異が問題になっていますが、静岡県内の陽性者由来のウイルスを対象にした変異株のゲノム解析など、私の研究グループは社会貢献につながる解析にも協力しています。

一方で、一見、役に立ちそうもないゲノム解析も基礎研究として非常に重要だと考えています。たとえば、ゼニゴケのゲノム情報は直接、産業応用には結びつかないかもしれませんが、陸上植物の基盤メカニズムを理解する上で欠かすことができません。学生への講義でよく話すのですが、その昔にハインリヒ・ヘルツが発見した「電波」は発見者本人が「何の役にも立たない」と言ったのですが、それなくして、今日、万人が持つ現在の携帯電話は存在しないわけです。「応用できそうになくとも、興味を抱いたことを納得するまで突き詰める」、研究者にはこのような姿勢も重要でしょう。



現在は、共同研究として「ゲノムの不安定さが引き起こす、遺伝子疾患」の解析も進めています。かなり大がかりなプロジェクトで遺伝子疾患の多くは非常に患者数の少ない孤発性です。「一人の患者で全ゲノムを読んでも得られる情報が乏しく、あまり意味がない」との見方もされるのですが、私はそうは思いません。全ゲノムを詳細に解読し、その情報を蓄積していけば、将来、病態に関わる重要な関係性などが見えてくると考えるからです。ゲノムを読むなら、「全部をきちんと解読し、そのデータをためておくこと」が何よりも重要だと思うのです。

総研大でバイオインフォマティクス やりませんか？

私が博士課程だった頃にくらべると、今の総研大の学生は非常に優秀だと思います。英語でのコミュニケーションは当たり前で、アカデミアの世界で生き残りたい、という気概がある学生も多い印象です。一方で、博士号取得後に民間企業に就職して活躍する卒業生もいます。みなさん目的意識が高く、大学のネーム

バリューではなく、研究者名を名指しし、その人の下で研究するために入ってきています。

実は、イエネコゲノムの解析は、私が博士論文の審査委員を務めた松本さんが、アニコム先進医療研究所に就職したことで共同研究をよびかけていただいたものです。

ひと昔前までは、博士号取得後のキャリアは限られていましたが、今はかなり自由度が大きくなっています。特に情報解析分野では、民間企業がデータサイエンティストを優遇していますので、環境も給与も恵まれています。若手研究者は、アカデミア、民間を問わず、いろいろ経験したら良いと思います。ただ、インフォマティシャンは待遇につられて民間企業に引っ張られる傾向が強いので、アカデミアに残ってもらえない傾向があるのは問題だと感じています。アカデミアにおいて、研究室を主宰するポストが少ない点も改善しないと、将来の展望という観点から残ってもらいづらいですね。

日本でバイオインフォマティクス研究というと、東京大学、東京工業大学あたりが非常に有名ですが、この場を借りて、総研大でも様々な研究ができることをアピールしておきたいと思います。コロナ禍が続いていますが、コンピュータさえあれば研究できる私たちは自宅でのテレワークで何の問題もなく、研究への影響はほぼありません。「我こそは」という方がいたら、ぜひ遺伝研の門を叩いてください。社会人学生や、私のように一旦就職してから再び学びたいといった方も大歓迎いたします。



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

2021年7月

■ データベースへのリンク
■ [MarpolBase: Genome Database for *Marchantia polymorpha*](#)

■ [Cats-I: Cats' genome Informatics](#)



■ 引用論文等

*1-----

■ [Insights into Land Plant Evolution
Garnered from the Marchantia
polymorpha Genome.](#)

Cell. 2017 Oct 5;171(2):287-304.e15.



*2-----

■ [Draft sequencing of the heterozygous
diploid genome of Satsuma \(Citrus unshiu
Marc.\) using a hybrid assembly approach.](#)

Front Genet. 2017 Dec 5;8:180.



*3-----

■ [AnAms1.0: A high-quality
chromosome-scale assembly of a domestic
cat Felis catus of American Shorthair breed.](#)

bioRxiv

doi: <https://doi.org/10.1101/2020.05.19.103788>

