

情報・システム研究機構ブックレット

SCIENCE REPORT

007-012

僕らは
ゲノムで
できている。



大学共同利用機関法人
情報・システム研究機構

情報・システム研究機構ブックレット

SCIENCE REPORT

007-012

僕らは



ゲノムで

できている。

大学共同利用機関法人

情報・システム研究機構

『サイエンスリポート』について

『サイエンスリポート(<https://sr.rois.ac.jp/>)』は、
大学共同利用機関法人情報・システム研究機構が
運営するウェブサイトです。

2016年12月公開以来、
当機構の大学共同利用機関法人としての役割を踏まえ、
自機構のみならず他機構および
広く大学・研究機関等の話題を採り上げ、
広く一般・マスコミの方々へ向けて、
学術の成果や取り組みに関する情報を提供しています。

その中心的なコンテンツとして、ウェブサイトと同名の連載記事を
定期的に(月1回程度)掲載し、
継続的に新しい記事を追加しています。
先端的な取り組みをわかりやすく紹介し、
人々の疑問に科学者コミュニティが答えるような構成で、
学術の役割と活動を広くご紹介しています。

本書は、その記事が身近にお手にとれるよう、
冊子にまとめました。

目次

03 『サイエンスリポート』について

08 Science Report 007

微生物のゲノムから何がわかる？

答える人 | 黒川顕教授

16 Science Report 008

ヒトゲノムの読み取り技術で医療が変わる？

答える人 | 鈴木穰教授

24 Science Report 009

ゲノムはどんな意味を持つのだろうか？

答える人 | 小原雄治特任教授

30 Science Report 010

生物学とデータベースの関係は？

答える人 | 五斗進教授

36 Science Report 011

植物ゲノムは暮らしのどこに役に立つ？

答える人 | 田畑哲之所長

42 Science Report 012

チンパンジーゲノムから何がわかったか？

答える人 | 松沢哲郎特別教授・藤山秋佐夫特任教授

51 大学共同利用機関法人 情報・システム研究機構について

機構長 藤井良一



SCIENCE REPORT

007-012

僕らは
ゲノムで
できている。

微生物の ゲノムから 何がわかる？

2003年のヒトゲノム解読で力を発揮した米国の生物学者クレイグ・ベンターの次の狙いは、海洋細菌群だった。翌2004年、航海から戻った彼は120万個もの新規遺伝子を発表して、世界を驚かせた。このように微生物のゲノム解析を通じて、その微生物が棲む環境についての情報を一挙に得る方法を「メタゲノム解析」という。微生物を実験室で培養する従来法と違って、採取してきた微生物群をそのまま読み込んでゲノム配列を決めるのが、大きな特徴だ。地球環境やヒトの体内環境は多くの微生物によって維持されているため、これら微生物群の把握には医療、創薬、環境、農業、建築など幅広い応用への期待がかかる。ゲノムを読む「次世代シーケンサー」の急速な発達を背景に発展しつつあるメタゲノミクス——そこから今、何を知ることができるのだろうか。
(写真は調査地のひとつ、鹿児島県にある霧島温泉。撮影：黒川顕)

答える人

黒川 顕教授

[国立遺伝学研究所]

国立遺伝学研究所教授、先端ゲノミクス推進センターセンター長。2014年より新学術領域「冥王代生命学の創成」領域代表者。東北大学卒、大阪大学大学院後期博士課程修了。奈良先端科学技術大学院大学准教授、東京工業大学教授などを経て、2016年より現職。2004年、日本で初めての本格的なメタゲノム解析に着手し、日本人13人の腸内細菌群のメタゲノム解析結果を公開。このメタゲノム解析手法を、土壌や温泉水など他の環境に応用した、微生物集団の変化の過程を探索の研究などで知られる。ちなみに修士課程までは地質学・プレートテクトニクスが専門。



そもそもゲノムって何だろう？

ゲノムとは、DNA(デオキシリボ核酸) でできた物質である。なかでもATGC(アデニン、チミン、グアニン、シトシン) の4つの塩基は、生物それぞれに固有の配列を持つ。ちなみにヒトゲノムは約30億塩基、これまでに見つかった一番小さいゲノムはカルソネラ・ルディアイという寄生性の微生物で15万9662塩基、こちらは文字数にしておよそ朝刊1部ぐらいに相当するそうだ。一方でゲノムとは、生物の全遺伝情報を指す。ATGCの塩基配列の3つの塩基の組み合わせは「コドン」あるいは「遺伝暗号」等と呼ばれ、20種類のアミノ酸のいずれかに対応している。遺伝暗号はDNAからmRNAへ「転写」され、続いてmRNAが「翻訳」されてタンパク質が合成される。このDNAからタンパク質合成までの一連の流れは、全生物に共通の原理となっている。「ゲノムは物質とデータの橋渡し役」と国立遺伝学研究所の黒川顕教授は言う。



写真は多摩川の下流付近。多摩川水系は、東京都の水道水源の約20%をまかなう。
ここでの微生物ゲノム解析は「学生たちががんばった成果」と黒川教授。

微生物は裏切らない生物である

ゲノムのおもしろいところは、小さな微生物から人類まで、客観的なデータで全生物の遺伝情報を決定できることだ。しかも「ATGCというたった4つの化学物質の並び方に、さまざまな意味が畳み込まれている」。つまりメタゲノム解析は、その解析結果を多次的に読み解き、あるいは他のデータと連携することで、さまざまな情報が引き出せるデータの宝庫なのだ。例えば、黒川教授が挑戦してきた微生物群のひとつに腸内微生物がある。「昨日と今日とで、私たちのおなかの様子は違いますよね。でも腸内環境のゲノム情報を調べれば、体温を測り忘れても、腸内の温度を知ることができます。pH(水素イオン指数)や湿度などの衛生データも把握できる。さらに時系列に沿ってデータをとることで、環境にどのような変化が起きたかも分かります」。しかし病院で1日ばかりで検査してきたようなデータが、なぜ微生物から得られるのだろうか? 「それは、微生物群集は環境に対して一意に決まる、言わば裏切らない生物だからです。その生物に聞けば、人為の及ばない、客観的な観察データが得られます」。また土壌を調べ

るのにも微生物が貢献する。「奥多摩湖から羽田空港まで、多摩川138キロの川沿いの微生物群集を調べました。するとこのデータに、例えば人口分布や犯罪率といった流域周辺の社会環境のデータを組み合わせることで、今度は地域環境を評価することができる。われわれは微生物の群集構造を見るだけで、周辺環境が推定できる技術を作り上げています」。実に、微生物メタゲノムの多層的な性質そのものが、さまざまな他分野や異業種との連携を可能にしているのである。「メタゲノムの本質は、たぶんそこなんです」と黒川教授は言う。

僕らの細胞も

微生物と共存している

このようなデータ連携の基盤となるのが、微生物の種類や数と環境をひも付けるデータベースだ。現在、世界で約17万サンプルの情報が公開されており、黒川教授の研究室では、これを集めて微生物環境を俯瞰できる「微生物GPS」を構築しているという。「僕らは地球上の微生物の全てのカタログを持っているので、これを使えば、微生物がこういう群集ならどういう環境であるかが、だいぶ見えてきます」。さらに解析の処理能力次第では、環境全体にあるゲノムを洗いざらい決めていく「ホログenom」も強力な手法だという。「僕らヒトの細胞も微生物もゲノムでできており、共存し合って生きている。そこでヒトゲノム、微生物一匹一匹のゲノム、環境データとしてのメタゲノムの全部を調べ、全体を1つのシステムとして捉えて、ゲノムレベル



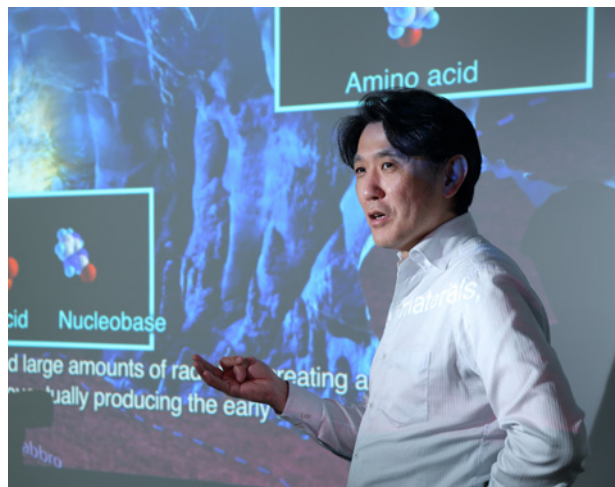
長野県白馬村にある白馬八方温泉は蛇紋岩体を貫く温泉であり、水温50度で水素含有量が極めて高く強アルカリ性(pH11)を示すことから、冥王代地球に類似した環境であると考えられている。この地下600mの蛇紋岩熱水に存在する微生物群集をメタゲノム解析することで、原始的な代謝システムが見えてきている。写真：黒川顕

で解いていこうという試みも始まりつつあります。

地球環境を理解して生命の起源に迫る

生命現象を考えるには、その環境を知らなければならない——黒川教授はこの考えを発展させ、46億年前から40億年前にあたる「冥王代」の地球環境を推定し、生命の起源に迫る研究にも取り組んでいる。「なぜ生命科学者が生命の起源を考えられないかといったら、当時どんな地球環境であったかを知らないからです」そこで地質学者、地球化学者、生命科学者、惑星科学者を集め、議論を進めていったところ、「当時の海はpHゼロの強酸で、岩石等が全部溶け込んだドロドロの溶鉱炉のような、たとえ生命が誕生したとしても持続的には生きられない過酷な環境であり、生命を育む母なる海、といった生易しいものではなかった」ことがわかってきた。原始大陸上の間欠泉のような場所で、強い放射能を受けて猛烈なスピードで化学反応が進展し、生命を形作るさまざまな物質が造り出され、やがて生命になった——そんな過程が「科学的な証拠付きで見えてきました」。それでもまだ、わからないところがある。「なぜゲノムが遺伝暗号という情報を担うようになったのか。それは依然として、謎です」と黒川教授は言う。

公開日：2017/06/12



黒川教授が代表を務める新学術領域では、科学的知見に基づいて、生命の誕生をCGやアニメーションで再現した映像YouTubeチャンネル「全球球史アトラス」も公開している。

●日本語版

<https://www.youtube.com/c/冥王代生命学の創成>

◎英語版

<https://www.youtube.com/>

channel/UCCToEEPIF7ur1m0efQTHG4g

ゲノム解析の学術連携を推進
「先端ゲノミクス推進センター」

先端ゲノミクス推進センターは、次世代型シーケンシングが生み出す大量のゲノムデータの最先端の研究基盤を大学・研究者コミュニティに提供し、他のゲノム研究機関との共同利用・共同研究を進めているセンターだ。日本に19ある大

学共同利用機関のひとつ、国立遺伝学研究所(静岡県三島市)内に2011年10月に設立された。設立以来、他の研究機関と連携した共同研究により、先端ゲノム科学の研究の成果を挙げている他、大学等からの大規模シーケンシングや情報解析の要請にも随時応える。



ヒトゲノムの 読み取り技術で 医療が変わる？

生物をその生物たらしめている塩基配列を、全遺伝情報すなわちゲノムという。現在、このゲノム情報の読み取りと解析に広く使われているのが「次世代シーケンサー」と呼ばれる装置だ。人類最初のヒトゲノム解読では莫大な時間と予算が投じられたが、その後シーケンサーは目覚ましい発達を遂げ、より速く、より正確に、より安くゲノムを読み取り、解析できるようになった。今や膨大な数のヒトゲノムが解読され、データベース等に登録され、そのようなビッグデータを駆使するゲノム科学が、人類の医療、健康、環境、開発、サービスなどに、いよいよ具体的に応用される段階に入っている。シーケンサーが読み取り、解析する情報は、医療をはじめとする応用分野を、これからどのように変えていくのだろうか。

答える人

鈴木 穰教授

[東京大学]

東京大学大学院新領域創成科学研究科 教授。1994年東京大学卒、1999年同大学博士課程修了(学術博士)。1999年理化学研究所ゲノムサイエンスセンター リサーチアソシエイト、2000年東京大学医科学研究所ゲノム構造解析分野 助手を経て、2004年現所属 准教授。恩師である菅野純夫東京大学教授とともに25年以上にわたりトランスクリプトームに注目した研究を展開。ヒトゲノムの変異がどのようにがん、免疫、神経疾患等の表現型を示すのか、システムレベルでの解明に取り組む。



先端医療を病院だけでなくジャングルでも。

次世代シーケンサーは、米国企業などが世界的に市場を寡占していることから、ゲノム科学の基盤技術を担うだけでなく、「結果の承認、追実験といった実際的な手順を画一化する効果を生んでいる」と東京大学の鈴木穰教授は言う。「よい面も悪い面もある。次のステップをどう展開するかが重要です」。そのひとつとして鈴木教授が取り組むのが、特に発展途上国における現地＝オンサイトでのゲノム診断だ。コンパクトなシーケンサーを用いて、高熱を発症した患者の血液をその場で採取し、高速でゲノム解析する。マラリア、デング熱等が見つければ、その場で治療につなげることができるほか、ゲノム診断では、同じ病原体でも薬に効果がある／ないタイプの判別や、予想外の感染源でも発見できるといった優位点がある。そこで熱帯の屋外のような医療・衛生設備のない場所でも迅速に結果が出せるポータブルな装置を実現しようと、一連の技術開発を行っているという。「自宅で糞便などから個人のゲノムを採取して健康診断したり、野外調査で新しい生物種を判定したり、ウナギなど食品の産地偽装から髪の毛で個人を特定する犯罪捜査まで、応用範囲は意外にも広い」と鈴木教授。特に空港など「水際」における検疫チェックは、ゲノム診断が大いに役立つ可能性があるそうだ。



ポータブルゲノム診断のデモ機はまさに「手のひらサイズ」。鈴木教授が手にしている基盤を専用の読み取り機にセットすれば、パソコンで読み取ることができる。

ゲノム固有の突然変異を識別して、がんと闘う

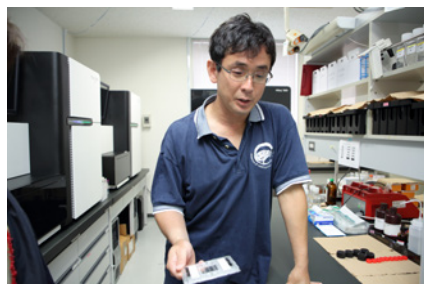
読み取ることから一歩進んで、ゲノム情報を入力、そこから形成されるものを出力として、その間にあるコード化／解読のシステムを解明し、予測や制御を行うという研究にも期待がかかる。DNA上の遺伝情報は1個体共通だが、DNAからタンパク質合成情報を読み取った「mRNA」階層の遺伝情報は体内の組織ごとに、また外部からの影響に呼応して同一個体内でも遺伝情報が異なる。長年にわたり、このmRNA階層の遺伝情報(トランスクリプトーム)を研究してきた鈴木教授は、がん注目する。「がんは、突然変異によってトランスクリプトームを変化させる疾病なんですね。ゲノムに変異が起こったら、遺伝子発現にどう影響するか、それを正常に近づけようとする薬剤に対してどういう変異が起こるか等を調べることができます」。もちろん、実際の治療に役立てたいという思いは、大きな動機だ。「がんは患者によって、言い換えればゲノムによって、原因や経過が非常に多様で、例えば肺がんだけでも世界に500とも1,000とも言われる種類があります。ある人に効いた薬が、別の人に効くかどうかは、なかなか言えない。まさに1人1人にストーリーがあり、そのような個別の突然変異に最適な治療が行えるとよいわけです」。



細胞ひとつひとつを選り分けて読み取る、シングルセル分離解析機。

方法を工夫して創薬をスピードアップ

そのような治療を実現するためにも、さまざまな工夫が必要だ。鈴木研究室では、ラボでの実験とともにコンピューター・シミュレーションによって実測と予測を掛け合わせる手法の開発も手がける。また、特にがん治療で問題となるのが、仮に99%のがん細胞を殺すことができても、残りのたった1%の抗がん剤耐性細胞こそが、再発を引き起こすという点だ。「その1%はどれなのか、ターゲットとなる細胞を高精度に検出し、ゲノムを読み出し、さらにその結果をシミュレーションなどに反映していかなければならない」。一方、近年、患者の持つ免疫力を活性化させてがんを攻撃する新しい治療薬「免疫チェックポイント阻害剤」が注目を集めているが、この薬の評価でも「どの免疫細胞が有効なのか、細胞一個一個について見ていく必要がある」と鈴木教授は言う。

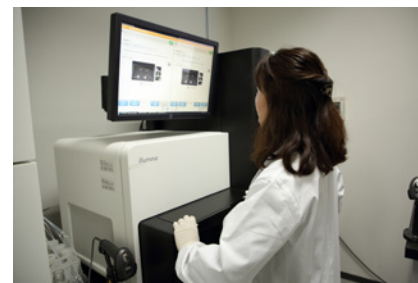


鈴木教授が手にしているのは、シングルセル分離解析機用のトレイ。トレイは微細な区画に仕切られており、機械にかけるとその1つ1つに各細胞が分離され、遺伝子情報を読み取ることができる。

シングルセル技術と次世代シーケンサーを駆使して

1つ1つの細胞を分離・分析する技術をシングルセル技術という。次世代シーケンサーの試料の前処理のひとつとして、一般に、対象となる組織をすりつぶしてから機械にかけるが、がんの場合、解析したいのは細胞の「かたまり」ではない。「シングルセル技術で細胞1個1個を取り出し、次世代シーケンサーで読み出しと解析を行う研究開発を、現在大腸がんを対象に、がん研究所細胞生物部の八尾良司部長と共同で研究を進めています。これは、科研費の研究者をゲノム解析で支援する「先進ゲノム支援」で行っているもので、大腸がんがどのような細胞組成を持ち、中でもがんの根源にあたるような細胞がどう発生し、突然変異を示すのかが追って明らかになるはずです」。

公開日：2017/07/10



次世代シーケンサーを操作する様子。研究員左手の下にある黒い部分の中に試料をセットし、画面操作で読み取りを開始する。

大学共同利用機関を中心にゲノム解析を支援 「先進ゲノム支援」

文部科学省科学研究費助成事業の新学術領域研究『学術研究支援基盤形成』先進ゲノム解析研究推進プラットフォーム「先進ゲノム支援」は、「科研費」に採択された研究を対象に、最先端のゲノム解析及び情報解析技術を提供する研究事業である。大学共同利用機関である国立遺伝学研究所(静岡県三島市)を中核機関として、多くの大学・研究機関が参加し、ネットワークを形成しており、旧事業の「ゲノム支援」以来、分野横断的な研究成果を挙げている点にも特色がある。支援活動を通じて、生命科学の最先端研究の推進、そしてゲノム解析の高度化や解析技術の発展に取り組んでいる。



●先進ゲノム支援ホームページ <https://www.genome-sci.jp/>



ゲノムは どんな意味を 持つのだろうか？

人間ひとりひとりや多種多様な生物種のゲノムが、ますます高速かつ大量に読み取られ、公開される時代になってきた。ゲノムはまさに生命科学のビッグデータであると同時に、ゲノムが決まるということは、人間個人や生物種を識別できることを意味する。ヒトとチンパンジーの塩基配列の違いがたった1.23%であるとか、これまでの進化の系統樹が書き替えられるといった新しい知見は、まさにゲノム解析の賜物と言えるだろう。ゲノムによってさまざまな科学的仮説が検証できることから、ゲノムそのものを対象としない研究においても、ますますゲノム解析が要請されるようになっている。——ゲノム解析が必要な研究は、どう進められているのだろうか。

答える人

小原雄治 特任教授
[国立遺伝学研究所]

国立遺伝学研究所 特任教授、ライフサイエンス統合データベースセンター (DBCLS) センター長、新学術領域研究「先進ゲノム支援」研究支援代表者。1974年京都大学理学部卒、1988年英国ケンブリッジMRC分子生物学研究所客員研究員等を経て、1996年より国立遺伝学研究所教授、2004～2012年 所長を務める。分子生物学の発展を背景に、生物の表現型とゲノムを結ぶメカニズムを追求する遺伝学の発展に尽力。また研究の発展には、これらのデータの収集・共有が必須であることから、統合データベースの構築・整備に注力する。



イノベーションは5年に2回起こる

ゲノムを研究するゲノミクスの分野は、国際的にも競争が激しい分野だ。研究を支える技術的な発展も目覚ましく、国立遺伝学研究所 小原雄治特任教授（DBCLSセンター長）によれば、「5年の間に2回ぐらい、イノベーションが起こる」という。ゲノムを読み取るために広く使われる世界標準的な装置にも変遷があり、「新しいマシンが利用可能になるが早いかな、即対応していかなければならない」。それはつまり、ラボでどんな実験を行い、どんな試料を用意し、どんな解析を行うか、という研究の一連のプロセスが再検討されなければならないことを意味し、しかも各段階に高度な専門性が要求されるという。一方で、ゲノムを必要とする研究領域は拡張の一途だ。「昨日までゲノムなんて関係なかった研究者が解析したいというケースが増えている」と、小原特任教授。「環境、医学、農学などの分野が多いですね。例えば土や川の微生物を調べたら、ゲノム解析によって面白いことが分かってくる。医学では病気遺伝子はもちろん腸内細菌など、農学では品種改良のために、栽培植物のゲノムを調べたい……」など、さまざまな要望があるという。シングルセルのゲノム解析（Science Report 008 参照）も急増しているようだ。

川や土壌の微生物を対象としたゲノム解析は、要望の多い分野のひとつ。



ゲノムという研究分野を全体として支援する

このようなゲノム研究の展開に対応するためには、個々のラボで対応するよりも、規模的にもまとまって専門性の高い機能を提供し、装置のアップデートも集中的に行えるような体制を組むことが望ましい。そこで、我が国の学術研究の中心的な競争的研究資金である「科研費」に採択された研究を対象とした公募により、そのゲノム解析の部分を支援する研究事業が始まった。「2000年に科研費の特定領域研究に支援班ができたのが最初です。その機能が新学術領域研究「ゲノム支援」となり、さらにリソースや技術をもっと分野横断的に提供できるよう組み替えられ、大学共同利用機関である国立遺伝学研究所を中核機関とした、現在の体制につながっています」。小原特任教授は現在、その新学術領域研究「先進ゲノム支援」で、研究支援代表者を務める。個々の大学等では整備・維持が困難な最新の設備・技術を常に整備し、国際競争の激しい最先端研究のゲノム解析を担うなど、メンバーとともに支援依頼者の要望に応えてきた。「新しい装置を試す中で、ゲノムを読み取るラボの技術もどんどん上がっている。またビッグデータ解析ができる情報系の研究者達が実データを扱うことで、情報解析のスキルが向上していく。支援を依頼した研究室の学生が1週間ぐらい滞在して、オンザジョブトレーニングを行う——と、さまざまなかたちでの好循環を目指しています」。



ライフサイエンスのデータベースの重要性

ところでゲノムは元来、個人を識別する生体情報という、個人情報保護や研究倫理の観点からも慎重な取り扱いを要する特殊なデータである。一方、ビッグ

データでもあることから「やっぱり集めたほうが勝ち」という側面もある。小原特任教授は言う―「米国だけでなく、例えば中国などにも大規模なゲノムセンターがあり、解析サービスも行っています。当然ながら、データはサービスの提供側に集まる。ここに、日本が学術としてデータを持たなければならない理由があります」。またデータは関連づけられていないと活用できない。「例えば医学なら、ゲノムとエピジェネティクス、あるいは医療記録との関連づけなどが必要です。『僕らはゲノムでできている』のだけど、ゲノムだけでは決まっていない。生き物も含めその内外全部のゲノムを取得・解析する分野を「ホログenom」と言いますが、そのように見ていくと、実際には寄生されていたり、腸内には細菌がいたりといった環境を含めて生き物は成立しているんですね。するとどこまでが「僕ら」なのか、生きていくための単位が変わっていく可能性がある。生命とは何か、個と種はどこで区別されるのかといった根源的な問いに答えられる日も来るのではないのでしょうか。

公開日：2017/08/10



原核生物からヒトまで 「ナショナルバイオリソースプロジェクト」

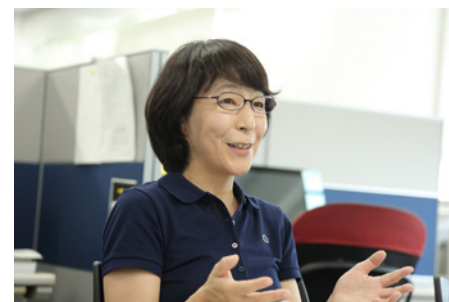
ライフサイエンスの基盤となる、研究開発の材料としての動物、植物、酵母や大腸菌などの微生物等をバイオリソースという。ナショナルバイオリソースプロジェクト(NBRP、プログラムスーパーバイザー：小原雄治)は、我が国として重要なバイオリソースを全国のNBRP中核的拠点(現在30)で収集・保存・提供するプロジェクトだ。この総合窓口&情報センター機能を担う国立遺伝学研究所 系統生物研究センターの川本祥子准教授は、本年度から前任者を引き継いで、「生物が違くと、同じ遺伝子名でも意味が違うといったように言葉や文化が違う。



<http://www.nbrp.jp/>
NBRPのリソースを利用した研究論文は、現在およそ2万8,000本にも及ぶ。

生物種を越えて検索できるようなデータベース」を目指す。「そのためにはすべてのゲノムが読まれ、比較ゲノム解析ができていなければならない」と川本准教授。遺伝子に横のつながりができれば、例えば植物では、稲・小麦・大麦などの单子葉の植物で、成功した育種を他の種でも成功させようという時などに威力を発揮する。また動物では特に「ヒトの疾患の解明に伝統的に使われてきたマウスやラットだけでなく、ゼブラフィッシュやショウジョウバエなどのモデル生物を用い、そのシンプルな型を活かして、ヒトの稀少疾患の病態解明を加速させることができるのでは」と期待がかかる。NBRPを通じて「例えば臨床の研究者とショウジョウバエの研究者のマッチングなどのお役に立てれば」と川本准教授は言う。

このほかNBRPは、生態系に関わる分布や環境などの「生物多様性情報」を集めようという国際的なネットワーク「GBIF(地球規模生物多様性情報機構) <http://www.gbif.jp/v2/>」の日本ノードとしても活動しており、「ある地域でどんな生物が生息しているかといったデータを登録いただいています。最近では屋久島のクモや近畿で営巣するツバメの調査などの例がある」という。また自身、遺伝医療分野の研究者経験も持つ川本准教授は「いかにデータベースが賢くても、それをどう使ってもらうかという問題意識は常



ナショナルバイオリソースプロジェクト(NBRP)を担当する、国立遺伝学研究所川本祥子准教授。

にある」という。特に今後の個別化医療へ向けて「患者さんなどエンドユーザのニーズに、システムとしてどのように応えられるか」といったオープンサイエンス的な視点も採り入れていく予定だ。

生物学と データベースの 関係は？

実験や観察で得られた個々ばらばらなデータを統合して、みんなで使えるようにする、というのは科学者の大きな夢のひとつだろう。学術のビッグデータが公開され、研究コミュニティが自由に利用できると、科学の発展が加速する。そこで学術データを標準化し、統合化しようとする動きが、世界的に進められている。そのためにはまずデータが公開される必要があるが、学術データのオープン化は、世界的にも特にヒトゲノムの公開で加速してきた経緯がある。生物の圧倒的に多様な世界を統合するために、実際にどんなふうにデータが収集・整理され、どう役立てられているのだろうか。21世紀の生物学とデータベースの切っても切れない関係をのぞいてみよう。

答える人

五斗 進教授

[ライフサイエンス統合データベースセンター (DBCLS)]

情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター (DBCLS) 教授(写真右)。1989年九州大学工学部 情報工学科、1994年、博士(工学)(九州大学)。1995年、京都大学化学研究所の金久實教授らとバイオインフォマティクス・データベース「KEGG」発足に参加、2017年より現職。大量のゲノム情報を整理して解析する「バイオインフォマティクス」を専門とする。写真左は、片山俊明特任助教(DBCLS)。今年10年目を迎えるライフサイエンスデータの標準化に関わる合宿形式の国際会議「BioHackathon」を担う。



生物学のデータベースは百科事典を目指す

ゲノム解析を始めとするデータや論文として発表された文献情報等が集められた生物学のデータベースは、日々の新しい発見によって改訂・編纂される巨大な百科事典(Encyclopedia)のようなものだ。その名の通り「Kyoto Encyclopedia of Genes and Genomes(KEGG) <http://www.genome.jp/kegg/>」は、京都大学の化学研究所で約20年の歴史を持つゲノムと遺伝子のデータベースである。発足から研究開発に取り組んできた、現在ライフサイエンス統合データベースセンター(DBCLS)の五斗進教授は「当時は、塩基配列やアミノ酸配列といった文字データがあり、一方教科書等を見ると、代謝経路やシグナル伝達系の図が書いてあるといった状況でした。この文字列と図がつながるといいなということで、ちょうど広まってきたウェブ上で実現しようということになりました」と振り返る。

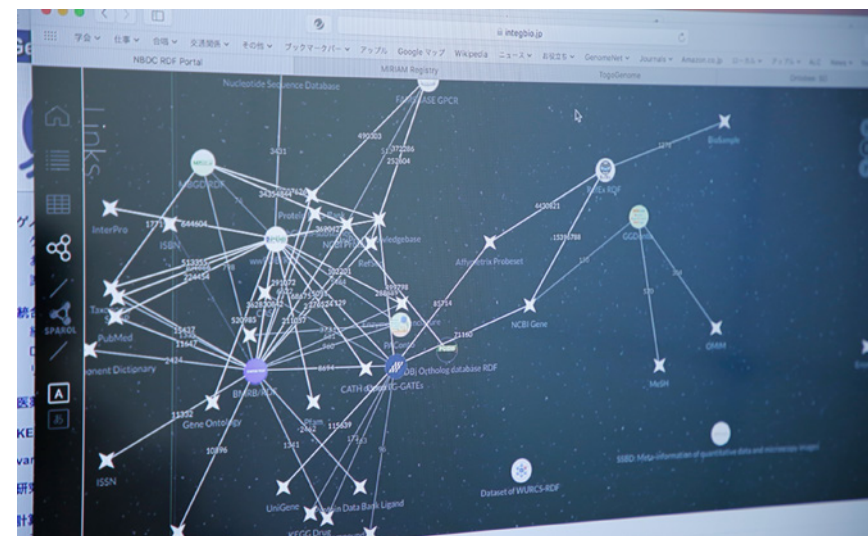


遺伝子とその機能の関連を他の生物種と比較する

KEGGは、ゲノムや遺伝子と、それがどう機能するか——つまり代謝経路やシグナル伝達の経路等の「パスウェイ」を結び付けるデータベースだ。「ヒトゲノムを含めて何干、何万種類という生物種のゲノムが決まってくると、その中にどんな遺伝子があるかが分かりますね。その遺伝子が例えば代謝系の中のどこで働いているかという流れ図を、KEGGでは「パスウェイマップ」で表示します」。生物種ごとの遺伝子と機能・反応プロセスの結びつきが分かるだけでなく、それぞれの機能を、他の生物種はどのように持っているかも見ることが出来るという。「海洋や土壌等の環境から採ったメタゲノム情報も集めているので、生物種間の関係を、機能の相互的な関係から見るといった新しい利用方法にも期待しています」。

データベースとデータベースを横断的に利用する

現在日本国内には、KEGGのパスウェイデータの他にも東北大学のヒトゲノム、かずさDNA研究所の植物ゲノムを始め、さまざまなデータベースが公開されている。そこでDBCLSでは微生物、真核生物、霊長類、鳥類、魚類、植物など数多の



データベースを横断的かつシームレスに利用できるようデータの標準化と統合化を進めている。「データをRDFと呼ばれる形式にして、セマンティック・ウェブの技術に基づいたデータの統合をするための仕組みを構築しています。この形式では、例えば遺伝子Aと疾患Bは関係があるというだけでなく、どんなタイプの関係があるのか、AとBのつながり方も指定して検索できるので、探しているものが発見しやすく、その後の解析・解釈も効率化します」。医療や生物の研究者が検索するだけでなく、開発用のインターフェースを提供してバイオインフォマティクスが自分でアプリケーションを作って解析したり、データサイエンティストがデータベースそのものから情報を取り出して大量のデータから知識発見したりと、利用階層に合わせて活用できるよう公開されているという。

統合化・標準化が進む

ゲノム医療のためのデータ群

DBCLSでは、ライフサイエンスデータの統合化と標準化に関わる具体的な機会として、合宿形式の国際会議「BioHackathon」を年1回開催している。10年にわたりこの会議をとりまとめる片山俊明特任助教によれば「近年特にヒトの病気の原因をゲノム変異から調べるといった、医療データベースの統合を推進している」という。また今回のBioHackathonでは、世界的にも新しい動きの1つ「ゲノムグラフ」も議論になりそうだ。「ヒトゲノムは個体ごとにほんの少しずつ違いがあるので、これまではその違いを別途管理し、後で参照できるようにしていました。これに代わって、膨大な数のゲノムを束ねて、違いも一括して扱おうという手法がゲノムグラフです。束ねた中から日本人だけに注目する使い方もできるので効果が高められるはず」と片山助教。ゲノムグラフから治療や投薬など一連の過程を標準化する「パイプライン」の開発も重要だ。「ゲノムを全部束ねたグラフを使って、疾患や人種ごとのサブセットを同じパイプラインで解析できるようにすれば、みんなが共通の最新技術を利用できる」。このような作業過程を記述する言語「CWL(Common Workflow Language)」に関する開発も今回のBioHackathonのテーマだという。国際連携による標準化がいつそう進展しそうだ。

公開日：2017/09/11

10周年を迎えた 「BioHackathon」



DBCLSが提供するサービスのひとつ「TogoGenome」を解説する片山特任助教。BioHackathon 2017については「10周年なので、今までの活動を総括するため、シンポジウムの日程を2日間にしました」とのこと。また「今年大きく変わったのは、日本医療研究開発機構(AMED)を始め医療・創薬関係の参加者が増えていること。今までの10年間はデータベース統合の基礎技術に注力してきましたが、これからはゲノムと健康に関わる国際連携「Global Alliance for Genomics & Health」などの課題を踏まえ、いっそう国際的な標準化に関わっていく」という。9月9～10日、JSTサイエンスプラザ(東京都千代田区)にてシンポジウムを、11～16日ホテル大観(岩手県)にて合宿形式の会議を開催。

植物ゲノムは 暮らしのどこに 役に立つ？

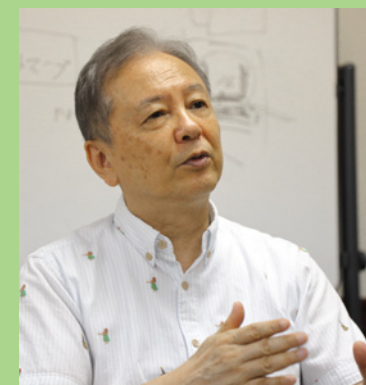
地球上の多様な生命活動の中で食物連鎖のてっぺんにいるのがヒトとすれば、植物は底辺を担う。しかも植物には光合成という重要な働きがあり、植物なしには、その上にいる動物もみんな生き延びることができない。光合成過程で重要な役割を担っているのが、シアノバクテリア(ラン藻)である。生物や地球環境の進化を考える上でも重要なこの生物は、1996年、世界で4番目に全ゲノムが明らかにされた。解読に成功したのは、日本の研究所——公益財団法人かずさDNA研究所である。千葉県にあるこの研究所は、特に植物ゲノムの研究やデータベースで知られ、シアノバクテリア解読以降も研究機関・大学との共同研究や産学連携によりシロイヌナズナ、トマト、ハクサイ、ユウカリ、食用イチゴ、カーネーション、ダイコン、サツマイモ、ソバ、ラッカセイ祖先種、イチジク、サクランボ「佐藤錦」等のゲノム解読を次々に共同発表してきた。野菜や果物が多く含まれていることからわかるように、植物のゲノムは農業や食物を通じて私たちの生活に直接関わるほか、地球温暖化による環境変化に適応するためにも欠かすことができない。今回は、そんな植物ゲノムの世界をのぞいてみることにしよう。

答える人

田畑哲之所長

[かずさDNA研究所]

かずさDNA研究所所長。1977年神戸大学理学部生物学科卒業、1979年京都大学大学院理学研究科博士前期課程修了、1983年理学博士(京都大学)。1982年米国カリフォルニア大学博士研究員、1985年京都大学化学研究所助手等を経て、1994年かずさDNA研究所主席研究員、2013年より現職。1996年のシアノバクテリア(ラン藻)ゲノム解読では、中心的役割を果たした。1997年東京テクノフォーラム21ゴールドメダル賞、2001年日本植物生理学会特別賞、同年Kumho Science International Award等を受賞。

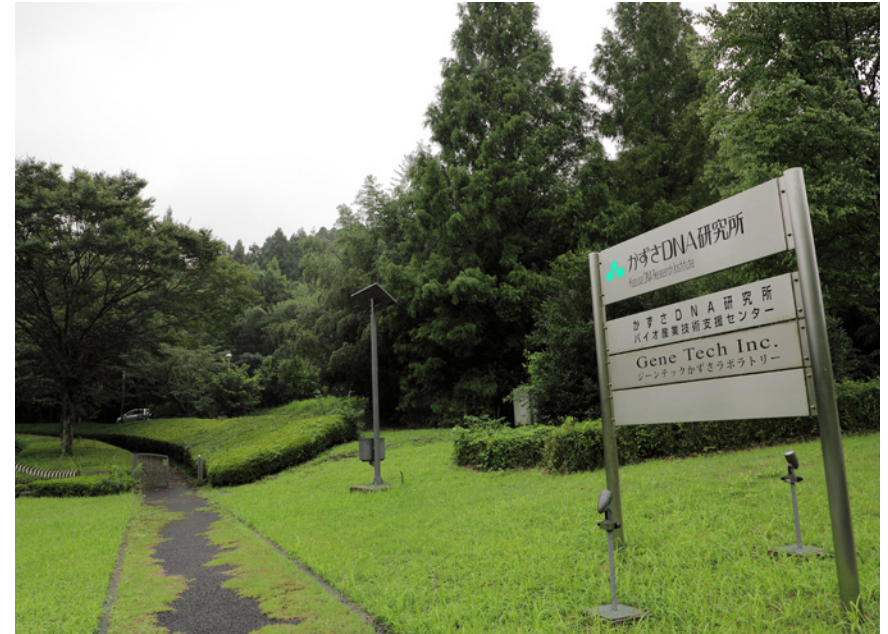


植物ゲノムは長くて遺伝子が詰まっている

総じて植物ゲノムは動物に比べて長く、遺伝子も多いのだそうだ。「ゲノムには遺伝子を成すひと続きの部分と、遺伝子と遺伝子の間をつないでいるだけでタンパク質の合成にあずからない部分とがありますが、植物ゲノムは遺伝子の部分が多く、全体的にも長い特徴があります。動物と違って動くことができない植物は、特に環境耐性に関する遺伝子をたくさん持っているんですね。一方動物は、どちらかと言うと遺伝子が少なめで、つなぎ方を変えて変化に対応する傾向があり、戦略も対照的」と言うのは、かずさDNA研究所の田畑哲之所長だ。さらに「倍数性といって、人を含め動物の多くは母親と父親の2系統から遺伝子が伝えられる2倍体ですが、例えば植物である小麦は6倍体で、A・B・Dという似ているけれども少しずつ異なる3種類のゲノムを2セット併せ持っています。私たちが食べるパンもこの6倍体の小麦でできているんですね」。しかし植物だと、なぜゲノムが増えてしまうのだろうか？「細胞分裂時等に起こる何らかの異変が原因でしょうが、動物に起こると、生き延びることができません。しかし植物は比較的それを許容する。また人の影響も大きくて、雑種強勢といって雑種になると勢いが強くなる現象があるのですが、ゲノムの倍加によっても植物体が大きくなることが知られています。このような「変異種」が人為的に選抜され、作物として現在まで生き残ってきたと考えられます」。ちなみにサツマイモは6倍体、イチゴは8倍体と長く、また、より複雑だ。「ゲノムを読み取ってもなかなか区別できないので、作物のゲノム解析は実はすごく難しいんです」。

植物のゲノミクスを実際の畑に活かす

田畑所長が植物ゲノムに取り組んでいるのには、いくつかの理由がある。「作物を使った育種技術を誰がやるのか」という問題意識もその一つだ。「植物の研究はこれまでどちらかというと実験室の中で、実験材料として確立された「モデル植物」を使って行われ、栽培の現場とは結び付いていませんでした。精密にコントロールされた環境と、いつ雨が降るやら病気が来るやら分からない環境を、つなぎようがなかったんですね。しかしフィールドで取ったデータとゲノミクスを結



かずさDNA研究所は、緑豊かな千葉県木更津市のかずさアカデミアパーク内にある。

び付けることに、われわれは何とか挑戦しなければいけない。そこで、そもそもフィールドを科学的にどう扱うかについての方法論を作り、次にその解析の手法を作る。さらにその結果として植物にどんな変化が生じたかを評価する、という3つの課題を立てて、研究を進めています」。この挑戦は現在CREST研究プログラム「環境変動に対する植物の頑健性の解明と応用に向けた基盤技術の創出」の中で推進されているそうだ。「ヒトゲノムが治療や創薬へ活かすステージに入っているように、植物もやはり社会実装のほうへ向かわなければならない。そこにある大きな困難を何とかクリアするのが、研究所の課題でもあるし、日本の課題でもあると考えています。特に近年は環境変動が激しく、東南アジアの害虫や病気が、九州周辺にまで北上してきていることがあります。新しい抵抗力を備えた品種を育てるためには、すぐにも育種スピードを速めなければなりません」。



研究所では植物ゲノムの他、個別化医療に関連して稀少性疾患のゲノム解析の受け入れも行っているという。画面背景に見えるのは千葉県の公式キャラクター「チーバくん」。

ゲノムは今後ますます社会に活かされるべき

このような問題に立ち向かえるようになったのには、ゲノミクスという科学そのものの発展が大きく寄与している。学生時代、生物学を専攻していた田畑所長は「いつになったら生物全体が分かるだろう、とずっと考えていた」と振り返る。「当時は個々の現象や個々の遺伝子に関する理解を1つ1つ積み上げるのが生物学でした。ゲノミクスが登場して、まず網羅的なリストを作り、そのリストから全体の理解へ向かうような新しい研究分野をもたらしました。それがデータ駆動型サイエンスとしての今日のゲノミクスだと言えるでしょう」。例えば、タンパク質の相互関係や、代謝によって生成される物質、その化学変化といったものを全ゲノム総当たりに調べる。かずさDNA研究所では、このような解析に用いられる「オーミクス解析」を備えたデータベースの拡充も行っているという。ただ難しいのは、仮にデータがうまく合っても、それがそのまま生物学の答えではないという点だ。「ゲノムデータは他の科学とも連携できるような共通のプラットホー

ムではあるのですが、仮にデータに同じ部分があっても、環境も違えば、他の遺伝子も違うので、異なる機能を発現するかもしれません。ゲノムという基盤の上に、生物学がどんな理解を積み増せるのかが勝負」と田畑所長は言う。このような観点から、田畑所長は、ゲノム解析で生命科学の進展を支援する「ゲノム支援(2010～2014年度)」の共同研究にも参加してきた。植物ゲノムの成果が、いよいよ実際の社会に活かされるステージへ向けて、科学者の挑戦が続いている。

公開日：2017/10/10



植物約250万種を検索 「ナショナルバイオリソースプロジェクト」

シロイヌナズナ、イネ、コムギ、オオムギ、広義キク属、アサガオ、ミヤコグサ・ダイズ、トマト、藻類等の植物は、国が戦略的に整備することが重要なものとして、「ナショナルバイオリソースプロジェクト(NBRP)」(Science Report 009参照)に登録されている。研究開発に不可欠な「モデル生物」であるこれらの植物



は、その収集・保存・提供が行われているほか、植物約250万種(2017年10月10日現在)の情報はウェブ上でも検索できる。画面は、イネのデータベース。

イネ(稲) データベース:OryzaBase(代表機関:国立遺伝学研究所)
<https://shigen.nig.ac.jp/rice/oryzabase/>

チンパンジーゲノム から 何がわかったか？

2017年10月30日、京都大学において、松沢哲郎 京都大学高等研究院副院長・特別教授、郷康広 自然科学研究機構新分野創成センター特任准教授、藤山秋佐夫 情報・システム研究機構国立遺伝学研究所特任教授が出席して記者発表が行われた。京都大学霊長類研究所のチンパンジーであるアキラ(父)、アイ(母)、アユム(息子) のゲノムを大規模に解析して、親子1世代で起こるゲノム変異を明らかにしたという研究報告だ。約30億塩基対あるチンパンジーのゲノムのうち、1億塩基対あたり平均1.48個の「新規1塩基突然変異」が生じていたというのだが……ではそこから、何がわかるのだろうか？

答える人

松沢哲郎 特別教授

[京都大学高等研究院]

京都大学高等研究院・特別教授、京都大学霊長類研究所・兼任教授、理学博士(京都大学)。1974年、京都大学文学部哲学科卒業。1977年11月から「アイ・プロジェクト」とよばれるチンパンジーの心の研究を始め、野生チンパンジーの生態調査も行う。チンパンジーの研究を通じて人間の心や行動の進化的起源を探り、「比較認知科学」とよばれる新しい研究領域を開拓した。



藤山秋佐夫 特任教授

[国立遺伝学研究所]

情報・システム研究機構データサイエンス共同利用基盤施設施設長、国立遺伝学研究所特任教授、理学博士(名古屋大学)。大阪大学医学部、大阪大学細胞工学センター、コールドスプリングハーバー研究所、シカゴ大学を経て、1989年国立遺伝学研究所助教授、2002年国立情報学研究所教授、2008年国立遺伝学研究所教授を経て、2016年より施設長。専門はゲノム生物学。ヒトゲノム、チンパンジーゲノムの比較解析等で知られる。



なぜチンパンジーゲノムなのか？

今回のゲノム解析の源となるチンパンジーの体毛や血液は、愛知県犬山市にある京都大学霊長類研究所で採取されたものだ。霊長研は、世界的にも著名な霊長類学の総合的研究拠点であり、松沢哲郎特別教授は1977年、ここで暮らす「アイ」という名のチンパンジーに出会い、以来40年にわたって「チンパンジーの心」を研究してきた。2000年4月24日、そのアイが23歳(推定)で出産したのが、息子のアユムである。松沢哲郎著『おかあさんになったアイ(講談社学術文庫)』には、出産直後のアユムの「手足がだらりと垂れ下がって」いたため心配された様子等が詳しく記されている。「僕らは学習行動を通じて、チンパンジー達の中に、生まれてから蓄積していく知識のような部分を研究してきました。氷山にたとえれば、海上に浮いている僕らの研究に対して、海面下にある見えない部分を大きく担っているのがゲノムです。なかでもゲノムの中の遺伝子の働きが詳細に分かってくると、もっとはっきりと氷山全体が見えてくるのではないかという期待があります」と松沢教授は言う。「たとえば攻撃性が高いという性格はアンドロゲンという男性ホルモンと関わっていますが、アンドロゲン受容体の遺伝子は既に特定されています。では今度は攻撃性を抑制する遺伝子はどこか……というふうにして、今後どんどん研究が進んでいくものと考えています」。



松沢哲郎教授(右)と藤山秋佐夫教授(左)、京都大学にて。



ヒトとチンパンジーのゲノムを比較する

一方、藤山秋佐夫特任教授は、ヒトゲノム計画に参加していた2003年頃から、チンパンジーゲノムに注目してきた。「進化的に最もヒトに近いチンパンジーを調べ、ヒトゲノムと比較することで、共通の祖先からそれぞれが進化するために重要な役割を果たした遺伝情報を明らかにすることができます。またチンパンジーとヒトではかかる感染症もほとんど共通であることなどから、チンパンジーゲノムの解明が、ヒト疾患の新たな治療法の開発にも役立つと期待されています」。藤山教授らはすでにヒトゲノムの21番染色体と、これに対応するチンパンジーのゲノムの22番染色体を決定・比較する成果等を挙げてきた。「2004年の研究では、共通の祖先からヒトとチンパンジーが分岐したのが600万年前、1世代の長さを20年と仮定してシミュレーションを行い、突然変異の確率を計算しました」と藤山教授は言う。2004年のある日、その共同研究のために藤山教授が犬山を訪ねた日のことを、松沢教授はよく覚えているという。「100人来て99人までアイちゃんはどれですかって聞くんだけど、藤山先生だけ、雄のチンパンジーの「ゴンはどうですか?」(笑)——2004年の段階で、すでにチンパンジーゲノムについて道をつけていた」と振り返る。



親子トリオのゲノム解析で何がわかる？

今回の成果はまず、ヒトゲノムではすでに進められている1個体ごとの遺伝情報「パーソナルゲノム」を、チンパンジーを対象として「親子トリオ」と呼ばれる父・母・子のセットで解析している点で注目に値する。そしてヒトを含めたパーソナルゲノム研究としても未踏の4500～5700億塩基対もの大規模データを使い、1世代にかかる実時間を計算に入れた高精度な解析であることが特徴だ。これにより「進化の駆動力である1世代に生じる新規1塩基突然変異は1億塩基対あたり平均1.48個で、これまでにヒトで報告された値(0.96～1.2個)より高く、その75%が父親(精子)由来であることが明らかになった」と郷康広特任准教授は説明する。つまり、親から子へ遺伝情報が受け継がれる際に、どんなエラーがどれだけ起こり、ゲノムが変化したのかが詳しく確かめられたのである。「親子トリオから一歩進んで、3世代や、兄弟のゲノム等が調べられれば、より詳しいことがわかってくるはず」と藤山教授は言う。

賢馬ハンス100年のギャップを越えて

ところで、霊長類をもう少し広く眺めると、その周りにはさまざまな哺乳類がいる。松沢教授は近年「馬」を対象として、新たな氷山に挑みつつあるという。「哺乳類は6,600万年前の恐竜が死に絶えた世界に、急速に適応・放散しました。哺乳類の共通祖先は夜行性の小型のネズミに似た生き物だったと言われます。他の多くの哺乳類もこれと同じように地上性であり、このように地上に満ち満ちた哺乳類を考える時、ピットフォールのように、誰も馬の研究をしていない」。一方、多くの哺乳類に反して、霊長類は樹上という道を選んだ。「チンパンジーを見てもニホンザルを見ても、四肢はすべて手のかたちをしています。四つ足動物が木に登って、4つの手になった。さらにヒトは、樹上から地上へと戻った非常にユニークな哺乳類であり、人間になって初めて足ができたと考えらることで、6,600万年前から今までの霊長類の歴史が、ひもとけるようになる。ここでもゲノムは、旅の友なんです」と、松沢教授は言う。チンパンジーの時と同様に馬においても、実験と野外の研究を並行して行う計画だ。

公開日：2017/11/16



馬の研究が進まなかった原因の1つは、クレバー・ハンスという馬の影響だと松沢教授は見る。「19世紀末から20世紀初頭にかけて、計算問題などに答える馬が登場して、その後トリックだったことが判明したんです。このため、その後100年にわたって研究が止まってしまった。たくさんの方が馬に関心を持っているのに、ギャンブルや乗馬など命令することにばかり使って、馬が何を考えているのかが明らかになっていない」。





大学共同利用機関法人

情報・システム研究機構について



平成16年、すでに大学共同利用機関として活動していた国立極地研究所、国立情報学研究所、統計数理研究所、国立遺伝学研究所の4研究所が結集し、大学共同利用機関法人情報・システム研究機構が誕生しました。全国の大学等の研究者コミュニティと連携して、極域科学、情報学、統計数理、遺伝学についての国際水準の総合研究を推進する中核的研究機関を担うとともに、21世紀の重要な課題である生命、地球、人間、社会など複雑な現象に関する問題を、情報とシステムという視点から捉え直すことによって、新たな研究パラダイムの構築及び新分野の開拓を目指しています。

平成28年度、当機構は4研究所に「横串」を貫く組織改革を行い、「データサイエンス共同利用基盤施設」を設置しました。これにより、データ共有支援、データ解析支援、データサイエンティスト育成の取り組みを一層強化し、社会のイノベーションにつながるデータ駆動型科学の発展を推進しています。そして研究者コミュニティの要請に応える共同利用・共同研究により、大学等における研究の発展に貢献するとともに、産業界との連携や、市民が参加するオープンサイエンスも進めています。また総合研究大学院大学の基盤機関として、もうひとつの重要な使命である人材育成にも取り組んでいます。

情報・システム研究機構は、各研究所の学理の追究に基づき、データサイエンス時代の新しい研究パラダイム構築を通じて、現代の課題解決や超スマート社会構築等の社会の要請に応えてまいります。皆様の一層のご支援、ご協力を心よりお願いいたします。

機構長 藤井良一
2018年3月

『情報・システム研究機構ブックレット』について

本機構は、新たな研究パラダイムの構築と新分野の開拓を推進し、
また大学共同利用機関法人として、大学等の研究の発展に貢献しています。

『情報・システム研究機構ブックレット』は、
その研究と貢献をわかりやすく紹介していく、シリーズ小冊子です。

情報・システム研究機構ブックレット2

SCIENCE REPORT 007-012

僕らはゲノムでできている。

著 者 大学共同利用機関法人 情報・システム研究機構

監 修 藤山秋佐夫(情報・システム研究機構
データサイエンス共同利用基盤施設 施設長)

取材・文 池谷瑠絵(情報・システム研究機構 URA広報)

写 真 飯島雄二(Science Report007-012)、黒川顕(p8)

ブックデザイン hata design

発 行 大学共同利用機関法人 情報・システム研究機構

〒105-0001 東京都港区虎ノ門4丁目3番13号

ヒューリック神谷町ビル2階

TEL：03-6402-6200 FAX：03-3431-3070

<http://www.rois.ac.jp/>

発行日 2018年3月26日

©Inter-University Research Institute Corporation

Research Organization of Information and Systems

ISBN978-4-909638-07-6 C0345

Printed in Japan