

本件の取り扱いについては、下記の解禁時間以降でお願い申し上げます。

2021年10月19日

TV・ラジオ・WEB … 日本時間 2021年10月22日(金)午前3時  
 新聞 … 日本時間 2021年10月22日(金)朝刊

## メタゲノムからウイルスゲノムを洗い出せ！ ～細胞の免疫記憶を使ったウイルスゲノムの網羅的検出～

### ■ 概要

ウイルス<sup>(1)</sup>の進化と起源を知るには地球上に存在する膨大かつ多様なウイルスゲノム<sup>(2)</sup>を網羅的に収集する必要があります。近年「メタゲノム<sup>(3)</sup>解析」を利用したウイルスゲノムの「発掘」が活発におこなわれています。メタゲノムには細胞とウイルスの遺伝情報<sup>(4)</sup>が混在していて、ここからウイルスゲノムを洗い出す必要があります。しかしながら、今までの方法では既知のウイルスに類似するウイルスにのみ注目しており、ウイルス全体の多様性を捉えるには不十分でした。

情報・システム研究機構 国立遺伝学研究所の杉本竜太研究員、井ノ上逸朗教授と東京大学、理化学研究所、佐々木研究所の共同研究グループは、原核生物の「免疫記憶(CRISPR)」を利用してメタゲノムデータからウイルスのゲノム情報を検出する方法を開発しました(図1)。この方法は、既知のウイルスのゲノム情報に依存しないでメタゲノムからウイルス由来と思われる配列を網羅的に検出し、さらにその宿主の推定を可能にします。この方法によってヒト腸内メタゲノムから未知の配列を含む多様なウイルスのゲノム情報を検出することに成功しました。

今後は広範なサンプルに本方法を適用し、地球上に存在するウイルスの網羅的データベースを構築することでウイルスの進化と起源を知るための基盤を構築していきます。

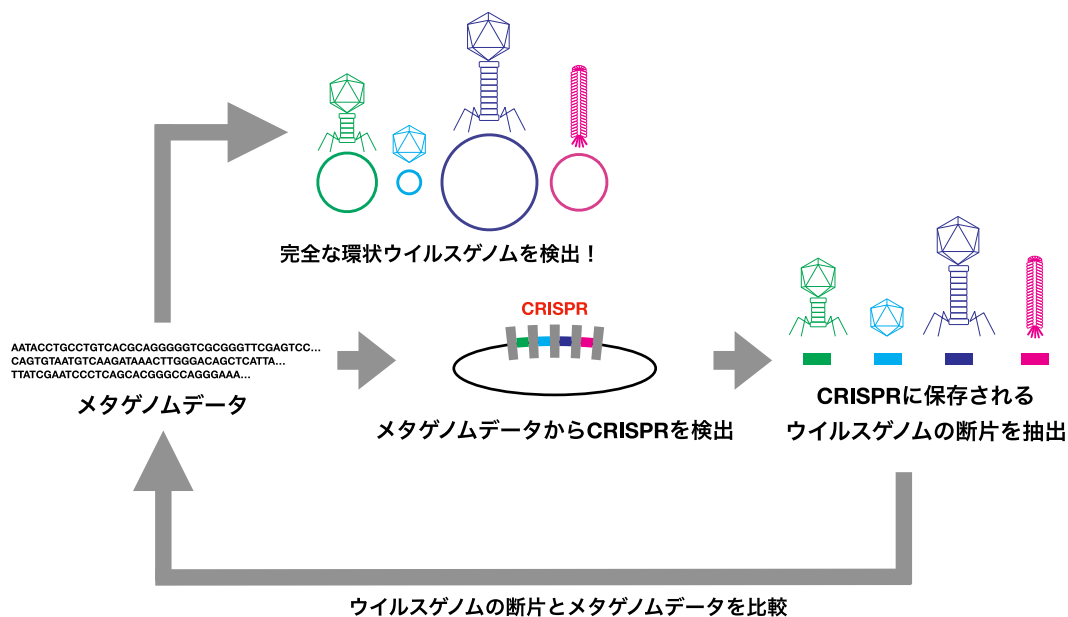


図1: CRISPRを使ったウイルスゲノムの検出法

## ■ 成果掲載誌

本研究成果は、米国科学雑誌「PLOS Computational Biology」に 2021 年 10 月 22 日(日本時間)に掲載されます。

論文タイトル: Comprehensive discovery of CRISPR-targeted terminally redundant sequences in the human gut metagenome: viruses, plasmids, and more

(CRISPR によってターゲットされるヒト腸内メタゲノム配列の網羅的検出: ウイルス、プラスミド及び未知のエLEMENT)

著者: R. Sugimoto, L. Nishimura, P. T. Nguyen, J. Ito, N. F. Parrish, H. Mori, K. Kurokawa, H. Nakaoka, I. Inoue

(杉本竜太, 西村瑠佳, Phuong Thanh Nguyen, 伊東潤平, Nicholas F. Parrish, 森宙史, 黒川顕, 中岡博史, 井ノ上逸朗)

## ■ 研究の詳細

### ● 研究の背景

ウイルスは地球上で最も多様で大量に存在しています。また、ウイルスはあらゆる環境に存在し、近辺の特定の細胞に感染して増殖します。ウイルスは身近なありふれた存在でありながら、その進化と起源についてはよくわかっていません。ウイルスは感染する細胞と共に長い時間をかけて進化してきたはずですが、不思議なことにウイルスの持つ遺伝子の大部分は我々の細胞の持つ遺伝子とは全く異なっています。ウイルスの起源を解き明かすことは、ウイルスの起源だけでなく、生命の起源とウイルスの関わりを解き明かす上での重要な意義があります。ウイルスの進化と起源を知るにはそのゲノム情報が必要になります。なぜなら進化や起源を推定するには異なる種間のゲノム情報を比較することが必要だからです。しかしながら、現在知られているウイルスのゲノムデータは、ヒトやモデル生物、家畜や作物に感染する病原性ウイルスのものに偏っていて、ウイルス全体の進化と起源を知るには不十分です。

近年では「メタゲノム解析」と呼ばれる手法がウイルスゲノム情報の収集に利用されるようになりました。例えばヒトの排泄物から DNA を取り出し、その DNA を次世代シーケンサー<sup>(5)</sup>で解析することで、そこに存在する細胞とウイルスの「全ての」遺伝情報を読み取ることができます(図2)。しかし、メタゲノムデータには細胞とウイルスの遺伝情報が混在しているので、メタゲノムデータからウイルスのゲノム情報を洗い出す必要があります。今までは既知のウイルスとの類似性に基づいてウイルスゲノム情報をメタゲノムデータから抽出していましたが、この方法では多岐にわたる新規ウイルスのゲノム情報を網羅的に抽出することは困難でした。



図 2: メタゲノム解析の概要

## ● 本研究の成果

本研究グループは、原核生物の「免疫記憶 (CRISPR<sup>(6)</sup>)」を使った、既知のウイルスゲノムデータに依存しないウイルスゲノム情報の抽出法を開発しました。

CRISPR には過去に感染したウイルスゲノムの断片が保存されています。このウイルスゲノム断片をタグとして使って検出すれば、メタゲノムデータからウイルスゲノム情報を抽出することができます。具体的には、メタゲノムデータから原核生物のゲノム情報を再構築し、そこから CRISPR に保存されているウイルスゲノム断片配列を収集します。収集したウイルスゲノム断片配列を再びメタゲノムデータと比較することで、これらの細胞に感染しているウイルスゲノムを網羅的に同定するのです(図1)。

ヒトの糞便メタゲノムデータにこの方法を適用したところ、CRISPR に標的とされるおよそ一万のゲノム配列を抽出することに成功しました。このうちおよそ三千のゲノムはウイルスの殻タンパク質の遺伝子を持っており、実際ウイルスであると考えられます。現在ゲノムデータベースにはおよそ四千の原核生物に感染するウイルスゲノムが登録されています。これらの既知のウイルスゲノム情報を全く使わずに、我々はスクラッチの状態からほぼ同数近いウイルスゲノムをヒトの腸内メタゲノムから検出できたのです。また、検出されたウイルスゲノムには既知のウイルスの系統とは異なる新規のウイルス系統も含まれていました(図3)。さらに抽出されたウイルスゲノムとそれを標的にする CRISPR の情報から、70%に及ぶウイルスの宿主を「門」のレベルで推定することにも成功しました。

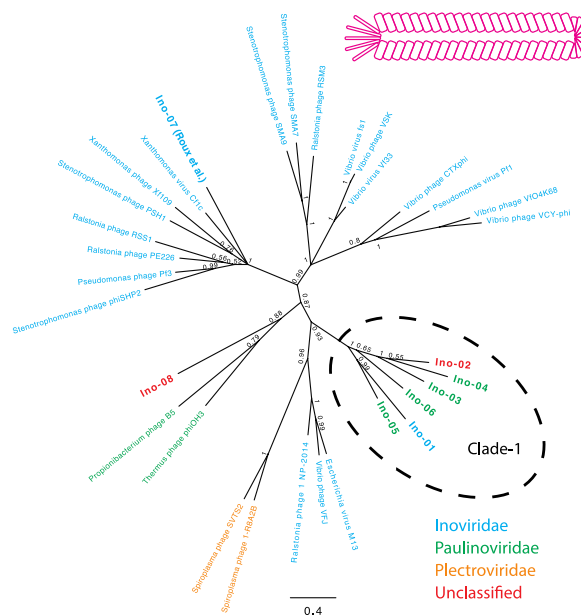


図 3: 検出された一部のウイルスの系統樹  
新規ウイルスの系統(Clade-1)を検出した。

以上の結果は本方法が既知のウイルスと近縁ではないウイルス種のゲノム情報も抽出でき、さらにそのウイルスの宿主も推定できることを示しました。また、抽出されたゲノム情報には既存のいかなるウイルス種にも含まれない全く未知のゲノム情報も含まれていたことから、既知のウイルスとは全く異なる新たな感染因子が存在することが示唆されました。

## ● 今後の期待

今後はこの方法を海水や土壌、熱水など、より広範なメタゲノムデータに適用して、地球上に存在するあら

ゆるウイルスゲノム情報を網羅的に検出収集することを目指します。特に古細菌ウイルスや RNA ウィルスなど、未解明の部分の非常に多いウイルスについて重点的に調査します。収集したウイルスゲノム情報を用いてウイルスの進化と起源にせまり、ウイルスと生命との関わりをも明らかにしていきます。

## ■ 用語解説

### (1) ウィルス

タンパク質の殻に覆われた遺伝物質で、細胞に感染して自己複製を行う存在。

### (2) ゲノム

ある細胞ないしウイルスの持つ全ての遺伝情報。

### (3) メタゲノム

ある環境中に存在するウイルス及び細胞全ての遺伝情報の集合。

### (4) 遺伝情報

DNA か RNA を担体として保存される情報で、特に次の世代に伝達する情報。

### (5) 次世代シーケンサー

DNA を読み取ってその遺伝情報を文字列として出力する機械。

### (6) CRISPR

Clustered Regularly Interspaced Short Palindromic Repeats の略で、原核生物の染色体の特定の部分をいう。CRISPR には細胞に過去に感染したウイルスゲノムの断片が保存されている。これを転写して特定のタンパクと複合体を作ることによって次に感染してきた同一種のウイルスを特異的に制限する獲得免疫記憶としての役割を果たす。

## ■ 研究体制と支援

本研究は、国立遺伝学研究所 人類遺伝研究室、東京大学、理化学研究所、国立遺伝学研究所 ゲノム多様性研究室、ゲノム進化研究室、佐々木研究所の共同研究グループにより遂行されました。

本研究は、情報システム研究機構の未来投資型プロジェクト及び、日本学術振興会の科研費挑戦的研究(萌芽)(20K21405)の支援を受けておこなわれました。

## ■ 問い合わせ先

<研究に関すること>

- 国立遺伝学研究所 人類遺伝研究室  
教授 井ノ上 逸朗 (いのうえ いつろう)

<報道担当>

- 国立遺伝学研究所 リサーチ・アドミニストレーター室 広報チーム

※時節柄、Zoom 会議での取材にも対応できますので、Zoom 会議をご希望の場合には、その旨お知らせください。