

# ゲノムデータから読み解くウイルス感染拡大の実態

## —どこからどこへ伝播したか？—

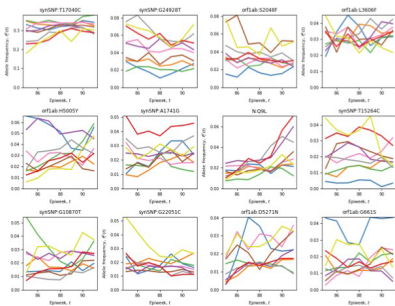
### 概要

COVID-19 のパンデミックは、病原体がどのように地域を越えて広がっていくのかという「地域間伝播」の重要性を再認識させました。従来は、人の移動データや接触履歴に基づいて伝播を推定してきましたが、社会的・地理的に離れた地域間で起きるような稀な伝播イベントを捉えることは困難でした。

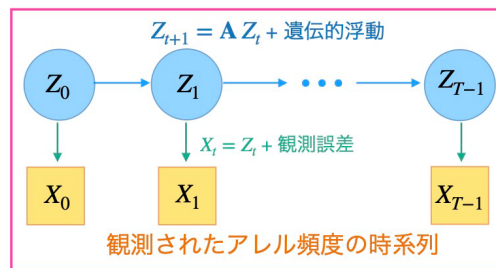
京都大学医生物学研究所 岡田崇 准教授、米国カリフォルニア大学バークレー校 Giulio Isacchini 研究員、米国ハーバード大学 T.H.チャン公衆衛生大学院 QinQin Yu 研究員、米国カリフォルニア大学バークレー校 Oskar Hallatschek 教授の研究グループは、SARS-CoV-2 の豊富なゲノムデータに着目し、各地域のアレル頻度の時系列データから感染伝播の流入を直接推定する数理手法を開発しました。観測誤差やアレル頻度の揺らぎも考慮するために隠れマルコフモデルを利用しています。この手法を英国および米国の SARS-CoV-2 データに適用した結果、変異株ごとに地域間の感染ネットワークがどのように変化するかに加えて、感染伝播がどの程度の時間スケールで起きているか、また、その速度が地域間でどのように異なるかといった動態も明らかになりました。本研究は、ゲノムデータに基づく新たな時系列解析の道を拓くものであり、今後の疫学モニタリングや病原体の拡散予測モデルへの応用だけでなく、さまざまなゲノムデータへの応用が期待されます。

本研究成果は、2025 年 11 月 26 日に米国の国際学術誌「*Proc. Natl. Acad. Sci. U.S.A.*」にオンライン掲載されました。

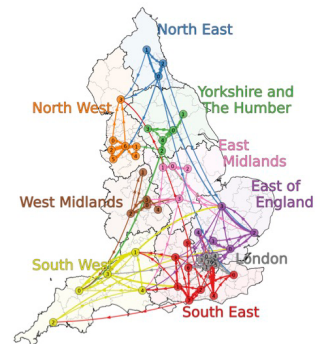
### アレル頻度の時空間データ



### 地域間感染率の推定



### 推定された感染ネットワーク



ウイルスのゲノムデータから地域間感染動態を推定する。

## 1. 背景

パンデミック期の COVID-19 は、病原体がどのようにして地域を越えて広がっていくのか、その地域間伝播（注1）の実態を正確に把握することの重要性を浮き彫りにしました。この問題に対する従来のアプローチでは、人の移動履歴、携帯電話の位置情報、接触調査など、感染の動きを間接的に推測するデータに依存せざるを得ませんでした。これらのデータはおおまかな人流を把握するには有用な一方で、社会的・地理的に離れたコミュニティ間で生じるような稀で予測しにくい伝播イベントを捉えるには限界があります。また、実際のウイルス伝播は「移動」だけでは説明しきれず、集団構造や人々の行動様式など、さまざまな要因が複雑に影響します。このため、間接データだけに頼った解析では、地域間伝播の全体像を定量的に把握することは困難でした。

一方、近年急速に発展したウイルスのゲノムシーケンシングによる監視体制は、こうした課題を克服する強力な可能性を秘めています。世界中で数百万規模のウイルスゲノムが収集されるようになったことで、ウイルスの変異頻度やその時間変化を高い解像度で追跡できるようになりました。特に、ある地域内における複数の変異の頻度がどのように増減するかを時系列で観察することで、その地域が他地域からどの程度ウイルスを「受け取っているか」という、伝播の流入の強さを直接的に推定できる可能性があります。本研究では、この膨大なゲノム頻度データに着目し、地域間の実際の感染伝播をゲノム情報から直接測定する新たな解析手法を開発しました。

## 2. 研究手法・成果

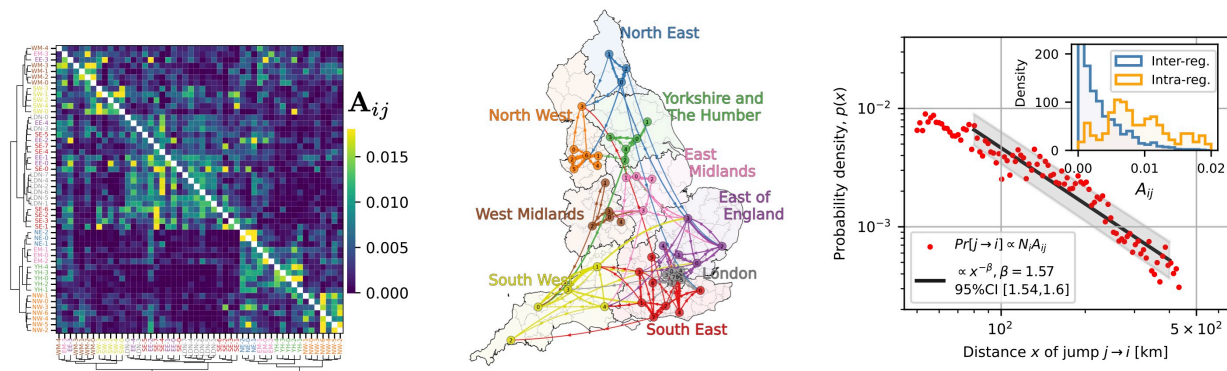


図 1. 英国デルタ株に対する解析結果。左図) 推定された地域間感染率の行列。中央図) 感染経路ネットワーク。右図) 地域間感染率の地理的距離に対する依存性。

本研究では、地域間の感染症伝播を定量的に把握するための新たな手法として、各地域における中立変異（注2）の頻度変化に着目した推定手法を開発しました。特に、各地域のアレル（注3）の頻度が時間とともにどのように推移し、他地域とどれだけ類似していくかを計測することで、感染伝播の流入の強さ、すなわち「他地域からどれだけウイルスが持ち込まれているか」を直接推定することを可能にした点が特徴です。この手法では、観測されるアレル頻度の変動を確率的に記述する「隠れマルコフモデル」（注4）を採用しており、揺らぎを含むデータから、背後にある伝播ネットワークの構造を正確に抽出できるよう設計されています。人流データなどの間接的データに頼るのではなく、ゲノム頻度の動態を通して「感染の痕跡」を捉えるというアプローチは、これまでのアプローチとは一線を画すものです。

この推定手法を用いて、イギリスおよびアメリカ合衆国の SARS-CoV-2 ゲノムデータを解析しました。その結果、地理的に近い地域同士で感染が頻繁に移動している傾向が確認された一方で、距離が離れている地域間でも明確な「長距離伝播」が存在することが明らかになりました（図 1）。これにより、従来の人流データに基づく予測では見逃されていた伝播経路が定量的に示されました。また、これらの流行波ごとに地域間の伝播ネットワークがどのように変化したかを定量的に捉えることに成功し、感染症のダイナミクスを時空間的に把握する新たな道を切り開きました。

### 3. 波及効果、今後の予定

本研究で開発された手法は、ウイルスのゲノム情報のみを用いて地域間の感染伝播を直接推定できるという点で、感染症疫学の新たな分析基盤となる可能性を秘めています。これにより、従来の移動データや人流モデルでは見落とされがちだった稀な伝播経路や、地理的距離を超えた結びつきを定量的に明らかにすることができます。

このような分析により、今後の感染症対策においても具体的な貢献が期待されます。たとえば、流行初期における「感染の持ち込み元」の特定や、ワクチン・治療薬の戦略的配布、変異株の早期検出といった分野で、より精度の高いリスク評価と資源配分が可能になります。感染症の伝播経路を地理的境界に限らず、年齢層や社会経済的背景などのコミュニティ単位でも応用できる点から、今後の拡張性も高いと考えられます。また、本手法は、人に感染するウイルスに限らず、微生物群集のメタゲノムデータなど、様々な遺伝的な時系列データへの応用も可能であり、さらなる応用に向けた取り組みを進めています。

### 4. 研究プロジェクトについて

本研究は、科学研究費助成事業（JP22K03453, JP22K06347, JP25K09723, JP25K07168）、理化学研究所数理創造研究センターの支援を受けて実施されました。

#### <用語解説>

- 注 1) 地域間伝播：感染症がある地域の集団から別の地域の集団へと広がる現象を指します。地理的・社会的に区切られた集団間でウイルスなどの病原体が移動する過程を表します。COVID-19 のようなパンデミックでは、この地域間の移動を正確に捉えることが感染拡大防止の鍵となります。
- 注 2) 中立変異：ウイルスや生物の遺伝子に生じる変異のうち、生存や増殖に有利・不利といった選択圧の影響を受けないものを指します。本研究では、この中立変異の頻度変化を手がかりに、地域間の感染伝播を推定する手法を構築しています。
- 注 3) アレル：ウイルスのゲノムにおける一つの塩基配列が変異すると、それが新しいアレルとして記録されます。集団内に注目するアレルが存在する割合を、アレル頻度と言います。
- 注 4) 隠れマルコフモデル：観測できない「隠れた状態」が時間とともに変化する過程を、確率的に記述する数理モデルです。本研究では、実際に観測されたゲノムのアレル頻度データから、偶然性による変動（遺伝的浮動）や測定誤差を考慮して感染伝播ネットワークを推定するために使用されています。

#### <研究者のコメント>

ゲノムデータは観測誤差やサンプルバイアスが無視できないため、プロジェクト当初は、「アレルの揺らぎが

ら地域のつながりが見えるなんて本当だろうか」と半信半疑でしたが、本手法がうまく働いたときは大きな喜びでした。本手法は、ウイルスだけでなく、さまざまなゲノムデータへの応用・展開があると期待しています。

**<論文タイトルと著者>**

タイトル：Uncovering heterogeneous intercommunity disease transmission from neutral allele frequency time series (中立アレル頻度の変動から解明する地域間感染動態)

著者：Takashi Okada, Giulio Isacchini, QinQin Yu, Oskar Hallatschek

掲載誌：*Proc. Natl. Acad. Sci. U.S.A* DOI：[doi.org/10.1073/pnas.2500663122](https://doi.org/10.1073/pnas.2500663122)