

# 研究報告書

## 「インフルエンザウイルスの遺伝子変異に内在する数学的構造の探求」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 伊藤 公人

### 1. 研究のねらい

インフルエンザウイルスの主要抗原はヘマグルチニン(HA)であり、感染およびワクチン接種によりHAに対する血中抗体の産生が誘導される。インフルエンザの予防にはワクチン接種が有効であるが、人の集団内で流行するうちに HA の抗原領域に変異を持つウイルスが選択されて流行するウイルスの抗原性が年々変化する。このため、流行株の抗原性に合わせてワクチン株を変更する必要がある。

WHO が推奨したワクチン株と実際にそのシーズンに流行した A 香港型 (H3N2) インフルエンザウイルスを比較すると、1997 年以降の 15 シーズンのうち 9 シーズンでワクチン株と流行株が一致しない。ワクチン株は、新しい変異株が流行する数ヶ月前に選定しなければならず、流行株を先回りして予測することが重要となる。

本研究では、インフルエンザのワクチンを先回りして準備するために、世界各国のサーベイランスで得られる大量の遺伝子データからウイルスの遺伝子変異に内在する数学的構造を発見し、将来の変異を予測することを目的とする。ウイルス学・計算機科学・数理科学の研究者が協働して、過去に流行したインフルエンザウイルスの遺伝子配列を大規模に解析し、ウイルス遺伝子の時間発展を説明する数理モデルおよびその解析手法を開発する。将来の変異ウイルスの遺伝子配列を実際に予測して、予測の精度および数理モデルの妥当性を評価する。具体的には、下記の3つの研究項目を達成する。

- (1) ウイルス遺伝子の次元圧縮写像の探索
- (2) ウイルス遺伝子の時間発展を説明する数理モデルの構築
- (3) インフルエンザウイルスの変異予測および予測精度の検証

これらの研究項目を通じ、インフルエンザウイルスの遺伝子変異に内在する数学的構造を探求し、過去・現在・未来における遺伝子変異のダイナミクスに迫る。

インフルエンザウイルスは、医学、疫学、獣医学等の分野で古くから広く研究されているが、将来の変異を高い精度で予測する手法は未だ存在しない。従来、生物の進化は現存する種の遺伝情報から過去に遡る方向で研究されてきた。しかし、ウイルスの変異予測では、過去から現在、そして未来に向う進化を研究しなければならない。ウイルス学・計算機科学・数学の研究者が一丸となって本課題に取り組むことにより、数学と生命科学の研究者の連携を促進すると同時に、数理科学分野における新たな方法論の開拓をねらう。

## 2. 研究成果

### (1) 概要

本研究では、インフルエンザウイルスの遺伝子変異を予測するために、ウイルス学・計算機科学・数理科学の研究者が協働し、過去に流行したインフルエンザウイルスの遺伝子配列を大規模に解析し、ウイルスの遺伝子変異に内在する数学的構造を探索した。

#### (1) ウイルス遺伝子の次元圧縮写像の探索

A 香港型ウイルス(H3N2)の流行におけるウイルス表面分子(ヘマグルチニン; HA)の継時的変化を把握するために、過去 40 年間に人から分離されたウイルスの HA アミノ酸配列を多次元尺度構成法(MDS)により解析し、HA の進化を表す三次元地図を構築した。その結果、H3N2 亜型ウイルスの HA アミノ酸配列の進化の軌跡は、MDS 空間上で一定の曲率をもつ曲線上の幹と連続的に現れる房から構成されることが判明した。この現象を盆栽様モデルと名付け、その数学的構造を解析した。

#### (2) ウイルス遺伝子の時間発展を説明する数理モデルの構築

MDS 空間における曲線的な進化は、同じ位置のアミノ酸が複数回置換していることを意味する。また、MDS 空間上で、ウイルスの遺伝子配列の時間発展が、一定の曲率を持つ曲線として観測されたことから、異なる年代の HA 間の相対的な距離に規則性があることが示唆された。HA アミノ酸配列の各位置における置換頻度がガンマ分布に従うと仮定すると、この相対的な距離をよく回帰できることが判明した。また、人の集団でのウイルス感染、免疫、および変異を表す非線形状態空間モデルを構築し、計算機シミュレーションで盆栽様進化が現れることを確認した。

#### (3) インフルエンザウイルスの変異予測および予測精度の検証

異なる年代の HA 間の相対的な距離から回帰したガンマ分布のパラメータを利用し、過去 10 年に溯ってそれぞれ翌年のアミノ酸置換を予測する試験を行った。1997 年から 2008 年の各年に対し、HA アミノ酸配列から翌年のアミノ酸置換を予測し、実際に起こったアミノ酸置換と予測結果を比較したところ、本手法は再現率=70%、適合率=45%で翌年のアミノ酸置換を予測することが判明した。この結果、サーベイランスで得られる HA の遺伝子情報から、翌年の抗原変異株が持つアミノ酸置換を比較的高い精度で予測できることが示唆された[PLoS one, 2011]。

### (2) 詳細

#### (1) ウイルス遺伝子の次元圧縮写像の探索

大量の遺伝子情報から抗原変異に伴うアミノ酸置換の規則性を探索するために、過去 40 年間に人から分離されたウイルスの HA アミノ酸配列を多次元尺度構成法(MDS)により解析した。MDS 法とは、対象間の相違度に基づいて対象物を空間上の点に配置する解析手法であり、対象間の相違度を空間上の点の間の距離として視覚化することができる。NCBI より、H3N2 亜型ウイルスの HA1 領域 の塩基配列 6,806 本を取得し、アミノ酸配列に翻訳した。同一配列を削除して得られたアミノ酸配列 2640 本を MDS 法により解析し、HA の進化を表す三次元地図を構築した結果、H3N2 亜型ウイルスの HA アミノ酸配列の進化の軌跡は、MDS 空間上で一定の曲率をもつ曲線上の幹と連続的に現れる房から構成されることが判明した(図 1)。

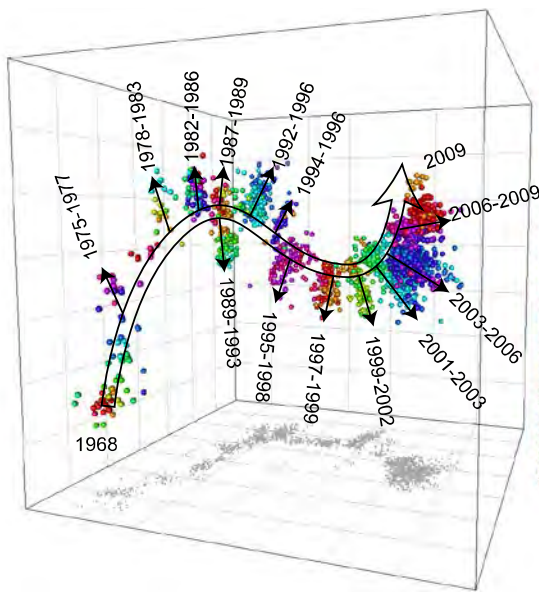


図 1. アミノ酸配列の時間発展  
各点はウイルス株を表し、分離年によって色分けしてある。ウイルス株同士の距離は HA 配列での異なるアミノ酸の数を表し、1 グリッドで 10 箇所のアミノ酸の違いを表す。三次元の樹形が一本の長い幹と連続した短い房から構成されることから、ある流行株から派生した変異ウイルスの中から、常に一種類のウイルス株が世界中でほぼ同時に選択淘汰されて流行を続けていることがわかる。

(2) ウイルス遺伝子の時間発展を説明する数理モデルの構築

MDS 空間における曲線的な進化は、同じの位置のアミノ酸が複数回置換していることを意味する。また、MDS 空間上で、ウイルスの遺伝子配列の時間発展が、一定の曲率を持つ曲線として観測されたことから、異なる年代の HA 間の相対的な距離に規則性があることが示唆された。A 香港型(H3N2) インフルエンザウイルスについて、1968 年のパンデミック時から現在までの HA のアミノ酸配列の変異を解析したところ、HA1 領域のアミノ酸置換の速度分布は、形状母数  $\alpha = 0.129$ 、平均置換速度  $\bar{r} = 0.0118$  のガンマ分布としてよく近似でき、分離年の差が  $t$  年である二つのウイルス株間におけるアミノ酸置換の数  $d$  は、

$$d = L \times (1 - (\alpha / (\alpha + \bar{r}t))^{\alpha}) \quad (1)$$

という  $t$  の関数でよく近似できることが判明した(図 2)。ここで、 $L$  は HA1 領域のアミノ酸の数 328 を表す。この性質は、HA1 領域の特定の アミノ酸がウイルスの抗原性に深く関与し、頻繁に置換することと合致する。

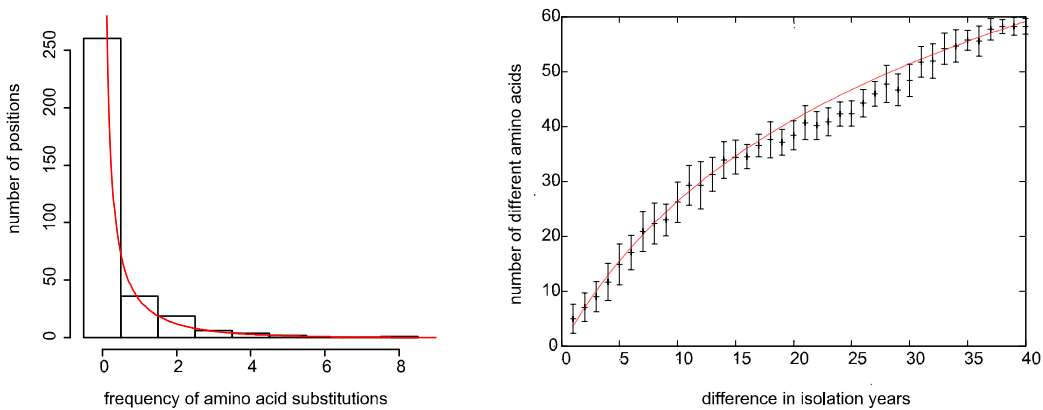


図 2. HA1 領域のアミノ酸置換の速度分布(左)とウイルス株間のアミノ酸置換数(右)

また、人の集団でのウイルス感染、免疫、および変異を表す非線形状態空間モデルを構築し、計算機シミュレーションにおいて盆栽様進化が現れることを確認した。

### (3) インフルエンザウイルスの変異予測および予測精度の検証

式(1)により、ある年のウイルスが他の年のウイルスと比べたときの異なるアミノ酸の個数を予測できる。サーベイランスで得られるウイルス株の中から、翌年についての予測値との二乗誤差が最小となる株を特定し、その株の持つアミノ酸置換を翌年の変異として提示する手法を考案した。本手法により翌年のアミノ酸置換を過去に溯って予測し、実際のアミノ酸置換を比較したところ、60%以上の再現率で翌年のアミノ酸置換を予測した。1997年から2008年の各年に対し、HAアミノ酸配列から翌年のアミノ酸置換を予測し、実際に起こったアミノ酸置換と予測結果を比較したところ、本手法は再現率=70%、適合率=45%で翌年のアミノ酸置換を予測することが判明した。この結果、サーベイランスで得られるHAの遺伝子情報から、翌年の抗原変異株が持つアミノ酸置換を比較的高い精度で予測できることが示唆された[PLoS one, 2011]。同手法を最新のデータに適用した予測結果を図3に示す。

Test Year	Predicted substitutions	Actual substitutions in the next year
2001-2002	<u>R50G</u> E83K <u>V106A</u> <u>D144N</u> V202I <u>W222R</u> <u>G225D</u> <u>V226I</u>	V106A D144N W222R G225D
2002-2003	L25I R50G H75Q E83K A131T H155T Q156H G186S V202I	L25I R50G H75Q E83K A131T H155T Q156H V202I
2003-2004	Q75H Y159F S189N S227P	K145N Y159F S189N <u>V226I</u> S227P
2004-2005	H156R P227S T248I K326T	S193F D225N
2005-2006	G50E K83E K173E L194P	
2006-2007	Q33R T48I G50E D85G K140I K207R	G50E K140I
2007-2008	L3F K83N L157S P169Q K173N K310R	K173Q
2008-2009	E62K L70I N144K K158N N189K I260M R261Q	E62K N144K K158N N189K
2009-2010	R150K K189R A272T C277S	K62E K144N T212A
2010-2011	N31D D53N Y94H S124N P221L I230V E280A	N145S A198S V223I N312S

図3. 予測と実際の変異。青と赤は的中を表し、下線は変異時期の外れた予測を表す。

### その他の研究成果

#### A. パンデミック(H1N1)2009 インフルエンザウイルスの表面分子の抗原構造予測

パンデミック(H1N1)2009 ウイルスとスペイン風邪の H1N1 ウイルスの HA の抗原構造を比較した。その結果、これら二つのウイルス HA の Sa および Sb 領域の構造は酷似していること、N結合型糖鎖付加モチーフの位置と数が一致していることを見いだした。これらの類似性に基づき、パンデミック(H1N1)2009 ウイルスの今後の変異を予測した[PLoS one, 2010]

#### B. 遺伝子配列のリサンプリングアルゴリズムの開発

DNA シークエンシング技術の発展により、世界各国で得られた病原微生物の遺伝子情報がインターネット上に公開されている。しかし、サーベイランスの度合いには、地域、年代、宿主の何れにおいても大きな偏りがあり、適切な標本抽出となっていないため、データ解析へのサンプリングバイアスの影響が懸念された。本研究では、塩基配列データからサンプリングバイアスを除去するリサンプリングアルゴリズムを提案した[PLoS one, 2013]。

### 3. 今後の展開

人の集団におけるウイルスの感染・免疫・変異を、非線形状態空間モデルで表すことにより、本研究の予測手法をデータ同化の枠組みとして一般化することができ、感染症研究分野に新たな統計的手法を導入するとともに、変異予測の精度向上が期待される。

### 4. 自己評価

本研究の主要成果は、計算機科学(伊藤)、ウイルス学(高田)、計算科学(五十嵐)、数理科学(宮崎)といった出身分野の異なる研究者の協働により達成された。感染症研究におけるこのような異分野連携の例は少なく、諸分野連携によってのみ達成できた成果であると考えられる。また、本研究テーマに関連して、「ウイルス」、「最新医学」、「応用数理」、「遺伝」、「感染・炎症・免疫」等の学術雑誌より依頼を受け、計5報の総説を執筆した。学術集会においては、「日本ウイルス学会」、「数理生物学会」、「分子生物学会」、「情報処理学会」、「電子情報通信学会」、「人工知能学会」で口演を行った。さらに、さきがけ「生命現象の革新モデルと展開」分野の研究者と交流を深め、1月には北海道大学において研究集会を催した。これらの共同研究、学会口演、研究集会、論文執筆を通し、諸分野の研究連携を大いに促進できたと思う。

### 5. 研究総括の見解

ウイルスの変異予測という困難な問題に現場のウイルス学を含む複数分野の研究者との協働による挑戦的なさきがけ課題である。変異の詳細が不明であるが故に、その解析は data-driven な手法となる。MDS 空間上で、ウイルスの遺伝子配列の時間発展をプロットすると、一定の曲率を持つ曲線として観測されることから、異なる年代の HA 間の相対的な距離に規則性があることが示唆された。このことから HA1 領域のアミノ酸置換の速度分布は、あるパラメータ値をもつガンマ分布としてよく近似でき、予測が可能となる。これはこの分野における大きな発見であると考えられ、伊藤氏により新たな道が開かれたと言える。今後予測精度の向上とその検証が進めば、インフルエンザ予防に大きな力を発揮することになるであろう。

### 6. 主な研究成果リスト

#### (1) 論文(原著論文)発表

- |   |
|---|
| 1. Ito, K., Igarashi, M., Miyazaki, Y., Murakami, T., Iida, S., Kida, H. and Takada, A. Gnarled-trunk evolutionary model of influenza A virus hemagglutinin. PLoS ONE. 2011. 6: e25953.                   |
| 2. Igarashi, M., Ito, K., Yoshida, R., Tomabechi, D., Kida, H. and Takada, A. Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PLoS ONE . 2010. 5(1): e8553. |
| 3. Yonezawa, K., Igarashi, M., Ueno, K., Takada, A., Ito, K. Resampling nucleotide sequences with closest-neighbor trimming and its comparison to other methods, PLoS one (in press)                      |
| 4. Ito, K., Zeugmann, T. and Zhu, Y. Clustering the normalized compression distance for influenza virus data. Lect. Notes Comput. Sc, 2010, 6060: 130-146.  |
| 5. Sata, K., Hirata, K., Ito, K. and Kuboyama, T. Discovering networks for global propagation of  |

(2)特許出願

研究期間累積件数:0件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

新聞報道

1. 2010年2月15日(月) 徳島新聞他, 遊歩道ーインフルエンザ, 「流行予測の手段探る」招待講演等

1. Ito, K. (2012) Can computers predict the mutation of influenza viruses? OMRON Satellite Events, The 6th International Conference on Soft Computing and Intelligent Systems, The 13th International Symposium on Advanced Intelligent Systems, 20 November 2012, Kobe, Japan.(チュートリアル講演)

市民講演会

1. 伊藤公人. コンピューターでインフルエンザウイルスの変異を予測する、2011年10月23日、岡山国際交流センター、岡山市
2. 伊藤公人. コンピューターでインフルエンザウイルスの変異を予測する、2011年11月20日、日本未来科学館、東京都

総説

1. 伊藤公人. インフルエンザウイルスの変異予測. 感染・炎症・免疫. 2012. 42巻4号, 54-56
2. 伊藤公人. 数理科学はインフルエンザウイルスの変異を予測できるか?. 応用数理. 2012. 22巻3号, 15-17
3. 伊藤公人, 高田礼人. ウイルスはどうやって生き延びているのか?ーインフルエンザウイルスの存続様式と進化. 遺伝. 2012. 66巻4号, 358-364
4. 伊藤公人. バイオインフォマティクスによるインフルエンザウイルスの変異予測. 最新医学. 2011. 66巻12号, 2632-2640
5. 伊藤公人. インフルエンザウイルスの抗原変異とバイオインフォマティクス. ウイルス. 2011. 61巻1号, 3-13