

# 新規情報学的手法を用いたインフルエンザウイルスのゲノム塩基配列の変化の方向性の予測

岩崎 裕貴<sup>1,2\*</sup>・阿部 貴志<sup>3</sup>・和田健之介<sup>1</sup>・池村 淑道<sup>1</sup>

## インフルエンザウイルスゲノム配列の解析

インフルエンザウイルスは、ヒト以外にもトリやブタなどの多くの動物種で流行を繰り返しており、それら動物種に由来する株がヒトへも感染が可能であることから、このウイルスの地球上からの撲滅は不可能と思われる。20世紀以降で1918年、1957年、1968年、2009年の4度のヒト集団内での大規模な流行（パンデミック）を引き起こし、多大な被害をもたらしてきた。

ウイルスは、自身の増殖に多くの宿主因子を必要とする。したがって、他の動物種（たとえばトリやブタ）からヒトへと感染を起こした場合、新たな宿主因子に依存して増殖する必要がある。今までのゲノム配列が、新宿主因子に最適な配列ではない可能性が考えられる。もしこの考えが正しいならば、インフルエンザウイルスは流行している宿主ごとに、ゲノム配列の特徴の一部を異にする可能性が考えられる。言い換えれば、インフルエンザウイルスの宿主特異的なゲノム配列の特徴を見いだすことにより、ゲノム配列変化の予測が可能になる。本稿では、インフルエンザウイルスゲノムの変化予測を行うために重要となる、ウイルスゲノム配列の宿主特異的な特徴を解析した例を紹介する。

## 膨大な量のインフルエンザウイルスを対象とした解析

現在、国際塩基データベース上には、1万株を越えるインフルエンザウイルスのゲノム塩基配列が登録されている。これらすべてのインフルエンザウイルス株を対象とした大規模解析を行うために、一括学習型自己組織化マップ：BLSOM解析を行った。BLSOMはアライメントフリーのクラスタリングツールの一種であり、膨大かつ複雑な情報からの能率的な特徴抽出が可能である。広範囲な生物種を対象にした、筆者らのBLSOM解析では、ゲノム断片配列中の連続塩基組成（たとえば10 kb中の4連続塩基組成）のみを用いても、断片配列は生物種ごとに明確に分離していた<sup>1)</sup>。2連や3連や4連続塩基組成は、genome signatureと呼ばれるように、生物種ごとに明瞭な特徴があり、10 kb程度に断片化しても、その

特徴は識別可能でありBLSOMはそのgenome signatureを適確に認識して、分離（自己組織化）させていた。

本研究では、データベース上に登録されている全インフルエンザAおよびB型株を対象に連続塩基組成に基づいたBLSOM解析を行った。計算機には連続塩基頻度のみを与えているが、由来する宿主（ヒト、トリ、ブタなど）に基づき、クラスタ化（自己組織化）していた（図1）。ウイルスの増殖が、多数の宿主因子に依存していることを反映して、宿主別に異なった連続塩基組成を持っていると考えられる。2009年4月頃からはブタを起源として流行を開始した新型株（図1B：濃緑）は、元々はトリ起源の株がブタを経由してヒトへと感染をした。興味深いことに、従来ヒトで流行している季節性株（図1A：薄緑）と離れた場所に位置しており、トリ（図1A：赤）・豚（図1A：青）由来株の領域に囲まれていることが見られた。いまだ十分には、新しい宿主であるヒト細胞環境での増殖に最適化していない可能性が考えられる。一方、古くからヒトのみで流行を繰り返しているB型株（図1A：水色）は、季節性株よりもトリ由来株からさらに遠く離れた位置に領域を形成していた。宿主を変えた際に、方向性のある変化が起きる可能性が示されており、以下の解析でその検証を行った。

## 季節性ヒト由来株の時系列変化

図2では、ヒト季節性株の時系列的な変化に注目している。ここでは、図1の4連続塩基BLSOMについて、1974年以前、1975年から1989年、1990年から1999年、2000年から2005年、2006年以降の各期間内に分離されたヒト季節性株H1N1（新型株除く）およびH3N2株の領域をそれぞれの図で、ピンクで表示した。流行初期である1974年以前に分離された季節性株は、由来として推定されているトリ株の近くに見られた。1975年以降、季節性ヒト由来株はBLSOM上において、時間とともにトリ由来株の領域から離れてB型株に近づいている方向性が見られた。B型株はヒト集団内のみで流行しており、現在知られているA型の季節性株よりも古くからヒト集団内で流行を繰り返していると考えられている。

\* 著者紹介 長浜バイオ大学大学院バイオサイエンス研究科（学術振興会特別研究員DC） E-mail: b105023@nagahama-i-bio.ac.jp

<sup>1</sup>長浜バイオ大学、<sup>2</sup>学術振興会特別研究員DC、<sup>3</sup>新潟大学

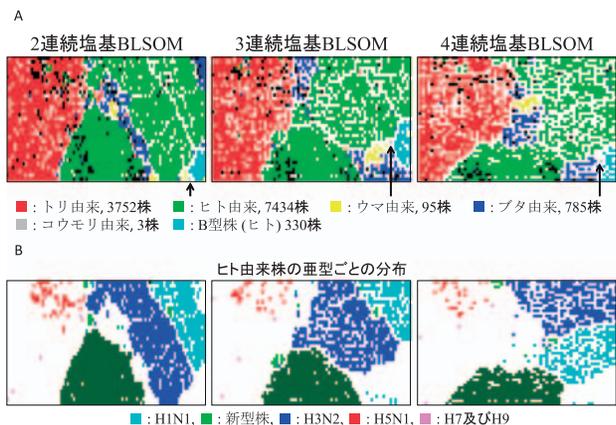


図1. 全インフルエンザウイルスを対象とした連続塩基組成のBLSOM解析. (A) 連続塩基組成に基づいたBLSOM. 各々の格子点に複数の宿主由来の株が分類されていた場合は黒で表示し, 単独の宿主由来株のみが分類されていた場合は宿主に対応した色で表示している. (B) ヒト由来株の亜型ごとの分布. 新型H1N1株を含むヒト由来株のみを表示し, ヒト由来株については亜型ごとに色分けを行った.

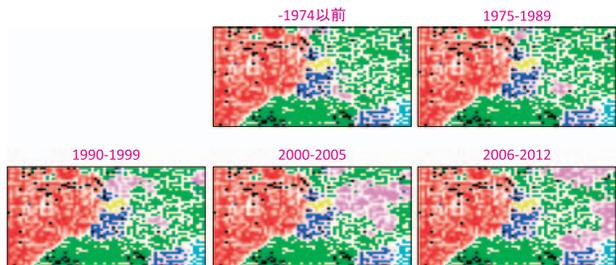


図2. A型ヒト季節性株の分離年代ごとの分布. 図1の4連続塩基BLSOMについて, 着目した年代に分離されたヒトH1N1およびH3N2株の領域を桃色で着色した.

ヒト季節性株の4連続塩基がB型株に近づいていることを検証するために, 個々のヒト季節性株とB型株の4連続塩基組成のユークリッド距離を算出した(図3). ユークリッド距離は多次元空間(4連続塩基組成では256次元空間)における距離を表し, ヒト季節性株H1N1およびH3N2株の両方も, 時間とともにB型株との距離を小さくしている. 4連続塩基がB型株に似てきていることは, ヒトを宿主として流行を開始し, ヒト集団内で流行を繰り返した場合, 方向性のある変化がゲノム配列に蓄積して行くことを示している.

### 宿主ごとの分離に寄与している連続塩基

BLSOMは, この宿主に依存した自己組織化に寄与する4連続塩基を特定することを可能にする(図4). B型

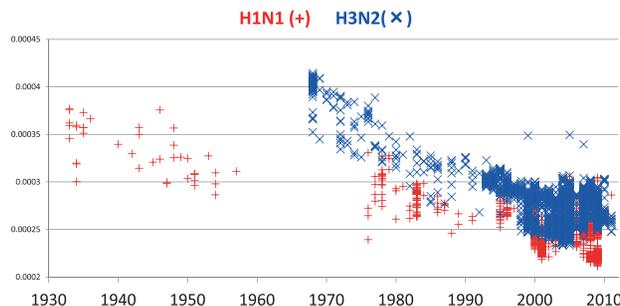


図3. B型株との4連続塩基組成のユークリッド距離. ヒト季節性A型H1N1(+)とH3N2(x)の各株とB型株のベクトルの平均値との間のユークリッド距離を, 各A株の分離年代ごとにプロットした.

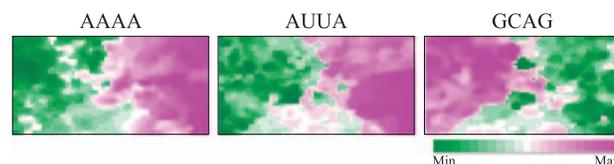


図4. 宿主ごとの分離に寄与している4連続塩基. 図1の4連続塩基BLSOMについて, それぞれの格子点における着目連続塩基の頻度を, マップ内で順位付けし, 頻度が高い格子点を紫, 低い格子点を緑で表示した.



図5. 新型株の分離された季節ごとの分布. 図1の4連続塩基BLSOMについて, 着目した時期に分離された新型株の領域を桃色で着色した.

株を含むヒト由来株は, AとU塩基に富む連続塩基を好む傾向が見られ, トリ由来株はGとC塩基に富む連続塩基を好む傾向が見られた. 新型株は, ヒト由来とトリ由来株の中間の性質があることが見られた. ヒト宿主に十分に適応するための移行的な期間にあると考えられる. ここでは示していないが, 2連や3連続塩基でも同様な結果を得ている<sup>2)</sup>.

### 新型株に見られる方向性のある変化

図5では, 新型株の分離された時期による違いに注目している. ヒトでの流行開始の直後である2009年4月, その後の2009年8月, 12月に分離された新型株の領域を, それぞれの月の図においてピンクで表示した. 4月に分離された新型株の主要部分は, トリ株領域の近くに分布

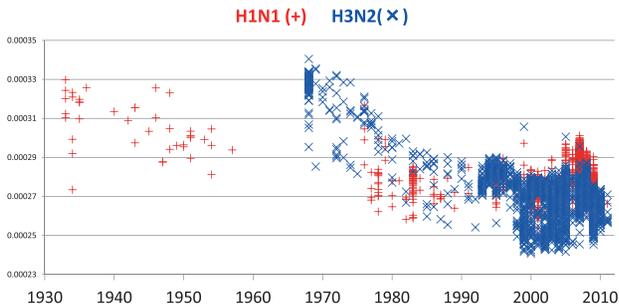


図6. コウモリ由来株とヒト季節性株との4連続塩基組成のユークリッド距離. ヒト季節性A型H1N1(+)とH3N2(×)の各株とコウモリ由来株間のユークリッド距離を, 各A型株の分離年代ごとにプロットした.

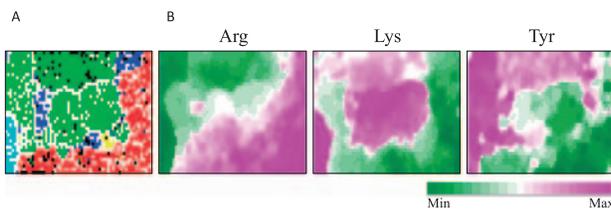


図7. 全インフルエンザウイルスを対象としたアミノ酸組成に基づくBLSOM解析. (A) アミノ酸組成に基づいたBLSOM. 色分けは図1Aと同様である. (B) 宿主ごとの分離に寄与しているアミノ酸. 色分けは図4と同様である.

しているのに対して, 12月の分離株の多くは, トリ領域から離れて, 主要部分はヒトの季節性株に近づく傾向を示していた. BLSOM上で確認できたこの方向性を生む主要因はGやCの数をへらし, AやUの数を増やしている傾向であったが, これ以外のより複雑な傾向も見えている<sup>2)</sup>. 季節性株で見られたインフルエンザウイルスのヒトを宿主とした場合のゲノム変化の方向性(図2)は, 新型株でも明瞭に見られた.

### コウモリ由来H17N10株

インフルエンザウイルスにおいて, もっとも恐れられているのは, 今までにヒトで流行したことのない抗原を持つ株, 言い換えればヒトが抗体を持っていない株のヒトへの感染である. 特に, プタのようなほ乳動物で流行している株の方が, トリ由来株よりもヒトへの感染の可能性が高いと考えられており, 注目されている. A型のインフルエンザウイルスは, 16種類のHAおよび9種類のNAの抗原型の組み合わせによって分類される. 自然宿主であるトリは, 上記の16×9の144種類すべてのインフルエンザウイルス株が感染可能である. 近年, グアテマラのコウモリから既存のA型株とは配列が大きく

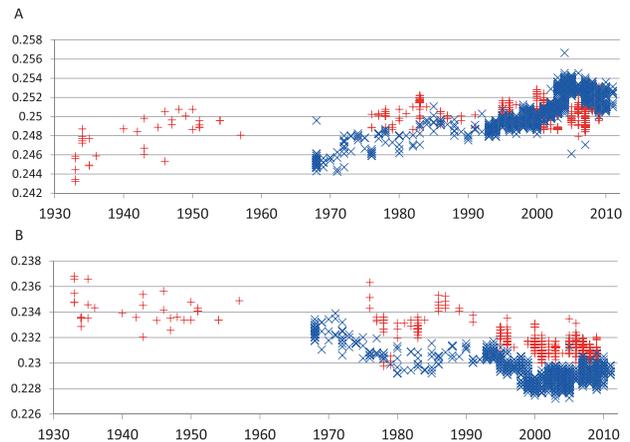


図8. AU-richおよびGC-richなアミノ酸頻度のヒト季節性株における変化. (A) AU-richアミノ酸頻度の時系列変化. ヒト季節性H1N1(+)とH3N2株(×)株におけるAU-richアミノ酸の頻度を年代ごとにプロットした. (B) GC-richアミノ酸の時系列変化.

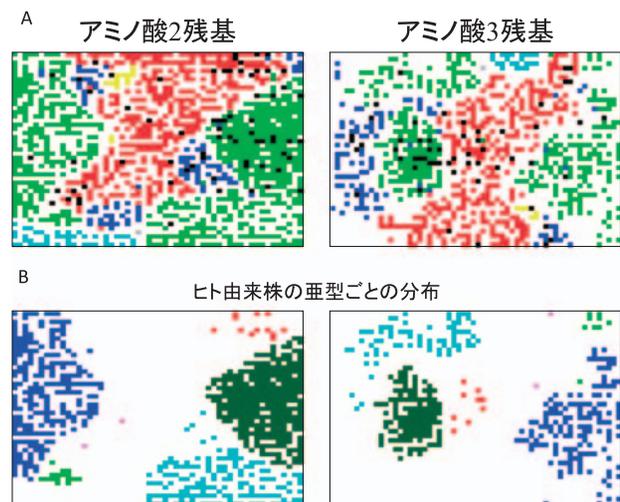


図9. 全インフルエンザウイルスを対象としたペプチド組成に基づくBLSOM解析. (A) 2連および3連アミノ酸組成に基づいたBLSOM解析. 色分けは図1Aと同様である. (B) ヒト由来株の亜型ごとの分布. 色分けは図1Bと同様である.

異なる新規性の高い株が分離された<sup>3)</sup>. この株は遙か昔にトリから分かれたと推定されており, H17N10と特徴付けられた. BLSOM上のコウモリ株(図1: 矢印)に注目すると, 直接の由来であるトリ株領域からは遠く離れて, 2006年以降に分離されたヒト季節性株およびB型株の近くに位置している. ほ乳類宿主に特異的なゲノム変化の方向性を検証するために, 個々のヒト季節性株とコウモリ株の4連続塩基組成のユークリッド距離を算出した(図6). 図3のB型株の解析と同様に, ヒト季節

性株のH1N1およびH3N2の両方とも、時間とともにコウモリ由来株とのユークリッド距離を小さくしており、4連続塩基組成が似てきている。インフルエンザウイルスのゲノム変化には、ヒトのみならず、ほ乳類宿主に共通の特徴があることが示されたので、ブタやウマで流行している株の変異予測も可能になる。危険な株の予測に情報を提供できる。

#### アミノ酸とペプチド組成に注目したBLSOM解析

インフルエンザウイルスについて、より機能に特化した特徴を知るためには、アミノ酸や連続アミノ酸レベルでの特徴抽出を行うことが重要である。実際に、ヒトを含むほ乳類宿主に適応するために重要と考えられるアミノ酸配列レベルでの変異が数多く報告されている<sup>4-6)</sup>。筆者らは、膨大な量のインフルエンザウイルス株から、宿主への適応に関連したアミノ酸組成の特徴抽出を行うためのBLSOM解析を行ったところ、この場合も宿主ごとの明確な分離がみられた(図7)。宿主ごとの分離に寄与するアミノ酸に注目すると、ヒト由来株はAおよびUを多く含むコドンにより翻訳されるアミノ酸(AU-rich amino acid)を好み、トリ由来株は、GおよびCを多く含むコドンにより翻訳されるアミノ酸(GC-rich amino acid)を好む傾向が見られた。塩基組成の影響を強く受けていると考えられる。図8では、季節性ヒト由来株

H1N1およびH3N2株のAU-richおよびGC-richアミノ酸の流行初期から2011年までの変化をプロットした。H1N1およびH3N2株の両方とも、AU-richアミノ酸は流行初期から徐々に増加しており、GC-richアミノ酸は減少している傾向がみられた。アミノ酸レベルでの変化の方向性が見られ、ウイルスゲノムの変化の方向性と一致していた。2連や3連アミノ酸組成へと解析を進めて行くことにより(図9)、ゲノム組成に依存しない傾向も見られており、タンパク質の機能と関係した変化を予測できる可能性が考えられる。ウイルスのゲノム配列の時系列的な変化は、完全にランダムと考えられる傾向にあるが、宿主の変換のような状況下では、塩基配列およびアミノ酸配列の両方のレベルで、一定の範囲内では予測可能な、方向性のある変化が起きている。予測能力のある情報解析はウイルスゲノムの解析において、実用性が高く重要である。

#### 文 献

- 1) Abe, T. *et al.*: *Genome Res.*, **13**, 693 (2003).
- 2) Iwasaki, Y. *et al.*: *DNA Res.*, **18**, 125 (2011).
- 3) Tong, S. *et al.*: *PNAS*, **109**, 4269 (2012).
- 4) Li, Z. *et al.*: *Virology*, **79**, 12058 (2005).
- 5) Massin, P. *et al.*: *Virology*, **75**, 5398 (2001).
- 6) Li, Z. *et al.*: *Plos Pathog.*, **5**, e1000525 (2009).