

1. SARS-CoV-2 ゲノミクスとサーベイランスへの応用

黒田 誠

国立感染症研究所 病原体ゲノム解析研究センター センター長

2019年末の中国・武漢を発端とする新型コロナウイルス (severe acute respiratory syndrome coronavirus 2: SARS-CoV-2) は2020年1月から2月にかけて国内に侵入し、現在に至るまで複数のクラスター (集団感染) を発生させている。2020年1月初旬から現在に至るまで行政対応として自治体から提供された検体RNAのSARS-CoV-2ゲノム情報の解析を実施してきた。部分的な封じ込め・クラスター叩きによる収束は可能であるが、世界的な発生動向を見ても終息には程遠い状況である。本稿では、行政対応におけるゲノム情報の利活用はもとより、SARS-CoV-2が他のコロナウイルスとどのような特徴 (相違) をもったゲノム情報を有するのか、これまでの知見をもとに概説したい。

新型コロナウイルス SARS-CoV-2 ゲノム情報

世界各地のSARS-CoV-2のゲノム配列が解読され、2020年10月20日現在で109,794患者のSARS-CoV-2ゲノム配列 (ゲノム分子疫学に適正な完全長配列; complete, high coverage) がGISAIDに登録されている¹⁾。SARS-CoV-2の変異速度は現在のところ25.6塩基変異/ゲノム/年 (つまり、1年間で平均25.6箇所の変異が見込まれる) と推定されており (図1: <https://nextstrain.org/ncov/global?l=clock>)、これはゲノム情報が収集され始めた2020年3月時点の数値から大きく変動はない。コロナウイルス全般でNsp-14エキソヌクレアーゼ活性によるproof-reading校正機能をもったRNA依存性RNAポリメラーゼ (RdRp) を有しており²⁾、他のRNAウイルスと比べても変異 (進化) 速度は遅いほうに分類される。また、 (+) 鎖RNAウイルスであり分節を有さないため、遺伝子再集合 (リアソータント) はなく組換え頻度も少ないと想定される³⁾。

1年も満たない短期間の世界拡散において10万件以上のゲノム情報が登録されていることにより、新興ウイルス感染症の初期に何が生じるのかゲノムレベルで詳細に理解できる非常に貴重な体験をしている。毎日のように登録情報が増え続けているため、リアルタイムな情報分析は必須である (つまり、一時期のスナップショットで軽々な判断をしないことも心がけなくてはいけない)。SARS-CoV-2ゲノムの変異情報をまとめたwebサイト (図2: <http://giorgilab.dyndns.org/coronapp/>) は遺伝子単位でアミノ酸置換の特徴がリスト化してあり有用である。また、国別にゲノム配列が分類され、国ごとの特徴も評価できるであろう。全世界の登録情報から、C>T変異が半数を超える特徴があり、これは感染増殖時にAPOBEC3ファミリーによる塩基の脱アミノ化による影響が示唆されている⁴⁾。C>Tが多数占めることから、ゲノム全般がATリッチ (GC%が低く) へ進むことが想定される。SARS-CoV-2と近縁のコロナウイルスにおけるゲノム進化を考察した論文が報告されているのでご参照いただきたい⁴⁾。中国・武漢Wuhan-Hu-1株 (GenBank ID: NC_045512) がGC38.0%のところ、ヒトコロナウイルス4種 (HKU1, NL63, OC-43, 229E) が32.0 - 38.2%を示しコロナウイルスの特徴として近似していた。一方、2002/2003 SARS-CoVは40.8%、中東呼吸器症候群コロナウイルス (MERS-CoV) が41.2%とGC%が高く、APOBEC3ファミリーによるC>T変異の影響が少ないことが示唆されている。

また、正の自然選択 (dN/dS 検定) [dN : 非同義置換率 (非

連絡先

〒162-8640

東京都新宿区戸山1-23-1

国立感染症研究所病原体ゲノム解析研究センター

TEL: 03-5285-1111

E-mail: makokuro@niid.go.jp

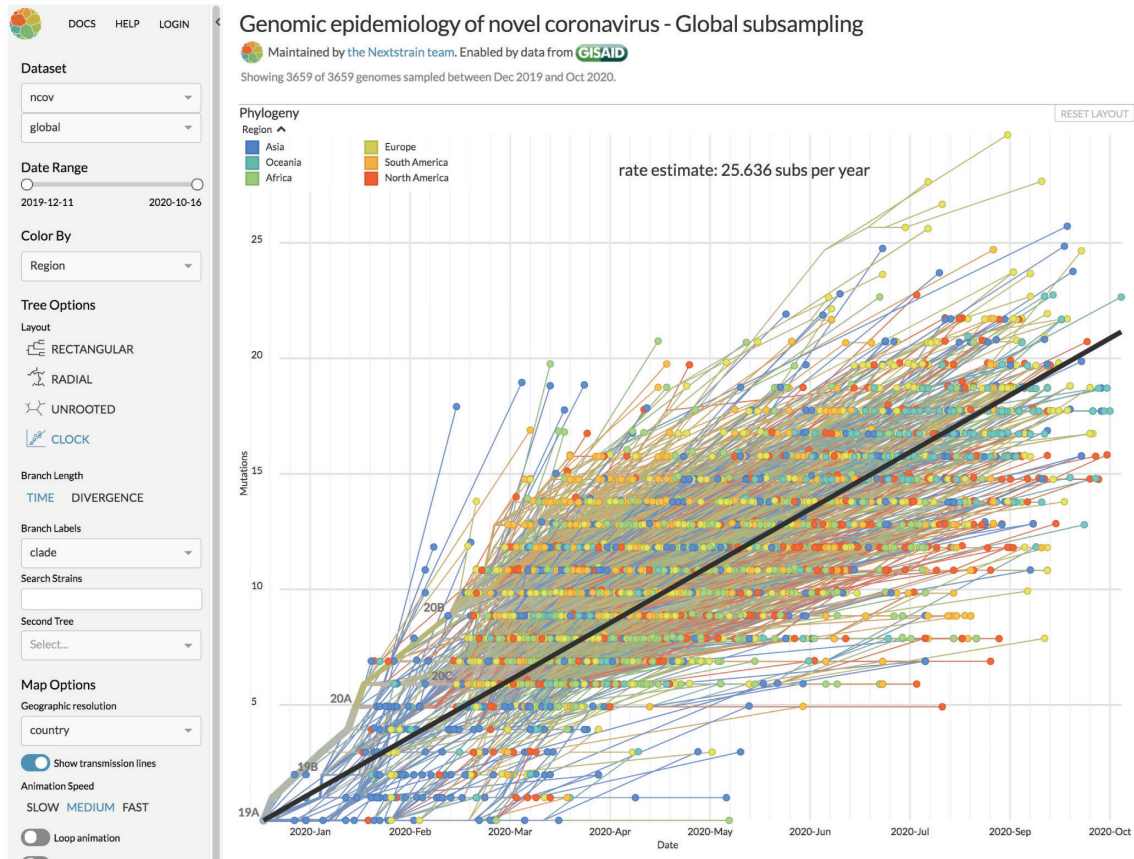


図1 SARS-CoV-2 ゲノム情報 (変異数) と時系列プロット (Nextstrain サイト). 2019/12 から 2020/10 までの代表 3,659 ゲノムデータを元に変異速度の平均がリアルタイムに示されている. (2020/10/23 イメージ取得).

同義座位あたりの非同義置換数), dS : 同義置換率 (同義座位あたりの同義置換数) の解析の結果, ヒトコロナウイルス 4 種の dN/dS が 0.13 - 0.22 とアミノ酸置換率が低いところ, SARS-CoV-2 の dN/dS は 0.73 とアミノ酸置換率が際立って高いことが分かっている. 一方, 2002/2003 年 SARS-CoV は 0.43, MERS-CoV が 0.23 であることから, SARS-CoV-2 の dN/dS (0.73) が示す高頻度のアミノ酸置換率は “ウイルス側で人・宿主への適応にまだ自信が持てずに迷っている?” ことを暗示しているかもしれない. 今はまだ我々人類との戦いに平衡関係に達していないがゆえ, “人・宿主への適応は初心者なのか?” とも想像している (根拠は薄い). もちろん, 全世界・全人類が SARS-CoV-2 に対する有効な既存免疫を有さないことから, 我々はただの培養器のような扱いを受け, 彼らの攻撃を受け入れているに過ぎず, その過程で多様な変異をそのまま放置しているのかもしれない.

さらに, SARS-CoV-2 ゲノムの領域固有の正 / 負の選択変異を評価した報告から⁵⁾, RdRp および Nucleocapsid 蛋白質には負の選択圧, そして Spike 蛋白質には正の選択圧がかかっている事がわかっている. コロナウイルス種全般でも Spike 蛋白質に正の選択圧が強いことが示されたが,

SARS-CoV-2 では Spike 蛋白質のレセプター (ACE2 に該当) 結合領域 RBD に正の選択圧が見られたのに対し, 2002/2003 SARS-CoV の RBD はむしろ変異は少ない (許容しない) という特徴が見えてきた. SARS-CoV-2 ではレセプター親和性の改善を模索する一方, 2002/2003 SARS-CoV では免疫逃避に関係していることを暗示した変異なのかもしれない.

変異 (mutation) だけでなくコドン・フレームをずらさない短めの塩基挿入・脱落 (inframe-insertion/deletion) や遺伝子単位の欠損 (deletion) も報告されるようになってきた. SARS-CoV-2 ゲノム上の ORF8 に 382 塩基が欠失した変異株 ($\Delta 382$) のウイルスに感染した人では, 重症化するリスクが低いとの報告がある⁶⁾. $\Delta 382$ では ORF7b の一部と ORF8 の大部分を失っており⁷⁾, この $\Delta 382$ と同様の欠損株は複数の国 (オーストラリア, バングラデシュ, スペイン) から別々に発見されていることから, homoplasy (独立に進化して同じ形質を獲得した現象) に該当するものと思われる. しかしながら, 野生型と $\Delta 382$ (相当の表現型) で, ウイルスの複製能力には差がないことは確認されており, 両者の具体的な相違について明らかになっていない.

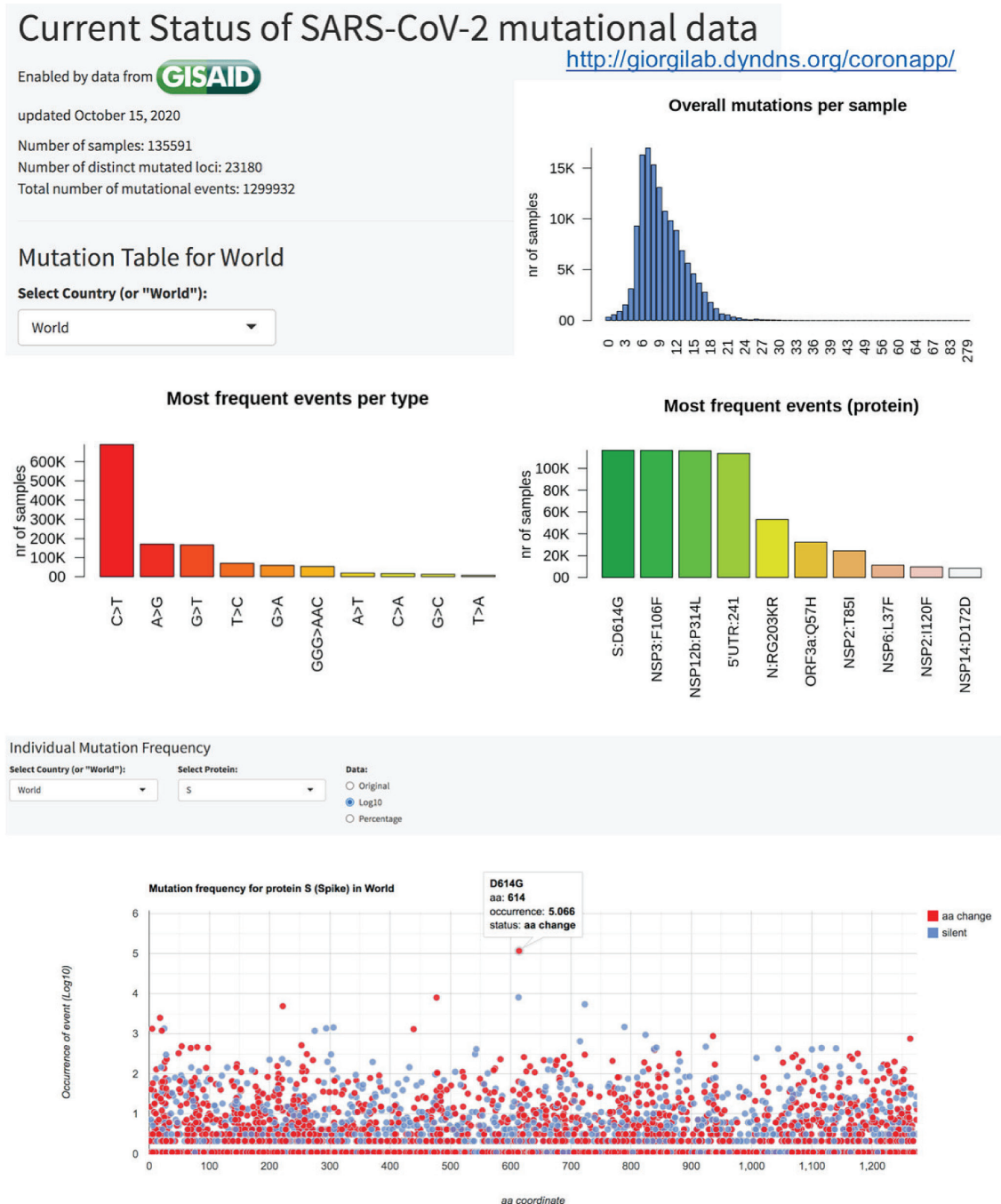


図2 SARS-CoV-2 変異データのリアルタイムモニタリング。2020/10/15までの計135,591 GISAID登録のうち、23,180変異サイト(29.9 kbゲノムの77.5%領域に相当)に生じた1,299,932変異パターンを集計しグラフ化している。検体あたり9箇所の変異が最も多く(パネル右上)、C>T変異が半数を超え(パネル中央左)、Spikeタンパク質のD614Gアミノ酸置換等の検出数が優勢(パネル中央右)になっている。遺伝子単位で変異箇所とアミノ酸置換(赤●:アミノ酸置換あり, 水色●:アミノ酸置換なし)を閲覧でき、Spike蛋白質ではD614G(パネル下)を有す登録が最も多い。

分子系統樹の分岐を元に SARS-CoV-2 の分類を提案している PANGOLIN (<https://cov-lineages.org/pangolin.html>) では、現在、313 系統(例: B.1.1)まで分類されている。武漢から発した1つのウイルスクローンから1年も満たな

い期間で生じた変異をもとに得られた分子系統関係である。この短期間の生物事象をそこまで詳細に分類すべきであるか若干の疑問を感じるが、2020/10/18 時点の114,801ゲノム情報の PANGOLIN 系統を地図上に図示化したサイト

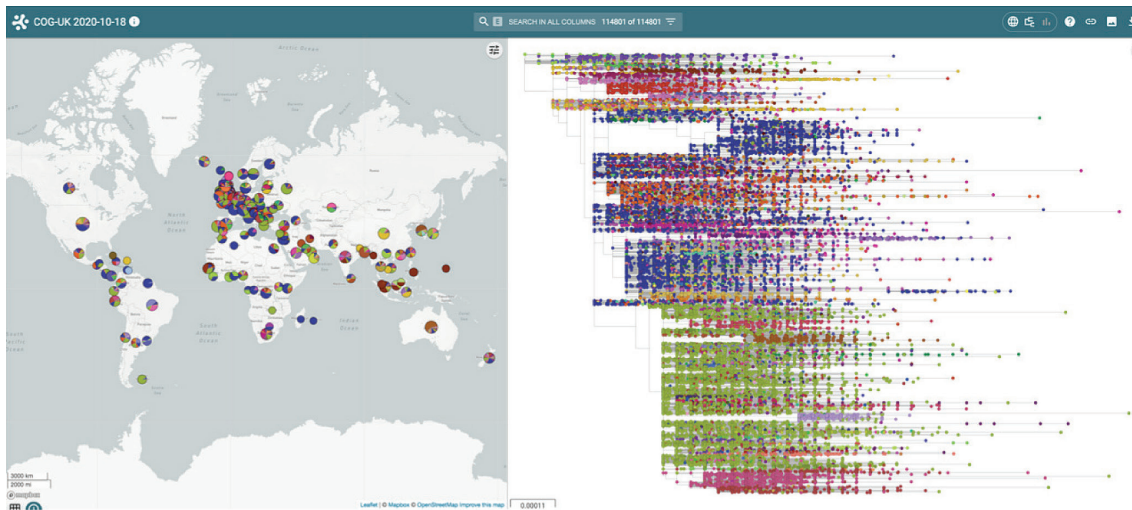


図3 SARS-CoV-2分子系統の世界的分布。

MicroReact (<https://microreact.org/project/cogconsortium-2020-10-18/3f7e33c9/>). 分子系統ごとに配色（パネル右）し、その検出割合を国ごとに円グラフ表示（パネル左）。

Microreact（図3：<https://microreact.org/project/cogconsortium-2020-10-18/3f7e33c9/>）を閲覧頂きたい。世界各国の分布状況が分子系統の検出割合として反映され、“分子系統”として各地域で今何が起きているのか が理解しやすくなっている。ただし、これは3月以降の各国の出入国規制による往来制限が主たる要因と思われるため（以下の“創始者効果”の項目で詳細を述べる）、公衆衛生学としては重要であるがウイルス学として特筆すべき分布ではないと考えられる。

コロナウイルス全般で、これまでに知られている遺伝子単位だけでなく、Overlapping gene (OLG) の存在が示唆されている。シクロヘキシミド等の翻訳阻害剤の併用によりリボソームが結合した領域を同定する Ribosome profiling (Ribo-seq) 法にて、翻訳産物（ペプチド、蛋白質）をコードする遺伝子を探索する方法である。SARS-CoV-2ではOLGの1つORF3dが新たに発見され、最も近縁のコウモリ Bat CoV (RaTG13/2013) ゲノムには見つからない一方、そのさらに上流の共通祖先型であるセンザンコウ Pangolin CoV (GX/P5L/2017) ゲノムと共有する特徴が見られている⁸⁾。COVID-19患者血清で抗体上昇がみられるため、アクセサリ蛋白質として何かしらの役割と特徴があるのかもしれない。

SARS-CoV-2ゲノム分子疫学解析によるクラスター対策

国内においても各地の自治体協力施設からPCR陽性検体を収集し、国内^{9,10)}、ダイヤモンド・プリセンス号の乗員乗客¹¹⁾、空港検疫所の陽性患者のSARS-CoV-2ゲノム配列を確定している。Wellcome Trust財団が支援するARTIC Networkが提唱したmultiplex PCR増幅法（49

plexを2チューブ¹²⁾）を基本にしている。Primer-dimer形成により増幅効率の悪い箇所が複数見つかったため、我々独自に改善した方法¹³⁾でゲノム確定を実施している。上述のSARS-CoV-2ゲノム情報が有する特徴を利活用し、感染伝播の追跡をウイルス株の親子関係として示すハプロタイプ・ネットワーク図を作成し、行政対応として自治体・保健所に還元して聞き取り・記述による積極的疫学調査を支援している。2020/08/06で公開したハプロタイプネットワーク図の概観は以下のサイトから閲覧願いたい (<https://www.niid.go.jp/niid/ja/basic-science/467-genome/9787-genome-2020-2.html>)。

ハプロタイプ・ネットワーク図を俯瞰的に概観してみると、2020年1-2月は中国・武漢由来が直接侵入し、ある一定のクラスターとして国内に拡散したが、それも一時的であった。3月中旬から欧州を經由して流入したSARS-CoV-2を発端に各地で感染拡大が進行し、4月において全国規模の緊急事態宣言に至った。（<https://www.niid.go.jp/niid/ja/basic-science/467-genome/9586-genome-2020-1.html>）。その後、感染は一旦収束傾向を見せ、緊急事態宣言の解除及び段階的に様々な活動を再開する中で、6月中旬から9月にかけて首都圏を中心に再び新規感染者数が増加し、緊急事態宣言下を上回る陽性者数が報告された。（<https://www.niid.go.jp/niid/ja/basic-science/467-genome/9787-genome-2020-2.html>）。ネットワーク図を俯瞰的に概観すると、日本では中国・武漢系統が流入し、そのヨーロッパ・欧州からの二次的な流入によって感染拡大が続いている。入国制限および都道府県の往来自粛そして行政によるクラスター対策によって日本には武漢系統の流入を収束させたものの、あらたに欧州系統が拡大し今も継続して進行中で

■ 創始者効果 (Founder effect)

最初に入ったものが勝ち！ (理論)

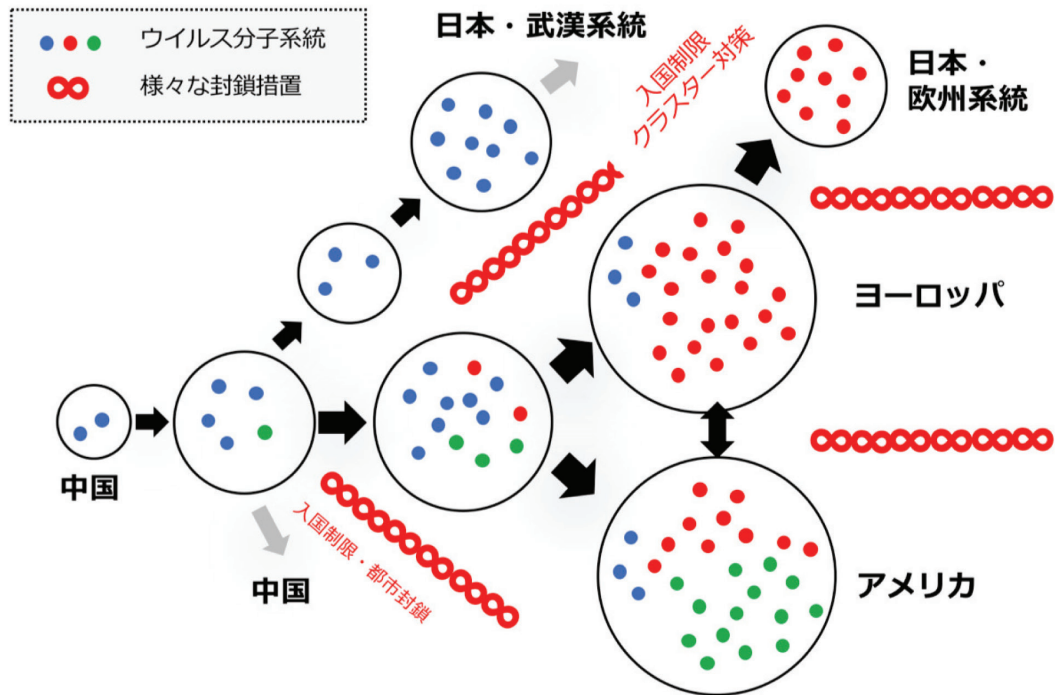


図4 創始者効果 (Founder effect)

生態学・集団遺伝学の用語。始祖効果 (しそこうか)、入植者効果 (にゅうしょくしゃこうか) とも。隔離された個体群が新しく作られるときに、新個体群の個体数が少ない場合 (SARS-CoV-2 においては全世界において新型かつ全世代に防御免疫がないと想定される)、元になった個体群とは異なった遺伝子頻度の個体群が出来ることを示す。(本イメージは概念図であり、実際の詳細とは一部異なる)

ある。一度国内に侵入し拡散したものは、それが起源となってその特徴的な親子関係 (系譜) をもって拡散するのである (図4)。

疫学調査によるクラスター集団の特定のみならず、さらにゲノム情報を配列指紋としてクラスター集団を特定し、積極的疫学調査を補完して感染源を特定するとともに、感染経緯の全体像を把握するために非常に重要な調査となる。今後、将来発生するクラスターを最小限に抑え込むためにも、クラスター発生に至る要因を特定し、地方自治体にゲノム情報を迅速に還元し、効果的な感染症対策の連携を図っていくことが肝要であろう。

SARS-CoV-2 ゲノム情報は我々に何を示してくれるのだろうか？

行政対応として SARS-CoV-2 ゲノム確定を集中的に遂行してきた。2020年5月までは、報道関係や研究者一般からのご意見で、“変異すると強毒になると言われていますが、ゲノム変異はあるのでしょうか？”とのご質問が多数

あった。当時、数ヶ月の知見で明言できるほど情報が収集されておらず、私見になるのだが、“黒田：その前提は正しいのでしょうか？ ウイルスは変異し続けます。病原性の変化はゲノム情報だけではわかりません。臨床情報と照らし合わせて総合的に判断することが必要です。現在のところ比較対照がありません”との回答であった。さらに時が過ぎ、2020年7月においては、報道関係や研究者一般から“最近、弱毒化したと言われていますが、ゲノム変異はあるのでしょうか？”とのご質問があり、非常に違和感を覚えた記憶がある。ほんの数ヶ月前とは逆の“前提仮説”の思い込みを提示されたわけだ。これも私見になるのだが、“黒田：その前提は正しいのでしょうか？ ウイルスは変異し続けます。病原性の変化はゲノム情報だけではわかりません。3-4月の欧州系統の初期と比較されてのご質問だと思われていますが、同一のゲノム配列を有するウイルスで感染したクラスター内 (ネットワーク図の●に集合する検体群) においても患者によって病状が異なるようです。との見解に至っている。

“コロナはコロナ” という感覚でいる。新型であるがゆえ、全世界・全世代において防御免疫が無く、世界的な拡散によってパンデミックに至ったと言える。ゲノム情報としてヒトコロナウイルス4種 (HKU1, NL63, OC-43, 229E) と基本的な特徴を共有する “コロナウイルス” でもある。しかしながら一方、2002/2003 SARS-CoV と同様にヒト ACE2 をレセプターにする結合特性を有することが特筆すべき病原性であると言えるだろう。

一般論としては、ウイルスは病原性をさげて広く深くウイルス種を残していく適応・潜伏の方向に向かうと推定される⁴⁾。新型コロナウイルスの病原性の変化については単にゲノム情報を確定しただけでは判定できるものではなく、患者の臨床所見、個別ウイルス株の細胞生物学・感染実験等を総合的に考慮する必要があると考えている。

さいごに (謝辞)

コロナウイルスが専門ではない私の本拙稿では物足りないところが多いことと思うため、より専門性の高い著者による総説・玉稿もご参照頂きたい。本稿では、自治体とともにコロナ対応している実情の一端でも読者に伝われば嬉しく思う。検体採取等調査にご協力いただきました医療機関、保健所および行政機関の関係者に深謝する。

参考：Global Initiative on Sharing All Influenza Data (GISAID)¹⁾

GISAID は、鳥インフルエンザが猛威をふるった2006年8月に医療分野の研究者たちによって設立されたインフルエンザウイルスの情報データベースである。新型コロナウイルス SARS-CoV-2 ゲノム情報も GISAIID が主体的に運用し登録・収集されている。

本稿に関連し、開示すべき利益相反状態にある企業等はありません。

I have no potential conflicts of interest to declare.

参考文献

- 1) <https://www.gisaid.org/epiflu-applications/next-hcov-19-app/>.
- 2) Becares M, Pascual-Iglesias A, Nogales A, Sola I, Enjuanes L, Zuniga S. Mutagenesis of Coronavirus nsp14 Reveals Its Potential Role in Modulation of the Innate Immune Response. *J Virol* 2016; **90**(11): 5399-414.
- 3) Hartenian E, Nandakumar D, Lari A, Ly M, Tucker JM, Glaunsinger BA. The molecular virology of coronaviruses. *J Biol Chem* 2020; **295**(37): 12910-34.
- 4) Simmonds P. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere* 2020; **5**(3).
- 5) Berrio A, Gartner V, Wray GA. Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. *PeerJ* 2020; **8**: e10234.
- 6) Young BE, Fong SW, Chan YH, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* 2020; **396**(10251): 603-11.
- 7) Su YCF, Anderson DE, Young BE, et al. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio* 2020; **11**(4).
- 8) Nelson CW, Ardern Z, Goldberg TL, et al. Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* 2020; **9**.
- 9) Sekizuka T, Kuramoto S, Nariai E, et al. SARS-CoV-2 Genome Analysis of Japanese Travelers in Nile River Cruise. *Front Microbiol* 2020; **11**: 1316.
- 10) Sekizuka T, Itokawa K, Hashino M, et al. A Genome Epidemiological Study of SARS-CoV-2 Introduction into Japan. *mSphere* 2020; **5**(6).
- 11) Sekizuka T, Itokawa K, Kageyama T, et al. Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak. *Proc Natl Acad Sci U S A* 2020.
- 12) ARTIC Network protocol. <https://artic.network/ncov-2019>
- 13) Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS One* 2020; **15**(9): e0239403.

SARS-CoV-2 genomics and its application to genome surveillance

Makoto KURODA

Ph.D., Director
Pathogen Genomics Center
National Institute of Infectious Diseases.

The initial coronavirus disease 2019 (COVID-19) outbreak occurred in Wuhan, China in late December 2019. After the first case of COVID-19 in Japan on January 15, 2020, multiple nationwide COVID-19 clusters were identified, the Japanese government focuses on mitigating the emerging COVID-19 clusters by conducting an active nationwide epidemiological surveillance. However, an increasing number of cases continued to appear even now, many cases exhibit unclear infection routes. We have evaluated the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome sequences from the COVID-19 cases to characterize their genealogical networks in order to demonstrate possible routes of spread in Japan. Here, this review summarized a recent action using the SARS-CoV-2 genome-wide surveillance in Japan, and the basic genome features as one of coronavirus species.

