

1. 大量シーケンス解析によって明らかになるウイルスの素顔 ～新型コロナウイルスを含めて～

中川 草

東海大学医学部分子生命科学

はじめに

2005年に454ライフサイエンス社が大規模並列型のDNA解読装置、いわゆる次世代シーケンサーを販売した。それからDNAシーケンス技術は様々な革新を経るに伴いライフサイエンス分野を大きく発展させた。ウイルス学もその例外ではない。大量シーケンスのデータ解析により、ヒトを含めた様々な生物のゲノムの中にウイルス由来する配列が存在し、そのうちの一部は遺伝子となっていること、そしてそのようなウイルス由来の遺伝子がダイナミックに進化していることがわかった。また、様々な環境・生体中からの大量シーケンス結果から、いまだ同定されていない多種多様のウイルスが存在することが明らかになってきた。そのような解析はバイローム (virome) 解析と言われる。新型コロナウイルス (SARS-CoV-2) 研究においても virome 解析が活用されている。本シンポジウムでは、大量シーケンスにより進展しているウイルス学研究について概説し、ウイルスの進化に迫った研究成果を報告した。

ゲノムに内在化するウイルス由来の配列の進化とデータベース

現在までに、様々な生物・ウイルスのゲノム配列が解読されてきた。これらのゲノムの比較解析の結果、真核生物のゲノムの中には、ウイルス (主にレトロウイルス) に由

来する配列が多数存在することが明らかになった。一般的に、ウイルス由来の配列は機能がなく、いわゆるジャンクDNAであると考えられてきたが、その配列の一部はタンパク質をコードし、宿主で遺伝子として機能することが分かってきた。一方で、そのようなウイルス由来の遺伝子は、多くの生物に保存されているとは限らない。例えば哺乳類の胎盤は、胎児と母体の境界となる合胞体性栄養膜と呼ばれる細胞があり、細胞栄養層が細胞融合して発生した多核の細胞である。この細胞融合に関して、ウイルス由来の遺伝子 (Syncytin など) が関与することが報告された¹⁾。また、そのような細胞融合に関与するウイルス由来の遺伝子は、生物系統ごとに異なったウイルスに由来する²⁾。ウイルス由来の遺伝子では、機能が類似していても、その遺伝子のもととなったウイルスが異なることがあることから、これらの遺伝子の進化は、嗅覚受容体やMHC遺伝子などの多重遺伝子族の進化で提唱されている「生と死の過程の進化モデル」³⁾という遺伝子進化モデルに合わせて説明できると我々は考えた⁴⁾。生と死の過程進化モデルとは、遺伝子重複によって、似たような機能をもつ遺伝子配列が冗長となった場合、機能する配列が進化的に偶然に他の類似配列に置き換わるものである。ウイルスに由来する遺伝子においても、ウイルスの内在化が継続的に起こることにより、類似した配列が冗長化し、機能する配列が進化するなかで偶然に置き換わる可能性があると考え、「バトンパス仮説」として報告した⁴⁾。実際に、このような遺伝子の置き換わりは Syncytin だけではなく、Arc というレトロウイルスの Gag 様の遺伝子でも見られた⁵⁾。

以上のように、ウイルスに由来する遺伝子は、たとえ宿主で機能を獲得した場合でも、進化的に置き換わり、また、生物種特異的であることが多いため、ゲノム配列のみから調べることは困難である。一方で、近年はゲノム配列情報だけではなく、さまざまな組織・細胞のトランスクリプトーム (RNA-seq) データが入手しやすくなり、ウイルスに由来する遺伝子も、宿主で機能を獲得したならば転写発現し

連絡先

〒259-1193

神奈川県伊勢原市下糟屋143

東海大学 医学部 分子生命科学

TEL: 0463-93-1121

FAX: 0463-93-5418

E-mail: so@tokai.ac.jp

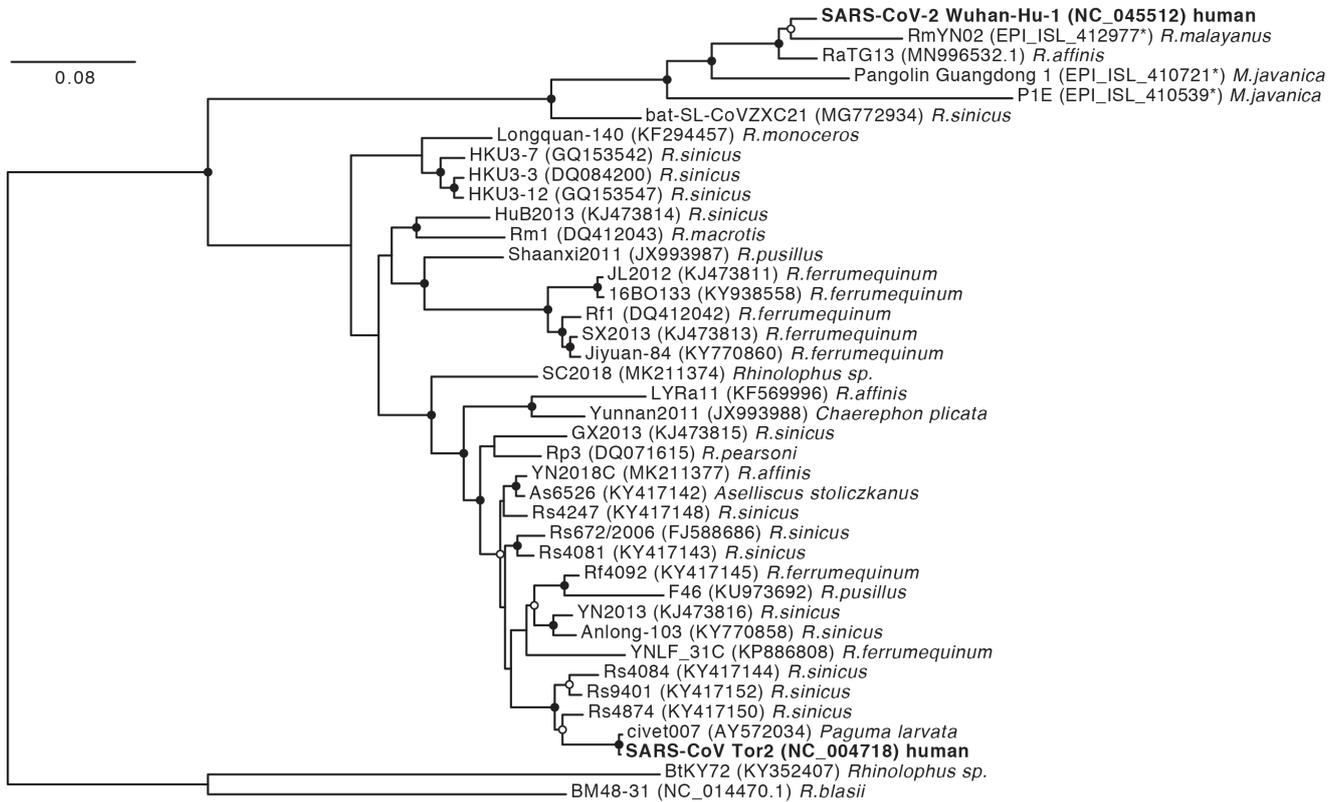


図1 Sarbecovirus 亜属に属するコロナウイルスの系統樹

SARS-CoV-2 と SARS-CoV, そしてそれに配列の類似度が高いコロナウイルスの 40 株 (すべて *Sarbecovirus* 亜属に属する) のゲノム配列を元にした系統樹を作成した. 塩基配列は MAFFT の L-INS-i プログラムを使用してマルチプルアライメント (多重整列) し, RAxML-NG を使って最尤系統樹を推定した (塩基置換モデルとして GTR+G+I を使用, ブートストラップ検定は 100 回). 株名, GenBank ID (* をつけたものは GISAID ID), ウイルスがみつかった生物種名をそれぞれ記した. また, 太字で表記したのはヒトから発見されたコロナウイルスである. 各分岐に示した黒丸は 95%, 白丸は 80% 以上, ブートストラップ検定でそれぞれ支持されたものである²⁰⁾. 本系統樹は Konno et al. を元に改変したものである²⁰⁾.

ているはずである. したがって, ゲノムの中からウイルスに由来するタンパク質をコードする可能性のある領域を予測し, そして RNA-seq データを併せて解析すればその発現を同定できるはずである. そこで我々は, ヒトを含めた 19 種の哺乳類の標準ゲノム配列を用いて, タンパク質をコードするウイルス由来の配列を網羅的に同定し, 遺伝子注釈したデータベース gEVE を構築した⁶⁾ (<http://geve.med.u-tokai.ac.jp>). 各ゲノム配列中のウイルスが内在化した配列を探索し, ウイルス由来のモチーフ配列をもつコード配列を同定した. その結果, 20 ゲノム配列 (ウシについては 2 つの異なるゲノムアセンブリを使用した) から 736,771 配列の蛋白質をコードする可能性のあるウイルス由来の配列を同定した. 本データベースを活用して, ウシの胎盤由来の RNA-seq データを解析し, 胎盤発生時に発現している BERV-K3 の Gag 由来に由来する配列を発見した⁷⁾. RNA-seq 以外にも様々なオミクスデータ, 例えばリボソームプロファイリングデータなども活用することが可能である.

一方で, gEVE データベースを作成する際に使用したウイルス由来の機能モチーフ配列はほとんどレトロウイルス由来で, すべてのウイルス様配列が網羅されていない. また, 解析の対象とした生物種は, 染色体レベルまでゲノムがアセンブルされている生物のみと非常に限られていた. そこで, そのような弱点を補強するため, すべての真核生物ゲノム (4,102 種) を対象に, 逆転写酵素を持たないすべてのウイルス (7,007 種) のゲノム配列に総当たりで配列類似性探索 (BLASTN) を行った. その結果, およそ半数 (3,856/7,007) のウイルスに, 少なくとも 1 つの真核生物ゲノムに類似配列が存在することがわかった. これらの配列を詳細に調べると, 1) ウイルスが宿主の真核生物のゲノムに内在化した場合, 2) 宿主の配列をウイルスが獲得した場合, 3) 宿主ゲノムの中にウイルス由来の配列が混入したことがあることが分かった. 以上の結果をまとめて, pEVE (<http://peve.med.u-tokai.ac.jp>) として一般公開した⁸⁾. 簡易的なスクリーニングであるが, 任意のウイルスに対して, 類似な配列がゲノムに存在しないか確認することがで

きる。

環境・生体に存在する様々なウイルス

現在、地球上の様々な環境、そして様々な生物から DNA・RNA 配列のシーケンシングが行われている。そのような大量シーケンズデータの中から、いままで報告されていない様々なウイルス由来の配列が同定されている。例えば、今年度のネオウイルス学シンポジウムの招待講演者のオハイオ州立大学の Matthew B. Sullivan 教授らが関わっているタラ海洋探査プロジェクト (<https://oceans.taraexpeditions.org>) では、世界中の海洋地域で継続的にサンプリングを行っている。そのようにサンプリングして集めた海水をウイルスが透過できるサイズのフィルターで通過させたのち、DNA を濃縮してメタゲノムシーケンズを行う。そのようなシーケンズデータをアセンブリ (集合) してコンティグ配列を作成し、既知のウイルスとの配列類似性や、遺伝子・塩基出現頻度の特徴量などをもとに、ウイルス同定を行っている。近年では北極海周辺の周辺データを大量に解析し、その海域のウイルスの多様性が高いことを示した⁹⁾。現在ではそのようなメタゲノムデータから同定されたウイルス配列のほうが、いままで登録されているウイルス配列よりも数百倍も多い¹⁰⁾。そのような大量シーケンズデータから同定されたウイルスは分離培養されていないが、ゲノム配列の平均塩基一致率 (Average Nucleotide Identity, ANI) という単位で種の定義を行っている。これは細菌のメタゲノムデータ解析でも使用されている単位であり、今後ウイルスの定義についても影響を与える可能性があると考えられる。

現在では、トランスクリプトーム (RNA メタゲノム) データから、網羅的に RNA ウイルスが同定されている。復旦大学 Yong-Zhen Zhang (張永振) 教授 (第 66 回日本ウイルス学会学術集会のネオウイルス学シンポジウムの招待講演者) とシドニー大学 Edward C. Holmes 教授らの研究グループは、蚊やダニなどの無脊椎動物、そして爬虫類、魚などの脊椎動物から大量トランスクリプトーム解析を行い、新規 RNA ウイルスを多数同定した^{11,12)}。近年ではそのようなメタゲノムから同定された新規ウイルスが実際にヒトに感染し、発熱などの症状を引き起こすことがわかってきた¹³⁾。私達もネコの尿サンプルから RNA を抽出し、イルミナ社の MiSeq でシーケンズを行い、その配列をアセンブリし、ネコパラミクソウイルス由来の配列を同定した¹⁴⁾。今後、このようなメタゲノム・メタトランスクリプトームからのウイルス同定がさらに進むと考えられる。

実際に、2019 年中国武漢で始まり世界中に感染が広がった、いわゆる新型コロナウイルス感染症 COVID-19 の原因となったコロナウイルス SARS-CoV-2 のゲノム配列も、

Zhang 教授と Holmes 教授らの研究グループによる virome 解析によって明らかになった¹⁵⁾。SARS-CoV-2 は重症急性呼吸器症候群 (SARS) ウイルスと類似していたことから命名されたが、それよりも類似度が高いコロナウイルスが、さまざまなコウモリ (主にキクガシラコウモリ属, *Rhinolophus*)¹⁶⁾、マレーセンザンコウ (*Manis javanica*)¹⁷⁻¹⁹⁾ から発見されており、それらのウイルス同定にも virome 解析が関わっている。これまでに見つかっている SARS-CoV-2、そして SARS-CoV と配列の類似度が高いコロナウイルスの系統樹を **図 1** に示した。このように、多くの類似するコロナウイルスが報告されている一方で、SARS-CoV-2 の宿主や中間宿主となった生物種は現在まで見つかっていない。それらを明らかにするためには、コウモリを中心とした様々な生物の virome 解析が重要であると考えられる。

おわりに

現在もシーケンズ技術の発展は続いている。例えば長鎖シーケンズ法の開発、一細胞シーケンズや染色体構造を見るための技術など、様々な技術革新が進んでいる。このようなシーケンズ技術の進歩がウイルス学の発展につながり、ウイルスの自然な状態、いわば素顔が明らかになるだろうと考える。

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。

文献

- 1) Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC Jr, McCoy JM. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*. 403(6771):785-789, 2000.
- 2) Imakawa K, Nakagawa S. The Phylogeny of Placental Evolution Through Dynamic Integrations of Retrotransposons. *Prog Mol Biol Transl Sci*. 145:89-109, 2017.
- 3) Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA*. 94(15):7799-7806, 1997.
- 4) Imakawa K, Nakagawa S, Miyazawa T. Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. *Genes Cells*. 20(10):771-788, 2015.
- 5) Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N, Belnap DM, Erlendsson S, Morado DR, Briggs JAG, Feschotte C, Shepherd JD. The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell*. 173(1):275,

- 2018.
- 6) Nakagawa S, Takahashi MU. gEVE: a genome-based endogenous viral element database provides comprehensive viral protein-coding sequences in mammalian genomes. *Database (Oxford)*. 2016. pii: baw087, 2016.
 - 7) Sakurai T, Nakagawa S, Bai H, Bai R, Kusama K, Ideta A, Aoyagi Y, Kaneko K, Iga K, Yasuda J, Miyazawa T, Imakawa K. Novel endogenous retrovirus-derived transcript expressed in the bovine placenta is regulated by WNT signaling. *Biochem J*. 474(20):3499-3512, 2017.
 - 8) Kryukov K, Ueda MT, Imanishi T, Nakagawa S. Systematic survey of non-retroviral virus-like elements in eukaryotic genomes. *Virus Res*. 262:30-36, 2019.
 - 9) Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Domínguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay JÉ, Vik D; Tara Oceans Coordinators, Babin M, Bowler C, Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*. 177(5):1109-1123, 2019.
 - 10) Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee KB, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit MA, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat Biotechnol*. 37(1):29-37, 2019.
 - 11) Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ. Redefining the invertebrate RNA virosphere. *Nature*. 540(7634):539-543, 2016.
 - 12) Shi M, Lin XD, Chen X, Tian JH, Chen LJ, Li K, Wang W, Eden JS, Shen JJ, Liu L, Holmes EC, Zhang YZ. The evolutionary history of vertebrate RNA viruses. *Nature*. 556(7700):197-202, 2018.
 - 13) Wang ZD, Wang B, Wei F, Han SZ, Zhang L, Yang ZT, Yan Y, Lv XL, Li L, Wang SC, Song MX, Zhang HJ, Huang SJ, Chen J, Huang FQ, Li S, Liu HH, Hong J, Jin YL, Wang W, Zhou JY, Liu Q. A New Segmented Virus Associated with Human Febrile Illness in China. *N Engl J Med*. 380(22):2116-2125, 2019.
 - 14) Sakaguchi S, Nakagawa S, Mitsuhashi S, Ogawa M, Sugiyama K, Tamukai K, Koide R, Katayama Y, Nakano T, Makino S, Imanishi T, Miyazawa T, Mizutani T. Molecular characterization of feline paramyxovirus in Japanese cat populations. *Arch Virol*. 165(2):413-418, 2020.
 - 15) Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. *Nature*. 579(7798):265-269, 2020.
 - 16) Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 579(7798):270-273, 2020.
 - 17) Liu P, Chen W, Chen JP. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (*Manis javanica*). *Viruses*. 11(11). pii: E979, 2019.
 - 18) Lam TT, Shum MH, Zhu HC, Tong YG, Ni XB, Liao YS, Wei W, Cheung WY, Li WJ, Li LF, Leung GM, Holmes EC, Hu YL, Guan Y. Identifying SARS-CoV-2 Related Coronaviruses in Malayan Pangolins. *Nature*. in press.
 - 19) Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol*. 30(7):1346-1351.e2, 2020.
 - 20) Konno Y, Kimura I, Uriu K, Fukushi M, Irie T, Koyanagi Y, Nakagawa S, Sato K. SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is further increased by a naturally occurring elongation variant. *bioRxiv* doi: <https://doi.org/10.1101/2020.05.11.088179>