

QUAN HỆ DI TRUYỀN CÁC LOÀI CORONAVIRUS DỰA TRÊN VÙNG PROTEIN KHÔNG CẤU TRÚC (NSP11) GENE ORF1ab

GENETIC RELATIONS OF CORONAVIRUS BASED ON NONSTRUCTURAL PROTEIN (NSP11) REGION OF ORF1ab GENE

TRƯƠNG THẾ QUANG^(*)

TÓM TẮT: Giải thành công bài toán sắp hàng nhiều trình tự NSP11 các loài coronavirus (CoV) thuộc chi Betacoronavirus. Phân chi Emboecovirus có khoảng cách di truyền xa đối với các phân chi Sarbecovirus, Nobecovirus, Merbecovirus và Hibecovirus với khoảng cách di truyền lần lượt là 1,097; 1,125; 1,113 và 1,141. Phân chi Sarbecovirus có khoảng cách di truyền gần với các phân chi Nobecovirus, Merbecovirus và Hibecovirus với khoảng cách di truyền lần lượt là 0,584; 0,563 và 0,461. Trong phân chi Sarbecovirus SARS-CoV-2 người có quan hệ di truyền gần gũi với CoV tê tê với tỷ lệ tương đồng 99,19 % và CoV dơi với tỷ lệ tương đồng 96,64 %. CoV tê tê và CoV dơi có mối quan hệ di truyền chặt chẽ với tỷ lệ tương đồng là 96,70 %. Có thể chứng minh giả thuyết SARS-CoV-2 người có nguồn gốc lây truyền từ CoV dơi hoặc tê tê.

Từ khóa: coronavirus 2 hội chứng hô hấp cấp tính nghiêm trọng; protein không cấu trúc 11; quan hệ di truyền.

ABSTRACT: We successfully solved the problem of multiple NSP11 sequences alignment of coronavirus (CoV) species of genus Betacoronavirus. Subgenus Emboecovirus has a long genetic distance for subgenus Sarbecovirus, Nobecovirus, Merbecovirus and Hibecovirus with a genetic distance of 1.097, 1.125, 1.113 and 1.141 respectively. Subgenus Sarbecovirus has a close genetic distance for subgenus Nobecovirus, Merbecovirus and Hibecovirus with a genetic distance of 0.584, 0.563 and 0.461 respectively. In the subgenus Sarbecovirus, human-SARS-CoV-2 had a close genetic relationship for pangolin CoV with a identity of 99.19% and bat CoV with a identity of 96.64%. Both pangolin CoV and bat CoV have a close genetic relationship with a identity of 96.70%. It is possible to prove hypothesis SARS-CoV-2 is from bat CoV or pangolin CoV.

Key words: severe acute respiratory syndrome coronavirus 2; nonstructural protein 11; genetic relations.

1. ĐẶT VẤN ĐỀ

Coronavirus (CoV) là một nhóm virus gây nhiễm trùng đường hô hấp ở người, từ các triệu chứng nhẹ đến kết quả gây chết người. Cho đến nay, có bảy chi CoV được biết là có thể lây nhiễm sang người. Bốn trong số các chi này, bao gồm human coronavirus 229E (HCoV-229E), human coronavirus OC43 (HCoV-OC43), human coronavirus NL63 (HCoV-NL63) và human coronavirus HKU1 (HCoV-HKU1). Bốn chi CoV này chỉ gây

ra bệnh tương đối nhẹ và các triệu chứng hô hấp hạn chế. Còn ba chi CoV khác, coronavirus hội chứng hô hấp cấp tính nghiêm trọng SARS-CoV (severe acute respiratory syndrome coronavirus), coronavirus hội chứng hô hấp Trung Đông MERS-CoV (Middle East respiratory syndrome coronavirus) và coronavirus 2 hội chứng hô hấp cấp tính nghiêm trọng SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) có khả năng gây bệnh cao và có thể dẫn đến

^(*) TS. Trường Đại học Văn Lang, quangtruongthe@gmail.com, Mã số: TCKH27-02-2021

các bệnh đường hô hấp nghiêm trọng và dẫn đến tử vong ở những bệnh nhân bị nhiễm bệnh.

SARS-CoV gây chết người đầu tiên xuất hiện vào năm 2002 tại tỉnh Quảng Đông, Trung Quốc. Năm 2002-2004, SARS-CoV đã lây nhiễm cho 8.098 người và dẫn đến 774 người tử vong ở 29 quốc gia trước khi nó biến mất. Năm 2012, MERS-CoV nổi lên ở Ả Rập Xê-út. Nó đã gây ra hai đợt bùng phát ở Hàn Quốc vào năm 2015 và ở Ả Rập Xê-út vào năm 2018 và vẫn có những báo cáo liên tục về các trường hợp riêng lẻ hiện nay. Tính đến tháng 01-2020, đã có 2.519 người bị lây nhiễm bệnh MERS được xác nhận và 866 người tử vong ở 27 quốc gia [9]. Vào tháng 12-2019, một loại CoV mới có thể gây bệnh hô hấp nghiêm trọng đã xuất hiện ở Vũ Hán, Trung Quốc. Tổ chức Y tế Thế giới WHO (World Health Organization) đã đặt tên cho loại virus mới này là SARS-CoV-2 gây ra bệnh coronavirus năm 2019 (COVID-19). Biểu hiện lâm sàng của COVID-19 có thể thay đổi từ không có triệu chứng và các triệu chứng giống cúm nhẹ đến hội chứng suy hô hấp cấp tính và tử vong. Các biến chứng dài hạn về phổi, tim mạch và thần kinh cũng đã được báo cáo trong các trường hợp COVID-19. So với SARS-CoV và MERS-CoV, SARS-CoV-2 rất dễ lây lan với hệ số sinh sản cơ bản (hệ số lây nhiễm cơ bản) R_0 ước tính là từ 5 đến 10 tức một ca COVID-19 hiện có thể gây ra trung bình từ 5 đến 10 ca nhiễm mới [6]. Đại dịch COVID-19 ở nhiều quốc gia, vùng lãnh thổ trên thế giới và kéo dài cho đến nay vẫn chưa được kiểm soát, mặc dù các nỗ lực sâu rộng tìm kiếm vaccine đang được thực hiện để phòng chống lại loại virus này [5, tr.3-5]. Theo cập nhật của Worldometers về đại dịch COVID-19 trên toàn thế giới tính đến 22:45 GMT ngày 20-03-2021 đã có 221 quốc gia với 123.404.656 người bị nhiễm bệnh và số người chết lên đến 2.720.991 [11]. Đại dịch COVID-19 đã trở thành cuộc khủng hoảng sức khỏe cộng đồng nghiêm trọng nhất trong thế hệ

chúng ta và có tác động sâu sắc đến nền kinh tế toàn cầu và địa chính trị. Vì vậy, việc xác định quan hệ di truyền giữa các loài CoV giúp nhận biết được con đường khuếch tán lây nhiễm CoV từ động vật sang động vật và sang người là cần thiết, từ đó có thể bổ sung kiến thức góp phần tìm ra các phương pháp chẩn đoán và vaccine có hiệu quả hơn trong việc phòng ngừa lây nhiễm CoV ở người để kiểm soát được đại dịch này.

2. NỘI DUNG

2.1. Mục tiêu và nội dung nghiên cứu

Tìm hiểu quan hệ di truyền của các loài CoV, các phân chi CoV thuộc chi *Betacoronavirus* gồm các loài CoV gây bệnh đường hô hấp cho người. Từ đó, tìm ra mối quan hệ gần gũi giữa các CoV có vật chủ là các động vật và CoV có vật chủ là người. Xác định nguồn gốc và con đường lây truyền CoV từ một số động vật sang người. Xác định quan hệ di truyền của các loài CoV thuộc chi *Betacoronavirus* bằng các tham số khoảng cách di truyền D_{xy} , tỷ lệ tương đồng I_{xy} (%). Giải bài toán sắp hàng nhiều trình tự NSP11 của các loài CoV để tìm hiểu quan hệ di truyền của các phân chi CoV, các loài CoV thuộc phân chi *Sarbecovirus* đã có trên các vật chủ là một số động vật như dơi, tê tê và người.

2.2. Phương pháp nghiên cứu

2.2.1. Phân tích phát sinh chủng loài

Phân tích phát sinh chủng loài dựa trên sắp hàng 23 trình tự NSP11 là vùng mã hóa cho protein không cấu trúc (nonstructural protein) nằm trên gene ORF1ab của các loài CoV thuộc chi *Betacoronavirus*. Các trình tự NSP11 nêu trên được thu thập từ cơ sở dữ liệu GenBank, USA (Bảng 1). Giải bài toán sắp hàng nhiều trình tự NSP11 theo thuật toán ClustalW [2, tr.162-164], ứng dụng phần mềm Mega X phiên bản 10.2.4.

2.2.2. Ước lượng khoảng cách di truyền

Ước lượng khoảng cách di truyền giữa các loài CoV hoặc giữa các phân chi CoV bằng mô hình khoảng cách Kimura 2 - parameter [2, tr.215-216], ứng dụng phần mềm Mega X phiên bản 10.2.4.

Chọn bootstrap tính lặp lại 1.000 vòng lặp lại để tăng độ tin cậy tính toán.

2.2.3. Xây dựng cây phát sinh loài

Cây phát sinh loài là sơ đồ thể hiện mức độ tương đồng giữa các loài CoV qua quá trình tiến hóa di truyền dựa trên vùng chỉ thị (marker) NSP11. Qua đó, phân loại các loài CoV thuộc chi *Betacoronavirus* thành các phân chi khác nhau, cũng như tìm hiểu mối quan hệ họ hàng gần gũi với nhau và có cùng nguồn gốc từ một tổ tiên chung trong quá khứ. Cây phát sinh loài được xây dựng theo thuật toán Neighbor - Joining Tree [2, tr.201-213], ứng dụng phần mềm Mega X phiên bản 10.2.4. Chọn khởi tạo với bootstrap 1.000 lần lặp lại để tăng độ tin cậy tính toán.

2.2.4. Ước lượng tỷ lệ tương đồng

Phân tích quan hệ di truyền của các loài CoV thuộc phân chi *Sarbecovirus* dựa trên tham số tỷ lệ tương đồng (Identity) của hai loài coronavirus I_{xy} (%) và được tính theo công thức (1).

$$I_{xy}(\%) = \frac{100 \cdot (D_{\max} - D_{xy})}{D_{\max}} \quad (1)$$

Trong đó: I_{xy} (%) là tỷ lệ tương đồng của hai loài x và y, $D_{\max}$ là khoảng cách di truyền của hai loài lớn nhất, D_{xy} là khoảng cách di truyền của hai loài coronavirus x và y được tính theo mô hình khoảng cách Kimura 2 - parameter.

2.2.5. Phân tích thống kê

Kết quả ước lượng các tham số khoảng cách di truyền, tỷ lệ tương đồng được trình bày dưới dạng giá trị trung bình mẫu. Xử lý số liệu, so sánh các giá trị trung bình mẫu bằng phương pháp phân tích phương sai (Anova) với xác suất tin cậy 95 %.

2.3. Kết quả và thảo luận

2.3.1. Coronavirus

Theo hệ thống phân loại ICTV (International Committee on Taxonomy of Viruses) của Ủy ban Quốc tế về phân loại virus, nhóm coronavirus gồm các loài CoV, MERS-CoV, SARS-CoV, SARS-CoV-2 và các biến chủng của chúng đều

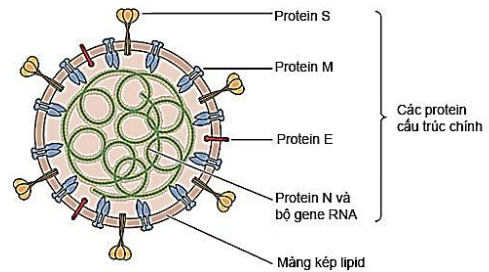
thuộc vực Riboviria, giới Orthornavirae, ngành Pisuviricota, lớp Pisoniviricetes, bộ Nidovirales, phân bộ Cornidovirineae, họ Coronaviridae, phân họ Orthocoronavirinae, được chia thành bốn chi *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus* và *Gammacoronavirus*. Các chi *Alphacoronavirus*, *Betacoronavirus* có nguồn gốc từ loài dơi. Các chi *Gammacoronavirus*, *Deltacoronavirus* đã tiến hóa từ nguồn gene của loài chim và lợn. Các coronavirus đều thuộc chi *Betacoronavirus*. Chi *Betacoronavirus* được chia thành năm phân chi là *Sarbecovirus*, *Embecovirus*, *Merbecovirus*, *Nobecovirus* và *Hibecovirus*. Các loài SARS-CoV-2, SARS-CoV, một số loài CoV và các biến chủng của chúng thuộc phân chi *Sarbecovirus*. Loài MERS-CoV thuộc phân chi *Merbecovirus*.

Hệ thống phân loại Baltimore là một hệ thống được sử dụng để phân loại virus dựa trên cách thức chúng tổng hợp RNA thông tin (mRNA), bằng cách phân loại virus dựa trên cách thức chúng sản xuất mRNA [10] để sao chép bộ gene của virus bên trong tế bào vật chủ. Theo hệ thống phân loại Baltimore, coronavirus thuộc nhóm IV là nhóm virus có bộ gene sợi đơn RNA dương tính (+ssRNA). Sợi đơn RNA dương tính của virus hoạt động như mRNA và có thể dịch trực tiếp thành các protein virus bởi các ribosome của tế bào vật chủ.

Coronavirus, có tên gọi xuất phát từ vẻ ngoài giống chiếc vương miện đặc trưng của chúng dưới kính hiển vi điện tử, là những virus RNA có đường kính khoảng 80-160 nm. Bộ gene của CoV là một phân tử RNA không phân đoạn, sợi đơn dương tính, có kích thước khoảng 30.000 nt (nucleotide), là bộ gene lớn nhất của tất cả các virus RNA đã biết. Đầu 5' của bộ gene CoV chứa hai khung đọc mở chồng lên nhau ORF1a và ORF1ab, có kích thước khoảng hai phần ba chiều dài bộ gene. ORF1a và ORF1ab có thể được dịch mã thành hai polyprotein (pp) là pp1a và pp1ab, chúng tiếp tục được phân cắt thành 16 protein không cấu trúc (NSP) tham gia vào quá trình sao chép bộ gene của virus và

tổng hợp mRNA dưới hệ gene. Đầu 3' của bộ gene CoV mã hóa bốn protein cấu trúc chính theo thứ tự là protein gai (S), protein vỏ bọc (E), protein màng (M) và nucleocapsid (N). Protein S, E, M tạo nên vỏ bọc của CoV, trong khi protein N tạo thành capsid để đóng gói bộ gene RNA (Hình 1). Đầu 3' của bộ gene cũng mã hóa nhiều protein phụ, thường là đặc trưng cho từng phân chi và có thể giúp CoV trốn tránh hệ thống miễn dịch hoặc tăng độc lực [4, tr.193-292]. Ví dụ: SARS-CoV chứa protein phụ ORF 3a, 3b, 6, 7a, 7b, 8a, 8b và 9b, MERS-CoV chứa ORF 3, 4a, 4b, 5, 8b và SARS-CoV-2 chứa ORF 3a, 6, 7a, 7b, 8, 10 [8], [9], [11].

Vòng đời của CoV được duy trì nhờ vào một số protein của virus. Để xâm nhập vào tế bào vật chủ, trước tiên protein S liên kết với các thụ thể receptor-binding domain (RBD) của CoV thông qua miền liên kết thụ thể của nó, và phức hợp virus-thụ thể sau đó được chuyển vị trí đến các endosome. Cả hai protein S của SARS-CoV và SARS-CoV-2 đều liên kết với enzyme chuyển đổi angiotensin-converting enzyme 2 (ACE2), trong khi protein S của MERS-CoV sử dụng dipeptidyl peptidase-4 (DPP4) làm thụ thể tế bào của nó. Tại endosome, protein S tiếp tục được phân cắt thành các tiểu đơn vị S1 (chứa RBD) và S2 (không chứa RBD), và tiểu đơn vị S2 làm trung gian dung hợp giữa vỏ virus và màng tế bào vật chủ. Sau khi xâm nhập vào tế bào, một số protein không cấu trúc NSP, đặc biệt là RNA polymerase phụ thuộc RNA (NSP11, NSP12) và helicase (NSP13), làm trung gian cho việc sao chép bộ gene CoV và phiên mã mRNA CoV, mRNA CoV tiếp tục được dịch mã thành các protein cấu trúc và phi cấu trúc khác nhau. Các protein N liên kết với RNA bộ gene CoV để tạo thành các nucleocapsid của virus, và các protein S, E, M tạo nên vỏ bọc của CoV. Sau khi lắp ráp, các hạt virus nảy chồi thông qua mạng lưới nội chất endoplasmic reticulum-Golgi (ER-Golgi) và thoát ra khỏi tế bào bằng cách xuất bào [3, tr.126-132].



Hình 1. Cấu trúc của coronavirus

2.3.2. Vùng chỉ thị NSP11

Ở đầu 5' bộ gene RNA của CoV, gene ORF1ab có chiều dài chiếm đến 2/3 bộ gene mã hóa cho polyprotein 1ab (pp1ab), pp1ab tiếp tục được phân cắt thành 16 protein không cấu trúc (NSP) tham gia vào quá trình sao chép bộ gene của virus và tổng hợp mRNA. Vùng chỉ thị (marker region) là một phần trình tự mã hóa cho protein không cấu trúc NSP11 (nonstructural protein 11) có kích thước khoảng 726 nt và nằm trên gene ORF1ab.

2.3.3. Sắp hàng nhiều trình tự NSP11 của các loài CoV

Sắp hàng nhiều trình tự NSP11 của các loài CoV dựa trên vùng chỉ thị NSP11 bằng thuật toán ClustalW, ứng dụng phần mềm Mega X. Các trình tự NSP11 của các loài CoV được thu thập từ GenBank (NCBI, USA), xem Bảng 1.

2.3.4. Cây phát sinh loài và phân chi CoV

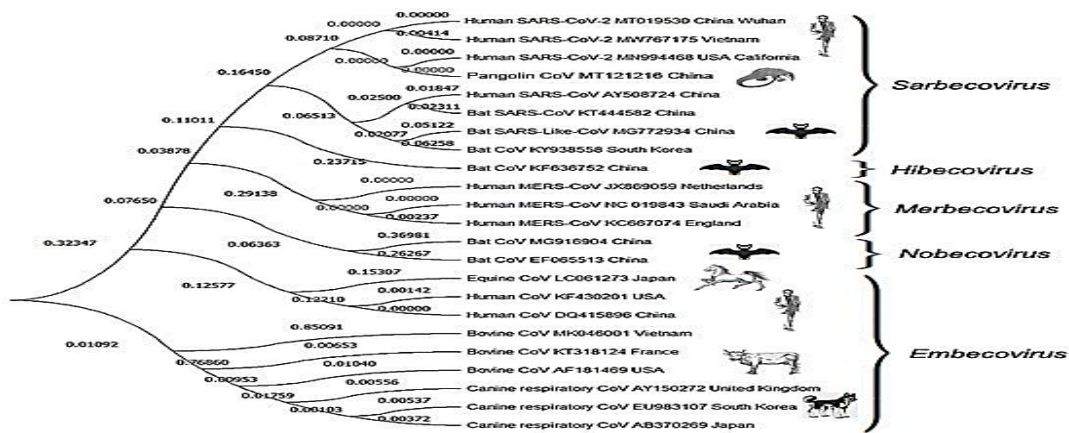
Kết quả sắp hàng nhiều trình tự NSP11 của các loài CoV thuộc chi *Betacoronavirus* (Bảng 2), bằng các công cụ Distance, Phylogeny của phần mềm Mega X đã thiết lập được ma trận khoảng cách di truyền và cây phát sinh loài (Hình 2). Với điều kiện khoảng cách di truyền giữa các nhóm $D_{xy} \geq 0,461$, các loài CoV thuộc chi *Betacoronavirus* được phân thành năm phân chi (Bảng 1, Hình 2).

2.3.5. Phân tích quan hệ di truyền giữa các phân chi CoV

Khoảng cách di truyền và sai số chuẩn tương ứng giữa các phân chi coronavirus thuộc chi *Betacoronavirus* dựa trên vùng chỉ thị NSP11 được nêu trong Bảng 2.

Bảng 1. Các loài CoV trong sắp hàng nhiều trình tự dựa trên vùng NSP11

STT	Tên loài coronavirus	Vật chủ	Quốc gia	Accession	Phân chi
1	SARS-CoV-2	Người	Việt Nam	MW767175 [7], [8]	Sarbecovirus
2	SARS-CoV-2	Người	Mỹ	MN994468	
3	SARS-CoV-2	Người	Trung Quốc	MT019530	
4	CoV	Tê tê	Trung Quốc	MT121216	
5	SARS-CoV	Người	Trung Quốc	AY508724	
6	SARS-Like-CoV	Dơi	Trung Quốc	MG772934	
7	SARS-CoV	Dơi	Trung Quốc	KT444582	
8	CoV	Dơi	Hàn Quốc	KY938558	
9	CoV	Dơi	Trung Quốc	KF636752	Hibecovirus
10	MERS-CoV	Người	Hà Lan	JX869059	Merbecovirus
11	MERS-CoV	Người	Ả Rập Xê Út	NC_019843	
12	MERS-CoV	Người	Anh	KC667074	
13	CoV	Dơi	Trung Quốc	MG916904	Nobecovirus
14	CoV	Dơi	Trung Quốc	EF065513	
15	CoV	Người	Mỹ	KF430201	Embecovirus
16	CoV	Người	Trung Quốc	DQ415896	
17	CoV	Ngựa	Nhật Bản	LC061273	
18	CoV	Bò	Việt Nam	MK046001	
19	CoV	Bò	Pháp	KT318124	
20	CoV	Bò	Mỹ	AF181469	
21	CoV	Chó	Hàn Quốc	EU983107	
22	CoV	Chó	Anh	AY150272	
23	CoV	Chó	Nhật Bản	AB370269	



Hình 2. Cây phát sinh loài và phân chi của coronavirus dựa trên vùng NSP11

Bảng 2. Khoảng cách di truyền và sai số chuẩn giữa các phân chi CoV dựa trên NSP11

STT	Phân chi	1	2	3	4	5
1	<i>Sarbecovirus</i>		0,019	0,021	0,018	0,041
2	<i>Nobecovirus</i>	0,584		0,019	0,021	0,038
3	<i>Merbecovirus</i>	0,563	0,586		0,023	0,046
4	<i>Hibecovirus</i>	0,461	0,599	0,565		0,047
5	<i>Embecovirus</i>	1,097	1,125	1,113	1,141	

Căn cứ khoảng cách di truyền của các phân chi CoV (Bảng 2), phân chi *Embercovirus* đều có khoảng cách di truyền xa đối với các phân chi còn lại. Cụ thể, khoảng cách di truyền của *Embercovirus* đối với *Sarbecovirus* là 1,097; đối với *Nobecovirus* là 1,125; đối với *Merbecovirus* là 1,113 và đối với *Hibecovirus* là 1,141. Phân chi *Sarbecovirus* có khoảng cách di truyền gần với các phân chi còn lại như *Nobecovirus* (khoảng cách di truyền 0,584),

Merbecovirus (khoảng cách di truyền 0,563), *Hibecovirus* (khoảng cách di truyền 0,461).

2.3.6. Phân tích quan hệ di truyền giữa các loài CoV thuộc phân chi *Sarbecovirus*

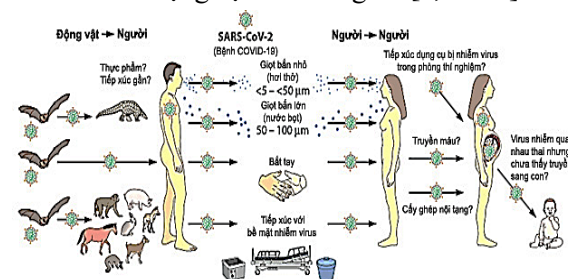
Khoảng cách di truyền của các loài CoV thuộc phân chi *Sarbecovirus* được ước lượng theo mô hình khoảng cách Kimura 2 - parameter. Từ đó ước lượng được tỷ lệ tương đồng I_{xy} (%) của các loài CoV thuộc phân chi *Sarbecovirus* (Bảng 3).

Bảng 3. Tỷ lệ tương đồng của các loài CoV thuộc phân chi *Sarbecovirus*

STT	Tên CoV	Tỷ lệ tương đồng I_{xy} (%)				
		1	2	3	4	5
1	Human SARS-CoV-2	100,00				
2	Pangolin CoV	99,19	100,00			
3	Bat CoV	96,64	96,70	100,00		
4	Human SARS-CoV	87,95	88,32	92,89	100,00	
5	Bat SARS-CoV	87,91	88,02	93,02	97,42	100,00

Phân tích quan hệ di truyền của các loài CoV thuộc phân chi *Sarbecovirus* dựa trên vùng NSP11 cho thấy SARS-CoV-2 người (Human SARS-CoV-2) có quan hệ di truyền gần gũi với CoV tê tê (Pangolin CoV) với tỷ lệ tương đồng 99,19 % và CoV dơi (Bat CoV) với tỷ lệ tương đồng 96,64 %. Mặt khác, CoV tê tê và CoV dơi có mối quan hệ di truyền chặt chẽ với tỷ lệ tương đồng là 96,70 % (Bảng 3). Có thể chứng minh giả định SARS-CoV-2 người có nguồn gốc từ CoV dơi và tê tê là vật chủ trung gian lây truyền (Hình 3). Kết quả nghiên cứu này cũng phù hợp với kết quả của các nghiên cứu phân tích thống kê Bayes sử dụng tái tạo địa lý thực vật học đã xác định virus SARS-CoV-2 mới xuất hiện thường lưu hành như một bầy không đồng nhất trong các ổ chứa động vật hoang dã trước khi chúng xuất hiện trong quần thể người, tổ tiên gây bệnh từ động vật sang người đã được tìm thấy trong ổ chứa động vật giả định là dơi móng ngựa (*Rhinolophus affinis*) thuộc họ Rhinolophidae, cũng như ở các loài động vật hoang dã khác được bán để làm thực

phẩm như tê tê (*Manis javanica*), cầy vòi mốc (*Paguma larvata*), lửng chó (*Nyctereutes procyonoides*), chồn bạc má (*Melogale moschata*)... Các đột biến trên các thụ thể receptor-binding domain (RBD) của CoV đã tạo điều kiện cho chúng có khả năng lây nhiễm sang các vật chủ mới hơn, do đó mở rộng phạm vi tiếp cận của chúng đến tất cả các nơi trên thế giới. Đây là mối đe dọa tiềm ẩn đối với sức khỏe của cả động vật và con người [1, tr.3-9].



Hình 3. Các đường lây truyền tiềm năng của SARS-CoV-2

3. KẾT LUẬN VÀ KIẾN NGHỊ

3.1. Kết quả nghiên cứu

Thu thập 23 trình tự NSP11 của các loài CoV, MERS-CoV, SARS-Like-CoV, SARS-CoV, SARS-CoV-2 thuộc chi *Betacoronavirus* trên các vật chủ dơi, tê tê, ngựa, bò, chó và

người. Giải bài toán sắp hàng nhiều trình tự NSP11 các loài CoV thuộc chi *Betacoronavirus*, kết quả các CoV được phân thành năm phân chi *Sarbecovirus*, *Embecovirus*, *Merbecovirus*, *Nobecovirus* và *Hibecovirus*. Phân chi *Embecovirus* đều có khoảng cách di truyền xa đối với các phân chi còn lại. Cụ thể, khoảng cách di truyền của *Embecovirus* đối với *Sarbecovirus* là 1,097; đối với *Nobecovirus* là 1,125; đối với *Merbecovirus* là 1,113 và đối với *Hibecovirus* là 1,141. Phân chi *Sarbecovirus* có khoảng cách di truyền gần với các phân chi còn lại như *Nobecovirus* (khoảng cách di truyền 0,584), *Merbecovirus* (khoảng cách di truyền 0,563), *Hibecovirus* (khoảng cách di truyền 0,461). Phân tích quan hệ di truyền của các loài CoV thuộc phân chi *Sarbecovirus* dựa trên vùng NSP11 cho thấy SARS-CoV-2 người có quan

hệ di truyền gần gũi với CoV tê tê với tỷ lệ tương đồng 99,19% và CoV dơi với tỷ lệ tương đồng 96,64%. Mặt khác, CoV tê tê và CoV dơi có mối quan hệ di truyền chặt chẽ với tỷ lệ tương đồng là 96,70%. Như vậy, có thể chứng minh giả thuyết SARS-CoV-2 người có nguồn gốc lây truyền từ CoV dơi hoặc tê tê.

3.2. Kiến nghị

Mở rộng phạm vi nghiên cứu thu thập thêm genome của nhiều loài CoV trên nhiều vật chủ là dơi, các động vật trung gian và người. Cần sử dụng thêm một số chỉ thị khác trong genome hoặc toàn bộ genome để nghiên cứu về quan hệ di truyền và nhận diện các đột biến SNP các loài CoV. Nghiên cứu về cơ chế CoV xâm nhập vào tế bào vật chủ và vòng đời của CoV, cũng như nguồn gốc và các con đường lây truyền khác nhau của CoV.

TÀI LIỆU THAM KHẢO

- [1] Humane society international (2020), *Mối liên hệ giữa các chợ động vật hoang dã và COVID-19*, Washington, D.C.
- [2] Trương Thế Quang (2018), *Tin sinh học (Bioinformatics)*, Nxb Đại học Quốc gia Thành phố Hồ Chí Minh.
- [3] Li J., Ulitzky L., Silberstein E., Taylor D.R., Viscidi R. (2013), *Immunogenicity and protection efficacy of monomeric and trimeric recombinant SARS coronavirus spike protein subunit vaccine candidates*, *Viral Immunol*, 26(2).
- [4] Masters P.S. (2006), *The molecular biology of coronaviruses*, *Adv Virus Res*, 66.
- [5] Rodriguez-Morales A.J., Bonilla-Aldana D.K., Balbin-Ramon G.J., Rabaan A.A., Sah R., Paniz-Mondolfi A., Pagliano P., Esposito S. (2020), *History is repeating itself: probable zoonotic spillover as the cause of the 2019 novel coronavirus epidemic*, *Infez Med* 28.
- [6] Shi Zhao, et al. (2020), *Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak*, *bioRxiv*, <https://www.biorxiv.org/content/10.1101/2020.01.23.916395v1.full>.
- [7] Trương Thế Quang (2021), *Analyze genetic relations of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) subspecies based on ORF1ab gene*, Accession MW767175, GenBank, National Center for Biotechnology Information (NCBI), USA, <https://www.ncbi.nlm.nih.gov/nuccore/MW767175>.
- [8] Trương Thế Quang (2021), *Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/VNM/VLU/2020 ORF1ab polyprotein, 3'-to-5' exonuclease region, (ORF1ab) gene, partial cds*, European Nucleotide Archive, London, England, <https://www.ebi.ac.uk/ena/data/view/MW767175>.
- [9] Wikipedia (2021), *Hội chứng hô hấp cấp tính nặng*, <https://vi.wikipedia.org>.
- [10] Wikipedia (2020), *Hệ thống phân loại Baltimore*, <https://vi.wikipedia.org>.
- [11] Worldometer (2021), *COVID-19 Coronavirus pandemic last updated: March 20, 2021, 22:45 GMT*, <https://www.worldometers.info/coronavirus/>.

Ngày nhận bài: 24-3-2021. Ngày biên tập xong: 24-4-2021. Duyệt đăng: 28-5-2021