

Đa dạng di truyền SARS-CoV-2

Đỗ Hoàng Đăng Khoa, Hồ Tá Giáp, Phan Ngọc Hân

Viện Kỹ thuật Công nghệ cao, Đại học Nguyễn Tất Thành
dhdkhoa@ntt.edu.vn

Tóm tắt

SARS-CoV-2 đang lây lan nhanh chóng gây nguy hại về tính mạng, sức khỏe và kinh tế toàn cầu. Đến đầu năm 2021, các chiến lược tiêm vaccine COVID-19 và thuốc hỗ trợ điều trị đã được nhiều quốc gia thực hiện và nghiên cứu, song song đó là sự biến đổi di truyền để kháng lại thuốc và vaccine của SARS-CoV-2 nhằm tồn tại và lây nhiễm. Các biến chủng SARS-CoV-2 mới và nguy hiểm liên tục được công bố bởi WHO như Alpha, Beta, Gamma, Delta, Eta, Iota, Kappa, Lambda, Mu và nhiều biến chủng chưa được đặt tên. Do đó, nghiên cứu về phân tích đa dạng di truyền là rất cần thiết để có bức tranh toàn cảnh về sự lây lan của các biến chủng, nhằm phục vụ cho nghiên cứu dịch tễ và các nghiên cứu di truyền SARS-CoV-2 khác. Kết quả nghiên cứu này cho thấy biến chủng A có thể là khởi nguồn của các biến chủng khác. Chúng tôi cũng đã vẽ ra bốn nhánh phát sinh cho bốn nhóm biến thể nguy hiểm. Biến thể Gamma có sự tương đồng về trình tự với biến thể Eta. Biến thể Beta thì có sự tương đồng cao với biến thể Iota và Mu. Biến thể Delta có quan hệ họ hàng với biến thể Kappa. Biến thể Alpha có quan hệ gần gũi với biến thể Zeta và Lambda.

Nhận 17.02.2022

Được duyệt 28.03.2022

Công bố 06.04.2022

Keywords

SARS-CoV-2,
COVID-19, Biến thể,
đa dạng di truyền

© 2022 Journal of Science and Technology - NTTU

1 Mở đầu

Cuối năm 2019, xuất hiện một virus mới với tên gọi là Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), hay còn được biết đến với tên khác là 2019 novel coronavirus (2019-nCoV) [1]. Virus mới này nhanh chóng được giải trình tự bộ gene và cho thấy tương đồng 79,6 % với coronavirus gây hội chứng hô hấp cấp nghiêm trọng (SARS-CoV) [2]. SARS-CoV-2 gây ra bệnh viêm hô hấp cấp kèm theo nhiều triệu chứng lâm sàng khác nhau với mức độ từ không triệu chứng đến hội chứng suy hô hấp cấp, căn bệnh này được gọi là “bệnh do coronavirus 2019” Coronavirus disease 2019 (COVID-19) [3]. Vào ngày 11 tháng 3 năm 2020, COVID-19 được WHO thông báo là đại dịch toàn cầu và y tế cộng đồng khẩn cấp thứ 6. Tính đến đầu tháng 9/2021, trên thế giới có khoảng 219 triệu người nhiễm SARS-CoV-2 và khoảng 4,5 triệu người đã tử vong do COVID-19. Riêng tại Việt Nam, từ ngày 3/1/2020 đến ngày

9/9/2021, đã có 563 676 trường hợp được xác nhận nhiễm COVID-19 với 14 135 trường hợp tử vong [4]. Vào cuối năm 2020, các nhà khoa học trên khắp thế giới đã có những kết quả đầu tiên về vaccine ngừa COVID-19, tính đến tháng 8/2021 có một số vaccine điển hình và đã được cấp phép ở Việt Nam như Vaccine COVID-19 (Vaccine AstraZeneca), Vaccine Gam-COVID-Vac (Sputnik V), SARS-CoV-2 Vaccine (Vero Cell), Vaccine Spikevax (Covid-19 Vaccine Moderna) và Pfizer-BioNTech vaccine. Ngoài vaccine, thuốc hỗ trợ điều trị bệnh nhân mắc COVID-19 cũng cho thấy hiệu quả điều trị tích cực như thuốc kháng virus remdesivir và thuốc kháng viêm tocilizumab [5-7].

Trình tự bộ gene SARS-CoV-2 được công bố lần đầu vào 11/1/2020 trên cơ sở dữ liệu Genbank của NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>) với mã số truy cập là MN908947.3 [8]. Dựa trên phân tích phả hệ bằng các công cụ tin sinh học, SARS-CoV-2 lúc

bây giờ được định danh là thành viên mới của dòng B của chi *Betacoronavirus*, bộ *Nidovirales* và họ *Coronaviridae* [1]. Kết quả so sánh trình tự gene của SARS-CoV-2 là tương đồng 79 % với SARS-CoV và 50 % với MERS [2,9]. Kết quả dựng cây phả hệ cũng cho thấy SARS-CoV-2 gần gũi với SARS-CoV hơn MERS [10]. Một thông tin quan trọng khi nghiên cứu mô hình tương đồng, SARS-CoV-2 cho thấy 96,2 % sự tương đồng với BatCoV RaTG13, một chủng coronavirus từ dơi *Rhinolophus affinis* [2].

Cấu trúc bộ gene của SARS-CoV-2 là sợi RNA mạch đơn, có mũ đầu 5' và đuôi poly-A đầu 3', cho phép hoạt động như một mRNA để thực hiện chức năng dịch mã [11]. Dựa trên phân tích toàn bộ bộ gene, bộ gene của các β -CoV mã hóa ra một số protein phi cấu trúc (non-structural protein, nsp) và 4 protein có cấu trúc bao gồm spike (S), envelope (E), membrane (M) và nucleocapsid (N) [12]. Kích thước hệ gene SARS-CoV-2 dao động từ 29,8 kb đến 29,9 kb với 12 khung đọc mở (ORFs) mã hóa 27 protein [13]. Hai gene *orf1ab* và *orf1a* định vị tại đầu 5' của bộ gene và mã hóa cho 2 protein tương ứng là pp1ab và pp1a. Hai protein này bao gồm 15 nsp là nsp1 đến nsp10 và nsp12 đến nsp16. Đầu 3' của bộ gene mã hóa cho 4 protein cấu trúc (S, E, M và N) và 8 protein phụ (3a,

3b, p6, 7a, 7b, 8b, 9b và orf14). Xét cấp độ acid amin, SARS-CoV-2 khá giống với SARS-CoV nhưng có một số điểm khác nhau đáng quan tâm. Ví dụ như protein 8a có ở SARS-CoV nhưng không có ở SARS-CoV-2; protein 8b có 84 acid amin trong SARS-CoV, nhưng dài hơn ở SARS-CoV-2, với 121 acid amin. Protein 3b có 154 acid amin trong SARS-CoV, nhưng ngắn hơn ở SARS-CoV-2, chỉ có 22 acid amin [10]. Tốc độ thay đổi bộ gene nhanh ở SARS-CoV-2 chưa từng thấy ở các loài virus khác [14]. Cơ sở dữ liệu về trình tự gene của SARS-CoV-2 liên tục được cập nhật về số lượng, do đó cần một hệ thống tên khoa học để phục vụ cho nghiên cứu. Ngày 29/1/2021, công trình nghiên cứu danh pháp SARS-CoV-2 trên tạp chí Nature Microbiology đã được xuất bản và nhiều hệ thống dữ liệu lớn về SARS-CoV-2 như GISAID, Nextstrain và Pango đã sử dụng [14]. Theo đó, các biến chủng của SARS-CoV-2 sẽ được đặt tên theo chữ cái và số như B.1.351 hay C.37 dựa vào nguyên tắc đặt tên của công trình nêu trên. Một hệ thống tên gọi dành cho SARS-CoV-2 thông dụng thứ hai hiện nay theo bảng chữ cái Hi Lạp được phát triển bởi WHO. Theo WHO, các biến chủng SARS-CoV-2 cần lưu ý sẽ được phân nhóm và đặt tên (Bảng 1).

Bảng 1 Danh sách các biến chủng SARS-CoV-2 theo hệ thống phân loại của WHO với các đột biến đặc trưng trên hệ gen. [15].

Tên (WHO)	Pango lineages	Điểm đặc trưng	Nguồn gốc	Ngày công bố	Accession number (GISAID)	Phân loại
Alpha	B.1.1.7	Spike A570D, Spike D614G, Spike D1118H, Spike H69del , Spike N501Y , Spike P681H , Spike S982A, Spike T716I, Spike V70del, Spike Y144del , N D3L , N G204R, N R203K, N S235F , NS8 Q27stop , NS8 R52I, NS8 Y73C, NSP3 A890D, NSP3 I1412T, NSP3 T183I, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323L	United Kingdom	Sep-2020	EPI_ISL_2738568	Variants of Concern (VOC)
Beta	B.1.351	Spike A243del, Spike A701V, Spike D80A , Spike D215G, Spike D614G, Spike E484K , Spike K417N , Spike L18F , Spike L242del, Spike L244del, Spike N501Y , E P71L, M H155Y, N Q9H, N T205I, NS3 A72S, NS3 L147F, NS3 Q57H, NS3 S171L, NS7a T115I, NSP2 T85I, NSP3 K837N, NSP5 K90R, NSP5 K100R, NSP6 F108del, NSP6 G107del, NSP6 M164V, NSP6 S106del, NSP12 P323L	South Africa	May-2020	EPI_ISL_2433571	

	B.1.351.2	Spike A243del, Spike A701V, Spike D80A , Spike D215G, Spike D614G, Spike E484K , Spike I934L, Spike K417N , Spike K1181N, Spike L18F , Spike L242del, Spike L244del, Spike N501Y , E P71L, N T205I, NS3 Q57H, NS3 S171L, NSP2 T85I, NSP3 K837N, NSP3 N1778S, NSP5 A193V, NSP5 K90R, NSP6 F108del, NSP6 G107del, NSP6 K61R, NSP6 L142F, NSP6 S106del, NSP12 P323L, NSP14 H455Y, NSP14 T113I			EPI_ISL_2 426605
	B.1.351.3	Spike A243del, Spike A701V, Spike D80A , Spike D215G, Spike D614G, Spike E484K , Spike K417N , Spike L18F , Spike L242del, Spike L244del, Spike N501Y , Spike P1143L, E P71L, N T205I, NS3 Q57H, NS3 S171L, NSP2 T85I, NSP3 K837N, NSP3 N1778S, NSP3 P1228L, NSP5 K90R, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323L, NSP13 I151V			EPI_ISL_2 433265
Gamma	P.1	Spike D138Y, Spike D614G, Spike E484K , Spike H655Y , Spike K417T , Spike L18F , Spike N501Y , Spike P26S, Spike R190S, Spike T20N, Spike T1027I, Spike V1176F, N G204R, N P80R, N S202C, NS3 S253P, NS8 E92K, NSP3 K977Q, NSP3 S370L, NSP4 S184N, NSP6 D267G, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323L, NSP13 E341D	Brazil	Nov-2020	EPI_ISL_2 466968
	P.1.1	Spike D138Y, Spike D614G, Spike E484K , Spike H655Y , Spike K417T , Spike L18F , Spike N501Y , Spike P26S, Spike P681H, Spike R190S, Spike T20N, Spike T1027I, Spike V1176F, N G204R, N P80R, N R203K, NS3 S253P, NS8 E92K, NSP1 M85del, NSP3 K977Q, NSP3 S370L, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323L, NSP13 E341D			EPI_ISL_2 466333
	P.1.2	Spike D138Y, Spike D614G, Spike E484K , Spike H655Y , Spike K417T , Spike L18F , Spike N501Y , Spike P26S, Spike R190S, Spike T20N, Spike T1027I, Spike V1176F, N G204R, N P80R, N R203K, NS3 S253P, NS8 E92K, NSP3 G1043V, NSP3 K977Q, NSP3 P1200S, NSP3 S370L, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323L, NSP13 E341D			EPI_ISL_2 466757
Delta	B.1.617.2, AY.1, AY.2, AY.3	Spike D614G, Spike D950N , Spike G142D, Spike L452R, Spike P681R, Spike S494L, Spike T478K , M I82T, N D63G, N D377Y, N R203M, N R385K, NS3	India	Oct-2020	EPI_ISL_2 988197

		S26L, NS7a G38R, NS7a L116F, NS7a T120I, NSP2 E167V, NSP3 C104F, NSP3 P822L, NSP4 A446V, NSP6 V149A, NSP12 G671S, NSP12 P323L, NSP13 P77L, NSP15 H234Y				
Eta	B.1.525	Spike A67V, Spike D215Y, Spike D614G, Spike D1084G, Spike E484K , Spike F888L, Spike H69del , Spike K150N, Spike Q52R, Spike Q677H, Spike Q1180L, Spike V70del , Spike Y144del , M I82T, N A12G, N T205I, NSP2 K67N, NSP3 T1189I, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323F	Multiple Country	Dec-2020	EPI_ISL_2 453388	Variants of Interest (VOI)
Iota	B.1.526	Spike A701V , Spike D253G, Spike D614G, Spike E484K , Spike L5F , Spike T95I , Spike T1027I, N M234I, N P199L, NS3 D2A, NS3 Q57H, NS8 T11I, NSP2 T85I, NSP4 L438P, NSP4 T73I, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 P323L, NSP13 Q88H	USA	Nov-2020	EPI_ISL_2 454354	
Kappa	B.1.617.1	Spike D614G, Spike E154K, Spike E484Q, Spike G142D , Spike L452R , Spike P681R , Spike Q1071H, Spike T95I, M I82S, N D377Y, N R203M, NS3 R134L, NS3 S26L, NS7a V82A, NSP3 T749I, NSP6 T77A, NSP12 P323L, NSP12 T26I, NSP13 G206C, NSP13 M429I, NSP15 K259R, NSP15 S261A	India	Oct-2020	EPI_ISL_2 455044	
Lambda	C.36	Spike D614G, Spike F490S, Spike G75V, Spike I1216V, Spike L452Q , Spike N148T, Spike T76I, Spike T859N, N G204R, N G214C, N P13L, N R203K, NS8 P36L, NSP2 T85I, NSP3 F1569V, NSP3 P1469S, NSP3 T428I, NSP4 L438P, NSP4 T492I, NSP5 G15S, NSP6 F108L, NSP12 P323L	Peru	Dec-2019	EPI_ISL_2 439187	
Mu	B.1.621	Spike D614G, Spike D950N , Spike E484K , Spike ins145N, Spike N501Y , Spike P681H , Spike R346K, Spike T95I , Spike Y144T , Spike Y145S , N T205I, NS3 A99V, NS3 I20M, NS3 Q57H, NS3 V256I , NS8 P38S, NS8 S67F, NS8 T11K, NSP3 S1285F, NSP3 T237A, NSP3 T720I, NSP4 L447F, NSP4 T492I, NSP6 Q160R, NSP12 P323L, NSP12 Y521C, NSP13 E261D, NSP13 P419S	Colombia	Jan-21	EPI_ISL_2 454576	
	B.1.427	Spike D614G, Spike L452R, Spike S13I, Spike W152C, N D144H, N T205I, NS3 Q57H, NS3 S40L, NS8 S67F, NSP2 L448I, NSP2 S32L, NSP2 T85I, NSP3 V253I, NSP4 S395T, NSP5 S121L, NSP12 P323L, NSP13 D260Y, NSP13 P53L, NSP14 V83I, NSP16 A121V	USA	Mar-2020	EPI_ISL_2 443292	Reclassifying VOIs/ VOCs

	B.1.429	Spike D614G, Spike L452R, Spike S13I, Spike W152C, N D377Y, N M234I, N T205I, NS3 Q57H, NS8 V32L, NS8 V100L, NSP2 T85I, NSP3 A85V, NSP3 E1846D, NSP3 V253I, NSP9 I65V, NSP12 P323L, NSP12 V174I, NSP13 D260Y	USA	Mar-2020	EPI_ISL_2 443854
Zeta	P.2	Spike D614G, Spike E484K, Spike L5F, Spike V1176F, M K166R, N A119S, N G204R, N M234I, N R203K, NSP5 L205V, NSP7 L71F, NSP12 P323L, NSP16 T151I	Brazil	Apr-2020	EPI_ISL_2 466228
	P.3	Spike A243del, Spike D614G, Spike E484K, Spike E1092K, Spike G142del, Spike H1101Y, Spike L141del, Spike L244del, Spike N501Y, Spike P681H, Spike V143del, Spike V1176F, Spike Y265C, N G204R, N R203K, NS8 K2Q, NSP3 D736G, NSP3 S1807F, NSP4 D217N, NSP4 L438P, NSP6 D112E, NSP7 L71F, NSP12 P323L, NSP13 A368V, NSP13 L280F	Philippines	Jan-2020	EPI_ISL_2 423558
	R.1	Spike D614G, Spike E484K, Spike G769V, Spike W152L, M F28L, N G204R, N Q418H, N R203K, N S187L, NSP7 D77N, NSP12 P323L, NSP13 G439R, NSP13 M576I, NSP14 P412H	Multiple Country	Jan-2020	EPI_ISL_2 462878
	R.2	Spike D614G, Spike E484K, Spike E1202Q, Spike Q677H, Spike T732S, N G204R, N R203K, NSP3 A231V, NSP3 K384N, NSP3 Q774R, NSP8 D143G, NSP8 R75K, NSP12 A311S, NSP12 P323L	Multiple Country	Jan-2020	EPI_ISL_2 284073
	B.1.466.2	Spike D614G, Spike N439K, Spike P26S, Spike P681R, N H300Q, N T205I, NS3 Q57H, NSP1 H110Y, NSP3 P822L, NSP3 S126L, NSP3 T350I, NSP6 L75F, NSP6 V149F, NSP12 P323L, NSP13 S259L	Indonesia	Nov-2020	EPI_ISL_2 429145

Chữ in đậm thể hiện đột biến đặc trưng ở các biến chủng.

WHO đã chia một số biến thể thành các nhóm khác nhau dựa trên đặc trưng của chúng. Đầu tiên là nhóm biến thể đáng lo ngại (Variants of Concern - VOC): là các biến thể làm tăng khả năng lây truyền hoặc thay đổi bất lợi trong dịch tễ học COVID-19; hoặc tăng độc lực, thay đổi biểu hiện bệnh lâm sàng; hoặc giảm hiệu quả của các biện pháp cộng đồng, chẩn đoán, vaccine, phương pháp điều trị có sẵn [15]. Nhóm biến thể đáng quan tâm (Variants of Interest - VOI): là các biến thể với những thay đổi di truyền được dự đoán hoặc biết là có ảnh hưởng đến các đặc điểm của virus như khả năng lây truyền, mức độ nghiêm trọng của

bệnh, vượt hàng rào miễn dịch, vượt khỏi chẩn đoán hoặc điều trị; các biến chủng này phải được xác định là nguyên nhân gây ra lây truyền cộng đồng đáng kể hoặc nhiều ổ dịch COVID-19, ở nhiều quốc gia với tỷ lệ lưu hành tương đối ngày càng tăng cùng với số trường hợp gia tăng theo thời gian, hoặc các tác động dịch tễ học rõ ràng khác cho thấy một nguy cơ mới đối với sức khỏe cộng đồng trên toàn cầu [15]. Cuối cùng là nhóm biến chủng cần phân loại lại (Reclassifying VOIs/ VOCs): là nhóm các biến chủng đang được theo dõi và đánh giá chính xác mức độ nguy hiểm [15].

Bảng 2 Các đột biến giống nhau của một số biến thể SARS-CoV-2.

20I (Alpha, V1) (B.1.1.7)	20H (Beta, V2) (B.1.351)	20J (Gamma, V3) (P.1)	21A (Delta) (B.1.617.2)	21B (Kappa) (B.1.617.1)	21D (Eta) (B.1.525)	21F (Iota) (B.1.526)	21G (Lambda) (C.37)	21H (Mu) (B.1.621)	20A/S:126A (B.1.620)
S:D614G	S:D614G	S:D614G	S:D614G	S:D614G	S:D614G	S:D614G	S:D614G	S:D614G	S:D614G
	S:E484K	S:E484K		S:E484Q	S:E484K	S:E484K		S:E484K	S:E484K
S:P681H			S:P681R	S:P681R				S:P681H	S:P681H
S:Y144-					S:Y144-			S:Y144S	S:Y144-
S:N501Y	S:N501Y	S:N501Y						S:N501Y	
			S:L452R	S:L452R			S:L452Q		
S:H69-					S:H69-				S:H69-
S:V70-					S:V70-				S:V70-
	S:L18F	S:L18F							
		S:T1027I							S:T1027I
			S:D950N					S:D950N	
	S:A701V					S:A701V			
	S:K417N	S:K417T							
		S:P26S							S:P26S
						S:D253G	S:D253N		
	S:A243-								S:A243-
	S:L242-								S:L242-
	S:L241-								S:L241-
						S:T95I		S:T95I	
S:D1118H									S:D1118H

Khảo sát sự lặp lại của các đột biến trên các biến thể nguy hiểm được phân loại bởi WHO, chúng tôi nhận thấy có một biến thể là B.1.620 chưa được WHO đặt tên nhưng lại chứa rất nhiều đột biến xuất hiện trên 9 biến thể nguy hiểm. Đột biến Spike D614G phổ biến trên cả 10 biến thể, đột biến Spike E848K xuất hiện trên 7 biến thể, còn lại các đột biến khác xuất hiện từ 2 đến 4 trên các biến thể (Bảng 2). Trong thời điểm đại dịch SARS-CoV-2 đang diễn ra trên toàn thế giới, các thông tin về các biến chủng rất quan trọng. Vì vậy, nghiên cứu được thực hiện nhằm tổng hợp các thông tin di truyền của các biến chủng cũng như xây dựng cây phát sinh loài để tìm hiểu mối liên hệ giữa các biến chủng khác nhau. Kết quả của nghiên cứu sẽ cung cấp các thông tin di truyền cần thiết cho các nghiên cứu tiếp theo về quá trình biến đổi của SARS-CoV-2 cũng như các mối liên hệ giữa các biến chủng để có kế hoạch phát hiện và tầm soát hiệu quả dịch bệnh không chỉ ở Việt Nam mà còn ở các nước khác.

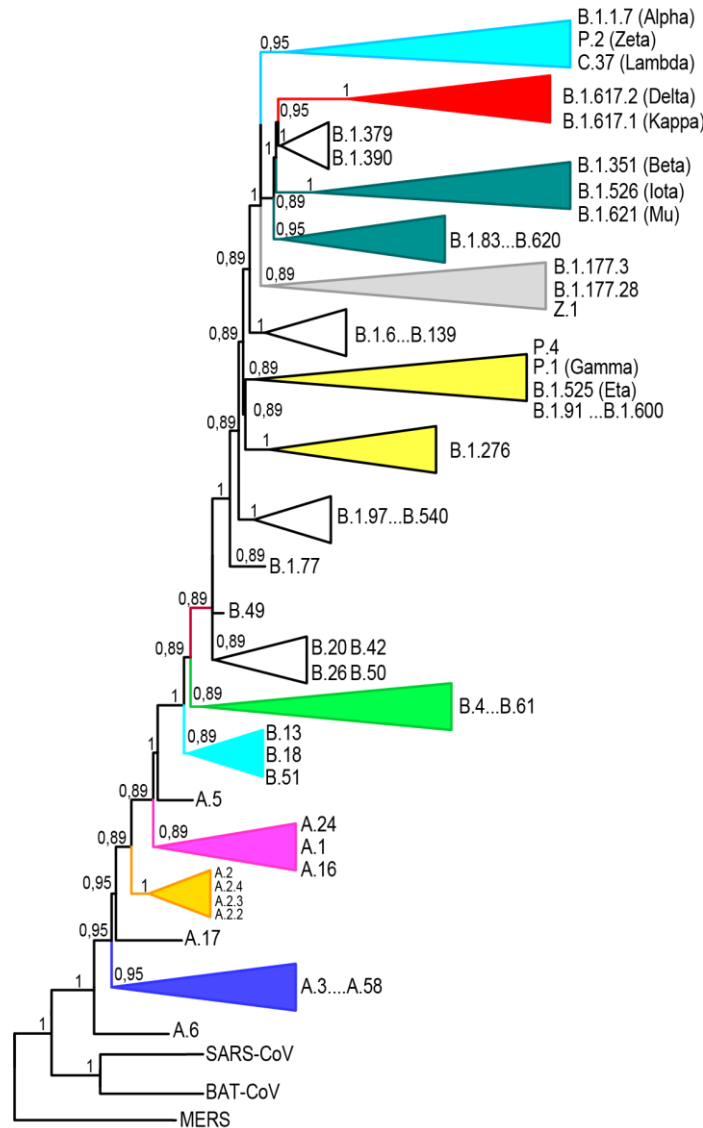
2 Vật liệu và phương pháp

Sử dụng dữ liệu từ GISAID (<https://www.gisaid.org/>), NCBI (<https://www.ncbi.nlm.nih.gov/>) và CNCB/NGDC (<https://ngdc.cncb.ac.cn/ncov/>), chúng tôi thu nhận được 1284 trình tự genome của chủng MERS, BAT-CoV, SARS-CoV và các biến thể SARS-CoV-2 đã được ghi nhận tính đến tháng 8/2021. Các trình tự thu nhận trên chỉ tiêu là trình tự genome hoàn chỉnh và có độ bao phủ cao. Các trình tự mã hóa trong bộ gen virus đã được trích xuất và sắp xếp thẳng hàng bằng chương trình MUSCLE trong phần mềm Geneious Prime. Sau đó, cơ sở dữ liệu đã được xử lý bởi chương trình jModelTest để tìm ra mô hình tiến hóa phù hợp nhất [16-17]. Phương pháp Bayesian Inference đã được thực hiện bằng mô hình thay thế với mô hình tỉ lệ phân phối gamma (gamma-distributed-rates) và 10 000 000 lặp lại bằng phần mềm MrBayes [18]. Hình cây phát sinh loài được hiệu chỉnh và minh họa bằng chương trình Figtree.

3 Kết quả và thảo luận

Dựa vào kết quả phân tích phả hệ, cây phát sinh loài chia dữ liệu trình tự đầu ra thành 24 nhóm bao gồm: nhóm ngoại là chủng MERS, nhóm BAT-CoV, nhóm SARS-CoV, nhóm A.6, nhóm {A.3 đến A.58}, nhóm A.17, nhóm các biến chủng A.2, nhóm {A.1, A.16, A.24}, nhóm A.5, nhóm {B.13, B.18, B.51}, nhóm {B.4 đến B.61}, nhóm {B.20, B.26, B.42, B.50},

nhóm B.49, nhóm B.1.77, nhóm {B.1.97 đến B.1.540}, nhóm B.1.276, nhóm biến chủng Eta-Gamma, nhóm {B.1.6 đến B.139}, nhóm B.1.177, nhóm {B.1.83 đến B.620}, nhóm có các biến chủng Beta-Iota-Mu, nhóm {B.1.379, B.1.390}, nhóm bao gồm biến chủng Delta-Kappa và nhóm của các biến chủng Alpha-Zeta-Lambda (Hình 3). Hình cây phả hệ gốc với chi tiết các nhóm sẽ được gửi cho các độc giả có yêu cầu qua thư điện tử cho tác giả liên hệ.



Hình 3 Cây phả hệ các biến chủng của SARS-CoV-2 với SARS-CoV, BAT-CoV và MERS đã được hiệu chỉnh.

Các nhánh bao gồm hai biến chủng trở lên được thể hiện bằng hình tam giác với các màu sắc khác nhau để dễ quan sát mối quan hệ di truyền. Các số ở các nhánh thể hiện giá trị xác suất hậu nghiệm.

Nghiên cứu này tập trung phân tích sự đa dạng của SARS-CoV-2 với dữ liệu phân tích là hầu hết trình tự của các biến thể SARS-CoV-2. Biến thể A.6 có trình tự hệ gene giống với dạng tổ tiên chung gần nhất

(most recent common ancestor) vì nằm ở cuối của cây phả hệ [14]. Dòng A.6 là biến chủng được biết đến là biến chủng Thái Lan mang đột biến điển hình là A829T. Đột biến A829T và D614G là hai đột biến

quan trọng và xuất hiện sớm nhất được ghi nhận [19]. Ngoài ra, biến chủng A.6 cũng được ghi nhận từ Trung Quốc theo một thống kê trên 194 mẫu, bao gồm Thái Lan (183), Trung Quốc (3), Singapore (3), Nhật Bản (3) và Campuchia (2). Qua số liệu này có thể thấy dòng A.6 có thể đã lây lan từ sớm qua các nước khu vực châu Á [19]. Các thế hệ tiếp là các biến thể thuộc dòng A. Cho đến nay, các biến chủng thuộc dòng A chưa được ghi nhận làm tăng sự nguy hiểm của virus như lây lan nhanh hơn, gây các triệu chứng bệnh nặng hơn hay làm giảm và kháng hiệu quả của vaccine như các biến thể thuộc hai nhóm VOC (Variants of Concern) và VOI (Variants of Interest). Hậu duệ xa nhất của biến chủng A là chủng được gán nhãn A.5 với 267 biến đổi trong trình tự hệ gene và 93 biến đổi acid amin so với chủng gốc (Lineage Browse - Resource for Coronavirus 2019 (cncb.ac.cn)). Biến thể A được cho là là căn nguyên của đại dịch, gây bùng phát toàn cầu. Trong giai đoạn khởi đầu đại dịch là tháng 12/2019, người ta đã ghi nhận được biến thể A xuất hiện nhiều trên mẫu thu nhận từ Trung Quốc. Ngoài ra, nhiều mặt hàng xuất khẩu trên toàn thế giới, bao gồm Mỹ, Úc, Đông Nam Á, Nhật Bản và châu Âu cũng ghi nhận xuất hiện biến thể A [19]. Biến thể A.5, sau này đã gộp chung biến thể A.10 là biến thể đại diện của Tây Ban Nha và được ghi nhận tại một chuỗi bùng phát ở Scotland [21]. Biến thể A.5 cũng được ghi nhận tại Trung Quốc [21]. Ở Việt Nam, người ta đã ghi nhận hai biến thể trong giai đoạn khởi phát dịch là A.1 và A.23.1 [21].

Đối với dòng B, sự biến đổi genome của SARS-CoV-2 là rất đa dạng, được thể hiện trên cây phát sinh loài là phân nhánh rất nhiều (Hình 1). Bên cạnh đột biến D614G trên gen mã hóa protein gai giống biến chủng A, các biến chủng dòng B có thêm nhiều đột biến trên protein gai (Bảng 1 và Bảng 2). Một đặc điểm chung có thể thấy là các biến thể nguy hiểm được WHO xếp vào hai nhóm VOC và VOI đều nằm ở những nhánh cuối, xa gốc cây phát sinh loài nhất. Kết quả này cho thấy trong quá trình bùng phát dịch, SARS-CoV-2 đã tích lũy đột biến liên tục, vượt qua nhiều hàng rào miễn dịch trên nhiều cộng đồng và kết quả là SARS-CoV-2 đã tích lũy nhiều đột biến trên gene mã hóa protein gai (Bảng 1, Bảng 2) giúp cho việc gây bệnh COVID-19 trở nên nhanh và nguy hiểm hơn. Ví dụ cụ thể là bên cạnh các đột biến giống các biến thể khác thì các biến đổi đặc trưng của D950N và T478K trong

gene mã hóa protein gai đã làm cho biến chủng Delta tăng khả năng lây nhiễm và trở thành biến chủng phổ biến ở nhiều quốc gia trên thế giới (Bảng 1 và Bảng 2). Trong nghiên cứu này, các biến thể thuộc hai nhóm nhóm VOC và VOI đã được chúng tôi xây dựng cây phả hệ cùng các biến chủng khác để có được bức tranh toàn cảnh về sự phát triển và đa dạng di truyền của SARS-CoV-2. Biến thể Gamma hay còn gọi là biến chủng Brazil có sự tương đồng cao về trình tự nucleotide với biến thể Eta (97,1 %). Biến thể Nam Phi là Beta thì có sự tương đồng cao với biến thể Iota (99,4 %) và Mu (99,3 %). Biến thể rất nguy hiểm và đang gây lo ngại cho các cơ quan y tế trên toàn thế giới là biến thể Delta có quan hệ họ hàng với biến thể Kappa (99,6 %). Biến thể Alpha có quan hệ gần gũi với biến thể Zeta (99,8 %) và Lambda (99,8 %). Zeta và Lambda là hai biến thể đại diện cho châu Âu và Nam Mỹ. Như vậy, bốn nhóm là bốn nhánh phát sinh các biến thể nguy hiểm đã được xác định.

Nhóm biến thể Gamma và Eta: biến thể Eta được phát hiện lần đầu ở Anh và Nigeria trong khi biến thể Gamma được phát hiện trên bốn người Nhật trở về từ Brazil, sau đó biến thể này được phát hiện rộng rãi trên Brazil. Hai biến thể này được phân nhánh chung từ một nhánh tuy nhiên chỉ số xác suất hậu nghiệm là 0,89. Trong nghiên cứu này, hai biến chủng Gamma và Eta của SARS-CoV-2 có sự tương đồng nucleotide chỉ có 97,1 % nên có thể kết luận hai biến chủng này là hai hậu duệ không quá gần gũi như các biến chủng khác với độ tương đồng trên 99 %. Tuy nhiên, hai biến chủng này xuất phát từ một nhánh nên chúng có cùng tổ tiên.

Nhóm biến thể Gamma, Iota và Mu: biến thể Beta gây ra đợt bùng phát dịch nghiêm trọng ở Nam Phi trong khi Iota được tìm thấy lần đầu tại New York còn biến thể Mu tìm thấy tại các nước Nam Mỹ và châu Âu. Chỉ số xác suất hậu nghiệm là 1 thể hiện mức độ rất gần gũi của ba biến thể này. Cả ba biến thể này đều được phát hiện vào cuối năm 2020 và đầu năm 2021, có nghĩa là vào cuối năm 2020 thì tổ tiên của chúng đã được lây lan rất nhanh đến ba châu lục khác nhau và các hậu duệ đã tích lũy đột biến và sinh ra ba chủng rất nguy hiểm cho con người.

Nhóm biến thể Delta và Kappa: hai biến thể này xuất hiện tại Ấn Độ vào cuối năm 2020 và thể hiện trên cây phát sinh loài là gần gũi với nhau, có chỉ số xác suất hậu nghiệm là 1. Đây là nhóm biến thể thể hiện rõ nhất sự di truyền liên quan đến địa lí.

Nhóm biến thể Alpha, Zeta và Lambda: biến thể Alpha rất nổi tiếng với tên gọi khác là biến thể Anh còn Zeta và Lambda thì được phát hiện ở Nam Mỹ, cụ thể lần lượt là Brazil và Peru. Như vậy, có thể thấy sự tự do di lại tại châu Âu và Nam Mỹ có thể là nguyên nhân phát tán nhanh SARS-CoV-2.

4 Kết luận

Nghiên cứu này sử dụng các trình tự đã công bố trên các cơ sở dữ liệu của thế giới như GISAID, NCBI và CNCB/NGDC. Chúng tôi đã vẽ ra bức tranh toàn cảnh về sự phát triển, tiến hóa của SARS-CoV-2 trong đại dịch toàn cầu. Nghiên cứu này đã phần nào cập nhật được sự đa dạng di truyền của SARS-CoV-2, là tiền đề cho các nghiên cứu sau này liên quan đến SARS-CoV-2. Dựa vào kết quả phân tích phả hệ, có thể kết luận biến chủng A là nguồn gốc phát sinh các biến chủng khác. Sau đó sự biến đổi của SARS-CoV-2

diễn ra theo chọn lọc tự nhiên và kết quả là sự xuất hiện của 4 biến chủng được WHO xếp vào nhóm VOC, 5 biến chủng được WHO xếp nhóm VOI và 12 biến chủng nguy hiểm đang được nghiên cứu phân loại. Các nhóm biến thể VOC và VOI đã được chúng tôi xây dựng phả hệ nhằm xác định nguồn gốc. Kết quả là bốn nhánh phát sinh cho bốn nhóm biến thể nguy hiểm được xác định cụ thể là: biến thể Gamma có sự tương đồng về trình tự với biến thể Eta. Biến thể Beta thì có sự tương đồng cao với biến thể Iota và Mu. Biến thể Delta có quan hệ họ hàng với biến thể Kappa. Biến thể Alpha có quan hệ gần gũi với biến thể Zeta và Lambda.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Quỹ Phát triển Khoa học và Công nghệ - Đại học Nguyễn Tất Thành, đề tài mã số 2021.01.121/HĐ-NCKH.

Tài liệu tham khảo

1. N. Zhu et al., “A Novel Coronavirus from Patients with Pneumonia in China, 2019,” *N. Engl. J. Med.*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.
2. P. Zhou et al., “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *Nature*, vol. 579, no. 7798, pp. 270–273, 2020, doi: 10.1038/s41586-020-2012-7.
3. A. Rahimi, A. Mirzazadeh, and S. Tavakolpour, “Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection,” *Genomics*, vol. 113, no. 1, pp. 1221–1232, Jan. 2021, doi: 10.1016/j.ygeno.2020.09.059.
4. “Viet Nam: WHO Coronavirus Disease (COVID-19) Dashboard With Vaccination Data | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data.” <https://covid19.who.int/region/wpro/country/vn> (accessed Sep. 10, 2021).
5. X. Xu et al., “Effective treatment of severe COVID-19 patients with tocilizumab,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 20, pp. 10970–10975, May 2020, doi: 10.1073/pnas.2005615117.
6. J. Grein et al., “Compassionate Use of Remdesivir for Patients with Severe Covid-19,” *N. Engl. J. Med.*, p. NEJMoa2007016, Apr. 2020, doi: 10.1056/NEJMoa2007016.
7. A. Nili, A. Farbod, A. Neishabouri, M. Mozafarhashjin, S. Tavakolpour, and H. Mahmoudi, “Remdesivir: A beacon of hope from Ebola virus disease to COVID-19,” *Rev. Med. Virol.*, vol. 30, no. 6, pp. 1–13, Nov. 2020, doi: 10.1002/rmv.2133.
8. F. Wu et al., “A new coronavirus associated with human respiratory disease in China,” *Nature*, vol. 579, no. 7798, pp. 265–269, Mar. 2020, doi: 10.1038/s41586-020-2008-3.
9. R. Lu et al., “Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding,” *Lancet Lond. Engl.*, vol. 395, no. 10224, pp. 565–574, Feb. 2020, doi: 10.1016/S0140-6736(20)30251-8.
10. A. Wu et al., “Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China,” *Cell Host Microbe*, vol. 27, no. 3, pp. 325–328, Mar. 2020, doi: 10.1016/j.chom.2020.02.001.

11. A. R. Fehr and S. Perlman, “Coronaviruses: An Overview of Their Replication and Pathogenesis,” *Coronaviruses*, vol. 1282, pp. 1–23, Feb. 2015, doi: 10.1007/978-1-4939-2438-7_1.
12. Y. Ruan et al., “Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection,” *Lancet Lond. Engl.*, vol. 361, no. 9371, pp. 1779–1785, May 2003, doi: 10.1016/S0140-6736(03)13414-9.
13. J. F.-W. Chan et al., “Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan,” *Emerg. Microbes Infect.*, vol. 9, no. 1, pp. 221–236, Jan. 2020, doi: 10.1080/22221751.2020.1719902.
14. A. Rambaut et al., “A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology,” *Nat. Microbiol.*, vol. 5, no. 11, pp. 1403–1407, Nov. 2020, doi: 10.1038/s41564-020-0770-5.
15. “Tracking SARS-CoV-2 variants.” <https://www.who.int/activities/tracking-SARS-CoV-2-variants> (accessed Sep. 10, 2021).
16. R. C. Edgar, “MUSCLE: a multiple sequence alignment method with reduced time and space complexity,” *BMC Bioinformatics*, vol. 5, no. 1, p. 113, Aug. 2004, doi: 10.1186/1471-2105-5-113.
17. D. Darriba, G. L. Taboada, R. Doallo, and D. Posada, “jModelTest 2: more models, new heuristics and parallel computing,” *Nat. Methods*, vol. 9, no. 8, pp. 772–772, Aug. 2012, doi: 10.1038/nmeth.2109.
18. J. P. Huelsenbeck and F. Ronquist, “MRBAYES: Bayesian inference of phylogenetic trees,” *Bioinforma. Oxf. Engl.*, vol. 17, no. 8, pp. 754–755, Aug. 2001, doi: 10.1093/bioinformatics/17.8.754.
19. A. K. Banerjee, F. Begum, and U. Ray, “Mutation Hot Spots in Spike Protein of COVID-19,” Apr. 2020, doi: 10.20944/preprints202004.0281.v1.
20. G. Z et al., “An online coronavirus analysis platform from the National Genomics Data Center,” *Zool. Res.*, vol. 41, no. 6, Nov. 2020, doi: 10.24272/j.issn.2095-8137.2020.065.
21. “Lineage Browse - 2019 Novel Coronavirus Resource.” <https://ngdc.cncb.ac.cn/ncov/lineage?lang=en> (accessed Sep. 07, 2021).

Genetic diversity of SARS-CoV-2

Do Hoang Dang Khoa, Ho Ta Giap, Phan Ngoc Han
Hi-Tech Institute, Nguyen Tat Thanh University
dhdtkhoa@ntt.edu.vn

Abstract The rapidly spreading disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is endangering people’s health, lives and the entire global economy. By the beginning of 2021, the strategies of COVID-19 vaccination and method of treatment with drugs supplement have been implemented and researched in many countries. The genetic modifications of SARS-CoV-2 that give them resistance to drugs and vaccines for the virus survival and infection ability have also been studied. At the same time, new dangerous SARS-CoV-2 strains are continuously being announced by WHO such as Alpha, Beta, Gamma, Delta, Eta, Iota, Kappa, Lambda, Mu and many unnamed strains. Therefore, genetic diversity analysis is essential to provide an overview of the spread of SARS-CoV-2 variables, providing benefit information for epidemiology researches and other genetic studies. The results of this study suggest that the Alpha variant could be the source of the global pandemic. We also identified four groups of dangerous variants: The Gamma variant shares sequence similarity with the Eta variant, the Beta variant is highly similar to the Iota and Mu variants, the Delta variant is related to the Kappa variant, and the Alpha variant is closely related to the Zeta and Lambda variants.

Keywords COVID-19, SARS-CoV-2, Genetic diversity, variant

