

XÁC ĐỊNH BIẾN THỂ SARS-COV-2 TRÊN MẪU BỆNH NHÂN BẮC GIANG TRONG VỤ DỊCH 2021 BẰNG CÔNG NGHỆ GIẢI TRÌNH TỰ THỂ HỆ MỚI

Nguyễn Thị Hải Hà¹, Phạm Lê Anh Tuấn², Nguyễn Thị Tâm³, Nguyễn Thị Ngọc Lan²
Hoàng Hồng Thắm⁴, Tạ Thành Văn² và Trần Văn Khánh^{2,✉}

¹Trường Đại học Khoa học tự nhiên, Đại học Quốc gia Hà Nội

²Trường Đại học Y Hà Nội

³Đơn vị nghiên cứu lâm sàng Đại học Oxford Hà Nội

⁴Viện nghiên cứu dữ liệu lớn VinBigData - Tập đoàn VinGroup

Đại dịch của hội chứng viêm đường hô hấp cấp (COVID-19) do SARS-CoV-2 là mối đe dọa toàn cầu và đã gây ra những tổn thất nghiêm trọng về kinh tế và xã hội ở Việt Nam cũng như trên toàn thế giới. Việc xác định các biến thể SARS-CoV-2 đang lưu hành có ý nghĩa quan trọng trong việc theo dõi sự tiến hóa của virus, góp phần thúc đẩy nghiên cứu vắc xin và sinh phẩm chẩn đoán. Nghiên cứu này được thực hiện trên 38 bệnh nhân được chẩn đoán dương tính với SARS-CoV-2 từ khu công nghiệp tỉnh Bắc Giang với mục tiêu: Xác định biến thể SARS-CoV-2 và các biến đổi đặc thù trong bộ gen virus bằng phương pháp giải trình tự gen thế hệ mới (NGS). Độ tuổi trung bình tham gia nghiên cứu là $40,16 \pm 14,07$. Kết quả đã phát hiện được 38 mẫu virus SARS-CoV-2 thuộc biến chủng Delta, 16 biến đổi xuất hiện ở tất cả các mẫu bệnh phẩm, 27 biến đổi đặc thù chỉ xuất hiện ở một số mẫu trong đó có 3 biến đổi trên gen cấu trúc bao gồm S: Leu5Phe, Pro819Ser, His1101Tyr.

Từ khóa: SARS-CoV-2, NGS, đột biến.

I. ĐẶT VẤN ĐỀ

Bệnh viêm đường hô hấp cấp tính được gây ra bởi một loại virus có tên virus corona gây hội chứng hô hấp cấp tính nặng 2 (SARS-CoV-2), lần đầu xuất hiện vào cuối tháng 12 năm 2019 ở Vũ Hán, tỉnh Hồ Bắc, Trung Quốc. Sau đó, dịch bệnh đã bùng phát toàn cầu với tốc độ lây lan nhanh chóng và gây ra rất nhiều hậu quả nghiêm trọng. Tổ chức Y tế Thế giới (WHO) chính thức đặt tên cho căn bệnh này là "COVID-19". COVID-19 ảnh hưởng đến ít nhất 216 quốc gia trên thế giới. Theo số liệu của WHO, đã có khoảng 590.366.194 ca mắc và 6.438.794 trường hợp tử vong được xác nhận

trên toàn cầu tính đến ngày 09/08/2022.

Coronavirus (CoVs) là thành viên của chi Betacoronavirus, họ Nidovirales. SARS-CoV-2 sở hữu 80% đặc điểm phát sinh loài với coronavirus gây hội chứng hô hấp cấp tính nghiêm trọng (SARS-CoV) và 50% giống với coronavirus gây hội chứng hô hấp Trung Đông (MERS-CoV). Các betacoronavirus đã gây ra hai đại dịch SARS-CoV năm 2002-2003 và MERS-CoV năm 2012 - 2013 với khả năng truyền bệnh từ động vật sang người. Hệ gen RNA đơn của các coronavirus có kích thước lớn, cỡ 26 - 32 kb trong đó SARS-CoV-2 có kích thước khoảng 29.9 kb. Hệ gen chính của nhóm virus này chứa các gen cơ bản, có thứ tự bất biến là 5'-*replicase*-S-E-M-N-3' trong đó riêng vùng *replicase* chiếm hơn 2/3 kích thước vùng mã hóa. Replicase - transcriptase cũng là protein duy nhất được dịch mã trực tiếp từ

Tác giả liên hệ: Trần Văn Khánh

Trường Đại học Y Hà Nội

Email: tranvankhanh@hmu.edu.vn

Ngày nhận: 27/12/2022

Ngày được chấp nhận: 17/01/2023

gRNA. Bên cạnh các gen cơ bản, SARS-CoV-2 vẫn bảo thủ mang khung đọc mở ORF1ab, theo sau là các gen S, E, M, N; các gen phụ trợ bao gồm ORF3a, 6, 7a, 8 và 10.¹

Tính đến năm 2020, số ca mắc COVID-19 được ghi nhận tại Việt Nam là dưới 1500. Tuy nhiên, đợt bùng phát COVID-19 thứ tư tại Việt Nam bắt đầu vào ngày 27 tháng 4 năm 2021 là làn sóng dịch mang đến nhiều tổn hại nghiêm trọng. Cụ thể, Bắc Ninh, Bắc Giang từng là tâm dịch lớn nhất cả nước, kéo dài từ đầu tháng 5 đến đầu tháng 7 năm 2021, do bùng phát dịch trong các khu công nghiệp. Tất cả các trường hợp được xác định kể từ ngày 19 tháng 5 năm 2021, đều bắt nguồn từ biến thể B.1.617.2 (Delta) mới của SARS-CoV-2. Biến thể Delta ban đầu được tìm thấy ở Ấn Độ vào tháng 12 năm 2020, sau đó được phân loại là biến thể đáng quan tâm VOI (Variant of Interest) vào ngày 04/04/2021 và biến thể đáng quan ngại VOC (Variant of Concern) vào ngày 11/05.

Ngày nay, với kỹ thuật giải trình tự thế hệ mới (Next Generation Sequencing - NGS), việc giải mã bộ gen virus đã trở nên đơn giản và nhanh chóng hơn nhiều. Kỹ thuật NGS cho phép các nhà khoa học giải trình tự toàn bộ hệ gen của nhiều mẫu bệnh phẩm virus cùng lúc với độ bao phủ cao, cho phép phát hiện nhiều biến đổi dù là nhỏ nhất. Trong nghiên cứu này, sử dụng kỹ thuật NGS trên các mẫu bệnh phẩm bệnh nhân nhiễm SARS-CoV-2 được thu thập tại Bắc Giang, chúng tôi đã xác định được hệ gen hoàn chỉnh của quần thể virus tại đây cũng như các biến đổi đặc thù đáng lưu ý.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

Mẫu dịch tỵ hầu từ 38 bệnh nhân được thu thập tại khu công nghiệp Tiên Sơn, tỉnh Bắc Giang. Các bệnh nhân được chẩn đoán mắc hội chứng viêm đường hô hấp cấp do virus

SARS-Cov-2 và nhập viện điều trị vào tháng 5 và tháng 6/2021.

2. Phương pháp

Tách chiết RNA và xác định chu kỳ ngưỡng (Ct) của mẫu bệnh phẩm

RNA của virus SARS-CoV-2 được tách chiết từ dịch tỵ hầu của bệnh nhân được chẩn đoán dương tính với SARS-CoV-2 theo kit QIAamp Viral RNA (QIAGEN, Đức). Sau đó tiến hành phản ứng Real-time PCR để xác định chu kỳ ngưỡng (Ct). Mẫu có Ct ≤ 30, nồng độ đạt trên 100 ng/μL được sử dụng cho giải trình tự gen thế hệ mới.

Tạo lập thư viện giải trình tự

Nghiên cứu đã tiến hành tạo lập thư viện giải trình tự bằng cách sử dụng quy trình của Illumina COVIDseq (Illumina, Mỹ). Tổng hợp cDNA từ mẫu RNA bằng enzyme phiên mã ngược sử dụng hexamers. Tiếp theo, genome virus SARS-CoV-2 được khuếch đại bằng PCR. Sau đó các đoạn gen được cắt thành các đoạn nhỏ và gắn đuôi A vào đầu phân đoạn DNA. Tinh sạch sản phẩm sau khi phân mảnh DNA. Tiếp theo, gắn index vào các phân đoạn DNA và khuếch đại bằng PCR. Tinh sạch sản phẩm PCR và định lượng bằng kit Qubit dsDNA HS Assay Kit. Tính toán nồng độ của từng mẫu dựa vào kích thước và nồng độ đo Qubit và pha loãng từng mẫu để đạt cùng một nồng độ (2nM). Các mẫu được trộn vào nhau (Pool), biến tính và tách mạch DNA bằng NaOH. Pha loãng mẫu Pool sau khi tách mạch về nồng độ 10 - 12pM và biến tính ở 96°C trong 2 phút.

Giải trình tự thế hệ mới

Mẫu Pool (600μL) đã biến tính được cho vào Cartridge hóa chất giải trình tự. Chất lượng mẻ chạy được đánh giá sơ bộ thông qua các chỉ số Cluster Density từ 1000 - 1200 K/mm² đối với kit 300V2 trên hệ thống Miseq, Passing Filter ≥ 80%, Q30 > 70%. Kết quả cho base đầu

tiên sẽ đạt nếu các thông số Q30 > 70%, BIC > 80%, FIT > 10%. Tiến hành giải trình tự lại nếu không đạt được các thông số trên.

Xử lý số liệu

Xử lý số liệu thu thập được bằng phần mềm Trimgalore để cắt các chuỗi tiếp hợp (adapter), kiểm tra chất lượng giải trình tự bằng FastQC, giống hàng (align) các trình tự SARS-CoV-2 với trình tự gốc của virus bằng phần mềm BWA. Tiếp theo, những biến đổi nhỏ được phân tích

bằng công cụ phân tích biến thể iVar (Intrahost variant analysis of replicates) để loại bỏ các trình tự môi khỏi dữ liệu giải trình tự dựa trên amplicon. Biến chủng được định danh bằng phần mềm Pangolin và NextClade.

3. Đạo đức nghiên cứu

Nghiên cứu này đã được chấp thuận bởi Hội đồng Đạo đức trong nghiên cứu Y sinh học của Trường Đại học Y Hà Nội số 753/QĐ-BKHCN. Các thông tin cá nhân hoàn toàn được bảo mật.

III. KẾT QUẢ

1. Đặc điểm của đối tượng nghiên cứu

Bảng 1. Đặc điểm của bệnh nhân trong nghiên cứu

Giới tính	Tổng số (n = 38)	Tỷ lệ (%)
Nam	18	47,37
Nữ	20	52,63
Tuổi trung bình	40,16 ± 14,07	

Nghiên cứu thực hiện trên 38 mẫu bệnh phẩm được chẩn đoán mắc hội chứng viêm đường hô hấp cấp do virus SARS-CoV-2, trong đó tỉ lệ nam giới và nữ giới lần lượt là 47,37%

và 52,63%. Độ tuổi trung bình tham gia nghiên cứu là 40,16 ± 14,07 (Bảng 1). Bệnh nhân được lựa chọn có độ tuổi nằm trong khoảng từ 21 đến 86 tuổi.

Bảng 2. Số lượng đột biến chỉ điểm cho biến thể Delta được phát hiện

Tổng số (n = 38)	Số đột biến chỉ điểm cho biến thể Delta được phát hiện		
	Toàn bộ 16 đột biến (xác nhận)	11 đến 15 đột biến (khả năng cao)	Dưới 11 đột biến (khả năng thấp)
	37/38 (97,37%)	1/38 (3,63%)	0/38 (0,00%)

Tại thời điểm thu thập mẫu nghiên cứu tại tháng 5/2021, biến chủng Delta được công nhận là biến chủng phổ biến nhất trên thế giới nói chung và tại Việt Nam nói riêng. Theo CDC Hoa Kỳ, những bệnh phẩm mang toàn bộ 16/16 đột biến chỉ điểm cho biến chủng Delta được xác nhận là mẫu bệnh phẩm của biến chủng này, trong khi những mẫu mang từ 11 đến 15

đột biến chỉ điểm được coi là mẫu có khả năng cao và dưới 11 đột biến được coi là mẫu có khả năng thấp. Trong số 38 mẫu nghiên cứu, 37/38 mẫu mang toàn bộ 16 đột biến chỉ điểm và 1/38 mẫu mang 15/16 đột biến (Bảng 2). Qua đó, có thể xác nhận là toàn bộ 38 mẫu virus SARS-CoV-2 trong nghiên cứu thuộc biến chủng Delta.

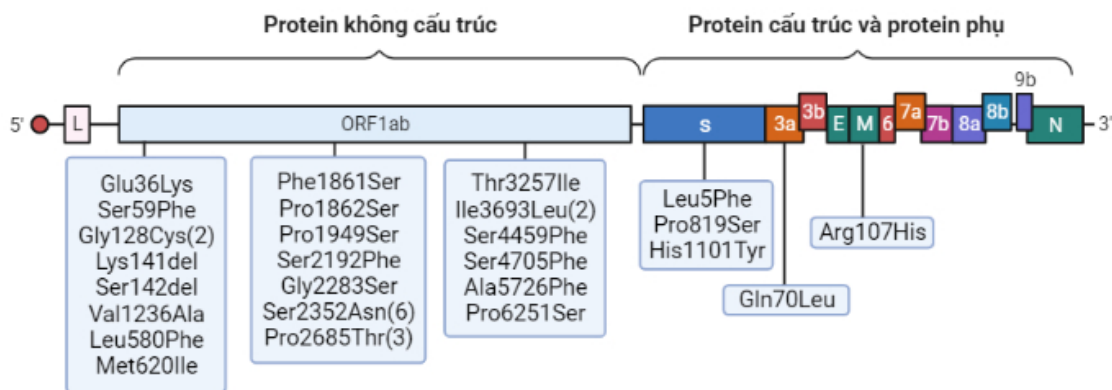
2. Các biến đổi đặc thù trong bộ gen virus

Ngoài những đột biến được coi là chỉ điểm cho các biến chủng kể trên, những biến đổi khác trong bộ gen (bao gồm cả đột biến và đa hình đơn - SNP), tuy không ảnh hưởng lớn tới những khả năng sinh trưởng, lây nhiễm hay né tránh vắc xin của virus, nhưng chúng lại là chỉ báo quan trọng để các nhà nghiên cứu có thể xác định con đường lây nhiễm cũng như dự

đoán xu hướng biến đổi của virus trong tương lai. Ở mỗi mẫu bệnh phẩm trong nghiên cứu này, chúng tôi đã xác định được nhiều đột biến khác nhau: Bao gồm cả những đột biến tương đồng trong tập hợp mẫu, và những đột biến đặc trưng của từng bệnh phẩm. Tổng hợp lại, chúng tôi tìm thấy 16 biến đổi xuất hiện phổ biến ở các mẫu bệnh phẩm (Bảng 3).

Bảng 3. Những biến đổi xuất hiện ở các mẫu bệnh phẩm

STT	Thay đổi	Gen	Tần suất (%)	STT	Thay đổi	Gen	Tần suất (%)
1	c.241C>T	5'-UTR	97,36	9	Val82Ala	ORF7a	100
2	Pro380Ser	ORF1ab	97,36	10	Thr120Ile	ORF7a	100
3	Gly662Ser	ORF1ab	97,36	11	Asp119del	ORF8	100
4	Pro1640Leu	ORF1ab	97,36	12	Phe120del	ORF8	100
5	Ala3209Val	ORF1ab	97,36	13	Thr60Ala	ORF9b	100
6	Val3718Ala	ORF1ab	97,36	14	Ala222Val	S	100
7	Thr3750Ile	ORF1ab	97,36	15	Gly671Ser	RdRp	97,36
8	Ser26Leu	ORF3a	100	16	c.29742G>T	3'-UTR	100



Hình 1. Những biến đổi đặc thù xuất hiện ở tất cả các mẫu bệnh phẩm

Những biến đổi xuất hiện ở một hoặc một số mẫu bệnh phẩm trên vùng gen tương ứng của chúng được biểu thị trên hình. Những biến đổi xuất hiện nhiều hơn một lần được đánh dấu với số trong ngoặc đơn là số lần xuất hiện của chúng

Tiếp theo, chúng tôi cũng xác định được 27 biến đổi mang tính đặc thù ở một hoặc một số mẫu bệnh phẩm. Trong đó, 23 biến đổi chỉ xuất hiện trên một mẫu bệnh phẩm duy nhất trong nghiên cứu và 4 biến đổi được tìm thấy ở từ hai mẫu trở lên. 27 biến đổi nói trên chủ yếu được tìm thấy trên gen ORF1ab của bộ gen virus SARS-CoV-2 (21/27 biến đổi), 1 biến đổi trên gen ORF3a, 1 biến đổi trên gen M và 3 biến đổi trên gen S (Hình 1). Đánh giá trên phương diện tần suất xuất hiện của đột biến, 24/27 biến đổi chỉ xuất hiện trên một mẫu bệnh phẩm nhất định, nhưng cũng có 4 biến đổi được tìm thấy trên nhiều hơn một mẫu bệnh phẩm. Cụ thể là: Biến đổi Gly128Cys được phát hiện trên mẫu S11 và S19; Ser2352Asn xuất hiện trên 6 mẫu S5, S6, S23, S25, S31, S34; Pro2685Thr xuất hiện trên các mẫu S8, S9 và S16; và cuối cùng là biến đổi Ile3693Leu xác định được trên các mẫu S12 và S18. Tuy nhiên, các biến đổi này chỉ phát sinh trên 21/38 mẫu bệnh phẩm (tương đương 55,26%), trong đó 16/38 mẫu chỉ phát sinh 01 biến đổi (42,11%), 2 mẫu bệnh phẩm mang 2 biến đổi (5,26%), 1 mẫu bệnh phẩm mang 3 biến đổi (2,63%), 1 mẫu mang 4 biến đổi (2,63%) và 1 mẫu phát sinh đến 6 biến đổi (2,63%).

Trong số này, một số biến đổi nhận được sự quan tâm, đặc biệt là những biến đổi trên gen cấu trúc, như các đột biến Leu5Phe, Pro819Ser, His1101Tyr trên gen S mã hóa protein gai (Spike) hay đột biến Arg107His trên gen M mã hóa protein màng (Membrane).

IV. BÀN LUẬN

Nghiên cứu giải trình tự bộ gen virus SARS-CoV-2 từ 38 mẫu bệnh phẩm bệnh nhân ở tỉnh Bắc Giang đem lại lợi ích cho cộng đồng và ngành Y tế Việt Nam nói riêng và toàn cầu nói chung trong bối cảnh hiểu biết về dịch bệnh SARS-CoV-2 vẫn còn nhiều hạn chế. Công nghệ giải trình tự thế hệ mới (NGS) có những ưu điểm tiến bộ về tốc độ giải trình tự rất nhanh, độ phân

giải cao và độ chính xác trong phân tích bộ gen, cho phép giải mã hàng triệu trình tự DNA cùng lúc. Nhờ vậy, nâng cao hiệu suất của quá trình giải mã bộ gen sinh vật. Nghiên cứu đóng góp tới sự hiểu biết toàn cầu về đột biến gen có liên quan tới tính miễn dịch cũng như điều trị bệnh lý do SARS-CoV-2, góp phần thúc đẩy nghiên cứu vắc xin, thuốc, sinh phẩm chẩn đoán.

Nghiên cứu cung cấp bằng chứng khoa học rất cụ thể về những biến đổi trong bộ gen virus SARS-CoV-2 xuất hiện ở tỉnh Bắc Giang. Đối tượng nghiên cứu là những bệnh nhân có độ tuổi trung bình $40,16 \pm 14,07$. Độ tuổi trung bình của nghiên cứu này tương đồng với nghiên cứu của Kushwaha Savitesh (2021).² Độ tuổi càng cao thì khả năng bị nhiễm và tiến triển bệnh theo hướng nặng càng tăng. Về giới tính, tỷ lệ phân bố bệnh giữa nam và nữ xấp xỉ 1:1. Tỷ lệ mắc bệnh ở nữ (52,63%) cao hơn nam (47,37%), tuy nhiên chênh lệch không đáng kể. Địa điểm nghiên cứu ở tỉnh Bắc Giang là một nơi tập trung nhiều khu công nghiệp và công nhân từ nhiều nơi đến làm việc và sinh sống. Khi đợt dịch bùng phát làm các ca lây nhiễm tăng nhanh và lan rộng ra cộng đồng, tình hình dịch bệnh hết sức phức tạp và khó lường khiến Bắc Giang trở thành một trong những tâm dịch lớn nhất cả nước.

Ở 38 bệnh nhân nghiên cứu đều xuất hiện biến thể Delta (B.1.617.2). Biến thể Delta làm gia tăng nhanh chóng các ca nhiễm trên toàn thế giới và có tốc độ lây lan cao hơn 60% so với biến thể Alpha. Vào thời điểm tháng 5 năm 2021, biến thể Delta là biến thể chiếm ưu thế. Delta của SARS-CoV-2 (B.1.617.2) chứa 16 đột biến, 10 trong số những đột biến đó nằm trên protein gai (S).³ Các đột biến trên protein gai (S) làm tăng khả năng bám vào thụ thể tế bào vật chủ. Với hệ số lây nhiễm $R_0 = 4.9$ và tải lượng virus cao hơn nhiều so với chủng gốc làm cho biến thể Delta tốc độ lây lan cao hơn và chiếm 98% ca bệnh mới toàn cầu.⁴

Về các đột biến xuất hiện phổ biến trên SARS-CoV-2, đột biến c.241C>T ở vùng 5'UTR là đột biến phổ biến ở tất cả các châu lục trên thế giới trong khi đột biến P380S ở vùng ORF1ab phổ biến ở Bắc Mỹ. A3209V, V3718A, T3750I là những đột biến xuất hiện ở châu Á.⁵ G662S là đột biến chỉ xuất hiện duy nhất ở Delta, các đột biến ở vùng ORF1ab có thể làm tăng khả năng nhân lên và kháng thuốc. Các đột biến S26L và V82A xuất hiện chủ yếu ở biến thể Delta (B.1.617.2).⁶ A222V là đột biến ở protein gai S được phát hiện ở Tây Ban Nha vào mùa hè năm 2020, S: A222V thúc đẩy tăng độ mở của vùng gắn thụ thể (RBD) và tăng ái lực liên kết ACE2 so với S: D614G.⁷ Đột biến G671S ở vùng RdRp của biến thể Delta là nguyên nhân gây ra tải lượng virus cao. Đột biến c.29742G>T là đột biến ở vùng 3'UTR xuất hiện phổ biến ở Thái Lan, Đức, Úc và Canada, Trung Quốc và Anh cũng có tỷ lệ tương tự.⁸ Đột biến điểm trên bộ gen là những yếu tố đóng góp lớn vào sự tiến hóa của SARS-CoV-2. Tỷ lệ cao các đột biến được phát hiện trong gen S, đặc biệt là trong vùng liên kết thụ thể (RBD) và vùng đầu cuối N (NTD), có khả năng là do sự chọn lọc thay thế để cải thiện đặc tính của virus, thông qua những thay đổi trong cấu trúc của protein đột biến dẫn đến cải thiện liên kết với thụ thể vật chủ hoặc thoát khỏi sự nhận biết kháng thể.⁹

Về các đột biến đặc thù của SARS-CoV-2, đột biến L5F có thể tạo điều kiện thuận lợi cho việc lắp ráp protein và giải phóng virus khỏi tế bào vật chủ. L5F kết hợp với D614G làm tăng khả năng lây nhiễm so với chủng gốc.¹⁰ Nghiên cứu của Guo (2020), L5F tăng khả năng bám dính epitope cho 37 allele HLA khác nhau.¹¹ H1101Y là đột biến chỉ điểm ở biến thể Theta, H1101Y làm thay đổi vùng RBD và FP (Fusion Protein).¹² P809S là vị trí đột biến nằm cách vị trí S2' 6 amino acid, P809S có thể giúp tăng khả năng phân cắt ở vị trí S2', P809S xuất hiện ở USA, Scotland, Nam Phi, Đức.¹³ M: R107H mặc

dù là đột biến có tần số tương đối thấp nhưng nó xuất hiện ở rất nhiều các biến thể VOC như alpha, beta, gamma và delta.¹⁴

Đây là nghiên cứu đầu tiên xác nhận nguyên nhân bùng phát dịch ở Bắc Giang do biến thể Delta gây ra. Các phát hiện của nghiên cứu này cho thấy rằng với sự tiến triển của đại dịch, số lượng đột biến đã tăng lên đáng kể, điều này cho thấy sự tiến hóa thích nghi của SARS-CoV-2 ở người để tăng khả năng lây nhiễm và có thể là nguyên nhân gây giảm hiệu quả của vắc xin. Ngoài những yếu tố trên, các yếu tố về môi trường sống cũng có thể góp phần vào sự lây lan của COVID-19. Vì vậy, chúng tôi cần thêm các nghiên cứu sâu hơn, cỡ mẫu lớn hơn để chứng minh đầy đủ ý nghĩa của những đột biến này và tìm thêm những đột biến mới trên quần thể người Việt Nam.

V. KẾT LUẬN

Bằng kỹ thuật giải trình tự NGS ở 38 mẫu bệnh nhân thu thập tại khu công nghiệp tỉnh Bắc Giang. Nghiên cứu đã phát hiện 38/38 mẫu thuộc biến chủng Delta, xác định được 16 biến đổi xuất hiện ở tất cả các mẫu bệnh phẩm, 27 biến đổi đặc thù chỉ xuất hiện ở một số mẫu trong đó có 4 biến đổi trên gen cấu trúc bao gồm protein gai S: Leu5Phe, Pro819Ser, His1101Tyr và protein màng M: Arg107His.

LỜI CẢM ƠN

Nghiên cứu được thực hiện dưới sự hỗ trợ kinh phí của đề tài cấp Bộ Khoa học và Công nghệ "Nghiên cứu xây dựng cơ sở dữ liệu đột biến gene của SARS-CoV-2 tại Việt Nam" số quyết định phê duyệt ĐTDL.CN.102/21, thực hiện từ 7/2021 - 12/2023.

TÀI LIỆU THAM KHẢO

1. MacLean OA, Lytras S, Weaver S, et al. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and

highly capable human pathogen. *PLOS Biology*. 2021; 19(3). doi:10.1371/journal.pbio.3001115.

2. Suratekar R, Ghosh P, Niesen MJM, et al. High diversity in Delta variant across countries revealed by genome-wide analysis of SARS-CoV-2 beyond the Spike protein. *Molecular Systems Biology*. 2022; 18(2). doi:10.15252/msb.202110673.

3. Kushwaha S, Khanna P, Rajagopal V, Kiran T. Biological attributes of age and gender variations in Indian COVID-19 cases: A retrospective data analysis. *Clinical Epidemiology and Global Health*. 2021; 11. doi:10.1016/j.cegh.2021.100788.

4. Kang M, Xin H, Yuan J, et al. Transmission dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant infections in Guangdong, China, May to June 2021. *Eurosurveillance*. 2022; 27(10). doi:10.2807/1560-7917.ES.2022.27.10.2100815.

5. Guruprasad K. Geographical Distribution of Amino Acid Mutations in Human SARS-CoV-2 Orf1ab Poly-Proteins Compared to the Equivalent Reference Proteins from China. *ChemRxiv*. Published online 2020. doi:10.26434/chemrxiv-2021-lf2zd-v2.

6. Muttineni R, R.N. B, Putty K, et al. SARS-CoV-2 variants and spike mutations involved in second wave of COVID-19 pandemic in India. *Transboundary and Emerging Diseases*. Published online March 24, 2022. doi:10.1111/tbed.14508

7. Ginex T, Marco-Marín C, Wiczór M, et al. The structural role of SARS-CoV-2 genetic background in the emergence and success of spike mutations: The case of the spike A222V mutation. *PLOS Pathogens*. 2022; 18(7).

doi:10.1371/journal.ppat.1010631.

8. Mukherjee M, Goswami S. Global cataloging of variations in untranslated regions of viral genome and prediction of key host RNA binding protein-microRNA interactions modulating genome stability in SARS-CoV-2. *PLOS ONE*. 2020; 15(8). doi:10.1371/journal.pone.0237559.

9. McLean G, Kamil J, Lee B, et al. The Impact of Evolving SARS-CoV-2 Mutations and Variants on COVID-19 Vaccines. *mBio*. 2022; 13(2). doi:10.1128/mbio.02979-21.

10. Li Q, Wu J, Nie J, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*. 2020; 182(5). doi:10.1016/j.cell.2020.07.012.

11. Guo E, Guo H. CD8 T cell epitope generation toward the continually mutating SARS-CoV-2 spike protein in genetically diverse human population: Implications for disease control and prevention. *PLOS ONE*. 2020; 15(12). doi:10.1371/journal.pone.0239566.

12. Bascos NAD, Mirano-Bascos D, Saloma CP. Structural Analysis of Spike Protein Mutations in an Emergent SARS-CoV-2 Variant from the Philippines. *bioRxiv*. Published online 2021.

13. Pulakuntla S, Lokhande KB, Padmavathi P, et al. Mutational analysis in international isolates and drug repurposing against SARS-CoV-2 spike protein: molecular docking and simulation approach. *VirusDisease*. 2021; 32(4). doi:10.1007/s13337-021-00720-4.

14. Marques-Pereira C, Pires MN, Gouveia RP, et al. SARS-CoV-2 Membrane Protein: From Genomic Data to Structural New Insights. *International Journal of Molecular Sciences*. 2022; 23(6). doi:10.3390/ijms23062986

Summary

IDENTIFICATION OF SARS-COV-2 VARIANTS IN PATIENTS DURING THE 2021 COVID-19 WAVE IN BAC GIANG BY NEXT GENERATION SEQUENCING

The coronavirus disease of 2019 (COVID-19) pandemic caused by SARS-CoV-2 is a global threat and has severe impact on social and economic in Vietnam and around the world. The identification of circulating SARS-CoV-2 variants has important implications in monitoring the evolution of the virus, contributing to the promotion of vaccine research and diagnostic biologicals. This study was carried out on 38 patients diagnosed positive for SARS-CoV-2 from the industrial zones in Bac Giang province to identify SARS-CoV-2 variants and specific mutations in the genome viruses by next-generation sequencing (NGS). The mean age of participants in the study was 40.77 ± 14.63 . Research results have detected all 38 samples of SARS-CoV-2 virus belonging to the Delta variant, identified 16 mutations appearing in all patient samples, and 27 specific changes appearing only in some samples in the same period. There are 3 mutations on structural genes including S: Leu5Phe, Pro819Ser, and His1101Tyr.

Keywords: SARS-CoV-2, NGS, mutation.