

Análise de diversidade genética da mutação D614G em SARS-CoV-2

Dallyne Bárbara Ramos Venancio¹

Pierre Teodósio Felix²

Resumo

Os níveis de diversidade genética em 18 genomas de SARS-CoV-2 portadores da mutação D614G, provenientes da Malásia e da Venezuela e disponibilizados publicamente no National Center of Biotechnology and Information (NCBI). Esses haplótipos foram previamente utilizados para análise filogenética. Todas as lacunas e sítios não conservados foram extraídos para a construção de uma árvore filogenética. Como metodologias específicas para estimadores F_{ST} emparelhados, Variância Molecular (AMOVA), Distância Genética, incompatibilidade, análises demográficas e de expansão espacial, análise de diversidade molecular e tempo de divergência evolutiva, 20.000 permutações aleatórias foram sempre utilizadas. Os resultados revelaram a presença de apenas 57 sítios de polimórfico e parsimônio-informativo. As análises sugerem que possíveis variações nos produtos proteicos, do vírus selvagem em relação à sua forma mutante, devem ser mínimas, trazendo tranquilidade quanto ao aumento do risco de morte pela nova forma do vírus, bem como possíveis problemas de ajustes graduais em alguns alvos moleculares para vacinas.

Palavras-chave: Bioinformática; D614G mutação; Diversidade genética; SARS-CoV-2; Coronavírus

1 Introdução

A rápida taxa de transmissão do SARS-CoV-2, detectada em Wuhan (província de Hubei, China) no final de 2019, resultou na disseminação do vírus em todo o mundo e em poucos meses (MOHAMMAD *et al.*, 2020) e mesmo supondo

¹ Centro Universitário da Vitória de Santo Antão – UNIVISA. Discente do Curso de Bacharelado em Biomedicina do Centro Universitário da Vitória de Santo Antão – UNIVISA, dallynnebarbara@outlook.com

² Centro Universitário da Vitória de Santo Antão – UNIVISA. Docente do Curso de Bacharelado em Biomedicina do Centro Universitário da Vitória de Santo Antão – UNIVISA, pierrefelix@univisa.edu.br

que, apesar de sua alta taxa de transmissibilidade, o vírus tenha uma capacidade de replicação lenta de seu RNA, uma variante com uma única substituição residual denominada (D614G), surgiu como cepa dominante em todo o mundo (ZHANG *et al.*, 2020).

O sequenciamento do SARS-CoV-2 mostrou que esta mutação já é comum em vários continentes e que está sempre disponível, em apenas dois agrupados (S e G), tendo seus parâmetros clínicos de gravidade ou letalidade, sempre associados apenas a um deles (S) (ELIZONDO *et al.*, 2020). Análises filogenéticas de algumas Sequências de SARS-CoV-2 do Uruguai revelaram que há uma prevalência relativa da variante S sobre G e, neste momento, o S-clato já é encontrado com frequência na Europa, reforçando a indicação de que é ainda mais infeccioso do que WT (CASTILLO *et al.*, 2020).

A mutação D614G, que é uma mudança de um aspartato carregado negativamente (D) para glicina (G) na região de alça do domínio S1 da proteína S (MOHAMMAD *et al.*, 2020), à medida que progredia lentamente no Leste Asiático, só apareceria no final de abril em amostras virais de pacientes chineses (90-97%), então descrito como surgindo na posição 614 da proteína S e resultante de uma única alteração do nucleotídeo A a G na posição 23.403 no genoma de referência Wuhan-Hu-1, conferindo vantagens estruturais ao vírus e um aumento em sua transmissibilidade e cargas virais, sua transdução em células humanas e sua patogenicidade e letalidade, levando a especulações, inclusive, de que a eficácia das vacinas e das contramedidas dirigidas à proteína S pode ser adversamente afetada, exigindo uma correção frequente dessas vacinas ou dessas contramedidas (MCAULEY *et al.*, 2020).

Hoje, com um padrão de emergência específico além da China, os isolados virais iniciais de SARS-CoV-2 estão em grupos diferentes daqueles evoluídos posteriormente, como os isolados de SARS-CoV-2 da Índia, que estão claramente dispersos e que apenas as amostras iniciais de Kerala são semelhantes aos isolados de Wuhan, todos originados predominantemente de pacientes com histórico de viagens à Itália. Isso deixa claro que a mutação ocorreu principalmente quando o vírus começou a se espalhar durante a segunda infecção aguda e morbidade na Itália, Espanha e outros países europeus (GUPTA *et al.*, 2020).

O impacto exato da mutação no fenótipo da doença é uma questão crítica e ainda precisa ser elucidada com precisão. Dados genéticos, estruturais e epidemiológicos combinados sugerem que a chave D614G pode causar um aumento na prevalência de déficits quimiossensoriais, conforme observado durante a progressão da pandemia do Leste Asiático para os países ocidentais. O aumento da ligação da variante do pico G614 ao receptor do hospedeiro ACE2 e/ou o aumento da eficiência da entrada de células por meio da estabilização do pico da proteína e da clivagem reduzida podem ser o mecanismo subjacente (BUTOWT *et al.*, 2020).

A disseminação global e o aumento da infecção da variante SARS-CoV-2, D614G, levanta a questão de se essa mudança estrutural comprometeria a eficácia das terapias antivirais direcionadas à proteína S, especialmente se fossem projetadas para atingir o D614 (YURKOVETSKIY *et al.*, 2020).

Mais estudos são necessários para determinar como a crescente diversidade de padrões de mutação influencia a aptidão e reprodução das populações virais, e como ocorre sua suscetibilidade e evitamento às respostas imunológicas, bem como aos tratamentos.

2 Metodologia

1. Banco de dados: 18 sequências do genoma completo do vírus SARS-CoV-2 portador da mutação D614G e da Malásia e Venezuela (16 e 2 haplótipos respectivamente), todas com extensão de 29.827pb, foram recuperadas do GENBANK NCBI (National Center for Biotechnology Information). Uma vez alinhado usando o programa MEGA X (TAMURA *et al.*, 2018), sites ambíguos, dados perdidos e lacunas, foram excluídos resultando em um sequência analisável com extensão de apenas 57pb.

2. Para visualizar sites variáveis: A representação gráfica dos sites foi feita usando o software WEBLOGO v3.

3. Análises de Estruturação Genética: Estimadores F_{ST} pareados, Variância Molecular (AMOVA), Distância Genética, incompatibilidade, análises demográficas e de expansão espacial, diversidade molecular e tempo de divergência evolutiva foram obtidos com o Software Arlequin v. 3.5 (EXCOFFIER *et al.*, 2010) usando 1000 permutações aleatórias (NEI, 1979 e KUMAR, 2018). As matrizes F_{ST} e de distância geográfica não foram comparadas.

3 Resultados e Discussão

Todas as 18 sequências da mutação (D614G) do vírus SARS-CoV-2 da Malásia e Venezuela, revelaram a presença de apenas 57 sítios polimórficos e de passimônio informativo entre os 29.827 pb analisados. A representação gráfica desses sites pode ser visualizada em um logotipo construído com o programa weblogo v 3.7.4., onde o tamanho de cada nucleotídeo é proporcional à sua frequência para determinados sites.

Tabela 1 - Descrição de todas as sequências analisadas.

Acesso	Autores	Espécie	Título	Local	Data
MT 907515					
MT 907516					21 – AGO-2020
MT 907517					
MT 907518	Loureiro,	SARS-		Venezuela	
MT 907519	C.L. e cols	CoV-2	Diversidade genética		
MT 907521			SARS-CoV-2 em		
MT 907520			Venezuela: predominância		09 – SET- 2020
MW 015946			de D614G variantes e	Malásia	
MW 015947			análise de um surto		
MW 015948			Sequência completa do		18- SET - 2020
MW 015949			genoma de uma cepa do vírus		
MW 015950			SARS CoV-2 que abriga o		
MW 015951			D614G Mutação da Malásia		
MW 015952					
MW 015953					
MW 015954					
MW 079428	Ahmad,				07 – OUT-2021
MW 079429	H.F.				

Fonte: GENBANK, autores, local e data de arquivamento.

Utilizando o método UPGMA, para os 57 sítios parsimônio informativo, foi possível entender que os 18 haplótipos compunham dois grupos distintos, sendo até possível que haja compartilhamento de haplótipos entre os países estudados.

3.1 Análise de Variância Molecular (AMOVA) e Distância Genética

As análises baseadas nos valores de F_{ST} confirmaram a presença de duas entidades genéticas distintas com um índice de fixação de 22% e com um maior componente de variação dentro das populações (78,14%) menor que 0,05. Divergências evolutivas significativas foram observadas dentro e entre os grupos e uma alta similaridade genética entre as sequências que compunham o grupo

da Malásia, bem como uma maior divergência evolutiva entre as sequências que compunham o Grupo da Venezuela.

Tabela 2: Componentes de variação haplotípica e valor de F_{ST} .

FONTE DE VARIACÃO	D.F	SOMA DOS QUADRADOS	COMPONENTES DE VARIÂNCIA	PERCENTAGEM DE VARIACÃO
Entre as populações	1	26.410	3.70447 Va	21.86
Dentro das populações	16	211.812	13.23828Vb	78.14
Total	17	238.222	16.94275	

Índice de fixação F_{ST} : 0.21865
 Testes de significância (110 permutações)
 Va e F_{ST} : P (valor rand.> valor obs.) = 0,19091
 P (valor rand. = Valor obs.) = 0,02727
 valor P = 0,21818 + -0,05107

Tabela 3: Índices de diversidade molecular dos sítios informativos de parsimônio de portadores do SARS CoV-2 da mutação D614G.

ESTATÍSTICAS	MALASIA	VENEZUELA	MEAN	s.d.
No. of transitions	13	8	10.500	3.536
No. of transversions	4	3	3.500	0.707
No. of substitutions	17	11	14.000	4.243
No. of indels	0	49	24.500	34.648
No. of ts. sites	13	8	10.500	3.536
No. of tv. Sites	4	3	3.500	0.707
No. of subst. sites	17	11	14.000	4.243
No. private subst. Sites	16	10	13.000	4.243
No. of indel sites	0	49	24.500	34.648
Pi	17.000	27.308	22.15417	7.28909
Theta_k	N.A.	N.A.	N.A.	N.A.
Theta_k_lower	N.A.	N.A.	N.A.	N.A.
Theta_k_upper	N.A.	N.A.	N.A.	N.A.
Theta_H	N.A.	N.A.	N.A.	N.A.
s.d.Theta	N.A.	N.A.	N.A.	N.A.
Theta_S	17.00000	3.31502	10.15751	9.67674
s.d. Theta_S	12.36932	1.50101	6.93516	7.68505
Theta_pi	17.00000	27.30833	22.15417	7.28909
s.d. Theta_pi	17.49286	14.14939	15.82112	2.36419

Matriz de diferenças pareadas entre as populações estudadas: entre os grupos; dentro dos grupos; e distância de Nei para as sequências de portadores

SARS-CoV-2 da mutação D614G. As variações de Tau (relacionadas à ancestralidade dos dois grupos) revelaram um tempo significativo de divergência, suportado pela análise de mismatch da distribuição observada ($\tau = 42\%$) e com taxas de mutação constantes entre as localidades.

Matriz de tempo de divergência entre as sequências de portadores SARS-CoV-2 da mutação D614G. Em evidência o alto valor τ presente entre as sequências do Brasil e da Venezuela. Gerado pelo pacote estatístico em linguagem R utilizando os dados de saída do Software Arlequin versão 3.5.1.2.

Mostrou que as análises moleculares estimadas por θ refletiram um nível significativo de mutações entre todos os haplótipos (transições e transversões). Mutações em Indels (inserções ou acréscimos), foram encontradas apenas no grupo de haplótipos da Venezuela. Os testes Tajima e F_s de F_u mostraram divergências entre as estimativas de ϕ e π gerais, mas com valores negativos e altamente significativos, indicando, mais uma vez, a ausência de expansões populacionais no grupo Malásia. Tem o índice de irregularidade ($R = \text{Raggedness}$) com bootstrap paramétrico simulou novos valores ϕ para antes e depois de uma suposta expansão demográfica e, neste caso, assumiu um valor igual a zero apenas para o grupo da Malásia, provavelmente devido ao pequeno número de haplótipos analisados.

Os gráficos gerados pelo Arlequin versão 3.5.1.2. apresenta os índices de diversidade molecular para as sequências de portadores SARS-CoV-2 da mutação D614G. No gráfico os valores de θ : (θ_k) Relação entre o número esperado de allos (k) e o tamanho da amostra; (θ_H) Homozigossidade esperada em uma relação equilibrada entre deriva e mutação; (θ_S) Relação entre o número de locais de segregação (S), tamanho da amostra (n) e locais não recombinantes; ($\theta\pi$) Relação entre o número médio de diferenças emparelhadas (π) e θ .

Já a matriz de distância inter haplotípica nas sequências de portadores de SARS-CoV-2 da mutação D614G da Venezuela, apresenta uma similaridade significativa também foi evidenciada para o tempo de divergência evolutiva genética entre todas as populações; apoiado por variações τ , análises de mismatch e análises demográficas e de expansão espacial. Com uma exceção representativa para haplótipos da Venezuela.

4 Conclusão

Como o uso de metodologias de estudo de diversidade genética ainda não utilizadas neste PopSet customizado para genomas mutantes SARS-CoV-2 (D614G), foi possível detectar a existência de dois grupos distintos para os sítios parcimonioso-informativos de genomas virais da Malásia e Venezuela, contendo variações interhaplotípicas significativas apenas na população venezuelana. Talvez, quando outras sequências da Malásia estiverem disponíveis em bancos de dados, essa configuração possa mudar. Os grupos aqui descritos apresentaram padrões mínimos de estruturação sendo ligeiramente superiores para a população da Venezuela, igualando os resultados obtidos por (FELIX *et al.*, 2020) para populações selvagens de SARS-CoV-2 neste mesmo país. Esses dados sugerem que o grau relativo de estruturação presente na Venezuela pode estar relacionado ao fluxo genético. Esses níveis de estruturação também foram suportados por metodologias simples de emparelhamento filogenético, como o UPGMA, que neste caso, com um padrão descontínuo de divergência genética entre os grupos (dá suporte à ideia de possíveis isolamentos resultantes de eventos de fragmentação passados), especialmente quando observa uma quantidade não tão numerosa de ramos na árvore gerada e com muito poucos passos mutacionais.

Essas poucas mutações possivelmente ainda não foram corrigidas por deriva ou falta do efeito fundador, que acompanha o comportamento de dispersão e/ou perda de haplótipos intermediários ao longo das gerações. Os valores encontrados para a distância genética suportam a presença deste padrão contínuo de baixa divergência entre os grupos estudados, uma vez que consideraram importantes as diferenças mínimas entre os grupos, quando os haplótipos entre eles foram trocados, bem como a inferência de valores maiores que ou iguais aos observados na proporção dessas permutações, incluindo o p-valor do teste.

A discriminação das 18 entidades genéticas nas duas localidades também foi percebida por grandes variações interhaplotípicas, hierárquicas em todos os componentes de covariância: por suas diferenças intra e interindividuais ou por suas diferenças intra e intergrupos, gerando um dendrograma que sustenta a ideia de que as diferenças significativas encontradas nos dois países, por exemplo, foram compartilhadas mais na forma do que no número, uma vez que o resultado das

estimativas da divergência evolutiva média encontrada nesses e em outros países, mesmo que existam, são muito baixo.

Com base no alto nível de compartilhamento haplotípico percebido na Venezuela, foram dispensados testes que medem a relação entre distância genética e distância geográfica, como o teste de Mantel. O estimador ϕ , embora extremamente sensível a qualquer forma de variação molecular (FU, 1997), apoiou a uniformidade entre os resultados encontrados por todas as metodologias empregadas, e pode ser interpretada como uma confirmação filogenética de que há um consenso na conservação do genoma SARS-CoV-2 mutante (D614G), nos dois países analisados, e, portanto, é seguro afirmar que o pequeno número de polimorfismos existentes não deve refletir grandes mudanças nos produtos proteicos das populações virais nos dois países.

Essa consideração fornece a segurança de que, embora haja diferenças nos haplótipos estudados, essas diferenças são mínimas para ambas as regiões analisadas geograficamente e, portanto, parece seguro extrapolar os níveis de polimorfismo e diversidade molecular encontrados nas amostras deste estudo para outros genomas mutantes de SARS-CoV-2 em outros países, reduzindo as especulações sobre uma grande diferença entre o SARS-CoV-2 selvagem e o SARS-CoV-2 mutante (D614G), no nível de produtos proteicos. Embora a forma mutante tenha maior velocidade de transmissão e infecção, as análises feitas neste estudo sugerem pequenas variações nos produtos proteicos, principalmente naqueles que foram alvos deste estudo, trazendo tranquilidade quanto ao risco de morte também pelo COVID-19, como possíveis problemas de ajustes a alguns alvos moleculares para vacinas.

5 Referências

BUTOWT, R.; BILINSKA, K.; VON BARTHELD, C.S. Chemosensory Dysfunction in COVID-19: Integration of Genetic and Epidemiological Data Points to D614G Spike Protein Variant as a Contributing Factor. **ACS Chem Neurosci**, 2020.

CASTILLO, A.E.; PARRA, B.; TAPIA, P.; LAGOS, J.; ARATA, L.; ACEVEDO, A.; ANDRADE, W.; LEAL, G.; TAMBLEY, C.; BUSTOS, P.; FASCE, R.; FERNÁNDEZ, J. Geographical Distribution of Genetic Variants and Lineages of SARS-CoV-2 in Chile. **Front Public Health**. 2020.

ELIZONDO, V.; HARKINS, G.W.; MABVAKURE, B.; SMIDT, S.; ZAPPILE, P.; MARIER, C.; MAURANO, M.T.; PEREZ, V.; MAZZA, N.; BELOSO, C.; IFRAN, S.; FERNANDEZ, M.; SANTINI, A.; PEREZ, V.; ESTEVEZ, V.; NIN, M.; MANRIQUE, G.; PEREZ, L.; ROSS, F.; BOSCHI, S.; ZUBILLAGA, M.N.; BALLESTE, R.; DELLICOUR, S.; HEGUY, A.; DUERR, R. SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. **medRxiv**, 2020.

EXCOFFIER, L.; LISCHER, H.E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, v.10,p. 564-567, 2010.

FELIX, P.T.; FILHO, C.B.N.; RAMOS, R.S.; PAULINO, A.J.; VENANCIO, D.B.R. Levels of genetic diversity of SARS-CoV-2 virus: reducing speculations about the genetic variability of the virus in South America. **bioRxiv**, 2020. DOI: 10.1101/2020.09.14.296491

FU, Y.X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics**, p. 915–925, 1997.

GUPTA, A.M.; CHAKRABARTI, J.; MANDAL, S. Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants. **Microbes Infect**, 2020.

KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C.; TAMURA, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. **Molecular Biology and Evolution**, v.35, p.1547-1549, 2018.

MCAULEY, A.J.; KUIPER, M.J.; DURR, P.A.; BRUCE, M.P.; BARR, J.; TODD, S.; AU, G.G.; BLASDELL, K.; TACHEDJIAN, M.; LOWTHER, S.; MARSH, G.A.; EDWARDS, S.; POOLE, T.; LAYTON, R.; RIDDELL, S.J.; DREW, T.W.; DRUCE, J.D.; SMITH, T.R.F.; BRODERICK, K.E.; VASAN, S.S. Experimental and in silico evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. **NPJ Vaccines**, 2020.

MOHAMMAD, A.; ALSHAWAF,E.; MARAFIE, S.K.; ABU-FARHA, M.; ABUBAKER, J.; AL-MULLA, F. Higher binding affinity of Furin to SARS-CoV-2

spike (S) protein D614G could be associated with higher SARS-CoV-2 infectivity. **Int J Infect Dis.** Kuwait, P.611-616, 2020. DOI: 10.1016/j.ijid.2020.10.033.

NEI, M.; LI, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proc.Natl.Acad.Sci**, USA, v.76, p.5269-5273, 1979.

YURKOVETSKIY, L.; WANG, X.; PASCAL, K.E.; TOMKINS-TINCH, C.; NYALILE, T.P.; WANG, Y.; BAUM, A.; DIEHL, W.E.; DAUPHIN, A.; CARBONE, C.; VEINOTTE, K.; EGRI, S.B.; SCHAFFNER, S.F.; LEMIEUX, J.E.; MUNRO, J.B.; RAFIQUE, A.; BARVE, A.; SABETI, P.C.; KYRATSOUS, C.A.; DUDKINA, N.V.; SHEN, K.; LUBAN, J. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. **Cell**, 2020.

ZHANG, J.; CAI, Y.; XIAO, T.; LU, J.; PENG, H.; STERLING, S.M.; WALSH, R.M.; RITS-VOLLOCH, S.; SLIZ, P.; CHEN, B. Structural impact on SARS-CoV-2 spike protein by D614G substitution. **BioRxiv**, 2020. DOI:10.1101/2020.10.13.337980.