

과제 2. Seq DB 응용

생명공학과 2020080600 이민재

비교할 변이를 찾기 위해 치명률(CFR, Case Fatality Rate)을 조사하였다. 조사 결과, Beta 변이가 가장 큰 치명률을 가지고 있었고(4.2%) Omicron의 치명률이 가장 낮았다. 이 결과를 참고해서, 코로나바이러스의 원형과 Beta, Omicron 변이를 비교하여 치명률에 영향을 주는 요인을 분석하고자 한다.

GenBank에서 각 변이에 해당하는 ID를 복사하여 BLAST를 사용하여 분석을 시작하였다. Query Sequence로 Wuhan strain을 사용하였고, Subject Sequence로 Beta, Omicron 변이를 사용하였다. 변이 간 비교를 위해 'Program Selection'에서 'discontiguous megablast'를 선택하였다.

변이된 부분을 찾기 위해 'Alignment View'를 'Pairwise with dots for identities'로 설정하고, 'CDS feature' 체크박스를 체크하여 단순 codon redundancy를 배제하였다. 그러면 그림과 같이 변이로 인해 amino acid의 구성이 달라진 부분을 찾을 수 있다.

CDS:surface glycopro	137	N D P F L G V Y Y H K N N K S W M E S E
Query	21971	AATGATCCATTTTGGGTGTTTATTACCAAAAAACAAGTTGGATGAAAGTGAG
Sbjct	21943	N D P F L ----- D H K N N K S W M E S E
CDS:surface glycopro	135	N D P F L ----- D H K N N K S W M E S E
CDS:surface glycopro	157	F R V Y S S A N N C T F E Y V S Q P F L
Query	22031	TTGAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTT
Sbjct	21994
CDS:surface glycopro	152	F R V Y S S A N N C T F E Y V S Q P F L
CDS:surface glycopro	177	M D L E G K Q G N F K N L R E F V F K N
Query	22091	ATGGACCTTGAAGGAAAAACAGGGTAATTTCAAAATCTTAGGGAATTTGTGTTAAGAA
Sbjct	22054
CDS:surface glycopro	172	M D L E G K Q G N F K N L R E F V F K N
CDS:surface glycopro	197	I D G Y F K I Y S K H T P I N L V R
Query	22151	ATTGATGGTTATTTTAAAAATATATTCTAAGCACACGCCTATTAATTAGTGCGT-----
Sbjct	22114 GAGCCA
CDS:surface glycopro	192	I D G Y F K I Y S K H T P I I V R E P
CDS:surface glycopro	215	D L P Q G F S A L E P L V D L P I G I
Query	22205	---GATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGTAGATTGCCAATAGGTATT
Sbjct	22171	GAA
CDS:surface glycopro	211	E D L P Q G F S A L E P L V D L P I G I

변이가 확인된 부분을 조사하여 치명률에 영향을 주는 단백질을 알아보고자 한다. 변이가 발견된 위치가 어떤 단백질에 해당하는지를 찾아본 결과 다음의 결과를 얻을 수 있었다.

1 Wuhan vs. Beta

- 1.1 ORF1a : 1561, 4381, 8461, 9421 → nsp2, nsp3, nsp4
- 1.2 surface glycoprotein : 22861, 23401, 25021, 25201
- 1.3 nucleocapsid phosphoprotein : 28861

2 Wuhan vs. Omicron

- 2.1 ORF1a : 2821, 6481, 11270, 11509 → nsp3, nsp6
- 2.2 ORF1ab : 18109 → (3'→5') exonuclease
- 2.3 surface glycoprotein : 21709, 21763, 21943, 22114, 22594, 22654, 22774, 22834, 22954, 23014, 23134, 23494, 23554, 23794, 23914, 24094, 24394, 24454
- 2.4 envelope protein : 26194
- 2.5 membrane glycoprotein : 26494, 26674
- 2.6 nucleocapsid phosphoprotein : 28234, 28294, 28825

원형의 CFR 이(3.6%) Beta 변이의 CFR 에 가깝기 때문에(4.2%) 이에 맞추어 Omicron 변이와 비교해본다면, 다른 protein 보다 surface protein, 즉 Spike protein 의 변이가 많이 일어났음을 관찰할 수 있다. 따라서 Spike protein 이 CFR 에 영향을 주는 큰 요인 중 하나라고 분석할 수 있다. 추가적인 조사 결과, Beta 변이에서 receptor-binding domain(RBD) 내 변이가 항체 회피 능력을 증가시키고 중증 진행 및 사망률에 영향을 준다는 보고가 있으며, Omicron 변이에서도 RBD 내에 많은 변이가 존재하나 변이의 조합이 전파력을 높인 반면 병원성을 낮추는 방향으로 진행되어 CFR 이 낮게 관측된다는 결과를 얻을 수 있었다.

[참고자료]

- 1. NCBI GenBank <https://www.ncbi.nlm.nih.gov/genbank/>
- 2. Li J, Lai S, Gao GF, et al. Case-fatality rates of SARS-CoV-2 variants of concern: A global meta-analysis of 112 studies. Int J Infect Dis. 2024;140:1-10. doi:10.1016/j.ijid.2024.03.015.
- 3. NCBI BLAST <https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&RID=06H61B55114>