

TRABALHO PRÁTICO 1 - ALINHAMENTO DE SEQUÊNCIAS

Gabriel Quintanilha Peixoto, Hércules Batista de Oliveira, Joicymara Santos Xavier, Kaíssa Pereira Barbosa, Wellen Quézia Bernardes Durães.

2. Realizar análises detalhadas dos significados das conservações / mutações do arquivo fornecido com as sequências dos vírus. Outros conjuntos de dados de interesse também podem ser utilizados. Nesse caso, entregar um relatório de até 5 páginas com essas análises.

Foram fornecidas 11 sequências de aminoácido da proteína *spike* de diversas cepas de *Coronavirus*, parte das quais são pertencentes às cepas originárias de murinos (grupo que inclui ratos e camundongos). Para avaliar a conservação das sequências, inicialmente foi feito um alinhamento com BLASTP (Camacho et al., 2009), um algoritmo heurístico (Fig. 1). Também foi feito um alinhamento com ClustalW mais detalhado em sua implementação no software MEGA X (Kumar et al., 2018), e análise filogenética, avaliando manualmente regiões conservadas. Por fim, avaliamos a presença de domínios proteicos com hmmscan (Potter et al., 2018) (Tabela 1).

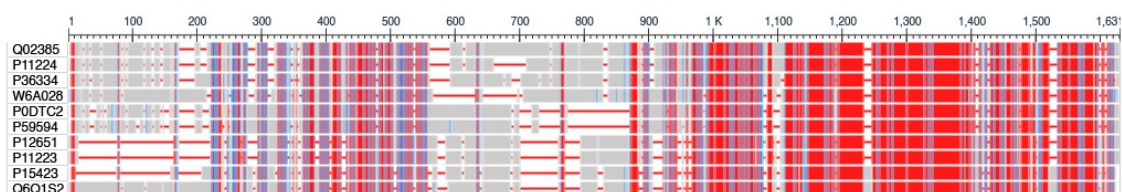


Figura 1: Visualização do alinhamento de 10 das 11 sequências da proteína *spike* de cepas de coronavírus (*Breda vírus* 1 não incluído). Regiões destacadas indicam conservação de sequências, enquanto as linhas vermelhas indicam regiões de *gaps*.

Ao alinhar as sequências com ClustalW no MEGA X (Kumar et al., 2018), percebemos que as sequências possuem vários blocos de conservação presentes em todas as 11 sequências fornecidas, com grande presença de substituições nas regiões alinhadas e também regiões com longos *gaps*. Ao remover essas regiões ricas em *gaps*, foi construída uma árvore filogenética de *Maximum Likelihood*, com valor de *bootstrap* 500.

Com hmmscan (Potter et al., 2018) podemos avaliar que essas regiões conservadas correspondem principalmente aos domínios proteicos detectados (Tabela 1), pertencentes a glicoproteínas presentes na superfície do capsídeo de coronavírus, que dá nome ao clado devido a seu formato tridimensional.

| Accession | Species | Detected domains |
|-----------|---|--|
| Q02385 | <i>Murine coronavirus</i> (strain JHMV/variant CL-2) | Corona_S2 Spike_NTD Spike_rec_bind |
| W6A028 | <i>Middle East respiratory syndrome-related coronavirus</i> | Corona_S2 Spike_rec_bind |
| P0DTC2 | <i>Severe acute respiratory syndrome coronavirus 2</i> | Corona_S2 Spike_rec_bind |
| P59594 | <i>Severe acute respiratory syndrome coronavirus</i> | Corona_S2 Spike_NTD Spike_rec_bind |
| P11224 | <i>Murine coronavirus</i> (strain A59) | Corona_S2 Spike_NTD Spike_rec_bind |
| P11223 | <i>Avian infectious bronchitis virus</i> (strain Beaudette) | Corona_S1 Corona_S2 |
| P36334 | <i>Human coronavirus</i> OC43 | Corona_S2 Spike_NTD Spike_rec_bind |
| P12651 | <i>Avian infectious bronchitis virus</i> (strain M41) | Corona_S1 Corona_S2 |
| P15423 | <i>Human coronavirus</i> 229E | Corona_S1 Corona_S2 |
| Q6Q1S2 | <i>Human coronavirus</i> NL63 | Corona_S1 Corona_S2 |
| O90304 | <i>Breda virus</i> 1 | Spike_torovirin |

A proteína spike é um determinante para a infecção das cepas de coronavírus em seus respectivos hospedeiros, visto que elas são reconhecidas por receptores na célula do hospedeiro e posteriormente internalizadas, dando início ao ciclo de replicação viral (Li, 2016). As mutações que ocorrem nas sequências desta glicoproteína alteram a estrutura tridimensional da cadeia polipeptídica, e podem ou inviabilizar o reconhecimento, ou aumentar a capacidade do vírus de infectar seu hospedeiro, ou mesmo sinalizar uma mudança de hospedeiro, o que se hipotetiza ter acontecido com a infecção em humanos que deu origem à atual pandemia.

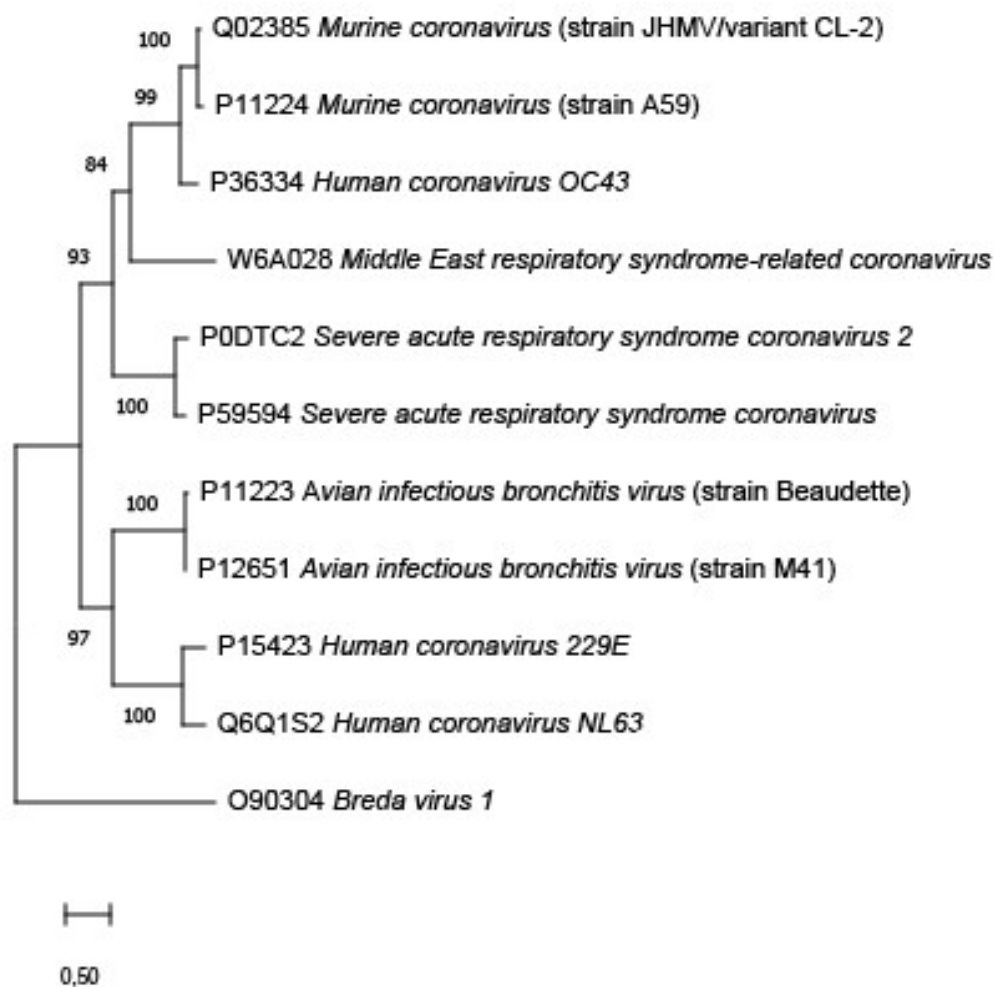


Figure 2: Árvore filogenética por *Maximum Likelihood* com bootstrap 500 para as sequências fornecidas.

A árvore filogenética obtida corrobora com essa hipótese, considerando que a espécie *Human coronavirus OC43* está no mesmo clado do coronavírus de roedores. No entanto, outras cepas de coronavírus humano são mais próximas do coronavírus que infecta aviários. Os trabalhos mais recentes, no entanto, indicam que a origem da cepa causadora da pandemia tenha surgido em morcegos (Shereen et al., 2020), que não estão presentes em nossas análises.

Referências

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 1–9. <https://doi.org/10.1186/1471-2105-10-421>

- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*, 3, 237–261. <https://doi.org/10.1146/annurev-virology-110615-042301>
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, 46(W1), W200–W204. <https://doi.org/10.1093/nar/gky448>
- Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, 24, 91–98. <https://doi.org/10.1016/j.jare.2020.03.005>