Plano de trabalho para relatório do projeto SARS –COV 2:

Os alunos devem colocar um título para o projeto (sugerimos que seja um título que indique em poucas palavras as principais conclusões do trabalho), indicar os autores e fazer um resumo (cerca de 200 palavras- que deve dizer basicamente o que foi feito e as conclusões.

Abaixo tem as sugestões de atividades que devem ser feitas e que perguntas devem ser respondidas:

1. Comparar 5 genomas inteiros identificando os genomas mais próximos ou distantes entre si, além das regiões do genoma mais próximas ou distantes, em relação ao SARS-Cov2.

Comparar os diferentes genomas estudados com comparações de nucleotídeos. É importante concluir a porcentagem de identidade entre os diferentes genomas e o genoma de SARS-Cov2. Com isso, será possível estabelecer quais são os genomas mais próximos evolutivamente. Discuta se isso está de acordo com as filogenias que tem sido publicadas.

1. Dividir os genomas em 10 segmentos de 3.000 bases (o genoma de coronavírus apresenta cerca de 30.000 bases no total) e comparar os genomas, em relação ao SARS-Cov2. Apresentar os resultados em gráficos.

A comparação também tem que ser só de nucleotídeos.

Mas cuidado:

Em todas as comparações realizadas neste e em outros itens, pode haver regiões que não tem alinhamento, e com isso o query cover é menos que 100%. Isso é bem normal de deve ser levado em conta na hora de calcular qual a % de nucleotídeos idênticos.

Exemplo: na comparação dos 3000 nucleotídeos, se você tem 50% de query coverage, significa que só 1500 bases alinharam. Os outros não tiveram identidade: portanto é zero identidade! E se a % de identidade deu 80%, então você teria 80% de identidade em 1500 bases, mas você tem que calcular em cima do total, ou seja, no caso 3000 bases.

Uma solução simples é multiplicar o % query coverage pelo % de identidade e obter o resultado correto: 40% no caso do exemplo.

Lembre-se que essa questão deve ser feita para todos os alinhamentos realizados, ok? Sobretudo pois raramente é 100%.

De qq forma, nesse item identificar quais as regiões que foram mais ou menos conservadas entre os diferentes coronavirus. O que isso significa? Você já pode localizar, pela posição no genoma, quais são os genes que podem ser mais conservados. Aqui você já pode sugerir que a análise gene a gene (próxima etapa) pode ajudar a confirmar as regiões mais conservadas.

1. Estudar 5 proteínas desses 5 genomas e realizar a comparação ao nível de nucleotídeo e de proteína, em relação ao SARS-Cov2. Especial atenção deve ser dada às proteínas SPIKE e Replicase.

No caso de genes você deve fazer comparações ao nível de nucleotídeo e de proteína. Na comparação de proteínas deve considerar identidade de aminoácidos e também similaridade. Explique a diferença entre esses dois valores (o que essa diferença significa). Não se esqueça de considerar sempre o Query Cover. Também deve explicar porque os valores são tão diferentes entre comparações de nucleotídeos e de proteínas (quem é mais conservado e por que?).

Com os dados deve responder: entre os genes analisados quais são mais e quais são menos conservados? Faça hipóteses que expliquem essas diferenças. O que isso implica para o uso de vacinas que usam apenas um gene?

1. Identificar INDELS em pelo menos dois dos genes analisados.

Para isso você deve usar o alinhamento múltiplo. Encontrando INDELS interessantes, tente refletir no que isso pode significar em termos de classificação e origem desses vírus (e mesmo de SARS-COv2).

1. Buscar os variantes e verificar quantas mutações os principais variantes considerados problemáticos (alfa, beta, gama e delta) apresentam no genoma e também nos genes analisados. Responder porque eles são considerados preocupantes.

Cuidado, essas mutações devem considerar aquelas que apresentam mudanças de aminoácidos apenas. É possível que você tenha que fazer experimentos de comparações de sequencias de genes entre SARS-COV2 e das variantes. Existe algum gene que sofre mais mutações que os outros (entre os genes analisados)? O que isso significa para as vacinas que estão sendo empregadas. Busquem informações de que tipo de alterações do vírus ocorrem com essas novas variantes.

Nessas variantes analisem com cuidado a região da SPIKE chamada RBD (Receptor Binding Domain). Como é a região que se liga no ACE2, essas mutações podem ser importantes.

Os alunos são estimulados a interpretarem os resultados de forma diferente das perguntas feitas aqui. É importante que você entenda os seus resultados e caso tenha novas ideias e interpretações, indique claramente no relatório.

Bom trabalho!

Sugestão de referências:

Grifoni A, Sidney J, Vita R, Peters B, Crotty S, Weiskopf D, Sette A. SARS-CoV-2 human T cell epitopes: adaptive immune response against COVID-19. Cell Host Microbe. 2021 May 21:S1931-3128(21)00238-9.

Becker M, Dulovic A, Junker D, Ruetalo N, Kaiser PD, Pinilla YT, Heinzel C, Haering J, Traenkle B, Wagner TR, Layer M, Mehrlaender M, Mirakaj V, Held J, Planatscher H, Schenke-Layland K, Krause G, Strengert M, Bakchoul T, Althaus K, Fendel R, Kreidenweiss A, Koeppen M, Rothbauer U, Schindler M, Schneiderhan-Marra N. Immune response to SARS-CoV-2 variants of concern in vaccinated individuals. Nat Commun. 2021 May 25;12(1):3109.

Zhou H, Ji J, Chen X, Bi Y, Li J, Wang Q, Hu T, Song H, Zhao R, Chen Y, Cui M, Zhang Y, Hughes AC, Holmes EC, Shi W. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. Cell. 2021 Jun 9:S0092-8674(21)00709-1.

Kwon SB, Ernst J. Single-nucleotide conservation state annotation of the SARS-CoV-2 genome. Commun Biol. 2021 Jun 3;4(1):698.