

Prof. João Carlos Setubal

**Tarefa Prática (TP): Análise de genomas de vírus**

Nesta tarefa prática os alunos farão análise de genomas de vírus. Os dois tipos de vírus que serão considerados serão SARS-CoV-2 (o agente causador da atual pandemia, abreviado por **SC2**) e o vírus da varíola dos macacos (Monkeypox virus, abreviado por **MPXV**). Nesta tarefa é importante o conceito de **genoma de referência**. Para SC2, o genoma de referência é o genoma da cepa Wuhan-Hu-1, com número de acesso NC\_045512.2. Para MPXV, o genoma de referência é o genoma da cepa MPXV\_USA\_2022\_MA001, número de acesso [ON563414](#).

Cada aluno deverá escolher um genoma de SC2 ou um genoma de MPXV, *diferente da referência respectiva*. **Não pode haver dois alunos com mesmo genoma**. Por esse motivo, escolha seu genoma e informe-o ao monitor Ignacio, por e-mail ([ignacio@iq.usp.br](mailto:ignacio@iq.usp.br)). **Informe o número de acesso (accession number) no GenBank**. Caso seu genoma já tenha sido escolhido por outro aluno, você será informado disso e terá que escolher outro genoma. O monitor Ignacio irá disponibilizar uma tabela com todos os números de acesso de genomas já escolhidos para facilitar este processo.

Cada aluno terá que realizar as seguintes análises sobre seu genoma e apresentar um relatório com os resultados:

SC2:

1. Informe o número de acesso do genoma, local e data de isolamento da cepa ao qual corresponde seu genoma. Informe a qual clado ou linhagem pertence seu genoma.
2. Obtenha um alinhamento do seu genoma contra a referência. Reporte o alinhamento completo, incluindo o cabeçalho onde são informadas as características do alinhamento. Você pode copiar e colar uma imagem, ou você pode copiar e colar os caracteres; neste último caso, use fonte mono espaçada (como courier).
3. Esse alinhamento deve ter aproximadamente 100% de cobertura de ambos os genomas (ou seja esse alinhamento deve ter cerca de 30 mil posições); caso não tenha, possivelmente seu alinhamento está errado. Supondo 100% de cobertura, você deve analisar o alinhamento, apontando as diferenças entre os 2 genomas e determinar se essas diferenças afetam aminoácidos de proteínas. Caso afetem aminoácidos (ou seja, mutação não sinônima) você deve identificar precisamente quais são os aminoácidos afetados, qual a mudança causada pela mutação, e quais são os genes afetados. Isto deve ser feito na forma de uma tabela. **Se seu alinhamento tiver menos do que 10 diferenças em relação à referência, você vai ter que necessariamente escolher outro genoma para seu relatório.**
4. Adicionalmente, apresente alinhamentos do seu genoma contra:
  - a. o genoma completo de um vírus SARS-CoV-1 (sua escolha)
  - b. o genoma de um betacoronavírus que tenha sido isolado de morcêgos (sua escolha)
5. Instruções sobre o passo 4. Você deve procurar os genomas estipulados, e esses genomas precisam ser completos (não podem ser fragmentos). Em cada caso, informe precisamente o genoma escolhido (por exemplo, transcreva o cabeçalho do registro do GenBank, incluindo a fonte e local de isolamento). Em cada caso, provavelmente o alinhamento estará dividido em

vários blocos, e deverá haver muitas diferenças em cada bloco. Por esse motivo, a análise destes alinhamentos também deve determinar e informar quais partes dos genomas não foram alinhadas. Isto deve ser feito na forma de uma tabela, e você deve indicar quais genes (do seu genoma) são afetados nas partes não alinhadas. Não é necessário analisar diferenças nos genes nas partes alinhadas.

6. Compare os 2 alinhamentos obtidos no passo 5 entre si. Você consegue afirmar qual é o genoma mais diferente do seu genoma? Justifique.

MPXV:

1. Informe número de acesso do genoma, local e data de isolamento da cepa ao qual corresponde seu genoma. Informe a qual clado ou linhagem pertence seu genoma.
2. Obtenha um alinhamento do seu genoma contra a referência. Reporte um resumo do alinhamento que mostre suas estatísticas gerais.
3. Esse alinhamento deve ter aproximadamente 100% de cobertura de ambos os genomas; caso não tenha, possivelmente seu alinhamento está errado. Supondo 100% de cobertura, você deve analisar o alinhamento, apontando as diferenças entre os 2 genomas e determinar se essas diferenças afetam aminoácidos de proteínas. Caso afetem aminoácidos (ou seja, mutação não sinônima) você deve identificar precisamente quais são os aminoácidos afetados, qual a mudança causada pela mutação, e quais são os genes afetados. Isto deve ser feito na forma de uma tabela. **Se seu alinhamento tiver menos do que 5 diferenças em relação à referência, você vai ter que necessariamente escolher outro genoma para seu relatório.**
4. Adicionalmente, apresente alinhamentos do seu genoma contra o genoma completo (não pode ser fragmento) de dois outros orthopox virus. Informe precisamente os genomas escolhidos (por exemplo, transcreva o cabeçalho do registro do GenBank, incluindo a fonte e local de isolamento). Não é necessário apresentar o alinhamento completo; basta um resumo. Caso um ou ambos os alinhamentos estejam divididos em vários blocos, apresente um resumo das diferenças em cada bloco; informe também quais partes dos genomas não foram alinhadas. Isto deve ser feito na forma de uma tabela, e você deve indicar quais genes (do seu genoma) são afetados nas partes não alinhadas. Não é necessário analisar diferenças nos genes nas partes alinhadas.
5. Compare os 2 alinhamentos obtidos no passo 4 entre si. Você consegue afirmar qual é o genoma mais diferente do seu genoma? Justifique.

Os passos acima são obrigatórios. Mas cada aluno é bem-vindo a fazer análises adicionais com seu genoma. A critério do professor, análises adicionais poderão aumentar a nota do projeto.

### Instruções de entrega do relatório

1. Relatórios são individuais
2. Seu relatório deve ser um único arquivo PDF
3. O nome desse arquivo deve ser o seu nome, sem mais nenhuma outra informação. Por exemplo: joasetubal.pdf
4. O arquivo deve ser enviado no prazo (28/10/2022, meio dia, hora de SP) para [setubal@iq.usp.br](mailto:setubal@iq.usp.br)
5. Use como **assunto** a seguinte cadeia de caracteres (use apenas minúsculas): **\*tarefapratica\***
6. Você receberá uma resposta automática, que será seu **recibo de entrega**. Sem esse recibo, não considerarei que você entregou o relatório

7. A primeira página do relatório deve informar seu nome, número USP, data, e nome da disciplina. O resto do relatório é formato livre. Entretanto, um dos critérios de correção será a legibilidade de seu relatório. Relatórios mal estruturados, confusos, com formatos inadequados dos alinhamentos, etc, serão penalizados.

### **Websites relevantes**

**NCBI virus:** <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>

### **Virus Pathogen Database and Analysis Resource:**

<https://www.viprbrc.org/brc/home.spg?decorator=vipt>

NCBI SARS-CoV-2 resources

<https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Outros websites úteis

<https://www.gisaid.org/>

<https://nextstrain.org/>