



Concurso Público Fiocruz 2023

Pesquisador em Saúde Pública

PE 24

Prova Discursiva

Biodiversidade e Vigilância Genômica no âmbito da Saúde Única

Espelho de Resposta

Pontuação de cada Questão Discursiva conforme Anexo II do Edital nº 3, de acordo com a Unidade detentora da vaga.

Espera-se que o candidato, no desenvolvimento do tema, tenha feito considerações técnicas adequadas sobre os seguintes pontos:

Questão 01

A) Para a investigação do surto em questão devem ser verificadas informações que podem ajudar no desenho dos experimentos, por exemplo:

I) Qual a faixa etária dos indivíduos acometidos? Existe concentração em alguma delas que não seja proporcional ao número de indivíduos do local?

II) Existe diferença significativa em relação ao número de casos em pacientes homens e mulheres?

III) Existe a informação da circulação na região de agentes patogênicos que causem os sintomas observados? Se não localmente, no país? Ou em países vizinhos?

IV) Que exames foram realizados nos pacientes? Quais os resultados disponíveis? Esses resultados apontam para patógeno viral, bacteriano, fúngico?

V) Existem casos graves? Óbito?

VI) Foi realizada discussão com a comunidade e autoridades de Saúde locais para levantamento de demais informações?

B) Nesse ponto o candidato deve propor experimentação laboratorial que possa ajudar na elucidação do surto, não se limitando à abordagem genômica, uma vez que é possível pensar em isolamento do agente infeccioso, ou testes sorológicos que podem contribuir para a identificação do surto. No caso o (a) candidato (a) deve mostrar seu conhecimento em diferentes técnicas laboratoriais que podem ser usadas em conjunto com as técnicas de sequenciamento propriamente dito.

C) Pontos principais que devem ser desenvolvidos um a um pelo candidato (a).

I) Identificação do patógeno;

II) Identificação dos possíveis reservatórios. Neste caso considerar que a fauna local foi amostrada para pesquisa de reservatórios, ou mesmo que o patógeno identificado possua relação com outros patógenos previamente identificados em animais;

III) Desenvolvimento de ensaio diagnóstico molecular e o monitoramento da necessidade de atualização do mesmo;

IV) Entendimento da dinâmica de transmissão, vetorial, direta;

V) Desenvolvimento de vacinas;

VI) Monitoramento da emergência de resistência seja para fármacos (caso sejam desenvolvidos), terapia com anticorpos monoclonais.

D) Pontos principais que devem ser desenvolvidos, um a um, pelo candidato (a).

I) As mutações nos genomas dos patógenos tendem a se acumular em razão do tempo. Ou seja, quanto maior o tempo de circulação de um patógeno em uma comunidade, maior será esse número de mutações. Assim, com base em um estudo filogenético calibrado pelo tempo é possível estimar o TMRCA, ou tempo do ancestral comum mais recente, que indica o surgimento da linhagem do patógeno.

II) Apesar do relógio molecular poder ser observado em qualquer patógeno que esteja replicando, é para os vírus de genoma RNA que esse conceito é mais facilmente observado, em função da rápida evolução desses patógenos. Assim, se for identificado um vírus RNA na investigação, em pouco tempo espera-se o surgimento de mutações que posteriormente funcionaram como uma assinatura daquela linhagem, facilitando o acompanhamento de sua dispersão. Existem diferentes modelos para análise de relógio molecular (estrito ou relaxado, por exemplo). O modelo ideal para um conjunto de dados específico pode ser encontrado utilizando diferentes ferramentas de Bioinformática tais como IQTREE2, BEAST, BEAST2.

E) Pontos principais que devem ser desenvolvidos, um a um, pelo candidato (a).

I) Representatividade das amostras analisadas de forma a equilibrar os genomas sequenciados por local de coleta e período (semana epidemiológica).

II) Protocolos que visem a redução de custos para a obtenção dos genomas, seja pela otimização de um número maior de amostras em uma mesma corrida, ou por otimização de procedimentos no preparo das bibliotecas.

III) Para patógenos sem informações genéticas *a priori*: necessidade de utilizar estratégias de metagenômica *shotgun* ou metagenômica dirigida (captura por sondas de enriquecimento) para a sua identificação. Em ambos os casos é importante ressaltar que um patógeno pode estar em quantidades ínfimas no tecido examinado. Assim, é necessário para ambas as estratégias diminuir a quantidade de material genético contaminante. No caso do material genético do hospedeiro podem ser utilizadas estratégias de degradação de ácidos nucleicos não encapsulados, depleção de RNA ribossomal, entre outras. No caso específico de investigação de um surto viral pode usar depleção de material microbiano.

IV) Necessidade de obtenção de um grande número de leituras (milhões de reads) de forma a remover sequências contaminantes utilizando programas de bioinformática.

V) Uma vez obtida uma sequência de alta qualidade do patógeno de interesse devesse mudar a abordagem de sequenciamento para outras com melhor custo-benefício, como o sequenciamento de *amplicons* (tiling PCR), para continuar a vigilância deste. No caso da utilização de estratégia baseada em *amplicons* deve-se preferir estratégias que utilizem fragmentos pequenos, em torno de 400 pares da base, para aumentar a sensibilidade e minimizar problemas com material genético degradado.

Questão 02

A) Pontos principais que devem ser desenvolvidos um a um pelo candidato (a).

A difusão do sequenciamento genético massivo, também conhecido como de nova geração, é certamente o principal avanço tecnológico com relação ao dito laboratório-molhado. A capacidade de gerar muitas bases sequenciadas em cada corrida permitiu o crescimento exponencial na geração de dados. Por outro lado, esse aumento na geração de dados de sequenciamento precisou ser acompanhado de melhoria na infraestrutura de computação para dar conta do grande volume de dados, além do desenvolvimento e aprimoramento de novas ferramentas de bioinformática.

B) Pontos principais que devem ser desenvolvidos um a um pelo candidato (a).

I) A montagem dos genomas é a etapa onde são removidos *reads*, ou regiões dos *reads*, de baixa qualidade, ou contendo sequências de adaptadores dos protocolos de sequenciamento. Após esse início, os *reads* são mapeados contra o genoma de referência viral (*template*) ou podem ser utilizados para montagem *De novo*, com posterior confirmação das sequências virais.

II) Uma vez finalizado o genoma consenso do SARS-CoV-2, esse deve ser analisado quanto à qualidade com ferramentas que indiquem, por exemplo, clusters de mutações, mutações de reversão, bases não ACTG, quantidade de bases não identificadas (N) entre outras análises. Assim, as amostras que porventura tenham problemas apontados na sequência de consenso devem ser reanalisadas.

III) Para facilitar a identificação de grupos de sequências mais próximas entre si, e assim facilitar o seu rastreo em diferentes locais do planeta, foram criados sistemas de classificação por linhagens como o *Pangolin* e o *Nextclade*.

IV) Devido ao enorme número de genomas gerados e depositados em bancos de dados públicos é muito comum utilizar estratégias que promovam *downsampling* de sequências idênticas ou com muito próximas entre si, de forma a diminuir o custo computacional com a análise do dataset. A Ferramenta CD-HIT é um exemplo que pode ser utilizada nessa abordagem. A redução de redundância do dataset ajuda principalmente se análises Bayesianas forem utilizadas de forma subsequente. Por fim, esse dataset não-redundante, formado por genomas que tenham passado nas etapas de avaliação de qualidade, seguirá para investigação das relações filogenômicas (ou filogeográficas/filodinâmicas, caso esse seja o interesse). Por fim a remoção de artefatos de sequenciamento, tais como mutações de reversão impactará positivamente na acurácia da reconstrução filogenética.

C) Pontos principais que devem ser desenvolvidos um a um pelo candidato (a).

O sequenciamento de patógenos pode contribuir também para identificação de marcadores genéticos de resistência (mutações ou elementos móveis) aos antimicrobianos, auxiliando dessa forma no tratamento de pacientes. Além das mutações de resistência, os resultados de sequenciamento promovem uma melhor identificação do patógeno em questão, seja por análise do genoma global, ou de regiões utilizadas na genotipagem (por exemplo, genes utilizados para o MLST de bactérias e fungos). Assim, essas informações facilitam o entendimento de como esse patógeno pode estar se dispersando em uma comunidade, sendo essa informação um importante subsídio para medidas de controle.

D) Pontos principais que devem ser desenvolvidos um a um pelo candidato (a).

I) Principalmente no início da pandemia, a VG estava limitada a poucos centros de pesquisa, a maioria no hemisfério norte. A distância dos centros de sequenciamento com os locais de ocorrência de casos afeta negativamente a velocidade dessa resposta. Em média, o período entre a coleta de amostras e a disponibilização dos genomas em bancos de dados públicos é superior a três semanas, que foi um tempo mais que suficiente para a dispersão das principais linhagens, em especial das VOC. Mesmo em locais com altas taxas de sequenciamento não foi possível impedir o espalhamento das VOC, vide VOC alpha na Inglaterra. Assim, o endurecimento das medidas de controle não poderia ser baseado somente na identificação das linhagens, mas sim na dinâmica de casos. Não obstante, a VG foi e continua sendo importante para o desenvolvimento de ensaios diagnósticos e de vacinas, em especial aquelas baseadas na incorporação de sequências virais, e no acompanhamento da evolução do vírus com o surgimento de mutações que aumentam a ligação ao receptor ou que promovem a evasão do sistema imune.

Como proposta para melhoria do processo é recomendado um design racional da amostragem para sequenciamento, levando em consideração local e períodos amostrados, além da descentralização racional da infraestrutura de sequenciamento de modo a atender diferentes regiões em um espaço de tempo adequado, considerando também questões logísticas.