

CONHECIMENTOS ESPECÍFICOS

Tendo em vista que a busca pela identificação de genes e o estudo das funções da *Arabidopsis thaliana* — que, no ano 2000, foi a primeira planta a ter seu genoma sequenciado — muito têm contribuído para a pesquisa na área de biologia vegetal, julgue os itens a seguir. Nesse sentido, considere que a sigla ORFs, sempre que empregada, se refere a *open read frames*.

- 51 O algoritmo de busca de ORFs no genoma de um organismo eucarioto é considerada uma tarefa de baixa complexidade.
- 52 Na anotação de genomas, a busca de domínios conservados é importante para o entendimento de suas funções e estudos evolucionários, sendo o programa Blast a ferramenta mais utilizada nesse processo.
- 53 Tamanho da planta, facilidade de multiplicação e genoma nuclear de tamanho consideravelmente grande são características que favorecem o uso da *A. thaliana* como organismo modelo para pesquisas em biologia molecular vegetal.

Com relação à análise de sequenciamento de genomas e ao GRC (*Genome Reference Consortium*), que é responsável por manter e melhorar a montagem do genoma de referência humano e do de outras espécies, julgue os itens seguintes.

- 54 Os elementos repetitivos de um genoma auxiliam em sua montagem, uma vez que há mais sobreposições nas sequências consenso.
- 55 O algoritmo do programa Blast produz a solução de um problema de grande processamento em um tempo razoável e com resultado aceitável, o que o torna mais rápido e mais indicado para alinhamento de grandes genomas.
- 56 Para acessar e utilizar os dados da nova montagem (*assembly*) do genoma humano (GRCh38), que foi liberado recentemente, o interessado deve pagar uma licença ao GRC.
- 57 O genoma é definido como uma completa amostragem dos genes expressos sob uma condição particular.

Julgue os próximos itens, relativos ao sequenciamento de fragmentos de cDNA (transcritos).

- 58 Se uma sequência proteica com 159 aminoácidos tiver 72% de homologia com outra sequência encontrada no banco de dados de proteínas, ela será caracterizada como conservada.
- 59 Para comprovar a existência de proteínas de interesse em sequências consenso, chamadas de *contigs*, recomenda-se utilizar o programa Blastp e um banco de dados de sequências de proteínas.
- 60 Alternativamente ao Blast, o programa Fasta pode ser utilizado para a busca de sequências proteicas em bancos de dados, visto que este proporciona maior precisão no resultado da busca.
- 61 Duas sequências proteicas com tamanho superior a 100 aminoácidos e que possuam 47% de identidade são consideradas homólogas.

No que se refere à linha de pesquisa denominada genômica comparativa, que possibilita entender a estrutura e a função dos genes envolvidos entre as diferentes espécies, bem como realizar estudos da evolução de genes e organismos, julgue os itens subsequentes.

- 62 A região da sequência de proteínas de diferentes espécies em que for encontrado um grupo de aminoácidos perfeitamente alinhados pode pertencer a um sítio ativo conservado.
- 63 Ao se comparar o genoma humano com o de um rato, observou-se a existência de grandes regiões de DNA em que os genes estão na mesma ordem, as quais são definidas como regiões sintênicas.
- 64 Dois genes são denominados ortólogos quando, em sua evolução, tiver havido divergência de uma espécie em duas, sendo mantida, porém, sua função biológica.
- 65 Dois genes são denominados parálogos quando exercem a mesma função biológica em espécies diferentes.

O vírus Ebola pertence à família *Filoviridae*, cujo gênero Ebolavirus apresenta atualmente cinco subtipos: *Zaire ebolavirus*, *Sudan ebolavirus*, *Tai Forest ebolavirus*, *Reston ebolavirus* e *Bundibugyo ebolavirus*. Acerca das ferramentas que podem auxiliar na identificação da presença desse vírus, julgue os itens que se seguem.

- 66 A técnica qRT-PCR é um tipo de PCR quantitativa, também chamada de PCR em tempo real, na qual é possível quantificar o número de moléculas amplificadas e corresponde a uma alternativa para detecção da presença do vírus Ebola.
- 67 A técnica convencional de sequenciamento Sanger, que produz até 384 sequências por corrida, pode ser utilizada para o sequenciamento do genoma do vírus Ebola.
- 68 A classificação filogenética das espécies ou subtipos do vírus Ebola pode ser realizada utilizando-se os programas de alinhamento Muscle ou Tcoffe.
- 69 Uma das técnicas utilizadas para detectar a presença do vírus Ebola é a RT-PCR: utiliza-se uma amostra de sangue do paciente da qual se isola o DNA viral.
- 70 Nas reações de RT-PCR para identificação do Ebola, utilizam-se *primers* (iniciadores), que são oligonucleotídeos complementares ao genoma viral.

Frederick Sanger, um dos cientistas mais importantes do século XX, foi agraciado com o prêmio Nobel de Química devido a sua contribuição para o desenvolvimento de um método de sequenciamento de DNA, batizado com seu nome, que possibilitou a revolução genômica da Biologia na última década daquele século.

Tendo o fragmento de texto acima como referência inicial, julgue os seguintes itens, acerca das ciências genômicas.

- 71 A bioinformática foi determinante para o sucesso da genômica: as ferramentas computacionais possibilitaram montar as sequências dos genomas. Atualmente, algoritmos avançados permitem prever a localização exata dos genes, bem como estabelecer sua função celular.
- 72 Atualmente, as sequências das regiões centroméricas e teloméricas dos cromossomos humanos não são completamente conhecidas, uma vez que a alta incidência de regiões repetitivas, que caracteriza o genoma humano, dificulta a montagem nesses segmentos.
- 73 Uma consequência da utilização dos microarranjos de DNA, a primeira forma experimental de medir níveis de expressão gênica de milhares de genes em paralelo, foi a disseminação, na bioinformática, da linguagem de programação R, que facilita a complexa análise dos dados gerados por essa metodologia.
- 74 As mobilizações de elementos transponíveis, que constituem aproximadamente metade do genoma humano, ocorrem tanto em células somáticas quanto em células germinativas, mas somente estas últimas podem ser herdadas, podendo ou não ter efeito fenotípico.
- 75 A predição computacional de genes de microRNAs envolve a busca no genoma de segmentos de dezenas de nucleotídeos capazes de formar uma estrutura terciária característica, na forma de grampo, que serve como precursor para a molécula madura.
- 76 A tecnologia Sanger para o sequenciamento de DNA baseia-se essencialmente nas distinções químicas entre as bases nitrogenadas de um nucleotídeo para decifrar a sequência do DNA.

A partir da metade da primeira década do século XXI, com o advento de novas tecnologias de sequenciamento de DNA (tecnologias NGS), ocorreu uma revolução nas ciências genômicas. A capacidade de geração de sequências foi incrementada, com a consequente queda nos custos. No que se refere a essas tecnologias, julgue os itens de 77 a 82.

- 77 Embora tenha custo elevado, o sequenciamento de um genoma pelo método Sanger gera uma montagem do genoma com maior tamanho médio de regiões contíguas de sequência (*contigs*) comparado com tecnologias NGS, dada a mesma cobertura média de bases.

- 78 Na etapa de pré-processamento de dados obtidos a partir de tecnologias NGS, recomenda-se remover os segmentos de vetores de clonagem e adaptadores, as sequências de baixa qualidade e as duplicatas de PCR (*polymerase chain reaction*).
- 79 O RNA-Seq, técnica baseada em tecnologias NGS para a análise da expressão gênica, apresenta como vantagem a possibilidade de resolução de múltiplas isoformas dos transcritos e como desvantagem a necessidade de um genoma de referência conhecido, o que limita a sua aplicação.
- 80 O formato FASTQ é o padrão de saída de dados dos novos sequenciadores automáticos, o qual fornece, além da sequência de cada fragmento, a probabilidade de erro na nomeação de cada base.
- 81 A epigenômica estuda fenômenos epigenéticos, como as modificações pós-traducionais das histonas e a metilação do DNA, as quais são identificadas a partir do emprego de técnicas como o ChIP-Seq e o sequenciamento bissulfeto, respectivamente.
- 82 O mapeamento genômico de sequências curtas oriundas de tecnologias NGS contra um genoma de referência utiliza algoritmos derivados do programa Blast, que faz uma busca eficiente por similaridade entre um grande número de sequências.

O prêmio Nobel de Química de 2013 foi concedido a cientistas pioneiros da bioinformática estrutural, que desenvolveram modelos computacionais para a simulação de dinâmica de proteínas. Acerca da bioinformática estrutural, julgue os itens que se seguem.

- 83 Bancos de dados como Pfam, SMART e CDD são importantes para auxiliar na anotação funcional de proteínas, uma vez que catalogam milhares de domínios proteicos, os quais são segmentos discretos e conservados que adotam uma estrutura tridimensional particular no contexto da proteína.
- 84 Programas de predição de estrutura que utilizam a modelagem por homologia, dependem de proteínas que possuam estrutura conhecida e identidade de sequência com a proteína predita.
- 85 Na dinâmica molecular, aproximações baseadas em modelos da física clássica são utilizadas para simular o movimento coordenado dos átomos e, com isso, capturar as flutuações conformacionais de moléculas, incluindo proteínas e ácidos nucleicos.
- 86 Com os avanços da simulação computacional, foi solucionado o problema da predição da estrutura tridimensional de proteínas a partir de sua estrutura primária, eliminando-se a lacuna, criada pela genômica, de milhões de sequências disponíveis em bancos de dados sem informação estrutural.

Na década de 70 do século passado, o cientista Carl Woese revolucionou o meio científico ao utilizar a sequência do RNA ribossômico (rRNA) para classificar os organismos. Com esse uso pioneiro da filogenia molecular, surgiu a classificação dos organismos em três domínios: eucariotos, bactérias e arquea. Considerando esse assunto, julgue os próximos itens, referentes ao papel da bioinformática no estudo da evolução de genes e de organismos.

87 O uso de sequências moleculares para estudos evolutivos requer como entrada um alinhamento múltiplo de sequências gerado por programas como *Clustalw* e *Muscle*.

88 A construção de árvores filogenéticas, que se baseia em sequências moleculares de proteínas ou de nucleotídeos, utiliza metodologias como máxima verossimilhança, máxima parcimônia e inferência bayesiana para fornecer representações exatas da taxonomia de um grupo de organismos.

89 Na análise de um alinhamento múltiplo de sequência de parálogos de uma família proteica, a observação de pares de resíduos (colunas do alinhamento) com frequência de substituição correlacionada sugere que estes estabelecem interações importantes para manutenção estrutural ou funcional (sítio ativo) na proteína e, por isso, observa-se a sua coevolução.

90 A metagenômica estuda as comunidades microbianas nos seus ambientes naturais. Técnicas de NGS são utilizadas para sequenciar regiões do gene de rRNA 16S. Computacionalmente, mesmo sequências que apresentem diferenças de até 3% das bases podem ser consideradas provenientes da mesma espécie.

Tendo em vista que a bioinformática possui estreita relação com o sistema operacional Linux, *softwares* livres e dependência de *hardwares* específicos para viabilizar as análises, julgue os itens de 91 a 95, relativos ao desenvolvimento de *software* na bioinformática.

91 O Linux, sistema operacional da família Unix, é amplamente utilizado na bioinformática por ser livre e possuir uma distribuição oficial centralizada.

92 Linguagens de programação compiladas, como C e FORTRAN, são direcionadas a aplicações que requerem alto desempenho de processamento, enquanto linguagens interpretadas, como PERL e Python, são utilizadas na maioria dos casos devido ao menor tempo de desenvolvimento e à necessidade de menos linhas de código-fonte.

93 Muitos servidores empregam o arranjo redundante de discos independentes (RAID), uma solução de virtualização de armazenamento que fornece ganhos de desempenho e de redundância. Nesse sentido, RAID 0 corresponde ao nível que oferece maior redundância e maior desempenho.

94 O modelo de computação em nuvem representa uma alternativa para redução de custos de implantação e manutenção de um laboratório de bioinformática, pois provê uma infraestrutura remota de virtualização de *software* e de *hardware* que permite o armazenamento e a análise de dados, baseado no sistema operacional Hadoop e no modelo de programação MapReduce.

95 A montagem de genomas requer grandes quantidades de memória RAM compartilhada (centenas de *gigabytes*), ao passo que uma simulação de dinâmica molecular exige uma arquitetura com múltiplos processadores e até mesmo unidades de processamento gráfico (GPUs).

Com o advento do sequenciamento de ácidos nucleicos em larga escala, tornou-se possível realizar a comparação de organismos em termos de suas sequências de DNA e das proteínas codificadas por essas sequências. No que diz respeito ao alinhamento de sequências, julgue os itens a seguir.

96 A matriz de regiões coincidentes entre duas sequências busca o alinhamento com menor similaridade calculada pelos algoritmos de programação dinâmica.

97 O alinhamento local permite que se preveja o melhor alinhamento possível, bem como o mais rápido, uma vez que ele busca somente alinhamentos entre regiões de alta diversidade entre as sequências.

98 Um alinhamento múltiplo auxilia no estabelecimento de uma relação evolutiva entre organismos e permite identificar sequências conservadas.

99 Os alinhamentos de genomas completos podem revelar deleções completas de blocos gênicos ou duplicações gênicas, as quais são importantes mecanismos de evolução.

Acerca do sequenciamento de genomas, julgue os itens seguintes.

- 100** O método de sequenciamento de Sanger demanda que um genoma seja fragmentado em partes, por meio da digestão com enzimas de restrição, para que se formem fragmentos contíguos (*contigs*).
- 101** O termo genoma é usado para designar o conjunto de proteínas de uma célula ou organismo.
- 102** A transcritômica visa sequenciar e analisar as sequências nucleotídicas a fim de identificar genes e outras sequências importantes, como as regiões reguladoras de genes.
- 103** Os fragmentos de DNA podem ser sequenciados por espectrometria de massa, que apresenta como vantagem a capacidade de distinguir resíduos específicos de citidina e timidina, além de resolver misturas isotópicas.
- 104** Faz parte da proteômica, subárea da genômica, a identificação do grupo de proteínas de uma célula ou organismo sob um conjunto de condições. Essa subárea, porém, não envolve os estudos de modificações pós-traducionais daquelas proteínas.
- 105** O método de Sanger utiliza a DNA polimerase para degradar uma fita de DNA, de modo a liberar nucleotídeos radiativos detectados pelo sensor.

A importância de se sequenciar genomas e transcritomas reside na informação que advém dessas sequências. Os bancos de dados de sequências organizam e catalogam as sequências de maneira lógica. A respeito desses bancos de dados, julgue os itens subsequentes.

- 106** O banco de dados CAZY armazena todas as proteínas e suas sequências degradadas por enzimas de restrição.
- 107** O Centro Nacional para Informação Biotecnológica contém o GenBank, que consiste no principal banco de dados de sequências de DNA, RNA e proteínas.
- 108** O número de acesso P00533 de uma proteína no *swissprot* corresponde a uma identificação única dessa sequência.

Acerca da anotação de sequências genômicas em bancos de dados, julgue os itens que se seguem.

- 109** O RefSeq representa a subdivisão do GenBank que contém sequências de processamento alternativo de forma redundante em determinado genoma.
- 110** O banco de dados GEO (*gene expression omnibus*) é o repositório público de dados de genômica relacionados ao sequenciamento de alto desempenho, genotipagem e arranjos de genes com expressão validada por curadoria.
- 111** A principal desvantagem do CHADO é a modulação por localização. Diferentemente dos esquemas de bancos de dados GFF e BioSQL, o CHADO é o único esquema que cria um subgrupo ou *rank*.
- 112** O banco de dados do GenBank, denominado OMIM, armazena os nucleotídeos (DNA e RNA) que estejam patenteados.

Julgue os próximos itens, relativos a algoritmos e bancos de dados aplicados à bioinformática.

- 113** InterproScan é um programa de busca que combina diferentes algoritmos para determinar assinaturas proteicas em várias bases de dados de estruturas de domínio e(ou) famílias de proteínas tridimensionais.
- 114** GENpept é um banco de dados de traduções automáticas das regiões repetitivas das sequências de nucleotídeos do GenBank.
- 115** O banco de dados HTC é o repositório de sequências genômicas geradas em larga escala e ainda incompletas.
- 116** As sequências armazenadas no GenBank são classificadas e organizadas usando-se termos taxonômicos compreensíveis desenvolvidos pelos mantenedores do banco de dados.
- 117** Em um programa de alinhamento múltiplo, o formato de submissão de sequências mais usado é o FASTA.
- 118** A programação dinâmica recursiva considera cada solução parcial no passo seguinte para que seja calculada com um número ilimitado de soluções parciais, de modo que o passo final contere a solução global.
- 119** A programação dinâmica consiste na busca de uma solução para um problema computacional, em um grande espaço de procura, por meio de cálculos iterativos.
- 120** O PDB (*protein dataBank*) é o principal repositório público devotado a estruturas tridimensionais de macromoléculas biológicas.

PROVA DISCURSIVA

- Nesta prova, faça o que se pede, usando o espaço para rascunho indicado no presente caderno. Em seguida, transcreva o texto para a **FOLHA DE TEXTO DEFINITIVO DA PROVA DISCURSIVA**, no local apropriado, pois **não será avaliado fragmento de texto escrito em local indevido**.
- Qualquer fragmento de texto além da extensão máxima de linhas disponibilizadas será desconsiderado.
- Na **folha de texto definitivo**, identifique-se apenas no cabeçalho da primeira página, pois **não será avaliado** texto que tenha qualquer assinatura ou marca identificadora fora do local apropriado.
- Nesta prova, ao domínio do conteúdo serão atribuídos até **40,00 pontos**, dos quais até **2,00 pontos** serão atribuídos ao quesito apresentação (legibilidade, respeito às margens e indicação de parágrafos) e estrutura textual (organização das ideias em texto estruturado).

O ano de 2014 pode ser o mais quente desde o início dos registros de temperatura no mundo, em 1880. O alerta veio da Administração Nacional de Oceanos e Atmosfera dos Estados Unidos da América, após a divulgação de que os meses de maio, junho, agosto e setembro bateram recordes de calor. Desde o início das medições, 2005 e 2010 foram os anos mais quentes da história. O pequeno intervalo entre os anos é um exemplo do efeito crescente das mudanças climáticas. Os dez anos mais quentes já registrados ocorreram nos últimos quinze anos e esta é a primeira vez em que o mês de setembro apresenta temperaturas tão altas sem a forte presença do fenômeno El Niño, que, no entanto, ainda pode manifestar-se este ano.

O Globo, 22/10/2014, p. 30 (com adaptações).

Considerando que o fragmento de texto acima tem caráter meramente motivador, redija um texto dissertativo acerca do seguinte tema.

MUDANÇAS CLIMÁTICAS E O DESAFIO DO DESENVOLVIMENTO SUSTENTÁVEL

Ao elaborar seu texto, aborde, necessariamente, os seguintes aspectos:

- ▶ fatores determinantes para a elevação da temperatura; [valor: 12,50 pontos]
- ▶ impacto das alterações do clima na vida das sociedades; [valor: 12,50 pontos]
- ▶ sustentabilidade como pressuposto para o desenvolvimento. [valor: 13,00 pontos]

RASCUNHO

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	



cespe

 Cebraspe

Centro Brasileiro de Pesquisa em
Avaliação e Seleção e de Promoção de Eventos