

FIOCRUZ

Concurso Público Fiocruz 2023

Tecnologista em Saúde Pública

Prova Objetiva e Discursiva

TE54 - Bioinformática



Prova Objetiva

01. Em sistemas operacionais como o Linux, o diretório raiz (/):

- (A) armazena os arquivos-folha temporários gerados dinamicamente durante o funcionamento do sistema operacional.
- (B) define o primeiro diretório na hierarquia de diretórios e arquivos.
- (C) contém os arquivos de configuração do ambiente de cada usuário, definindo o seu ambiente de trabalho.
- (D) contém os arquivos executáveis e as variáveis de ambiente de cada usuário.
- (E) define os caminhos relativos dos arquivos durante a inicialização do sistema operacional.

02. No sistema operacional Linux, a interface *shell* permite tanto executar programas distribuídos com o sistema operacional como desenvolver programas customizados. O comando *shell*, que permite iterar todos os arquivos *fastq* no diretório atual e mapeá-los no banco de dados denominado "referencia", o qual contém o genoma de referência indexado, também no diretório atual, produzindo arquivos *.sam* como saída, é:

- (A) `ls *.fastq | bwa mem referencia *fastq > *.sam`
- (B) `grep *.fastq | bwa mem referencia *fastq > *.sam`
- (C) `find *.fastq | bwa mem referencia *fastq > *.sam`
- (D) `search for *.fastq | bwa mem referencia *fastq > *.sam`
- (E) `for i in `ls *.fastq`; do bwa mem referencia $i > ${i}.fastq.sam; done`

03. Expressões regulares compreendem uma ferramenta de extrema importância para o processamento de arquivos de texto visando a detecção de padrões. Na bioinformática, expressões regulares são comumente utilizadas para detectar padrões de texto. A linguagem de programação python contém suas funcionalidades de expressão regular implementadas no módulo *re*. Considerando um arquivo texto contendo as sete linhas abaixo, a opção com o número de linhas distintas, que seriam selecionadas pela expressão regular `^ID\d+\.\d+$`, é:

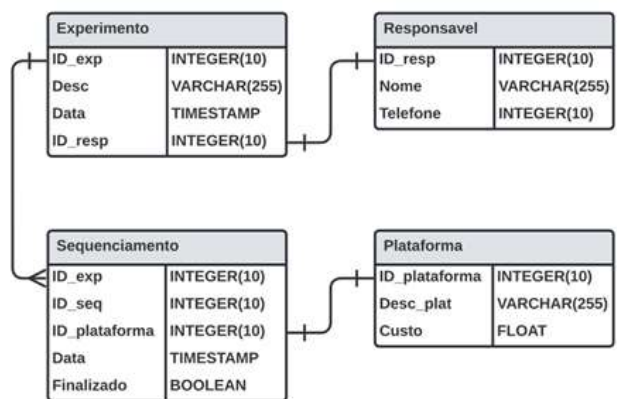
```
ID34343.A
ID34HN43.1
ID985433230
ID852495_23
_ID423243.1
ID2544343.97
ID1.1
```

- (A) sete linhas.
- (B) nenhuma linha.
- (C) cinco linhas.
- (D) seis linhas.
- (E) duas linhas.

04. No aprendizado de máquina, a etapa de validação cruzada (*cross-validation*) tem como objetivo principal:

- (A) treinar o modelo em um conjunto de dados desconhecido.
- (B) diminuir a complexidade dos modelos, aumentando sua explicabilidade.
- (C) avaliar a capacidade de generalização do modelo.
- (D) comparar o desempenho de modelos distintos.
- (E) diminuir a variância do modelo.

05. O modelo de banco de dados Entidade-Relacionamento (*Entity-relationship Model*) é utilizado há mais de 50 anos para descrever bancos de dados relacionais em termos de suas entidades e relacionamentos. Observe o esquema de banco de dados abaixo.



O esquema de banco de dados representado acima contém, respectivamente, ____ entidades e ____ relacionamentos.

A opção que preenche as lacunas da frase acima é:

- (A) 15 e 4.
- (B) 12 e 4.
- (C) 12 e 3.
- (D) 4 e 3.
- (E) 1 e 2.

06. O número de tipos de variáveis distintos representados no banco de dados apresentado na questão anterior é:

- (A) 3.
- (B) 5.
- (C) 7.
- (D) 12.
- (E) 15.

07. A abordagem de sequenciamento NGS comumente utilizada para avaliar interações proteína-DNA, tais como fatores de transcrição, é denominada:

- (A) ChIP-seq.
- (B) Bisulfite sequencing.
- (C) Methyl-seq.
- (D) RNA-seq.
- (E) Exome sequencing.

08. A tecnologia de sequenciamento de DNA conhecida pela sua capacidade de detectar modificações de nucleotídeos, tais como citosinas metiladas, durante o processo de sequenciamento é denominada:

- (A) Illumina.
- (B) SOLiD.
- (C) Ion Torrent.
- (D) Oxford Nanopore.
- (E) Sanger.

09. Algoritmos de alinhamento heurísticos como o BLAST visam:

- (A) reduzir o espaço de busca e acelerar o processo de alinhamento.
- (B) garantir que os alinhamentos ótimos serão sempre encontrados.
- (C) eliminar a necessidade do cálculo das matrizes de pontuação.
- (D) aumentar a sensibilidade do algoritmo para recuperar homólogos distantes.
- (E) permitir o alinhamento de múltiplas sequências.

10. O marcador mais adequado para estudos filogeográficos que visem inferir as relações evolutivas entre populações de uma mesma espécie compreende:

- (A) genes cópia simples.
- (B) variantes raras.
- (C) SNPs.
- (D) genes codificados de rRNA.
- (E) genes conservados em organismos filogeneticamente distantes.

11. Os vírus humanos de RNA destacam-se por sua:

- (A) alta capacidade de infectar espécies domésticas.
- (B) alta taxa de mutações.
- (C) alta infectividade.
- (D) baixa variabilidade fenotípica.
- (E) conservação em relação ao tamanho de seus genomas.

12. A propriedade genômica compartilhada entre os patógenos eucarióticos da espécie humana *Trypanosoma cruzi*, *Giardia lamblia* e *Plasmodium vivax* compreende:

- (A) a abundância de famílias multigênicas variáveis e imunogênicas.
- (B) o elevado número de genes codificadores de proteínas comparado ao genoma humano.
- (C) a abundância de genes conservados do metabolismo biossintético/energético.
- (D) a presença de proteínas para a invasão de células hospedeiras.
- (E) a presença de proteínas para a interação com o inseto vetor invertebrado.

13. Um desafio comumente encontrado na montagem de genomas complexos de eucariotos, utilizando tecnologias de leituras curtas (*short reads*), quando comparadas às tecnologias de leituras longas (*long reads*), é a:

- (A) dificuldade em obter grandes coberturas de sequenciamento.
- (B) dificuldade em obter modelos gênicos acurados.
- (C) elevação do custo por nucleotídeo sequenciado.
- (D) dificuldade no preparo da amostra.
- (E) dificuldade em resolver regiões repetitivas.

14. O processo de scaffolding na montagem de genoma é importante porque:

- (A) remove as leituras (reads) de baixa qualidade da montagem, melhorando a etapa de predição gênica.
- (B) permite o alinhamento das leituras (reads) ao genoma de referência.
- (C) auxilia no ordenamento e orientação dos contigs para reconstruir fragmentos genômicos maiores.
- (D) permite inferir o número de genes a partir dos dados de sequenciamento.
- (E) permite avaliar a qualidade da montagem em função da presença e ausência de marcadores.

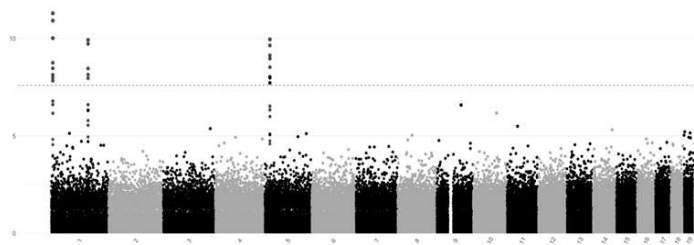
15. O propósito primário da normalização em dados de RNA-Seq visa:

- (A) ajustar para diferenças na profundidade de sequenciamento e no tamanho dos transcritos.
- (B) remover duplicatas de PCR que podem superestimar os valores de expressão de alguns genes.
- (C) identificar transcritos expressos diferencialmente.
- (D) alinhar os resultados de sequenciamento das bibliotecas ao genoma de referência.
- (E) identificar isoformas expressas diferencialmente.

16. Os resultados de expressão gênica diferencial resultam em listas que muitas vezes compreendem centenas de genes. A metodologia comumente utilizada para interpretar estas listas gênicas através da busca por funções biológicas e vias bioquímicas comuns aos genes da lista é denominada:

- (A) expressão em células únicas.
- (B) contextualização.
- (C) construção de redes de coexpressão.
- (D) comparação com bancos de dados curados de expressão gênica.
- (E) análise de enriquecimento.

17. Usualmente, resultados de GWAS são reportados na forma de gráficos do tipo Manhattan (vide abaixo). O valor estatístico do eixo Y, comumente utilizado para evidenciar associações interessantes, compreende:



- (A) Fold change.
- (B) Log10(fold change).
- (C) Valor-p.
- (D) $-\log_{10}(\text{valor-p})$.
- (E) F-score (média harmônica entre sensibilidade e especificidade).

18. O formato de configuração que corresponde ao comumente utilizado para definir parâmetros e configurações em linguagem de fluxo computacional como Snakemake e Nextflow é:

- (A) JSON.
- (B) YAML.
- (C) XML.
- (D) CSV.
- (E) INI.

19. A análise topológica de redes de interação proteína-proteína permite:

- (A) inferir as relações evolutivas das proteínas.
- (B) mensurar a força das interações proteína-proteína.
- (C) identificar proteínas regulatórias.
- (D) identificar proteínas conservadas entre espécies.
- (E) selecionar proteínas variáveis entre as condições experimentais.

20. A função de pontuação (*scoring function*), comumente utilizada em simulações de ancoragem molecular (*molecular docking*) para avaliar a afinidade de ligação entre um ligante e uma proteína, é denominada:

- (A) distância de Hamming.
- (B) distância Euclidiana.
- (C) RMSD (*Root Mean Square Deviation*) entre os carbonos-alfa.
- (D) energia livre de Gibbs.
- (E) coeficiente de correlação de Pearson entre as distâncias atômicas.

21. Os algoritmos de alinhamento de sequências são essenciais para a análise de sequências biológicas. Esses algoritmos são utilizados em diversas tarefas na Bioinformática, tais como montagem de genomas, análise filogenética e busca por similaridade. Com relação aos algoritmos de alinhamentos, analise as assertivas abaixo.

- I. O algoritmo de alinhamento global Needleman-Wunsch consome tempo $O(nm)$, onde n e m são os comprimentos das sequências que serão alinhadas.
- II. A matriz de programação dinâmica que o algoritmo Smith-Waterman calcula tem entradas negativas ao alinhar duas sequências de nucleotídeos no sistema de escore que fornece uma penalidade de -5 de abertura de lacuna.
- III. O e-value é o valor de probabilidade de encontrar, ao acaso, um hit com um escore maior que o escore calculado do alinhamento.
- IV. Dependendo do sistema de pontuação utilizado, o problema de alinhamento múltiplo é NP-hard.
- V. O algoritmo de alinhamento semi-global pode ser utilizado para ajudar na montagem de genomas.

Das assertivas acima, apenas:

- (A) I, IV e V são verdadeiras.
- (B) II é verdadeira.
- (C) III é verdadeira.
- (D) IV é verdadeira.
- (E) III e IV são verdadeiras.

22. O Bash é um interpretador de comandos do UNIX. Uma ferramenta disponível no Bash é o *pipe*, simbolizado por '|'. O *pipe* é utilizado para passar a saída de um comando como entrada para outro comando. Arquivos no formato FASTA são amplamente utilizados na bioinformática. Considerando o uso do *pipe*, a linha de comando que imprime apenas **um único número**, indicando quantas sequências o "arquivo.fasta" possui, é:

- (A) `grep ">" arquivo.fasta | wc -m`
- (B) `cat arquivo.fasta | grep ">" | cat -n | tail -n 1`
- (C) `cat arquivo.fasta | wc -l`
- (D) `grep -c ">" arquivo.fasta | wc`
- (E) `cat arquivo.fasta | grep ">" | wc -l`

23. Python é uma linguagem de programação amplamente utilizada na Bioinformática. Através do *lambda* é possível criar funções anônimas. O código que está sintaticamente correto, de acordo com a versão da linguagem Python superior a 3.10, é:

- (A) `print((lambda x: x + 1)10)`
- (B) `print(lambda x: x + 1)`
- (C) `print((lambda x: x + 1)(10))`
- (D) `print(lambda x: x + 1)(10)`
- (E) `print((lambda x:: x + 1)(10))`

24. Snakemake é um gerenciador de *workflows* baseado no paradigma do GNU Make. Neste paradigma, define-se um conjunto de regras; cada regra especifica como criar um arquivo de saída a partir de arquivos de entrada. O conjunto dessas regras e as dependências entre elas estabelecem um grafo de dependências entre as tarefas. É correto afirmar que o grafo é:

- (A) não dirigido.
- (B) dirigido, mas pode possuir ciclos.
- (C) dirigido e acíclico.
- (D) uma árvore.
- (E) uma floresta.

25. Os clusteres de alto desempenho focam em maximizar o desempenho para tarefas computacionais intensivas. Contudo, para se fazer um bom uso dos recursos de um cluster é necessário o conhecimento de modelos de programação paralela como o MPI (Message Passing Interface) e o OpenMP. Nesse contexto, é correto afirmar que:

- (A) o MPI é um modelo de programação paralela e que possui diversas implementações; um exemplo de implementação é o MPICH.
- (B) o OpenMP utiliza um modelo de memória compartilhada, em que múltiplas threads executam em paralelo e tem acesso a uma memória comum; nesse caso, não é necessário se preocupar com a sincronização entre as threads.
- (C) o OpenMP possui diretivas de pré-processador que podem ser inseridas no código-fonte, já o MPI também tem essa funcionalidade.
- (D) o MPI é baseado em troca de mensagens e não pode ser utilizado para aplicações distribuídas.
- (E) o MPI, assim como o OpenMP, utiliza memória compartilhada.

26. O Biopython é amplamente utilizado para realizar análises na área da Bioinformática. A interface Bio.SeqIO é utilizada para realizar a entrada e saída de arquivos suportando muitos formatos distintos. Considerando o código abaixo, é correto afirmar que:

```
from Bio import SeqIO

for record in SeqIO.parse("example.fasta", "fasta"):
    print(record.id)
```

- (A) o Bio.SeqIO.parse() está devolvendo apenas a próxima sequência.
- (B) o objeto record é do tipo Seq.
- (C) o Bio.SeqIO.parse() devolve uma lista de sequências.
- (D) o Bio.SeqIO.parse() recebe o nome do arquivo de entrada e o formato do arquivo, e devolve um SeqRecord iterator.
- (E) um objeto SeqRecord possui o campo length.

27. A escolha entre usar um teste de hipótese paramétrico ou não paramétrico depende das características dos dados e dos objetivos da análise. Por exemplo, se as suposições para um teste paramétrico são atendidas, prefere-se usar esses testes devido ao seu maior poder estatístico. Em relação aos testes de hipóteses, é INCORRETO afirmar que:

- (A) o teste de Shapiro-Wilk pode ser utilizado para testar a normalidade dos dados.
- (B) o teste de Mann-Whitney é um exemplo de teste não-paramétrico.
- (C) o teste t de Student deve assumir a normalidade dos dados.
- (D) o ANOVA é uma alternativa não-paramétrica para o teste de Kruskal-Wallis.
- (E) o teste de chi-quadrado se destina a encontrar um valor da dispersão para duas variáveis categóricas nominais e avaliar a associação existente entre variáveis qualitativas.

28. A seleção de características (feature selection) é uma etapa importante no contexto de aprendizado de máquina, principalmente quando há diversas dimensões que podem ser exploradas. O conjunto de características (features) selecionadas pode ser efetivamente utilizado para construir modelos preditivos ou realizar outras análises estatísticas. Dentro deste contexto, é correto afirmar que a seleção de característica:

- (A) não contribui para melhorar a precisão do modelo ao remover ruídos dos dados.
- (B) não pode ser feita utilizando algoritmos baseados em filtro, embutidos ou de wrapper.
- (C) é um processo que pode aumentar a velocidade de treinamento do modelo independentemente do tamanho do conjunto de dados.
- (D) é necessária apenas quando o modelo de aprendizado de máquina apresenta overfitting.
- (E) não tem capacidade de reduzir a complexidade do modelo, facilitando sua interpretação e reduzindo o risco de overfitting.

29. Para avaliar o erro de generalização de um classificador, são empregadas várias técnicas de validação. Uma delas divide o conjunto de teste em k segmentos, utilizando, em cada iteração, um segmento diferente para validação e os demais para treinamento. Outra técnica consiste em treinar o modelo com o conjunto completo, excluindo apenas um elemento, que é então usado para teste, e repetindo este processo para cada um dos elementos. Essas metodologias são reconhecidas na comunidade científica por nomes específicos. Neste contexto, são métodos de validação os abaixo relacionados, EXCETO:

- (A) k-fold cross validation.
- (B) leave-one-out.
- (C) bootstrapping.
- (D) hold-out.
- (E) split testing.

30. A normalização é um processo crucial no design de bancos de dados, visando organizar tabelas e suas inter-relações para reduzir redundâncias e dependências entre os dados. Este processo previne problemas comuns como anomalias de inserção, atualização e exclusão, ao mesmo tempo em que reforça a integridade e consistência dos dados. A normalização se concretiza pela adoção de um conjunto de regras denominadas formas normais, cada uma destinada a resolver questões específicas que podem resultar em ineficiências no armazenamento de dados e na realização de consultas. Neste contexto, é INCORRETO afirmar que:

- (A) na primeira forma normal (1NF), cada atributo da tabela deve ser atômico, significando que os campos não devem permitir conjuntos de valores ou listas, garantindo que cada campo contenha um único valor por registro.
- (B) a 2FN exige que a tabela esteja na 1FN e que todos os atributos não-chave dependam unicamente e completamente da chave primária, sem dependências parciais.
- (C) a 2FN não pode ter dependências parciais em chaves concatenadas.
- (D) uma tabela está na Terceira Forma Normal (3FN) se ela está na Segunda Forma Normal (2FN) e, além disso, não possui dependências transitivas entre os atributos não-chave e a chave primária.
- (E) a Quarta Forma Normal (4NF) ajuda a garantir que cada tipo de informação seja armazenado várias vezes, melhorando a eficiência e otimizando o uso de memória.

31. O Sequenciamento de Nova Geração (NGS, do inglês "Next-Generation Sequencing") refere-se a uma família de tecnologias de sequenciamento de DNA e RNA que permitiram avanços significativos na genômica e em áreas relacionadas, como a biologia molecular, a medicina personalizada e a microbiologia. Comparado às técnicas de sequenciamento de primeira geração, como o método de Sanger, o NGS é capaz de sequenciar bilhões de fragmentos de DNA simultaneamente, tornando-o mais rápido e mais barato. Essa capacidade revolucionou a pesquisa genética, possibilitando uma ampla gama de aplicações que vão desde o sequenciamento do genoma completo até a análise de expressão gênica, metagenômica e epigenômica. Em relação a este contexto, a alternativa que possui um conjunto de programas, todos específicos para análise de dados de NGS é:

- (A) HISAT2, Velvet, e Cap3.
- (B) Bowtie, SOAPdenovo, e Cap3.
- (C) Fastqc, Bowtie, e BLAST.
- (D) ABySS, Velvet, e SOAPdenovo.
- (E) Mummer, Minia, e AbySS.

32. Uma forma de identificar genes que são significativamente regulados em diferentes condições, como em tecidos saudáveis versus patológicos, é através da análise diferencial de expressão gênica. A técnica RNA-Seq é a mais utilizada atualmente, sendo capaz de gerar dados de sequenciamento de alta vazão. Após o sequenciamento é importante aplicar métodos estatísticos para discriminar quais genes apresentam mudanças significativas na expressão. Um passo importante que deve ser realizado antes de qualquer análise diferencial de expressão é:

- (A) a normalização dos dados de expressão.
- (B) utilizar métodos independentes como qPCR.
- (C) filtrar genes de baixa expressão.
- (D) determinar que genes serão analisados.
- (E) pesquisar sobre quais genes serão estudados.

33. Um princípio importante para se realizar a reconstrução filogenética é o princípio da máxima parcimônia. Neste princípio, a árvore filogenética que minimiza a quantidade de mudanças evolutivas é preferida. É correto afirmar que a máxima parcimônia:

- (A) maximiza a quantidade de homoplasias.
- (B) baseia-se na premissa de que eventos evolutivos mais complexos são mais prováveis.
- (C) foca-se identificar a árvore com o menor comprimento total de ramo.
- (D) considera que as taxas de mutação são consistentes em todos os ramos da árvore.
- (E) é frequentemente utilizada para construir árvores que possuem maiores distâncias genéticas entre espécies.

34. Um componente importante na anotação de um genoma é a identificação precisa dos genes e sua estrutura no genoma. Os preditores de genes *ab initio* possuem um papel fundamental nesta etapa. Muitos desses programas utilizam modelos ocultos generalizados de Markov (GHMM – *Generalized Hidden Markov Model*). Neste contexto, para se utilizar um preditor de genes *ab initio* é importante:

- (A) realizar o alinhamento de sequências de transcritos no genoma, a fim de fornecer evidências adicionais sobre a localização dos genes.
- (B) empregar o alinhamento de sequências de proteínas ao genoma alvo, o que pode ajudar a identificar regiões codificantes.
- (C) selecionar um programa que contenha modelos específicos para o genoma em análise, visando a precisão das predições.
- (D) optar por um programa que ofereça a análise mais rápida do genoma, visando a eficiência no processamento de dados.
- (E) desenvolver um pipeline que incorpore várias evidências, incluindo dados experimentais e computacionais, para uma anotação genômica mais robusta.

35. Em estudos de associação genômica ampla (GWAS), é importante identificar polimorfismos de nucleotídeo único (SNPs) associados a fenótipos patológicos ou traços biológicos. PLINK é uma ferramenta de referência para o processamento e análise de dados de genética neste contexto. Dentro do *pipeline* do PLINK, uma etapa é o procedimento de controle de qualidade, que visa a remoção de SNPs que não atende a determinados parâmetros bioestatísticos. As alternativas abaixo são critérios ou medidas para filtragem de qualidade de SNP em análise GWAS, EXCETO:

- (A) Minor Allele Frequency.
- (B) Missing genotyping rate.
- (C) Hardy-Weinberg Equilibrium test.
- (D) Chromosomal Structural Integrity.
- (E) Mendel error rate.

36. No contexto de biologia sistêmica, as redes de interação proteína-proteína são ferramentas importantes para o entendimento do funcionamento celular. Essas redes podem ser consideradas “redes complexas” e, assim, ser analisadas utilizando ferramentas da teoria de grafos. Nesse contexto, a métrica que é mais relevante para identificar proteínas essenciais para a estabilidade ou a funcionalidade de uma rede de interação é:

- (A) número de componentes conectados.
- (B) grau de centralidade.
- (C) coeficiente de clustering.
- (D) comprimento do caminho mais curto.
- (E) centralidade de intermediação.

37. No contexto da análise de docking molecular, é correto afirmar que a caixa de docking (*grid*):

- (A) aumenta a precisão do docking ao permitir que o ligante interaja com toda a superfície da proteína.
- (B) delimita o espaço de busca, reduzindo o tempo computacional necessário para a análise.
- (C) ajuda na identificação de todos os possíveis sítios de ligação na proteína.
- (D) permite a análise de interações não específicas entre o ligante e a proteína.
- (E) facilita a visualização tridimensional da proteína ao destacar os sítios de ligação para análise visual.

38. Em sistemas operacionais, as condições descritas por Coffman são necessárias para que ocorra:

- (A) *deadlock*.
- (B) *starvation*.
- (C) *overloading*.
- (D) *swapping*.
- (E) *stack overflow*.

39. A montagem de genomas complexos representa ainda um dos desafios mais intrincados da Bioinformática. As sequências repetidas, em particular, fornecem uma dificuldade notável, pois essas podem confundir algoritmos de montagem, resultando em diversas e múltiplas possibilidades. Para mitigar esse problema e alcançar a montagem correta, a estratégia mais eficaz seria:

- (A) empregar técnicas de bioinformática estrutural para analisar a conformação tridimensional do DNA e resolver as repetições.
- (B) melhorar as heurísticas que percorrem o grafo de montagem, permitindo distinguir entre caminhos que representam repetições reais e artefatos.
- (C) aumentar a cobertura de sequenciamento, garantindo que as repetições possam ser atravessadas várias vezes para confirmar sua presença e posição.
- (D) utilizar informações de mapeamento de leituras contra um genoma de referência para desambiguar as regiões repetitivas.
- (E) sequenciar leituras que são longas o suficiente para englobar a região repetida inteira ou de modo que cada repetição tenha sua própria variante única.

40. Considerando o genoma de bactérias e o genoma de eucariotos, a diferença típica entre os dois é que bactérias:

- (A) têm cromossomos lineares, e eucariotos circulares.
- (B) têm cromossomos circulares, e eucariotos cromossomos lineares.
- (C) possuem DNA mitocondrial, enquanto eucariotos não.
- (D) utilizam histonas para superenrolar seu DNA; eucariotos não.
- (E) têm comumente genes com a presença de introns; eucariotos não.

Prova Discursiva

QUESTÃO

No início dos anos 2000, quando o genoma humano estava próximo de sua primeira versão pública, os cientistas envolvidos nesse projeto realizaram uma aposta informal sobre o número total de genes codificadores de proteínas no genoma humano, a qual ficou conhecida como gene sweep. Nessa época, estimava-se que o genoma humano poderia conter entre 20 e 100 mil genes. Ao final, o número de genes codificadores de proteínas no genoma humano foi estimado em aproximadamente 20 mil.

Desenvolva um texto com o mínimo de 50 linhas e o máximo de 150, respondendo às perguntas abaixo:

1. Como um genoma codificando tão poucos genes é capaz de produzir um organismo complexo?
2. Sabe-se que o corpo humano contém aproximadamente 70.000 proteínas distintas. Por que o número de genes é tão menor do que o número de proteínas?

RASCUNHO

RASCUNHO

RASCUNHO

RASCUNHO

RASCUNHO

INSTRUÇÕES

1. Por motivo de segurança, a Fiocruz solicita que o candidato transcreva em letra cursiva, em espaço próprio no Cartão de Respostas da Prova Objetiva, a frase abaixo apresentada:

“As melhores coisas da vida não podem ser vistas nem tocadas, mas sim sentidas pelo coração.” (Dalai Lama)

2. Para cada uma das questões da prova objetiva são apresentadas 5 (cinco) alternativas classificadas com as letras (A), (B), (C), (D) e (E), e só uma responde da melhor forma possível ao quesito proposto. Você só deve assinalar UMA RESPOSTA. A marcação de nenhuma ou de mais de uma alternativa anula a questão, MESMO QUE UMA DAS RESPOSTAS SEJA A CORRETA.

3. A duração da prova é de 4 (quatro) horas, considerando, inclusive, a marcação do Cartão de Respostas e a Prova Discursiva. Faça-a com tranquilidade, mas controle o seu tempo.

4. Verifique se a prova é para o **PERFIL** para o qual concorre.

5. Somente após autorizado o início da prova, verifique se este Caderno de Questões está completo e em ordem. Folhear o Caderno de Questões antes do início da prova implica na eliminação do candidato.

6. Verifique, no **Cartão de Respostas da Prova Objetiva**, se seu nome, número de inscrição, identidade e data de nascimento estão corretos. Caso contrário, comunique ao fiscal de sala.

7. O **Caderno de Questões** poderá ser utilizado para anotações, mas somente as respostas assinaladas no **Cartão de Respostas da Prova Objetiva** e no **Caderno de Respostas da Prova Discursiva** serão objeto de correção.

8. Observe as seguintes recomendações relativas ao **Cartão de Respostas da Prova Objetiva**:

. não haverá substituição por erro do candidato;

. não deixar de assinar no campo próprio;

. não pode ser dobrado, amassado, rasurado, manchado ou conter qualquer registro fora dos locais destinados às respostas;

. a maneira correta de marcação das respostas é cobrir, fortemente, com esferográfica de tinta azul ou preta, o espaço correspondente à letra a ser assinalada;

. outras formas de marcação diferentes da que foi determinada acima implicarão a rejeição do **Cartão de Respostas**;

9. O fiscal não está autorizado a alterar quaisquer dessas instruções.

10. Você só poderá retirar-se da sala após 60 minutos do início da prova.

11. Quaisquer anotações só serão permitidas se feitas no caderno de questões.

12. Você poderá anotar suas respostas da prova objetiva em área específica do Caderno de Questões, destacá-la e levar consigo.

13. Os três últimos candidatos deverão permanecer na sala até que o último candidato entregue ao fiscal todo o seu material de prova.

14. Ao terminar a prova, entregue ao fiscal de sala, obrigatoriamente, o **Cartão de Respostas da Prova Objetiva**, o **Caderno de Respostas da Prova Discursiva** e o **Caderno de Questões**.

15. Prova Discursiva:

- A questão discursiva deverá ter um limite mínimo de 50 linhas e máximo de 150 linhas.

- Transcreva sua resposta para a parte pautada do **Caderno de Respostas da Prova Discursiva**. Não assine, rubrique ou coloque qualquer marca que o identifique, sob pena de ser anulado. Assim, a detecção de qualquer marca identificadora no espaço destinado à transcrição do texto definitivo acarretará nota ZERO na respectiva prova discursiva.

- O tempo total de duração das provas será de 4 (quatro) horas, incluindo o tempo para o preenchimento da Resposta Definitiva da Questão Discursiva. Nenhum rascunho SERÁ LEVADO EM CONTA.

Boa Prova!



Ao término da prova, anote aqui suas respostas e destaque na linha pontilhada.

01	<input type="checkbox"/>	09	<input type="checkbox"/>	17	<input type="checkbox"/>	25	<input type="checkbox"/>	33	<input type="checkbox"/>
02	<input type="checkbox"/>	10	<input type="checkbox"/>	18	<input type="checkbox"/>	26	<input type="checkbox"/>	34	<input type="checkbox"/>
03	<input type="checkbox"/>	11	<input type="checkbox"/>	19	<input type="checkbox"/>	27	<input type="checkbox"/>	35	<input type="checkbox"/>
04	<input type="checkbox"/>	12	<input type="checkbox"/>	20	<input type="checkbox"/>	28	<input type="checkbox"/>	36	<input type="checkbox"/>
05	<input type="checkbox"/>	13	<input type="checkbox"/>	21	<input type="checkbox"/>	29	<input type="checkbox"/>	37	<input type="checkbox"/>
06	<input type="checkbox"/>	14	<input type="checkbox"/>	22	<input type="checkbox"/>	30	<input type="checkbox"/>	38	<input type="checkbox"/>
07	<input type="checkbox"/>	15	<input type="checkbox"/>	23	<input type="checkbox"/>	31	<input type="checkbox"/>	39	<input type="checkbox"/>
08	<input type="checkbox"/>	16	<input type="checkbox"/>	24	<input type="checkbox"/>	32	<input type="checkbox"/>	40	<input type="checkbox"/>