



UNIVERSIDAD  
DE MÁLAGA

**PROCEDIMIENTO SELECTIVO  
CONFIGURACIÓN DE BOLSA DE TRABAJO**

**TÉCNICO ESPECIALISTA DE  
LABORATORIO  
CENTRO DE BIOINNOVACIÓN  
Y SUPERCOMPUTACIÓN  
“BIOINFORMÁTICA”  
(BTL3BSBIO)**

(Res. 07 de octubre de 2021)

**Prueba específica**

10 de enero de 2022

# SCBI-BIOINFORMÁTICA

## Cod: BTL3BSBIO

1.- Margaret Oakley Dayhoff fue pionera en el desarrollo de algoritmos para la comparación de secuencias, desarrollando el primer Atlas de Secuencias de Proteínas. Cuántas secuencias contenía en su publicación:

- a) 10
- b) 65
- c) 650
- d) 1000

2.- En qué año se publicó la primera versión útil del genoma humano:

- a) 1995
- b) 2001
- c) 2003
- d) 2021

3.- Qué opción utilizarías para abrir una consola en la que ejecutar comandos en el supercomputador Picasso:

- a) `ssh -p 22 picasso.scbi.uma.es -l usuario`
- b) `http usuario@picasso.scbi.uma.es`
- c) `ftp -p 22 usuario@picasso.scbi.uma.es`
- d) `ssh://usuario@picasso.scbi.uma.es`

4.- Qué sistema de colas se utiliza para enviar trabajos en el supercomputador Picasso:

- a) Bash
- b) PBS queue system
- c) Slurm
- d) SQS queue system

5.- Cómo se carga el entorno para poder ejecutar un programa en el supercomputador Picasso:

- a) `module load PROGRAMA/VERSION`
- b) `module exec PROGRAMA/VERSION`
- c) `module open PROGRAMA/VERSION`
- d) `module environment PROGRAMA/VERSION`

6.- Para calcular la media aritmética de una matriz de expresión con R aplicarías:

- a) `tapply(M1, c(1, 2), sqrt)`
- b) `apply(M1, 2, mean)`
- c) `apply(M1, 1, mean)`
- d) `sapply(M1, c(1, 2), mean)`

7.- Qué elementos agrupa el objeto de R *ExpressionSets*:

- a) Sólo el AssayData
- b) El AssayData, MetaData y ExperimentDat
- c) El AssayData y el ExperimentDat
- d) Ninguno de los anteriores

- 8.- Qué librería de R utilizarías para normalizar un experimento de RNA-seq, sabiendo que puedes combinarlo con cualquier otra librería para la expresión diferencial:
- NOIseq porque no necesita normalizar
  - RPKM o FPKM
  - edgeR porque contiene el método TMM
  - La normalización depende del método de expresión diferencial que se quiera aplicar
- 9.- Dónde se almacenan los datos brutos de las carreras secuenciadas en el NextSeq 550:
- Todos los datos brutos se almacenan en la partición para datos del secuenciador.
  - En el secuenciador se queda una copia temporal de la última carrera y todas las carreras se copian en el repositorio de respaldo
  - Se mantiene una copia de todos los datos en el secuenciador, y otra en el repositorio de respaldo.
  - Todos los datos brutos se almacenan en el supercomputador Picasso y no en el secuenciador.
- 10.- Dónde hay que ubicar los datos del NextSeq 550 de la UMA cuando se quieren hacer los análisis en Picasso:
- En el disco duro de mi ordenador
  - En el disco SCRATCH para que nunca se borre
  - En mi \$HOME para que se hagan copias de seguridad
  - Si no tengo un disco NAS no puedo trabajar con ellos
- 11.- Cómo se define el valor de calidad de secuenciación según Phred:
- $Q = -20\log_{10}P$  (donde P es la probabilidad de error)
  - $Q = -10\log_{10}P$  (donde P es la probabilidad de error)
  - $Q = 10\log_{10}P$  (donde P es la probabilidad de error)
  - Phred no devuelve un índice de calidad
- 12.- En qué parte de las lecturas se suelen acumular los errores de secuenciación con mayor frecuencia:
- En el extremo 5'
  - En el extremo 3'
  - En ambos extremos
  - En ninguno, los errores se distribuyen de forma aleatoria a lo largo de toda la lectura.
- 13.- Cuando se realiza un ensamblaje con el algoritmo de de Bruijn, en cuántos k-meros se descomponen cada lectura, donde k es la longitud del *kmer* y L la longitud de la lectura:
- $k + 1$
  - $L + k - 1$
  - $L - k + 1$
  - $L - k$
- 14.- Qué algoritmo usan los mapeadores Bowtie y BWA:
- La transformada de Burrows-Wheeler
  - EL RSA (Rivest, Shamir y Adleman)
  - Un árbol de decisión binario
  - Ninguna de las anteriores

- 15.- Qué mide el índice MAPQ de los ficheros SAM:
- a) La calidad de las variaciones mutaciones
  - b) La calidad del alineamiento de lecturas a una referencia dada
  - c) La calidad de los nucleótidos
  - d) La calidad de los heatmap generados
- 16.- Con respecto a los valores RPKM o FPKM:
- a) Son datos de expresión normalizados
  - b) Son datos de expresión normalizados, pero solo uno de ellos tiene en cuenta la longitud del gen
  - c) Son datos de expresión normalizados, pero uno de ellos da prioridad a fragmentos del gen.
  - d) Devuelve la distancia entre los genes mapeados
- 17.- Cuando necesitas mapear lecturas de DNA-seq procedentes de un eucariota sobre su genoma, debes utilizar:
- a) STAR
  - b) Bowtie2
  - c) HISAT
  - d) Cualquiera de las anteriores
- 18.- Si quieres eliminar las lecturas o secuencias redundantes obtenidos en una secuenciación, utilizarás:
- a) viquas
  - b) casava
  - c) hydra
  - d) cdhit
- 19.- Qué es el formato MAF:
- a) Un formato de multialineamiento de secuencias
  - b) Un formato de medias normalizadas de expresión
  - c) Un formato de anotación de mutaciones
  - d) La primera y la tercera son correctas
- 20.- En un proceso de identificación de variantes sobre genoma, puede existir un sesgo de acumulación de variaciones en una hebra:
- a) No, porque siempre se secuencia la misma hebra
  - b) No, porque los errores se cometen en ambas cadenas por igual
  - c) Si, ocurre con la tecnología de Illumina
  - d) Si usamos una calidad mínima de 30 el problema desaparece
- 21.- En la identificación de variantes somáticas, Qué herramienta aplica un test de Fisher para determinar la probabilidad de variación:
- a) ClinVar
  - b) VarScan y VarDirect
  - c) SomaticSpimer
  - d) COSMIC

- 22.- Cuál es el número mínimo de réplicas sugerido en un análisis de expresión diferencial utilizando RNA-seq:
- Al menos 3 réplicas biológicas
  - Al menos 3 réplicas técnicas
  - No existe número mínimo de réplicas
  - Todas las réplicas que se puedan hacer con el dinero disponible
- 23.- Qué método de normalización es el más recomendado para datos de RNA-Seq:
- Quantile
  - TMM
  - RPKM
  - Median
- 24.- Qué parámetros estadísticos utilizarías para medir el nivel de significación estadística en un análisis de expresión diferencial:
- logFC y p-value
  - p-value y FDR
  - FDR y logFC
  - Basta con utilizar uno de ellos, logFC, p-value o FDR
- 25.- Los datos de RNA-Seq se ajustan a una distribución:
- de Poisson
  - Lineal
  - Binomial negativa
  - Gaussiana
- 26.- Para analizar las interacciones entre las proteínas o genes la herramienta más versátil es:
- GenBank y ENA
  - STRINGS
  - DeepView
  - PANTHER
- 27.- Qué base de datos consultarías para encontrar información sobre piRNA en humano:
- RNA Central
  - piRBase
  - DASHR
  - Todas las opciones son correctas
- 28.- Qué contiene la base de datos PDB:
- Motivos proteicos
  - Estructuras tridimensionales de proteínas y ácidos nucleicos
  - Información sobre disruptores proteicos
  - Interacciones entre proteínas
- 29.- BiNGO y AGriGO son herramientas para calcular:
- Enriquecimiento de términos GO en una lista de genes
  - Aparición de vías KEGG en una lista de genes
  - Creación de redes desde una lista de genes
  - Esas herramientas bioinformáticas no existen

- 30.- En el contexto de la metagenómica, qué es un OTU:
- Outlander Taxonomic Unit
  - Operational Taxonomic Unit
  - Overfitted Taxonomic Union
  - Overall Taxonomic Utility
- 31.- Qué es un archivo biom, muy usado en metagenómica:
- Un archivo de biomoléculas
  - Un matriz de observaciones biológicas
  - Un archivo estructuras moleculares
  - Ese formato no es estándar
- 32.- Qué marcador utilizarías para hacer una asignación taxonómica de hongos:
- Ribosomal 16S
  - Ribosomal 18S
  - Las zonas ITS
  - Los kits de barcoding de plantas
- 33.- Qué software utilizarías para realizar un análisis de predicción funcional a partir de amplicones 16S con asignación taxonómica:
- Picrust
  - Mothur
  - Quiime
  - A partir del amplicon 16S no se pueden realizar predicciones funcionales
- 34.- Qué tipo de librerías de secuenciación pair-end sería suficiente utilizar para un análisis de la región V4-V5 hipervariable del RNA ribosómico:
- Librerías de 2x74 pb
  - Librerías de 2x150 pb
  - Librerías de 2x300 pb
  - Librerías 2x500 pb
- 35.- Para qué se utiliza el índice de Chao1:
- Para calcular la dispersión entre muestras
  - Para calcular una media de abundancia
  - Para calcular la diversidad  $\alpha$
  - Para calcular la diversidad  $\beta$
- 36.- Cuáles de los siguientes requisitos hay que cumplir para solicitar servicios en el portal de peticiones del Centro de Supercomputación y Bioinnovación:
- Es obligatorio estar registrado como usuario validado en el centro
  - Se deben aportar el DNI del usuario que realiza la petición y del investigador principal que correrá con los gastos
  - Hay que especificar los posibles riesgos de manipulación para el servicio solicitado
  - Todas las opciones son correctas
- 37.- En qué áreas se divide el Centro de Supercomputación y Bioinnovación:
- Supercomputación, Ómicas y Nanotecnología
  - Supercomputación, Bioinformática y Nanotecnología
  - Supercomputación, Bioinformática, Ómicas y Nanotecnología.
  - Supercomputación, Bioinformática, Genómica y Nanotecnología

- 38.- Qué nivel de profundidad mínimo utilizarías para identificar variaciones somáticas de alta frecuencia en tejido tumoral:
- Una profundidad de 4X
  - Una profundidad de 60X
  - Una profundidad de 500X
  - Esto sólo se puede hacer por secuenciación de Sanger
- 39.- Un usuario quiere realizar un análisis de supervivencia en humanos para comprobar la fiabilidad pronóstico de un posible biomarcador, qué método le recomendarías:
- Un análisis supervisado
  - Un análisis de clasificación jerárquica con MANOVA
  - Un análisis Kaplan-Meier
  - Un ANOVA
- 40.- Un usuario necesita determinar el perfil taxonómico bacteriano de una muestra de suelo. Como el número de condiciones a analizar según su diseño es muy elevado, qué técnica le recomendarías para que pudiese analizar el mayor número de muestras posibles:
- Metataxonómica
  - Metagenómica
  - Metatranscriptómica
  - No es posible realizar un perfil taxonómico de muestras de suelo
- 41.- Según la Ley de Prevención de Riesgos Laborales, en las reuniones del comité de seguridad y salud los delegados sindicales:
- Tendrán voz y voto en los acuerdos que se tomen
  - Podrán participar con voz, pero sin voto
  - Sólo podrán participar como oyentes
  - Asistirán como órgano consultivo para las decisiones que les afecten
- 42.- El comité de seguridad y salud se reunirá:
- Trimestralmente y siempre que lo solicite una de las representaciones en el mismo
  - Al menos 2 veces al año
  - Al menos 1 vez al año
  - Obligatoriamente a instancias de la inspección de trabajo y seguridad social
- 43.- En una empresa con 3.500 trabajadores, cuántos representantes de la empresa habrá en el comité de seguridad y salud:
- Sólo 1, el propio empresario
  - 5
  - 6
  - 7
- 44.- Según el IV Convenio Colectivo del Personal Laboral de las Universidades Públicas de Andalucía, durante los meses de julio y agosto y en los días no lectivos que coincidan con los permisos de Navidad, Semana Santa, Feria o festividades equivalentes, el personal disfrutará de una reducción horaria de:
- Media hora
  - Hora y media
  - Dos horas
  - Dos horas y media

45.- Según el IV Convenio Colectivo del Personal Laboral de las Universidades Públicas de Andalucía, los procedimientos para la provisión de vacantes de personal laboral fijo de las respectivas Universidades se realizarán bajo los principios de publicidad, igualdad, mérito y capacidad, y conforme a la siguiente prelación:

- a) 1º proceso selectivo de promoción interna. 2º concurso de traslado. 3º proceso selectivo de nuevo ingreso.
- b) 1º concurso de traslado. 2º proceso selectivo de promoción interna. 3º proceso selectivo de nuevo ingreso.
- c) 1º proceso selectivo de nuevo ingreso. 2º proceso selectivo de promoción interna. 3º concurso de traslado.
- d) 1º proceso selectivo de promoción interna. 2º proceso selectivo de nuevo ingreso. 3º concurso de traslado.

### **Preguntas de Reserva**

46.- Cómo se envían trabajos con paralelismo compulsivo basados en índices al sistema de colas del supercomputador Picasso:

- a) Utilizando el comando `send_array` para enviar los trabajos
- b) Indicando la opción `—array`
- c) Mediante un bucle `‘for’`
- d) Ninguna de las anteriores

47.- Cuando se hace un análisis con un script de R es conveniente generar un informe para el usuario, para ello lo mejor es:

- a) Utilizar las librerías que permiten generar informes HTML o PDF con markdown
- b) Tener un modelo en MS-Word en el que incluiremos los resultados más relevantes
- c) Darle al usuario no solo los ficheros con los resultados, sino también el código de R
- d) No se puede automatizar la generación automática de un informe

48.- Según los Estatutos de la Universidad de Málaga, el órgano encargado de garantizar la transparencia, la objetividad y el principio de igualdad en los procesos electorales que se lleven a cabo en la Universidad es:

- a) El Comité de Empresa
- b) El Claustro
- c) El Consejo de Gobierno
- d) La Junta Electoral General