**OKANDEJI JASMINE OGHENEMARO**

**16/MHS01/175**

**BCH 416 MID SEMESTER TEST**

QUESTIONS

1(a) Describe BLAST variants for different searches

 (b) State types of sequences used in bioinformatics

 (c) Highlight major steps in DNA analysis

2(a) Write briefly on the following:

 i Molecular bioinformatics

 ii Functional genomics

 iii Structural genomics

 iv Comparative genomics

 v Medical informatics

 (b) Itemize different databases and features used for bioinformatics

ANSWERS

1. Basic Logical Alignment Search Tool (BLAST) variants for different searches

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S/N | Programme | Query | ComparisonDatabase | Common use  |
| 1. | BLASTp | Protein | Protein  | Seeks to align an amino acid query |
| 2. | BLASTn | Nucleotide | Nucleotide  | Align new DNA sequence to a nucleotide sequence database |
| 3. | BLASTx | Nucleotide (translated) | Protein  | Analyse new DNA sequence (translated) to find out potential coding regions. |
| 4. | BLASTn | Protein | Nucleotide (translated) | Useful for expresses sequence tags (EST) analysis |
| 5. | BLASTx | Nucleotide(translated) | Nucleotide (translated) | Useful for EST analysis |

(b) Types of sequences used in Bioinformatics

* **Genomic DNA:** The genomic DNA act as the reservoir of genetic information of all organisms. Genomic DNA in recent time sequenced in many molecular biology laboratories
* **cDNA:** The double stranded molecules prepared by using mRNA as template and reverse transcriptase are called cDNA. These are expressed genes of the genomic DNA. By using cDNA molecules, substantial number of sequences have been determined and deposited in database.
* **Organellar DNA:** Eukaryotic cells consist of different types of organelles such as
* Chloroplast
* Mitochondria
* Golgi complex
* Nucleus

In eukaryotes,

* Genomic DNA is found in nucleus
* Organellar DNA molecules are located in the mitochondria and the chloroplast

The DNA molecules of these organelles are usually circular and double stranded of varying sizes

* **Expressed Sequence Tags (ESTs):** The cDNA is cloned into a vector and each one has ESTs associated with it, thus ESTs representing only fragment of genes. The average length of sequence is about 400 bases. Many sequence centers have automated the EST production where ESTs are produced rapidly. The EST data are used to find out expression pattern using the following formula:

$$Expression pattern =\frac{No of ESTs corresponding to each gene}{Total No of ESTs}$$

* **Gene Sequencing Tags (GSTs):** It has been found that the enzyme Mungbean nuclease (Mnase) cleaves between the genes of *Plasmodium falciparum.* Therefore, by digesting *Plasmodium falciparum* genome, a genomic library was established. It helps in identifying the genes of *Plasmodium falciparum.*

The approach for construction of GSTs is similar to ESTs. It is constructed by isolating one read of sequence from any of the 5’or 3’. The sequences obtained through this approach are called GSTs

* **Other biomolecules:** The databases also consist of sequences of tRNA and small sized RNAs. For example, 16S rRNA sequencing is done in tracing phylogenetic relationship among the species. A similar approach can also be made by using the other molecules. Like mRNAs, rRNA can be copied into DNA.

(c) Major steps in DNA analysis

* DNA isolation and purification
* PCR analysis
* Gel electrophoresis of PCR products
* Gel staining and photography
* Scoring of DNA fragments
* Data analysis using Numerical Taxonomy Statistical System (NTSYS-pc)
* Results and interpretations
1. **Molecular Bioinformatics:** This involves the use of computational tools to discover new information in complex data sets (from the one-dimensional information of DNA through the two-dimensional information of RNA and the three dimensional information of proteins, to the four-dimensional information of evolving living systems).

**Functional genomics:** Identification of genes and their respective functions

**Structural genomics:** Predictions related to functions of proteins

**Comparative genomics:** For understanding the genomes of different species of organisms

**Medical informatics:** This involves the management of biomedical data with special reference to biomolecules, *in vitro* assays and clinical trials.

(b) Databases and features for Bioinformatics

* Biological database
* Nucleotide sequence databases
* Protein sequence databases
* Molecular structure of databases
* Other databases