COMPUTATIONAL THINKING TEST USING BIOLOGICAL DATABASE

Name	:	Major	:
Gender	:	Semester	:
Age	:	University	:

1. A researcher wants to know the NCBI database of BRCA genes for his research data. He then opens the https://www.ncbi.nlm.nih.gov/ page by typing "BRCA gene" and clicking search. Then various gene names and their IDs appeared. The researcher selected the BRCA2 gene with ID 675 as shown in the following image (https://www.ncbi.nlm.nih.gov/gene/675).

There are several statements related to the information in the image above:

1) BRCA2 mutation causes loss of wild-type homozygosity
2) BRCA1 and BRCA2 are at increased risk of breast cancer
3) The BRCA2 gene may inhibit cell over-proliferation
4) Exon 11, a mutated region that often occurs in breast cancer
5) BRC motif increases the negative effects of gene mutations

What information can be deciphered from the above information to help researchers understand mutations in BRCA2 ID 675 genes?

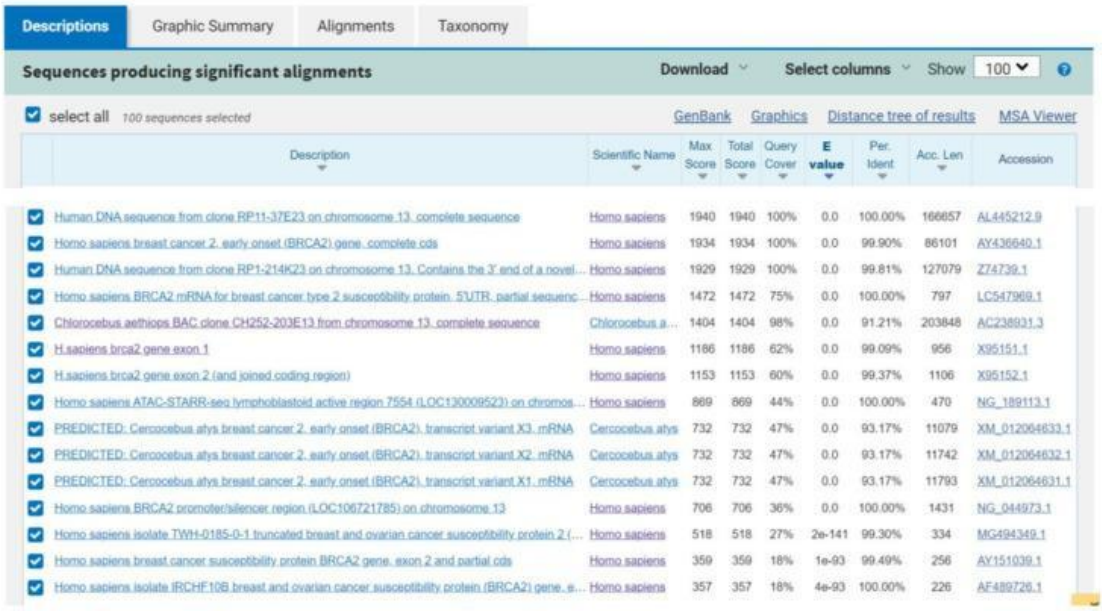
a. 1, 2, and 3
b. 3, 4, and 5
c. 2, 3, and 4
d. 2, 4, and 5

2. Still on the same page as the image above (question number 1), the researcher scrolled down the page and found the FASTA menu to find the nucleotide arrangement that makes

up the BRCA2 ID 675 gene. The researcher then copied a number of lines of nucleotide arrangement as shown below.

```
>NC_000013.11:sample
AGAGGCGGAGCCGCTGTGGCACTGCTGCGCCTCTGCTGCGCCTCGGGTGTCTTTTGCGGCGGTGGGTCGC
CGCCGGGAGAAGCGTGAGGGGACAGATTTGTGACCGGCGCGGTTTTTGTCAGCTTACTCCGGCCAAAAAA
GAACTGCACCTCTGGAGCGGGTTAGTGGTGGTGGTAGTGGGTTGGGACGAGCGCGTCTTCCGCAGTCCCA
GTCCAGCGTGGCGGGGGAGCGCCTCACGCCCCGGGTCGCTGCCGCGGCTTCTTGCCCTTTTGTCTCTGCC
AACCCCCACCCATGCCTGAGAGAAAGGTCCTTGCCCGAAGGCAGATTTTCGCCAAGCAAATTCGAGCCCC
GCCCCTTCCCTGGGTCTCCATTTCCCGCCTCCGGCCCGGCCTTTGGGCTCCGCCTTCAGCTCAAGACTTA
ACTTCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTTCTTGGAAACACGATCACTTTAACGGAATA
TTGCTGTTTTGGGGAAGTGTTTTACAGCTGCTGGGCACGCTGTATTTGCCTTACTTAAGCCCCTGGTAAT
TGCTGTATTCCGAAGACATGCTGATGGGAATTACCAGGCGGCGTTGGTCTCTAACTGGAGCCCTCTGTCC
CCACTAGCCACGCGTCACTGGTTAGCGTGATTGAAACTAAATCGTATGAAAATCCTCTTCTCTAGTCGCA
CTAGCCACGTTTCGAGTGCTTAATGTGGCTAGTGGCACCGGTTTGGACAGCACAGCTGTAAAATGTTCCC
ATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGACTGAATTAGAATTCAAACAAATTTTCCAGCGC
TTCTGAGTTTTACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTTTGGTCTTCTGTTTT
GCAGACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAAAAATGCCTATTGGATCCAAAGAGAGGCCA
ACATTTTTTGAAATTTTTAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTTATATAACTTTATAAA
```

A copy of the nucleotide array above is put into the *Nucleotide Blast* box to determine the kinship of the BRCA2 ID 675 gene with other genes. The result appears as shown in the image below.



Descriptions | Graphic Summary | Alignments | Taxonomy

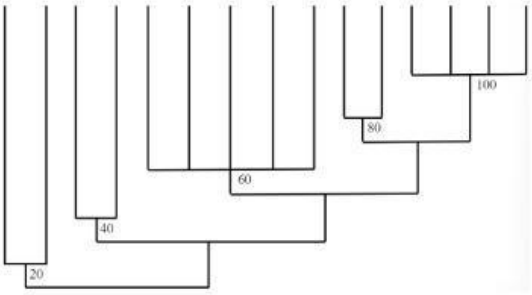
Sequences producing significant alignments — Download — Select columns — Show 100

select all 100 sequences selected — GenBank | Graphics | Distance tree of results | MSA Viewer

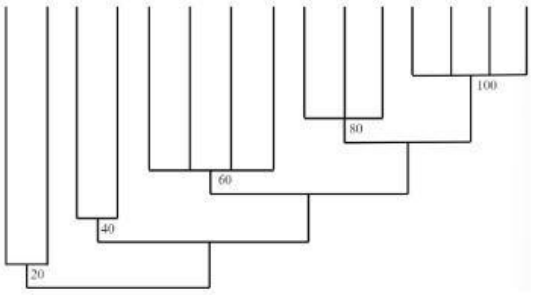
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Human DNA sequence from clone RP11-37E23 on chromosome 13, complete sequence	Homo sapiens	1940	1940	100%	0.0	100.00%	166657	AL445212.9
Homo sapiens breast cancer 2, early onset (BRCA2) gene, complete cds	Homo sapiens	1934	1934	100%	0.0	99.90%	86101	AY436640.1
Human DNA sequence from clone RP1-214K23 on chromosome 13, Contains the 3' end of a novel...	Homo sapiens	1929	1929	100%	0.0	99.81%	127079	Z74739.1
Homo sapiens BRCA2 mRNA for breast cancer type 2 susceptibility protein, 5'UTR, partial sequenc...	Homo sapiens	1472	1472	75%	0.0	100.00%	797	LC547969.1
Chlorocebus aethiops BAC clone CH252-203E13 from chromosome 13, complete sequence	Chlorocebus a...	1404	1404	98%	0.0	91.21%	203848	AC238931.3
H.sapiens brca2 gene exon 1	Homo sapiens	1186	1186	62%	0.0	99.09%	956	X95151.1
H.sapiens brca2 gene exon 2 (and joined coding region)	Homo sapiens	1153	1153	60%	0.0	99.37%	1106	X95152.1
Homo sapiens ATAC-STARR-seq lymphoblastoid active region 7554 (LOC130009523) on chromos...	Homo sapiens	869	869	44%	0.0	100.00%	470	NG_189113.1
PREDICTED: Cercocebus atys breast cancer 2, early onset (BRCA2), transcript variant X3, mRNA	Cercocebus atys	732	732	47%	0.0	93.17%	11079	XM_012064633.1
PREDICTED: Cercocebus atys breast cancer 2, early onset (BRCA2), transcript variant X2, mRNA	Cercocebus atys	732	732	47%	0.0	93.17%	11742	XM_012064632.1
PREDICTED: Cercocebus atys breast cancer 2, early onset (BRCA2), transcript variant X1, mRNA	Cercocebus atys	732	732	47%	0.0	93.17%	11793	XM_012064631.1
Homo sapiens BRCA2 promoter/silencer region (LOC106721785) on chromosome 13	Homo sapiens	706	706	36%	0.0	100.00%	1431	NG_044973.1
Homo sapiens isolate TWH-0185-0-1 truncated breast and ovarian cancer susceptibility protein 2 (...	Homo sapiens	518	518	27%	2e-141	99.30%	334	MG494349.1
Homo sapiens breast cancer susceptibility protein BRCA2 gene, exon 2 and partial cds	Homo sapiens	359	359	18%	1e-93	99.49%	256	AY151039.1
Homo sapiens isolate IRCHF10B breast and ovarian cancer susceptibility protein (BRCA2) gene, e...	Homo sapiens	357	357	18%	4e-93	100.00%	226	AF489726.1

From the results obtained, the researcher plans to construct a phylogenetic tree of several samples of search results based on the percentage of each gene on the *Query cover*. Select the image below to help researchers determine the phylogenetic tree pattern from the image above!

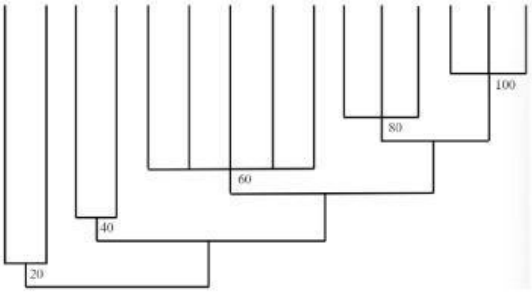
a. Picture A



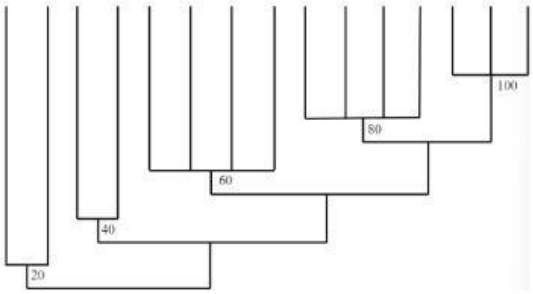
c. Picture C



b. Picture B



d. Picture D



3. From the nucleotide arrangement that the researcher copied and put into the *Nucleotide Blast* box (question number 2), data on the kinship of the BRCA2 gene ID 675 with other genes was generated. When the researcher clicked on the *Alignments* menu section, it showed a comparison of the nucleotide arrangement of the BRCA2 ID 675 gene with one of the organisms that had a similarity to it as shown in the image below.

The circled part is one of the types of mutations that can occur at the stage of DNA replication, analyze what type of mutation occurs in that part!

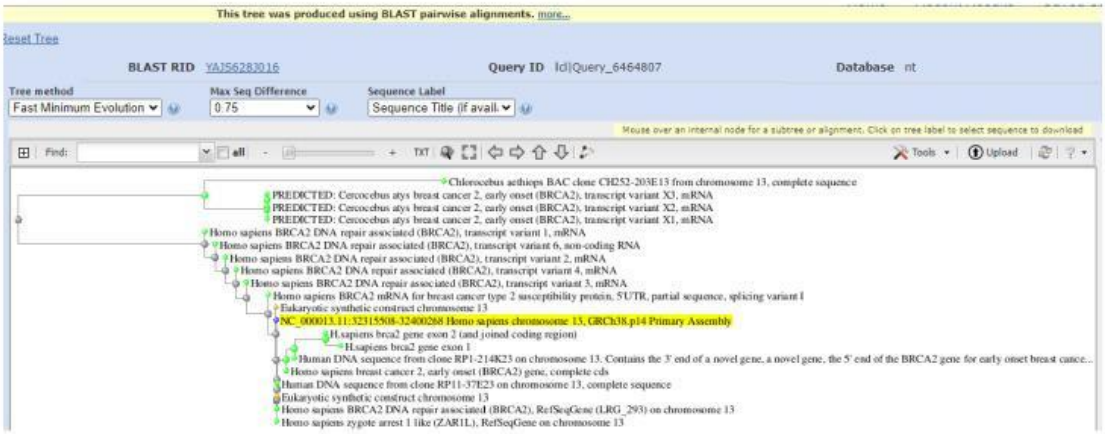
a. Transversion

b. Transition

c. Deletion

d. Insertion

4. Dimas, a researcher from BRIN, used the NCBI website to find out the kinship of the BRCA2 gene with other genes in certain organisms. On https://www.ncbi.nlm.nih.gov/gene/675 he clicks on FASTA (the array of nucleotides that make up the BRCA2 gene) and copies it. He then added the copy to the *Nucleotide Blast* menu(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). The results gave rise to several organisms that are closely related to the BRCA2 gene. Next, Dimas clicks on the *Distance Tree of Result* menu so that it displays the phylogenetic tree as shown in the image below.



Dimas understands the picture above and writes the results of his analysis below.

1) Write down the number of differences/analysis results for each character
2) Determine the character of each organism to be analyzed
3) Arrange phylogenetic trees according to kinship level
4) Determining the highest value (closest relatives) to the lowest
5) Analyze the differences in the character of each organism

Help Dimas arrange a systematic stage in arranging a phylogenetic tree as shown in the picture above!

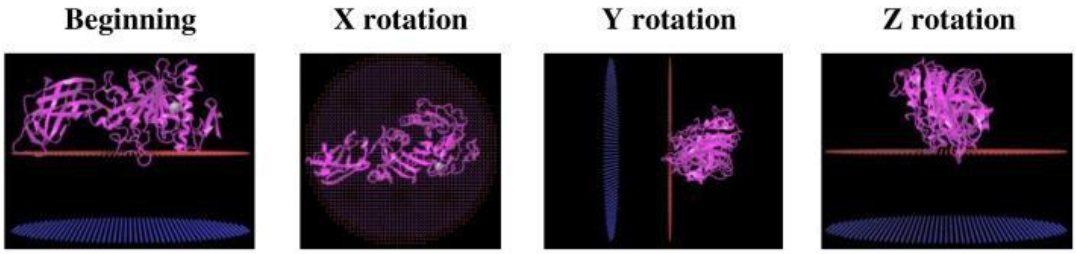
a. 1, 2, 5, 4, 3

b. 2, 5, 4, 1, 3

c. 2, 1, 5, 4, 3

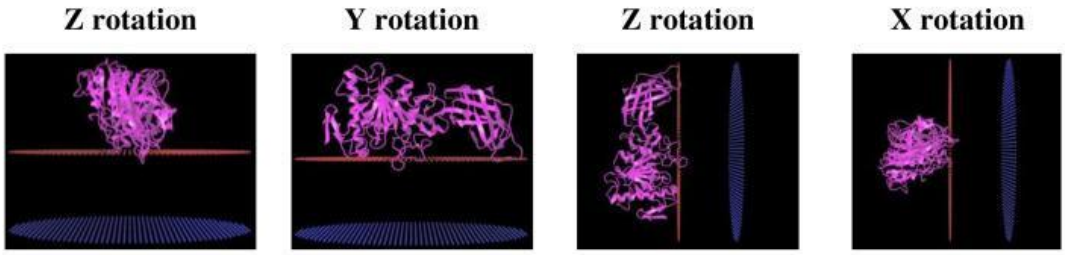
d. 2, 5, 1, 4, 3

5. The figure below shows the structure of Lipase Enzyme which can be accessed through the *search* menu on the NCBI website with the *structure* selection. Click search and many options appear, one of which is the image below (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?showanno=1&mmdbid=56257). On the menu, the *view* of the image can be changed to see other parts of the Lipase enzyme structure. If the initial *view* is changed to rotation x → rotation y → rotation z, then the change in the image can be seen in the image below.

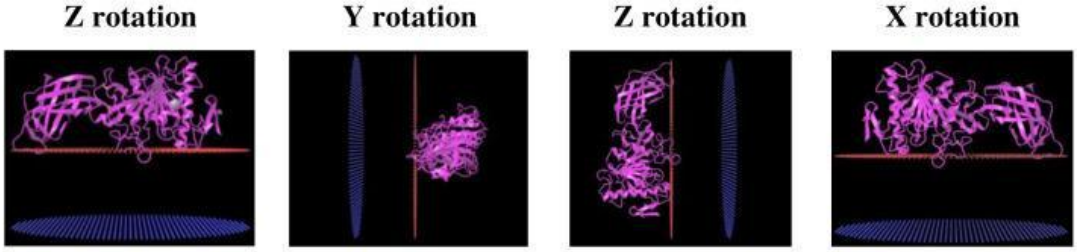


If we want to change the rotation starting from z rotation to y rotation → z rotation → x rotation, what is the form of the change!

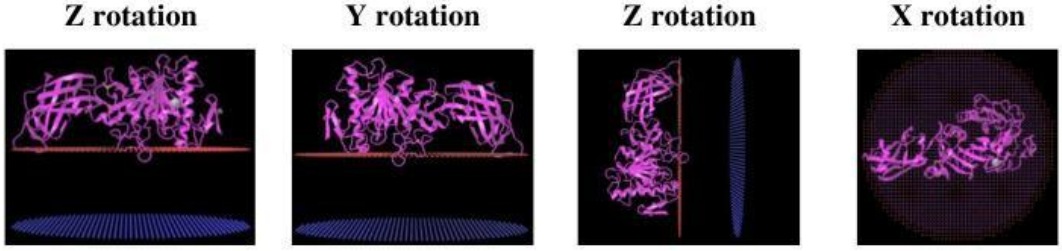
a. Picture A



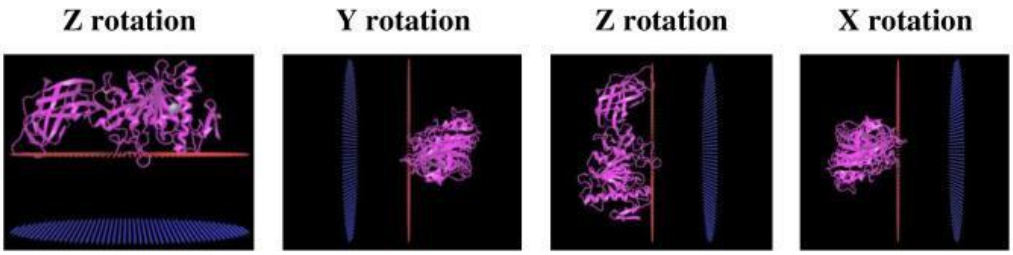
b. Picture B



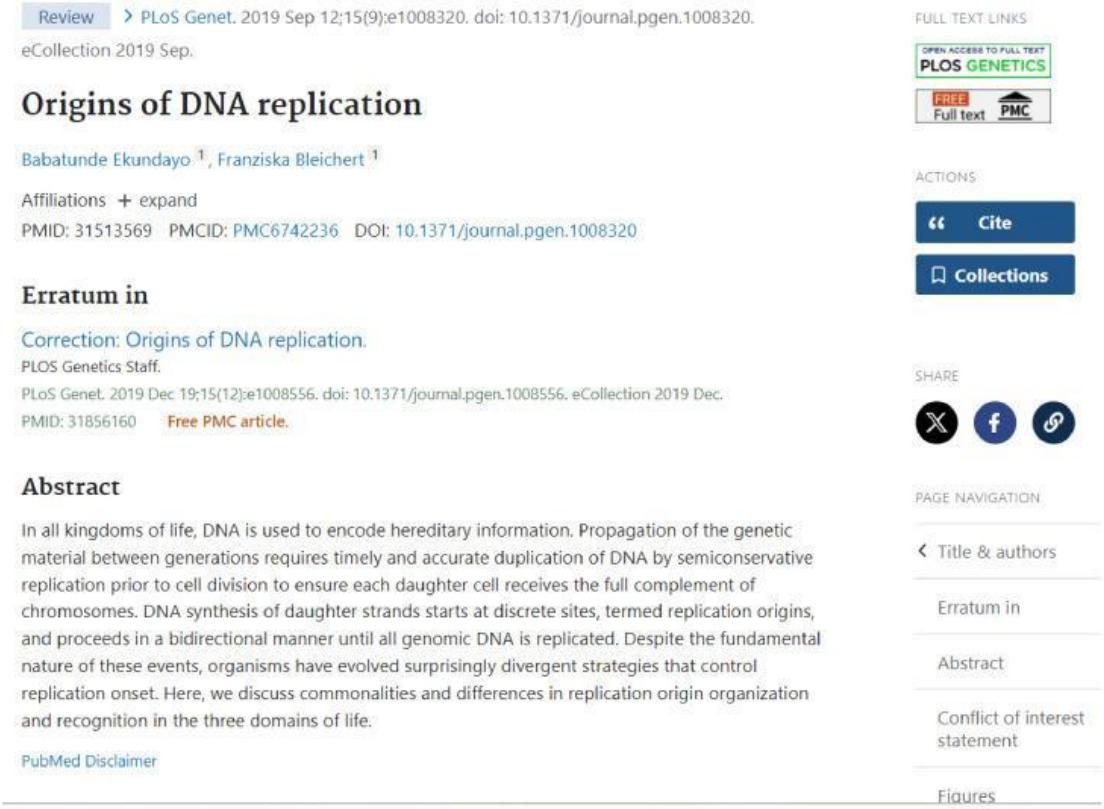
c. Picture C



d. Picture D



6. Indra is looking for literature on DNA replication. He opened the NCBI website which stores a database of DNA sequences and various genetic substance information and other health information. He then opens *the PubMed* (https://pubmed.ncbi.nlm.nih.gov/) menu, then types *DNA Replication* in the search field. The results show that many research articles discuss DNA replication. One of the materials that Indra reads is *Origins of DNA Replication* (https://pubmed.ncbi.nlm.nih.gov/31513569/). Indra read *the abstract* and found information that DNA replication was carried out in a semiconservative manner.



Review › PLoS Genet. 2019 Sep 12;15(9):e1008320. doi: 10.1371/journal.pgen.1008320. eCollection 2019 Sep.

Origins of DNA replication

Babatunde Ekundayo [1], Franziska Bleichert [1]

Affiliations + expand

PMID: 31513569 PMCID: PMC6742236 DOI: 10.1371/journal.pgen.1008320

Erratum in

Correction: Origins of DNA replication.
PLOS Genetics Staff.
PLoS Genet. 2019 Dec 19;15(12):e1008556. doi: 10.1371/journal.pgen.1008556. eCollection 2019 Dec.
PMID: 31856160 Free PMC article.

Abstract

In all kingdoms of life, DNA is used to encode hereditary information. Propagation of the genetic material between generations requires timely and accurate duplication of DNA by semiconservative replication prior to cell division to ensure each daughter cell receives the full complement of chromosomes. DNA synthesis of daughter strands starts at discrete sites, termed replication origins, and proceeds in a bidirectional manner until all genomic DNA is replicated. Despite the fundamental nature of these events, organisms have evolved surprisingly divergent strategies that control replication onset. Here, we discuss commonalities and differences in replication origin organization and recognition in the three domains of life.

PubMed Disclaimer

FULL TEXT LINKS
PLOS GENETICS
FREE Full text PMC

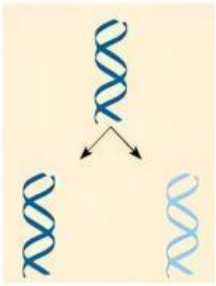
ACTIONS
Cite
Collections

SHARE

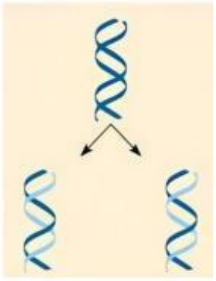
PAGE NAVIGATION
Title & authors
Erratum in
Abstract
Conflict of interest statement
Figures

Indra's search results regarding this information show that replication per DNA will produce one new strand and one old strand. The image below that exactly states the information is ....

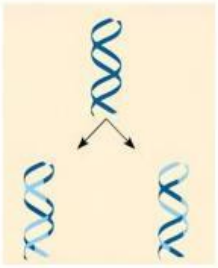
a. Picture A



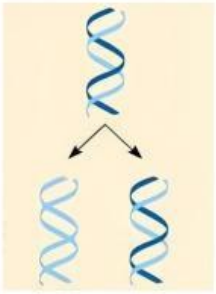
b. Picture B



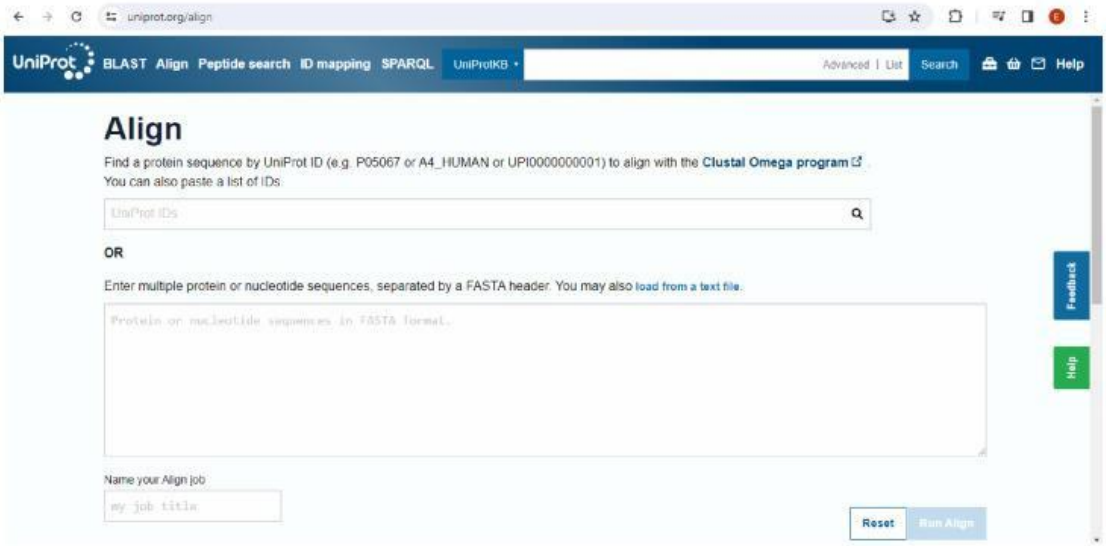
c. Picture C



d. Picture D



7. Andi has a number of DNA nucleotide arrangements. He wanted to know which parts had differences between one nucleotide and another. He then opened the UniProt website and visited the website address on *the Align* menu (https://www.uniprot.org/align) as shown in the image.

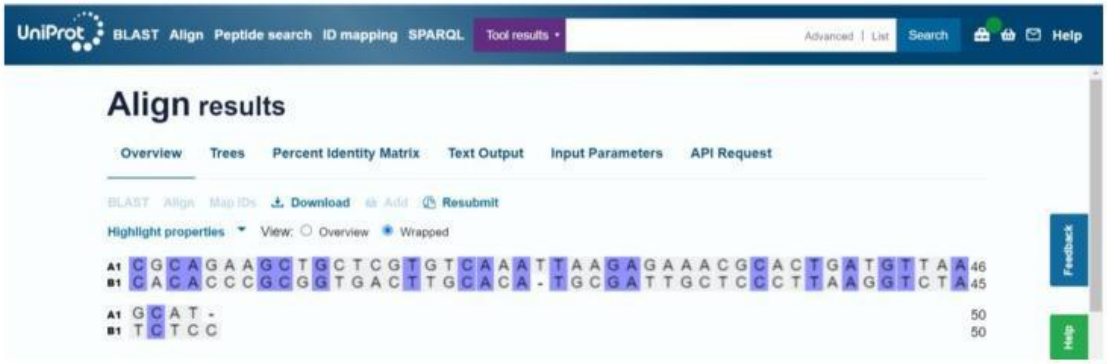


Help Andi to understand the instructions on the page so that Andi can know the difference from a number of nucleotide arrangements to be inserted. The statements below that *are not appropriate* are related to the components and what things must be considered in order to find a solution from the image above are ....

a. Inserting nucleotide arrays in the form of UniProt IDs

b. Inserting multiple nucleotide arrays more than one
c. Each nucleotide array is separated by a FASTA header
d. Nucleotide IDs do not have to be compatible with the Clustal Omega program

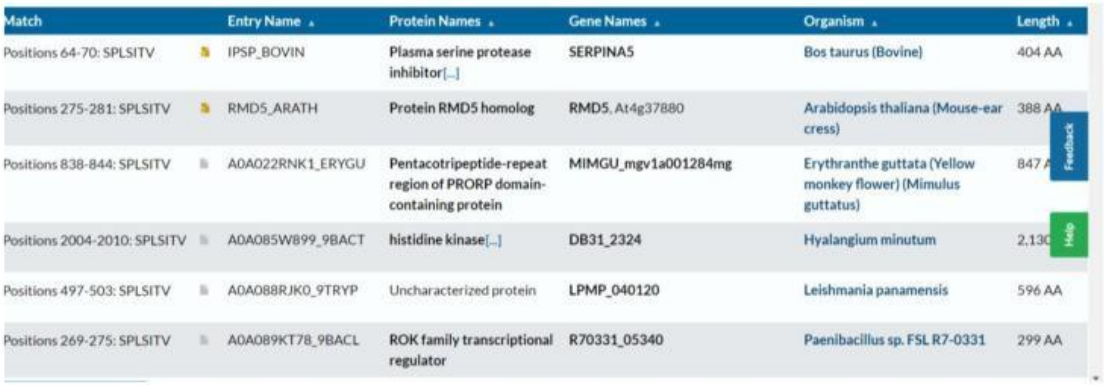
8. From question number 7, Andi tried to enter a number of nucleotide arrangements on the *Align* page according to the instructions, so he succeeded. The figure below is the *Align results* of the nucleotide arrangement that is inserted with a total of 50 nucleotide bases.



Andi tried to analyze the pattern of the website in aligning the nucleotide arrangement that was inserted. The statement below that is appropriate regarding the pattern in Align results is ....

a. Align result marks similar nitrogenous bases, not nitrogenous base pairs
b. Align result does not mark similar nitrogenous bases, but rather marks nitrogenous base pairs
c. Align result marks similar nitrogenous bases and marks nitrogenous base pairs
d. Align result does not mark similar nitrogenous bases and does not mark nitrogenous base pairs

9. *Peptide search* is part of the UniProt menu that functions to find protein names based on the array of amino acids inserted. Intan, a student of Biology Education tried to include the following seven amino acids: Serine, Phenylalanine, Leucine, Serine, Isoleucine, Tryptophan, and Valine. The results can be seen in the image below (https://www.uniprot.org/peptide-search/PM20240304e8795ad38346422b9c77cbd85352e938/overview).

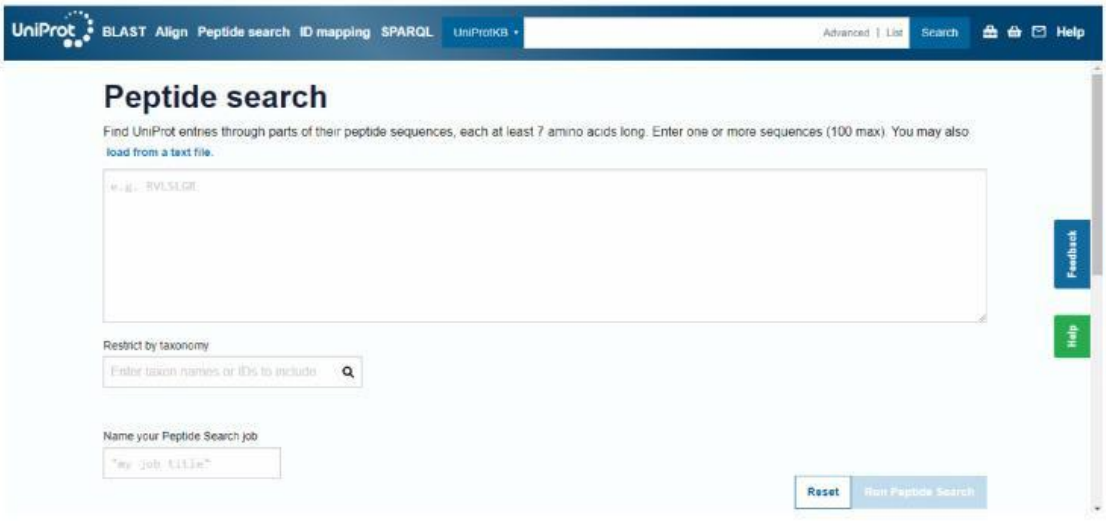


Match	Entry Name	Protein Names	Gene Names	Organism	Length
Positions 64-70: SPLSITV	IPSP_BOVIN	Plasma serine protease inhibitor[...]	SERPINA5	Bos taurus (Bovine)	404 AA
Positions 275-281: SPLSITV	RMD5_ARATH	Protein RMD5 homolog	RMD5, At4g37880	Arabidopsis thaliana (Mouse-ear cress)	388 AA
Positions 838-844: SPLSITV	A0A022RNK1_ERYGU	Pentacotripeptide-repeat region of PRORP domain-containing protein	MIMGU_mgv1a001284mg	Erythranthe guttata (Yellow monkey flower) (Mimulus guttatus)	847 A
Positions 2004-2010: SPLSITV	A0A085W899_9BACT	histidine kinase[...]	DB31_2324	Hyalangium minutum	2,130
Positions 497-503: SPLSITV	A0A088RJK0_9TRYP	Uncharacterized protein	LPMP_040120	Leishmania panamensis	596 AA
Positions 269-275: SPLSITV	A0A089KT78_9BACL	ROK family transcriptional regulator	R70331_05340	Paenibacillus sp. FSL R7-0331	299 AA

From the image above, there is one name for a protein that is *uncharacterized*. Analyze, the correct part of the statement based on the event is…

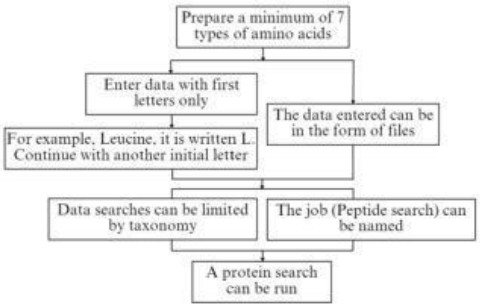
a. Proteins have not been identified in structure and function by UniProt
b. The sequence of amino acids inserted does not correspond to the protein
c. Proteins do not have a clear structure and their function is not yet known
d. Proteins are not yet clearly known for their function for other organisms

10. Just like Intan (question number 9), Nina also uses the UniProt website (https://www.uniprot.org/) to find out the name of the protein based on the type of amino acid produced from the protein synthesis process. Nina then visited the *Peptide Search menu* (https://www.uniprot.org/peptide-search), so that the page appears as shown in the image below.

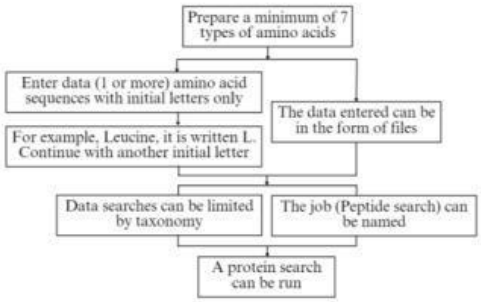


The protein synthesis stage carried out by Nina produces amino acids with the types of Serine, Phenylalanine, Leucine, Serine, Isoleucine, Tryptophan, and Valine. To find out the possible names of proteins, genes, and organisms in these amino acids, which of the following charts is appropriate related to the components and processes that must be included in the form according to the figure above!

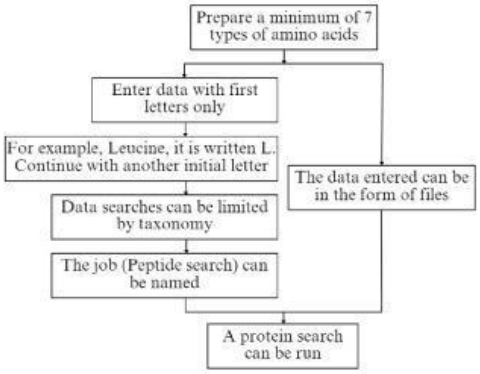
a. Chart A



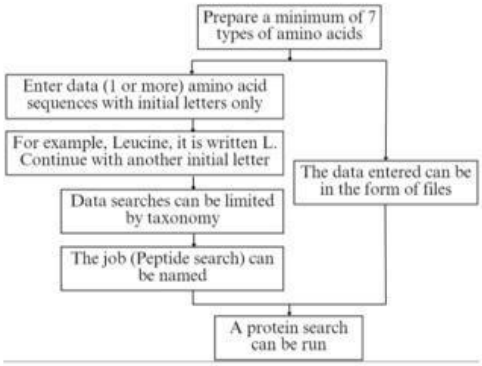
b. Chart B



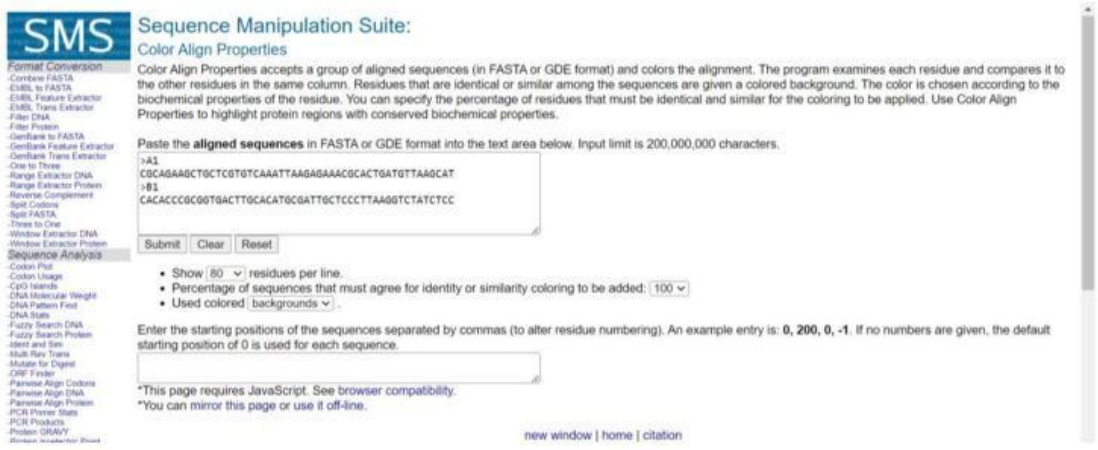
c. Chart C



d. Chart D



11. Sequence Manipulation Suite (SMS) is a collection of JavaScript programs designed to generate, format, and analyze short DNA and protein networks. One of the programs is *Color Align Properties* as shown in the image below (https://www.bioinformatics.org/sms2/color_align_cons.html). Anas, a Biology Education student, will use the website as a research medium in measuring students' computational thinking. Help Anas understand how to use the program.



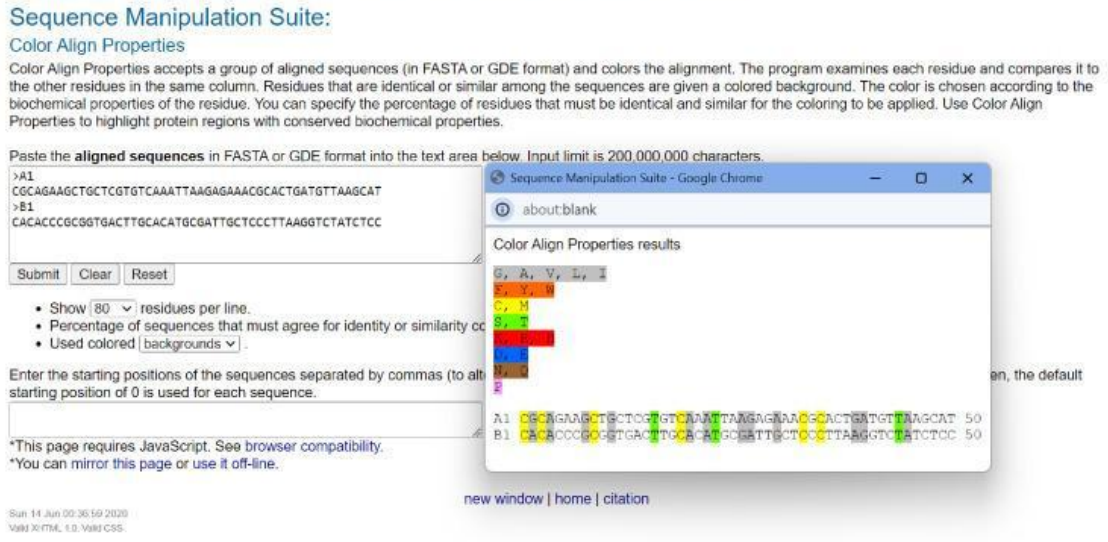
There are several statements related to the Sequence Manipulation Suite program, including:

1) Sequence of nucleotide bases
2) The nucleotide base sequence format is only deep FASTA/GDE
3) Inserted DNA sequence < 2
4) Identical nucleotide bases given a colored background
5) Color Align Properties compare residues with each other (biological properties)

From the statement outlined based on the image description, what information can be used to help Anas know the use of *Color Align Properties* in SMS?

a. 2, 3, dan 5
b. 1, 3, dan 5
c. 1, 2, dan 4
d. 2, 3, dan 4

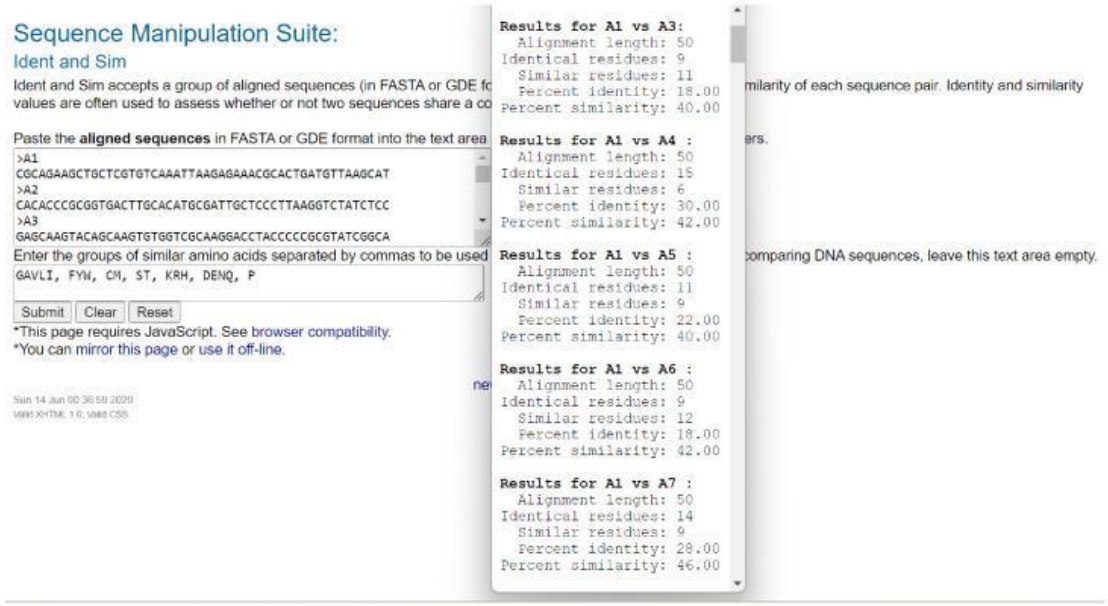
12. Thank you for helping Anas answer question number 11. Now he has successfully used *the Color Align Properties* program to see a comparison of two DNA sequence samples. The results are shown in the image below (https://www.bioinformatics.org/sms2/color_align_prop.html).



Based on the image, there are several nucleotide bases that are given and not colored. Identify the color pattern in the image, write down what nucleotide bases represent the DNA replication process!

a. A1: GGAAGATAG
b. A1: CCCTCTCCT
c. A1: GAGGAAAGAGAA
d. A1: GGAAATAGG

13. Another program from the Sequence Manipulation Suite (SMS) website is *Ident and Sim* (https://www.bioinformatics.org/sms2/ident_sim.html). Rani used the program to calculate the similarity of the DNA sequences that were aligned. The image below (right) is the result of several DNA sequences that Rani inserted into the program.



Analyze *the Ident and Sim results* above, which DNA groups have the closest relatives?

a. A1 vs A4
b. A1 vs A5
c. A1 vs A6
d. A1 vs A7

14. Irna did the same thing as Rani, namely taking advantage of the Ident and Sim (https://www.bioinformatics.org/sms2/ident_sim.html) to complete the final task. However, Irna did not understand how to use the program. Irna also asked Rani so that Rani gave several statements about the use of the following Ident and Sim program:

1) Enter DNA sequence data with a limit of 20 million characters
2) Jobs can be executed by clicking *submit*
3) Enter the amino acid group separated by a comma
4) If you want to compare amino acid groups, the area above is left blank
5) The lower area is filled with amino acid groups (formed by the initial letter only)
6) Prepare multiple DNA sequences in FASTA/GDE format

As far as Irna remembers, these six statements are the way that Rani taught. However, Irna forgot the order of the process. Help Irna arrange systematic stages and sequences so that Irna can use *the Ident and Sim* program to quickly calculate the similarity of a set of DNA/amino acid sequences!

a. 1, 3, 6, 4, 5, and 2
b. 6, 1, 4, 5, 3, and 2
c. 3, 1, 6, 4, 5, and 2
d. 6, 1, 3, 4, 5, and 2