lab5_Biology

Batuhan Seyhan S20343

2.03.2021

TASK 1

https://ibb.co/BzyWDfV

TASK 2

https://academic.oup.com/nar

TASK 3

https://academic.oup.com/nar/article/40/D1/D33/2902544?searchresult=1 According to the information provided on the GenBank website, a GenBank release occurs every two months and is available from the ftp site. The release notes for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Therefore, it can be concluded that the NCBI GenBank database is updated every two months. The last update was on December 8, 2022, according to the website.

TASK 4

Table S2 - Cancer databases MGDB http://bioinfo.ahu.edu.cn:8080/Melanoma/ Multiomics catalog of genes involved in melanoma Yes Online

MGDB: a comprehensive database of genes involved in melanoma

TASK 5

Based on the HLA-ADR database, there is evidence linking aspirin to a heightened risk of autoimmune diseases in individuals possessing specific HLA alleles. Notably, the HLA-DRB104 allele has been associated with an increased risk of autoimmune diseases when exposed to aspirin. Conversely, individuals carrying the HLA-DRB113 allele have been found to experience a decreased risk of autoimmune diseases with aspirin usage. While the database includes additional alleles that could potentially be linked to either an elevated or diminished risk of autoimmune diseases in aspirin users, further research is needed to establish clearer effects.

TASK 6

the species of microorganisms associated with the records deposited under the name "Kuzdraliński" on the NCBI GenBank website are as follows:

Fusarium poae - 155 records Rhizopus arrhizus - 96 records Rhizopus stolonifer - 54 records Blumeria graminis f. sp. tritici - 33 records Fusarium sp. - 30 records Puccinia triticina - 26 records Puccinia striiformis f. sp. tritici - 22 records Epicoccum nigrum - 18 records Fusarium sporotrichioides - 16 records Fusarium avenaceum - 14 records Lactobacillus helveticus - 9 records Fusarium culmorum - 6 records Pichia kudriavzevii - 4 records Fusarium graminearum - 3 records Aureobasidium pullulans - 3 records Rhizopus microsporus - 3 records Fusarium equiseti - 2 records Fusarium oxysporum - 2 records Alternaria sp. - 2 records Rhizoctonia solani - 2 records

TASK 7

Fasta format: >KJ195691.1 Pichia kudriavzevii isolate 4 26S ribosomal RNA gene, partial sequence GCGGCGAGTGAGCGGCAGAGCTCAGATTTGAAATCGTGCTTTGCGGCAC-GAGTTGTAGATTGCAGGTTGG AGTCTGTGTGGAAGGCGGTGTCCAAGTCCCTTG-GAACAGGGCGCCCAGGAGGGTGAGAGCCCCGTGGGAT GCCGGCGGAAGCAGTGAG-GCCCTTCTGACGAGTCGAGTTGTTTGGGAATGCAGCTCCAAGCGGGTGGTAA ATTC-CATCTAAGGCTAAATACTGGCGAGAGACCGATAGCGAACAAGTACTGTGAAGGAAA-GATGAAAAGC ACTTTGAAAAGAGAGTGAAACAGCACGTGAAATTGTTGAAAGGGAAGGGTATTC CCGACATGGGGAT TGCGCACCGCTGCCTCTCGTGGGCGGCGCTCTGGGCTTTCC-CTGGGCCAGCATCGGTTCTTGCTGCAGGA GAAGGGGTTCTGGAACGTGGCTCTTCG-GAGTGTTATAGCCAGGGCCAGATGCTGCGTGCGGGGACCGAGG ACTGCGGCCGTG-TAGGTCACGGATGCTGGCAGAACGGCGCAACACCGCCCGTCTTGAAAC

TASK 8

Accession: JX669571.1 GI: 425856826

TASK 9

Rozprawa doktorska:

Grzyby i tworzone przez nie mikotoksyny w owsie ekologicznym i jego produktach, 28/09/2011, Uniwersytet Przyrodniczy w Lublinie; Wydział Nauk o Żywności i Biotechnologii,

Uzyskany stopień doktor nauk rolniczych w zakresie technologii żywności i żywienia, specjalność: mikrobiologia żywności, Rozprawa habilitacyjna:

Identyfikacja specyficznych markerów DNA przydatnych do detekcji wybranych patogenów grzybowych pszenicy zwyczajnej i możliwości ich praktycznego wykorzystania w hodowli i uprawie tego zboża, 10/07/2019, Uniwersytet Przyrodniczy w Lublinie; Wydział Agrobioinżynierii,

Uzyskany stopień doktor habilitowany dziedzina nauk rolniczych w dyscyplinie rolnictwo i ogrodnictwo Promotor pomocniczy prac badawczych:

Doskonalenie procesu biosyntezy kwasu fumarowego przez grzyby z rodzaju Rhizopus 26/09/2019, autor rozprawy doktorskiej Elwira Komoń-Janczara , Publikacje:

2014, Early PCR-based detection of Fusarium culmorum, F. graminearum, F. sporotrichioides and F. poae on stem bases of winter wheat throughout Poland, European Journal of Plant Pathology, A. Kuzdraliński, H. Szczerba, K. Tofil, A. Filipiak, E. Garbarczyk, P. Dziadko, M. Muszyńska, E. Solarska | Artykuł

2013, Variability of S-layer proteins in Lactobacillus helveticus strains, Anaerobe, A. Waśko, M. Polak-Berecka, A. Kuzdraliński, T. Skrzypek | Artykuł

2013, Mycotoxin content of organic and conventional oats from southeastern Poland, Food Control, vol. 33, nr 1, pp. 68-72, Adam Kuzdraliński, Ewa Solarska, Jarosław Mazurkiewicz | Artykuł

2013, Deoxynivalenol and zearalenone occurence in beers analysed by an enzyme-linked immunosorbent assay method, Food Control, A. Kuzdraliński, E. Solarska, M. Muszyńska | Artykuł

TASK 10

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9198248/

Iłowiecka, K., Maślej, M., Czajka, M., Pawłowski, A., Więckowski, P., Styk, T., Gołkiewicz, M., Kuzdraliński, A., & Koch, W. (2022). Lifestyle, Eating Habits, and Health Behaviors Among Dietary Supplement Users in Three European Countries. Frontiers in public health, 10, 892233. https://doi.org/10.3389/fpubh.2022.892233

TASK 11

20,024

TASK 12

During the late 1990s and early 2000s, the Human Genome Project (HGP) was ongoing, aiming to sequence and map the entire human genome. At that time, estimates and discussions about the number of genes in the human genome varied.

Initially, it was widely believed that the human genome would contain around 100,000 to 140,000 genes. However, as the Human Genome Project progressed, researchers made new discoveries and revised their estimates.

TASK 13

The size of Homo sapiens chromosome 22, GRCh38.p14 Primary Assembly is 50,818,468 base pairs (bp)

TASK 14

The human chromosomes are arranged in a specific order, numbered from 1 to 22, and an additional pair of sex chromosomes, X and Y. Here is the order of the human chromosomes:

Chromosome 1 Chromosome 2 Chromosome 3 Chromosome 4 Chromosome 5 Chromosome 6 Chromosome 7 Chromosome 8 Chromosome 9 Chromosome 10 Chromosome 11 Chromosome 12 Chromosome 13 Chromosome 14 Chromosome 15 Chromosome 16 Chromosome 17 Chromosome 18

Chromosome 19 Chromosome 20 Chromosome 21 Chromosome 22 Sex Chromosome X Sex Chromosome Y

This is the standard numbering system for human chromosomes, based on their size and structure.

TASK 15

there are 282,666 publications available in the PubMed database that are affiliated with Poland

TASK 16

2,450

TASK 17

Here are the MAF values for the different alleles:

A allele: The MAF ranges from 0.28169 to 0.378357 across different populations or datasets. G allele: The MAF ranges from 0.310185 to 0.373416 across different populations or datasets.

TASK 18

Minor Allele Frequency (MAF) refers to the frequency at which the less common allele or variant occurs in a population. In genetics, a single nucleotide variant (SNV) can have two alleles, often denoted as the major allele (more common) and the minor allele (less common). The MAF is a measure of the occurrence of the minor allele within a population.

MAF is typically expressed as a decimal value ranging from 0 to 0.5, representing the proportion or percentage of individuals in the population who carry the minor allele. A MAF of 0.1 indicates that 10% of individuals in the population have the minor allele, while a MAF of 0.01 represents 1% frequency of the minor allele, and so on.

Knowing the MAF is important in genetic studies as it helps determine the prevalence and significance of genetic variants within a population. Higher MAFs indicate that a variant is more common, while lower MAFs suggest a rarer variant within the population.

Minor Allele Frequency (MAF) for the marker rs53576 in the human population is 0.39 (A). This means that the A allele, which is the minor allele, is present in approximately 39% of the population. It's worth noting that the highest population MAF listed for this variant is 0.49.

TASK 19

MAF: 0.39 (A)

Minor Allele Frequency (MAF) is a term used in population genetics to describe the frequency at which the less common allele of a genetic variant (typically a single nucleotide polymorphism, or SNP) occurs in a population. In a diploid organism like humans, where each individual has two copies of each gene, the MAF represents the frequency of the less common allele among all the copies present in the population.

For example, if a SNP has two alleles, A and T, and in a population of 100 individuals, 60 individuals have the AA genotype, 30 individuals have the AT genotype, and 10 individuals have the TT genotype, then the less common allele would be T. The MAF would be calculated as the frequency of the T allele, which in this case is $(30+10)/(2*100) = 0.2$, or 20%.

Task 20

The SNP is located on chromosome 3 at position 8762685 on the forward strand. The alleles associated with this SNP are A, G, and T, with the ancestral allele being G. The minor allele frequency (MAF) for allele A is 0.39 (or 39%) in the overall population. The highest population MAF is 0.49 (or 49%). The consequence of this SNP is described as an intron variant. The CADD (Combined Annotation Dependent Depletion) score for allele G is 5.949, and for allele T is 5.477. The GERP (Genomic Evolutionary Rate Profiling) score is -2.61. This variant is co-located with the HGMD-PUBLIC CR105047 variant. There are 6 HGVS names and 4 synonyms associated with this variant. Genotyping assays for this variant are available on 5 chips. The original source of this variant is dbSNP (release 154), and it is mentioned in 461 citations. The 1000 Genomes Project Phase 3 data shows the allele frequencies for different populations. Here are a few examples: ALL: A allele frequency = 39%, G allele frequency = 61% AFR: A allele frequency = 19%, G allele frequency = 81% AMR: A allele frequency = 36%, G allele frequency = 64% EAS: A allele frequency = 65%, G allele frequency = 35% EUR: A allele frequency = 35%, G allele frequency = 65% SAS: A allele frequency = 45%, G allele frequency = 55%

Task 21

3 dna:chromosome chromosome:GRCh38:3:8762285:8763085:1 GGTTTGGGGT-
TAGCTTTTGAGTTTTTTGATTTTGGATTTTTTGTCTTTTTAGCTGTTATTT
ATCAAACCTTTGGGGGGAAAAGAAGTGAAATCACCACAGGGCAGAAACCC-
TAAGGGAAAA CATTAACATTAGCTAAGAACATAAAAGAACACACAATTACT-
TAATCATATAAGTGTCTGA AGTTAACTGTCCATCTAATTGTGATTTGTAC-
CCAGAAGGGCCGAGCTTGTGCACTCTTCA TGGCCCAGAGTGAATATCCT-
GTCCAAGCTTCTCCTGCCGGCCCACCATGCTCTCCACATC ACTGGGTCAC-
CTCAAGAAAAAGCCCCTCCAAGGGGCCTGGTCCCCACACCTCGGGCACA
GCATTCATGGAAAGGAAAGGTGTACGGGACATGCCCGAGGATCCTCAGTC-
CCACAGAAAC AGGGAGGGGCTGGGAAGCTCATTCTACAGATGGGGAAACAGTTCAAACCAC
CTCCTGT GCCTGTCAGCCTTCCTCCCAGTCCAACGCTCCTGACAGAT-
GTTTGGTTGCCCCAGTGATG GGGCGCTCCTTTCTTTTCCAGGTTGCCAGTTCT-
GTTTCAGACAGCTGCTTAGAAAGTCCC TATTCCTCCTGAGTCCAGCTCTTCATG-
GCCATCCCTGCCCCGTCTCACTCACCTCTCTGC TCCCATTTTCCACG-
TATTTGGCAAGCACTGGTTGAGCTATCAATGACTGTGCAGCCTTGT GCCAGGCATC-
CCCTGGGGTAAAAGGCATCCCCTGGAGTTGTGCCACAAAACAGCCCACAC
ATTGGACTTGGGCTTAACAAG

Task 22

Identifying the Species of NRRL 26941 To determine the species of NRRL 26941, you would typi-cally need to consult the relevant databases or scientific literature that contain information about microbial species. The most common and comprehensive database for microbial classification is

the NCBI (National Center for Biotechnology Information) database, which includes the GenBank database for genetic sequences.

Task 23

Genome Size of NRRL 26941 The genome size of NRRL 26941 can be obtained from genomic sequencing data. Genomic sequencing is the process of determining the complete DNA sequence of an organism's genome. By analyzing the sequencing data, scientists can determine the size of the genome, which is typically measured in base pairs (bp) or kilobase pairs (kbp).

Task 24

Contig N50 and Its Meaning Contig N50 is a parameter used to assess the quality and completeness of a genome assembly. In the context of genome sequencing, contigs are contiguous DNA sequences obtained from the fragmented pieces of an organism's genome. The N50 value represents the length at which 50% of the genome assembly is contained in contigs of that length or longer. In other words, it is a measure of the average contig size, indicating the continuity and quality of the genome assembly. A higher N50 value suggests a more contiguous and complete assembly of the genome.

Task 25

Amylase Alpha-amylase 1C

Task 26

amylase assists the chemical process known as digestion. However, amylase does not have a coenzyme working with it. Amylase has a cofactor known as calcium to help catalyze the breakdown of food.

Task 27

AlphaFold

Task 28

 The reason the search engine might show records from species other than humans is that the PDB contains protein structures from various organisms. It serves as a resource for researchers working with proteins from different species and facilitates comparative studies.

Task 29

there are 202 genes that are only detected in the human brain.

Task 30

Disease related genes FDA approved drug targets G-protein coupled receptors Human disease related genes Transporters

Task 31

SNP rs53576. This SNP is associated with the oxytocin receptor and has been found to influence social behavior and personality

Task 32

The SARS-CoV-2 virus genome reference sequence designated as Wuhan-Hu-1 has a size of 29,903 base pairs. It is a single-stranded RNA (ss-RNA) molecule.

Task 33

The collection date of the sample designated as Wuhan-Hu-1 in the SARS-CoV-2 genome reference sequence is listed as "Dec-2019."

Task 34

The amino acid sequence of the spike protein targeted by the vaccines is a long sequence spanning multiple lines. Here is the complete amino acid sequence of the spike protein from the SARS-CoV-2 genome reference sequence Wuhan-Hu-1:

MESLVPGFNEKTHVQLSLPVLQVRDVLRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGV
LPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLPVYVFIKRSDARTA
PHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGI
QYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPV
YVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGH
VMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGR
SGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFI
KRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVE
LVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGEL
GVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSD
ARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAE
LEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLH
VLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTA
PHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGI
QYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPV
YVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGH
VMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGR
SGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFI
KRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVE
LVAELEGIQYGRSGELGVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGEL
GVLHVLPVYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGELGVLHVLPVYVFIKRSD
ARTAPH