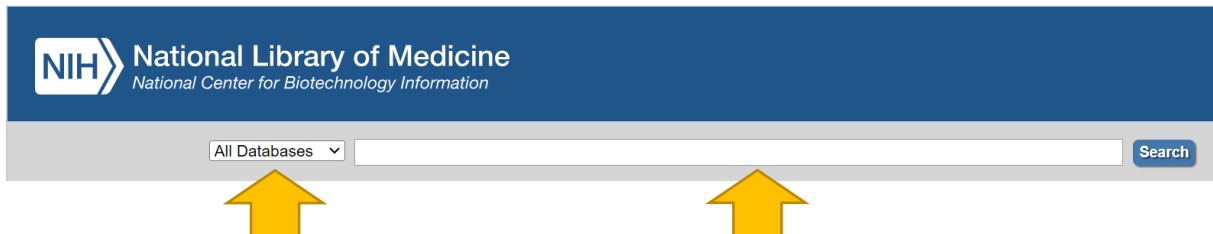


PODSTAWY BIOINFORMATYKI
LISTA ZADAŃ 1

BIOLOGICZNE BAZY DANYCH (1)

1. Wejdź na oficjalną stronę NCBI (<https://www.ncbi.nlm.nih.gov/>). Zapoznaj się z menu oferującym liczne bazy danych. Opisz dokładnie, jakiego rodzaju dane są przechowywane w 5 z nich.



2. Zapoznaj się z formatami zapisu sekwencji nukleotydowej FASTA oraz FASTQ. Który z tych formatów danych jest bardziej informatywny i dlaczego?
3. W bazie danych NCBI **Genome** znajdź genom referencyjny myszy domowej. Podaj:
 - a. nazwę genomu referencyjnego
 - b. unikalny numer identyfikacyjny dla chromosomu 1
 - c. długość chromosomu 19 (w pz).
4. W bazie **Nucleotide** (NCBI), wpisz nazwę organizmu *Fasciola hepatica*. Odpowiedz na poniższe pytania:
 - a. Ile rekordów zostało odnalezionych? Czy rekordy dotyczą tylko żądanego organizmu?
 - b. Dlaczego sekwencja o numerze AJ457042.1 została wyświetlona jako wynik poszukiwań? Jakiego organizmu dotyczy?Wykorzystaj zakładkę „Advanced” i ogranicz poszukiwania do żądanego organizmu.
 - c. Ile rekordów znaleziono? Czy wszystkie dotyczą motyli wątrobowej?
5. W bazie **RefSeq** (NCBI) wpisz ponownie nazwę gatunku *Fasciola hepatica*. Ile rekordów zostało odnalezionych? Czy rekordy dotyczą tylko żądanego organizmu?
6. W bazie danych **Ensembl** znajdź genom referencyjny człowieka i podaj jego nazwę.
7. Na podstawie identyfikatorów sekwencji określ jakiego gatunku oraz jakiego rodzaju cząsteczek genomu dotyczą
 - a. ENST00000471181.7
 - b. ENSFCAG00000032635
 - c. ENST00000505641
 - d. ENSBTAT00000067174
 - e. ENSRNOG00055033312
 - f. ENSVPAT00000002323.1
8. W bazie danych **Ensembl**, w sekcji Download znajdź:
 - a. sekwencje ncRNA dla *Homo sapiens*
 - b. wszystkie sekwencje białkowe dostępne dla *Sus scrofa*Zwróć szczególną uwagę na nazwy tych plików. Czym się charakteryzują?
9. W bazie danych **USCS** (Genomes) wybierz genom referencyjny człowieka. Znajdź gen kodujący enzym hydroksylazy fenyloalaninowej i odpowiedz na poniższe pytania:
 - a. Na którym chromosomie jest zlokalizowany?
 - b. Jakiej długości jest sekwencja?
 - c. Ile ma intronów?
 - d. W której tkance ulega najwyższej ekspresji?

10. Używając narzędzia **BioMart** znajdź wszystkie znane geny/transkrypty dla genomu *A. thaliana*. Poprzez zarządzanie atrybutami wyświetl ID genów, nazwy genów, ID transkryptów oraz koordynaty transkryptów.
11. Za pomocą narzędzia **BioMart** odnajdź polimorfizmy należące do kota domowego. Zawęż poszukiwania do chromosomu X i do pozycji pomiędzy 1 pz, a 2000 pz. Zadbaj o to, aby w pliku wynikowym pojawił się atrybut odnośnie wariantów alleli. Odpowiedz na poniższe pytania:
 - a. Ile polimorfizmów zostało znalezionych?
 - b. Jakie to polimorfizmy?
 - c. Jakie są warianty alleli?
 - d. Oryginalnie z której bazy danych pochodzi informacja o polimorfizmach?
 - e. Opisz strukturę pliku wynikowego (w domyślnym formacie tsv).
12. Kontynuując poprzednie zadanie, przyjrzyj się lokalizacji pierwszego ze znalezionych polimorfizmów. Przeglądanie fragmentu genomu, gdzie znajduje się polimorfizm rozszerz o 100 pz w obie strony. Odpowiedz na pytania:
 - a. Czy to jedyne polimorfizmy znane w tej lokalizacji?
 - b. Czy w pobliżu polimorfizmów znajdują się jakieś geny?Aby pozyskać powyższe informacje użyj „koła zębatego” służącego do modyfikowania wyświetlanych w przeglądarce informacji.

BIOLOGICZNE BAZY DANYCH (2)

13. W bazie danych **UniProt** odnajdź produkt ekspresji genu SRY dla człowieka. Jaka jest jego funkcja?
14. Ze strony głównej NCBI, przejdź do bazy **PubMed Central**. Korzystając z zaawansowanego sposobu wyszukiwania zapytań, odnajdź artykuły naukowe dotyczące:
 - a. CNV i napisanych przez autora Kyung-Won Hong
 - b. Ryjówki aksamitnej i opublikowanych w 2023 roku.Wykorzystaj opcję „Advance” lub wyrażenia logiczne. Jeśli to możliwe zacytuj po 2 artykuły. W jakich czasopismach zostały one opublikowane? Poświęć również czas na przejrzanie abstraktów zacytowanych przez Ciebie artykułów i zastanów się dobrze korespondują z Twoim zapytaniem.
15. W bazie danych **Gene** (NCBI) odnajdź gen SRY dla człowieka.
 - a. Na którym chromosomie jest zlokalizowany?
 - b. Podaj jego dokładne koordynaty.