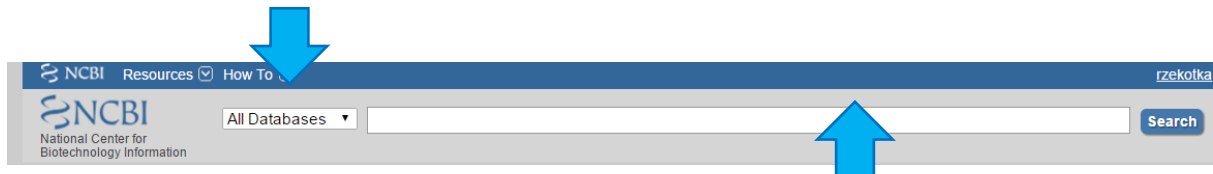


BIOINFORMATYKA
Lista 1

BIOLOGICZNE BAZY DANYCH – PRZYPOMNIENIE I AKTUALIZACJA

1. Wejdź na oficjalną stronę NCBI (www.ncbi.nlm.nih.gov). Przejrzyj menu oferujące liczne bazy danych. Szczególną uwagę zwróć na bazy **GeneBank**, **Nucleotide**, **Protein**, **SRA** oraz **PMC**.



2. Przypomnij sobie czym charakteryzowały się formaty zapisu sekwencji nukleotydowej FASTA oraz GeneBank. Który z tych formatów danych jest bardziej informatywny i dlaczego?
3. Ze strony głównej NCBI, przejdź do bazy **PubMed Central**. Korzystając z prostego sposobu wyszukiwania zapytań, odnajdź artykuły naukowe dotyczące:
 - a. Polimorfizmów liczby kopii u bydła domowego
 - b. Albinizmu i opublikowane przez autora Giorgio Brocco.Ile rekordów odnalazłeś? Jeśli to możliwe zacytuj po 2 artykuły. W jakich czasopismach zostały one opublikowane?
4. W bazie **PubMed** znajdź artykuł naukowy Proctor i wsp. z 2018 roku, który został opublikowany w czasopiśmie Toxicology Letters. Na jaki temat pisali autorzy tego artykułu? Do znalezienia odpowiedzi wykorzystaj opcję „Advance”.
5. W bazie **Nucleotide** odnajdź sekwencję o numerze D00265.1. Na podstawie informacji zawartych w formacie GenBank podaj odpowiedzi na poniższe pytania.
 - a. Jaka jest długość sekwencji?
 - b. Z jakiego organizmu pochodzi sekwencja?
 - c. Czy rekord zawiera dane na temat publikacji związanej z tą sekwencją? Zacytuj.
 - d. Co jest kodowane przez sekwencję?
 - e. Czy sekwencja pochodzi z genomu jądrowego?
6. W bazie **Nucleotide**, wpisz nazwę organizmu *Taenia solium*. Jakiego rodzaju cząsteczek dotyczą wyszukane rekordy?
7. Używając bazy **PMC**, znajdź publikację Sumaili i Mujamammi z 2017 roku. Na podstawie streszczenia, określ na temat jakiej choroby pisali autorzy oraz mutacja, którego genu jest odpowiedzialna za tę chorobę. Wymień 3 publikacje wymienione w sekcji „Similar articles in PubMed”, które odnoszą się do znalezionej publikacji.
8. Korzystając z bazy **Gene**, odnajdź podstawowe informacje na temat genu z zadania nr 7. Użyj opcji „Advanced”. W jedno pole wpisz nazwę genu (skrót), oraz zaznacz „Gene Name”, w drugie pole wpisz „human” i zaznacz „Organism”. Odpowiedz na pytania
 - a. Na którym chromosomie leży wyszukany gen?
 - b. Ile innych gatunków posiada homologii tego genu? Czym jest „homolog”?
 - c. Kiedy znaleziony wpis był ostatnio aktualizowany?
9. W bazie danych **Ensembl** znajdź i ściągnij:
 - a. adnotacje w formacie GTF dla *Bos taurus*
 - b. wszystkie sekwencje białkowe dostępne dla *Sus scrofa*
 - c. sekwencje ncRNA dla *Ovis aries*

10. W bazie **Genome** znajdź genom człowieka. Jaka jest długość chromosomu 10?

11. W bazie danych **USCS** odzyskaj narzędzie Genome Browser. 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 wariantów splicingowych?
- d. Ile ma egzonów?
- e. Czy zawiera elementy powtarzalne/transpozony?
- f. Ile w danym regionie zaobserwowano SNP?
- g. Gdzie ulega najwyższej ekspresji?