

PODSTAWY BIOINFORMATYKI

LISTA ZADAŃ 2

BIOLOGICZNE BAZY DANYCH

1. **Database of Genomic Variants archive** (DGVa) zawiera informacje na temat różnego rodzaju polimorfizmów DNA. Na stronie www.ebi.ac.uk/dgva znajdź sekcję „Data download” i dla gatunku makaka królewskiego wykonaj następujące polecenia:
 - a. sprawdź, ile jest dostępnych badań dotyczących tego gatunku
 - b. podaj ile polimorfizmów zostało opisanych w tych badaniach
 - c. podaj jak brzmią tytuły publikacji naukowych korespondujących z tymi badaniami
 - d. pobierz dostępne pliki i sprawdź jakiego rodzaju polimorfizmy zawierają
2. W bazie **Nucleotide (NCBI)**, wpisz nazwę organizmu *Fasciola hepatica*. Odpowiedz na poniższe pytania:
 - a. Ile rekordów zostało odnalezionych? Czy rekordy dotyczą tylko żądanego organizmu? W celu ułatwienia poszukiwań możesz przejść od razu do ostatniej strony
 - 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 motylicy wątrobowej?
3. Ze strony głównej NCBI, przejdź do bazy **PubMed Central**. Korzystając z zaawansowanego sposobu wyszukiwania zapytań, odnajdź artykuły naukowe dotyczące:
 - a. SNP i napisanych przez autora Kyung-Won Hong
 - b. niedźwiedzia brunatnego i opublikowanych po 2014Wykorzystaj opcję „Advance” lub wyrażenia logiczne. Jeśli to możliwe zacytuj po 2 artykuły. W jakich czasopismach zostały one opublikowane?
4. W bazie **PubMed** znajdź publikację Edith T. Zemanick i Michael W. Konstan z 2010 roku. Na temat jakiej choroby pisali autorzy tej publikacji? Mutacja, którego genu jest odpowiedzialna za tę chorobę?
5. Korzystając z bazy **Gene**, odnajdź podstawowe informacje na temat genu z zadania nr 4. Użyj prostego wyszukiwania poprzez wpisanie nazwy genu w wyszukiwarce.
 - a. Ile rekordów uzyskałeś?Następnie 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 dotyczące genu:
 - b. na którym chromosomie leży?
 - c. jaka jest jego pełna nazwa?
 - d. kiedy po raz ostatni zaktualizowano wpis na temat tego genu?
 - e. ile publikacji w PubMed jest z nim powiązanych?

- f. czy odnajdujesz powiązane z nim pseudogeny? Ile ich jest? Na których chromosomach są położone?
 - g. ile innych gatunków posiada homologii tego genu? Czym jest „homolog”?
6. W bazie **taksonomicznej** NCBI wyszukaj zapytanie *Megaloceros giganteus* i odpowiedz na pytania dotyczące tego gatunku:
- a. Do jakiej rodziny należy?
 - b. Ile sekwencji białkowych *Megaloceros giganteus* znajduje się w bazie danych?
 - c. Czy na podstawie wyszukanego rekordu, jesteś w stanie stwierdzić, ile osobników tego gatunku żyje do dziś? Uzasadnij odpowiedź.
7. W celu uzyskania odpowiedzi na poniższe pytania, przeszukaj bazę **UniProt**. W wyszukiwarce „Advanced Search” wpisz FOXP2 i wybierz “Gene Name”.
- a. Ile rekordów pasuje do zapytania?
- Ponownie użyj zaawansowanego wyszukiwania. Wpisz „human” i wybierz „Organism”.
- b. Ile tym razem wyszukano rekordów?
- Wybierz odpowiedni wpis i uzasadnij
- c. dlaczego wybrałaś/eś właśnie ten
- Następnie zapoznaj się z menu dostępnym po lewej stronie oraz:
- d. podaj pełną nazwę kodowanego przez gen FOXP2 białka.
 - e. jaką funkcję ono pełni?
 - f. jaka jest jego lokalizacja w komórce?
 - g. ile jest powiązanych z nim publikacji?
 - h. czy mutacja genu kodującego dane białko jest związana z jakąś chorobą? Z jaką? Jaki rodzaj mutacji ją powoduje?
8. Pozyskiwanie całych genomów. Używając opcji „ENA Advanced Search” w bazie danych **European Nucleotide Archive** (ENA), odnajdź wszystkie badania dotyczące krótkich sekwencji nukleotydowych (tzw. odczytów = reads), wygenerowanych dla całego genomu szczura wędrownego (strategia WGS – whole genome sequencing) przez platformę Illumina. Ile badań udało Ci się znaleźć? Podaj numery identyfikacyjne oraz tytuły pierwszych trzech.