

Ćwiczenie 3 – Biologiczne bazy danych: przeszukiwanie zasobów

1. Zadanie ma na celu przećwiczenie korzystania z programu przeszukiwania baz w systemie Entrez na przykładzie sekwencji aminokwasowej ludzkiego czynnika stresu cieplnego (HSF, heat shock factor). Liczba rekordów jaka odpowiada tej frazie zostanie w odpowiedni sposób ograniczona przy pomocy opcji systemu Entrez.

W przeglądarce otwórz stronę serwisu Entrez i wybierz z listy bazę danych 'Protein'.

- a. W polu zapytań wprowadź heat shock factor i naciśnij przycisk GO. Ile rekordów zostało znalezionych?
 - b. Zapisz zapytanie jakie zostało stworzone przez system Entrez w ramce 'Search details'.
 - c. Ogranicz przeszukiwanie poprzez zamknięcie szukanej frazy w podwójny cudzysłów. Co powoduje zamknięcie frazy w cudzysłów? Ile rekordów uzyskałeś/aś tym razem? Porównaj zapytanie z zakładki 'Details' z punktem poprzednim – na czym polega różnica?
 - d. Zawęż zakres przeszukiwania poprzez dodanie opcji ograniczającej uzyskane wyniki do rekordów pochodzących z człowieka. Zapisz skonstruowane zapytanie oraz liczbę uzyskanych rekordów.
 - e. Ogranicz zakres przeszukiwania do bazy danych 'GenBank'. Ile rekordów tego typu znajduje się w tej bazie a ile w 'RefSeq'?
 - f. Ile rekordów można zidentyfikować w bazie 'Gene' przy pomocy frazy z punktu (d)?
2.
 - a. Ile sekwencji nukleotydowych oraz białkowych pochodzących z dinozaura Tyrannosaurus rex znajduje się w bazach NCBI?
 - b. Jakie sekwencje zawiera dział nukleotydowej bazy danych 'EST'?
 3. Jaki typ danych zawiera baza UniGene? Korzystając z opcji znajdujących się w zakładce 'Advanced search' należy tak sformułować zapytanie skierowane do bazy UniGene aby odpowiedzieć na następujące pytania:
 - a. Ile rekordów w bazie danych reprezentują sekwencje znajdujące się na chromosomie 2 ryżu?
 - b. Ile rekordów w bazie danych reprezentują sekwencje z ryżu (bez ograniczenia do określonego chromosomu) dla których istnieje co najmniej 5 sekwencji EST?
 - c. Ile rekordów w bazie danych spełnia oba kryteria podane w poprzednich punktach (zastosuj funkcje łączenia zapytań poprzez 'History')?
 4. Ostatni znany wilk workowaty zmarł w ZOO w Hobart w 1936r (Thylacinus cynocephalus = wilk workowaty, wilk tasmański, tygrys tasmański). Z okazów muzealnych tego zwierzęcia uzyskano sekwencje DNA, z którymi można się zapoznać korzystając z 'Taxonomy Browser' (NCBI). Przeszukaj taksonomiczną bazę danych w poszukiwaniu wilka workowatego i powiedz, ile sekwencji DNA oraz białek się tam znajduje? Które geny zostały sklonowane?

5. Wyszukaj wszystkie artykuły poświęcone osteoporozie u kobiet powyżej 45 roku życia opublikowanych od 1 stycznia 1983 roku do chwili obecnej w języku polskim (wpisz w polu wyszukiwarki PubMedu osteoporosis, a następnie zastosuj odpowiednie opcje by zawęzić poszukiwanie do określonej daty i języka). Ile jest artykułów odpowiadających zadanemu kryterium wyszukiwania?
6. Skorzystaj z zakładki 'Advanced Search' i odszukaj artykułu opublikowany w czasopiśmie Cell vol. 147 na stronach 358–369.
 - a. Podaj jego autorów, tytuł oraz datę publikacji.
 - b. Ile artykułów o podobnej treści ('Similar articles') znajduje się w bazie PubMed?
7. Na głównej stronie bazy Pubmed wejdź w zakładkę 'Pubmed Tutorials'. Obejrzyj tutorial 'Use MeSH to Build a Better PubMed Query'. W oparciu o terminy MeSH znajdź artykuły związane z przykrym zapachem z ust (bad breath).
 - a. Jaki termin MeSH odpowiada temu zapytaniu? Ile artykułów znaleziono?
 - b. Ogranicz wyniki wyszukiwania do artykułów związanych z przykrym zapachem z ust, jego przyczyną ('etiology') i możliwym leczeniem farmakologicznym ('drug therapy') i dietą ('diet therapy'). Ile znaleziono publikacji?
 - c. Z jakiego czasopisma pochodzi najstarsza publikacja?
 - d. Kiedy po raz pierwszy wydano czasopismo?
8. Jesteś naukowcem badającym stadia rozwoju Danio pręgowanego. Twój opiekun naukowy poprosił Cię o sporządzenie listy wszystkich genów odgrywających rolę w rozwoju tego organizmu. Najszybszym sposobem na znalezienie tej informacji jest skorzystanie z narzędzia QuickGO w bazie ebi <https://www.ebi.ac.uk/QuickGO-Old/>.

W celu otrzymania listy adnotacji, w pierwszej kolejności w polu QuickGO odszukaj i zanotuj identyfikator GO dla procesu rozwoju (developmental process). Następnie, wejdź w zakładkę 'Filter', wybierz opcję 'Taxon' i wejdź w link 'UniProt'. W bazie UniProt wpisz zapytanie o Danio rerio.

- a. Jaki jest identyfikator GO dla procesu rozwoju?
- b. Jaki jest identyfikator dla tego organizmu?

Wróć do strony QuickGo i dokonaj filtrowania używając odnalezionych identyfikatorów GO i organizmu, upewniając się, że opcja 'Find annotations to descendants of these terms' jest zaznaczona. Korzystając z zakładki 'Statistics', znajdź informację na temat otrzymanych adnotacji:

- c. Ile jest wszystkich wyników adnotacji?
- d. Ile jest unikalnych białek?

Listę wszystkich adnotacji możesz pobrać poprzez zakładkę 'Download'.

- e. Czy aby pobrać wszystkie wyniki należy zmienić limit rezultatów?