

3. BIOLOGICKÉ DATABÁZY

3.2 NIE JE VÍRUS AKO VÍRUS

VEĽKÉ DATABÁZY

Veľké databázy sú mimoriadne veľké súbory údajov, ktoré možno výpočtovo analyzovať s cieľom odhaliť v nich vzory, pravidlá, súvislosti, trendy a zoskupenia. Biologické disciplíny sa dnes nezaobídu bez práce v takýchto obrovských dátových súboroch.

Otázka 1:

Ako vznikajú veľké biologické databázy?

Na evidovaní dát sa podieľajú výskumníci z celého sveta. Dáta sa neustále znásobujú, sú uložené a spracovávané na mnohých serveroch. Pri ich používaní sa stierajú hranice medzi štátmi. Ich bezpečné spravovanie vyžaduje rozsiahle digitálne prostredie so zložitou funkcionalitou a špecialistov, ktorí sa jeho rozvoju a ochrane neustále venujú.

Bioinformatika vznikla ako nový odbor na rozhraní biológie, informatiky, matematickej analýzy.

Bioinformatiku možno v širšom zmysle slova chápať ako akékoľvek použitie výpočtovej techniky pri riešení biologických problémov.

V užšom zmysle slova je bioinformatika spájaná iba s vývojom softvéru na automatizovanú analýzu biologických dát. Prvú „pracovnú verziu“ definície bioinformatiky zverejnil Národný inštitút zdravia USA (NIH) v roku 2000.

Vo všeobecnosti môžeme definovať tri hlavné úlohy bioinformatiky:

1. **Tvorba a správa databáz** biologických dát rôzneho charakteru (DNA, RNA alebo proteínové sekvencie, proteínové štruktúry, génové expresné profily, atď.).
2. **Tvorba algoritmov a matematických modelov** určených na analýzu biologických dát.
3. **Aplikácia bioinformatických nástrojov** za účelom interpretácie týchto dát vo svetle biologických súvislostí.

Úloha 1:

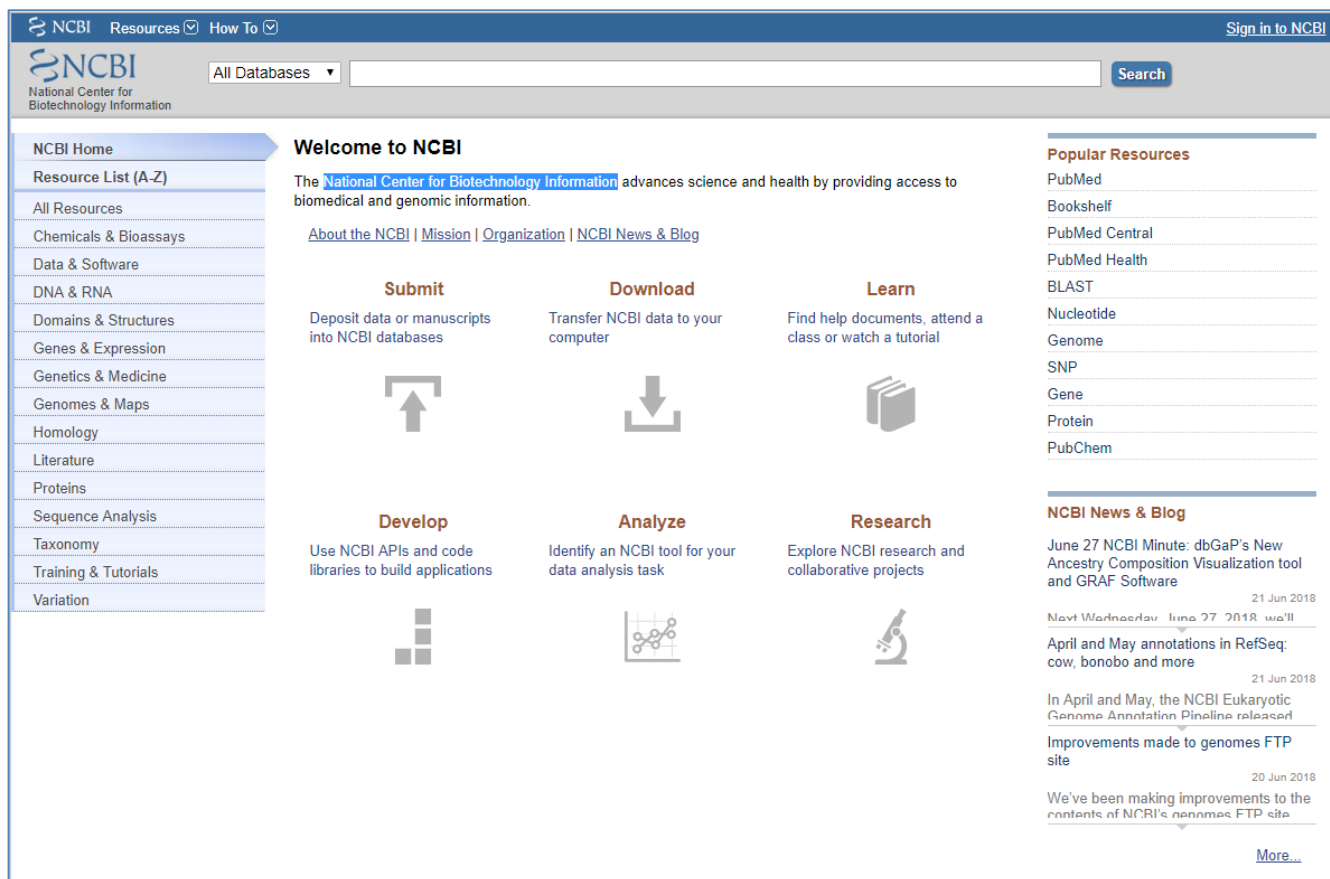
Porozmýšľajte, ako pomáhajú biológom, zdravotníkom, poľnohospodárom alebo ekológom rozsiahle databázy.

Navrhnite problém, ktorý by sa dal riešiť pomocou veľkého množstva reálnych dát:

Databáza NCBI

NCBI je národné centrum pre biotechnologické informácie v USA. Prostredie, ktorého úvodnú stránku vidíte, sa člení na niekoľko oblastí, vidíte ich v zozname vľavo. NCBI vzájomne prepája niekoľko genetických, biochemických, taxonomických a publikačných databáz.

Otvorte databázu na adrese <https://www.ncbi.nlm.nih.gov/>.



Titulná stránka databáz NCBI.

Vpravo sú uvedené najpoužívanější zdroje (databázy), s ktorými je NCBI prepojená. V databázach sa dá pracovať pomocou nástrojov, ktoré umožňujú údaje rôznym spôsobom vyhľadávať, triediť a analyzovať.

Po kliknutí na niektorú oblasť vidíte ponuku súvisiacich databáz a nástrojov, ktoré môžete využiť. Obsahuje návody na učenie sa ako s nástrojmi pracovať (tutorial) - a príručku pomoci (help manual).

Úloha 2:

Vyhľadajte v databáze NCBI **Taxonomy** záznam pre **púpavu lekársku** (*Taraxacum officinale*) a pokúste sa zistiť:

1. Koľko záznamov v databáze NCBI **Nucleotide** existuje pre túto rastlinu?
2. Koľko záznamov je pre púpavu lekársku aktuálne v databáze NCBI **Protein**?
3. Kliknite na názov a nájdite obrázky púpavy v niektorom externom informačnom zdroji (sú v tabuľke na konci stránky).
4. Akú ďalšiu informáciu a púpave sa vám podarilo z databáz získať?

Otvorte niektorý ďalší odkaz a chodte hlbšie do databázy.

Vyskúšajte postup hľadania aj s iným organizmom. Musíte poznať alebo najprv vyhľadať jeho vedecký názov.

Keď vedci a odborníci objavujú nové súvislosti, nevystačia si s jednoduchým vyhľadávaním. Na vyhľadávanie a porovnávanie vzorov v databázach NCBI používajú programovací jazyk Python.

Veľmi rozšíreným v bioinformatike je Python prispôsobený potrebám biológie.

Biopython je súbor voľne dostupných nástrojov na biologické výpočty napísaný v Pythone medzinárodným tímom vývojárov. Riešia potreby súčasnej a budúcej práce v oblasti bioinformatiky. Zdrojový kód je dostupný pod licenciou Biopython, ktorá je mimoriadne liberálna a kompatibilná s takmer každou licenciou na svete.

Aplikáciu nájdete na adrese <http://www.biopython.org>.

Základným zámerom biopythonu je umožniť čo najjednoduchšie používanie Pythonu v bioinformatike tým, že sa vytvorí vysoko kvalitné moduly a opakovane využiteľné triedy modulov. Patrí k nim napríklad FASTA na zobrazovanie a vyhľadávanie sekvencií nukleotidov a aminokyselín.

Tieto moduly sa používajú na prácu s veľkými databázami. Skúsenejší užívateľ, ktorý ovláda python, si môže niektoré z nich samostatne prispôbovať podľa vlastných požiadaviek.

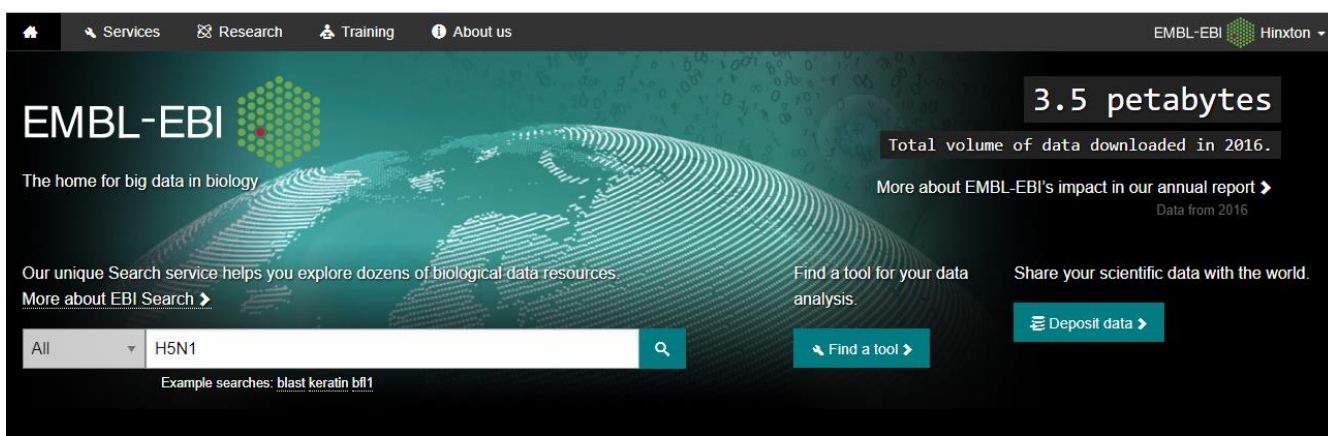
Databáza EBI

Databáza Európskeho bioinformatického inštitútu <https://www.ebi.ac.uk/> je ďalší príklad veľkej databázy.

Európsky inštitút pre bioinformatiku (EMBL-EBI) zdieľa údaje z experimentov v oblasti prírodných vied, vykonáva základný výskum v oblasti výpočtovej biológie a ponúka rozsiahly program školení používateľov, ktorý podporuje výskumníkov z akademickej obce a priemyslu.

Dáta a nástroje sú voľne dostupné bez obmedzenia. Je možné ich stiahnuť a nainštalovať lokálne. Jedinou výnimkou sú potenciálne identifikovateľné ľudské genetické informácie.

Zdroje EMBL-EBI sú komplexné a aktuálne. Biologické údaje sa musia umiestniť do verejného archívu a odkazovať sa v príslušnej publikácii.



Explore EMBL-EBI and our mission

The European Bioinformatics Institute (EMBL-EBI) shares [data from life science experiments](#), performs [basic research](#) in computational biology and offers an extensive [user training programme](#), supporting researchers in academia and industry. We are part of [EMBL](#), Europe's flagship laboratory for the life sciences. [More about EMBL-EBI and our impact](#)

Services

We provide freely available data and bioinformatics services to all facets of the scientific community ➤

Research

We contribute to the advancement of biology through basic investigator-driven research ➤

Training

We provide advanced bioinformatics training to scientists at all levels ➤

Industry

We help disseminate cutting-edge technologies to industry ➤

ELIXIR

We support, as an ELIXIR node, the coordination of biological data provision throughout Europe ➤

Vyhľadajme v databázach niektoré informácie o vírusoch.

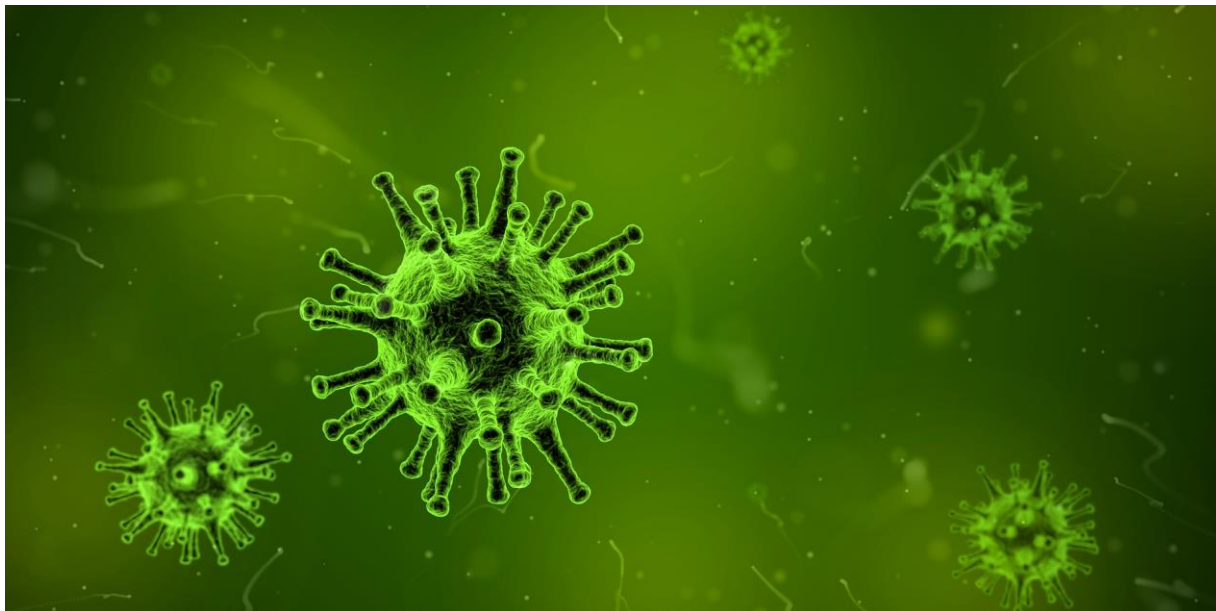
Chrípkový vírus

Chrípkových vírusov je veľa a sú veľmi variabilné, ľahko podliehajú mutáciám. To je dôvod, prečo sa venujú veľké prostriedky každý rok na rozpoznanie najnebezpečnejšieho typu, ktorý sa rýchlo šíri a na vývoj novej očkovacej látky.

Chrípkové vírusy typu A sa rozdeľujú na podtypy na základe prítomnosti povrchových glykoproteínov na povrchu vírusovej častice, ktoré majú antigénne vlastnosti. Hlavnými antigénnymi proteínmi vírusovej častice sú

- hemaglutinín (H)
- neuraminidáza (N)

Tieto proteíny sú veľmi premenlivé. Rozoznávame 16 typov proteínu hemaglutinínu (H1–H16) a 9 typov neuraminidázy (N1–N9). Ich kombinácia je charakteristická pre daný podtyp vírusu. Je známych pätnásť podtypov vírusu chrípky, ktoré sú schopné nakaziť vtáky. Podtypy H5 a H7 sú vysoko nákazlivé, napríklad hromadné rozšírenie vtácej chrípky v rokoch 2005 – 2006 spôsobil podtyp H5N1.



Vírus

Úloha 3:

Otvorte si stránku **EMBL-EBI**. Zadaťte do vyhľadávacieho okienka označenie vírusu **H5N1**. Zvoľte si makromolekulárne štruktúry (**Macromolecular structures**) a pozrite si niektoré zobrazenia proteínov tohto vírusu.

Nájdite informácie o tom istom víruse v databáze NCBI. Zvoľte si v rozbaľovacom políčku databázu nukleotidov **Nucleotide**. Do vyhľadávacieho políčka napíšte **H5N1**. Opäť skúste nájsť zobrazenie proteínu vírusu.

Vírus osýpok

Ďalšie ochorenie, ktoré sa začalo na Slovensku nečakane šíriť, sú osýpky. Súvisí to s poklesom miery preventívnej zaočkovanosti populácie. Osýpky sú veľmi nákazlivou vírusovou chorobou, prenášanou vzdušnou cestou z iných infikovaných ľudí. Vedecký názov osýpok je **Morbilli**.

Vírusy osýpok patria do skupiny paramyxovírusov. Známe sú predovšetkým gény paramyxovírusov hlodavcov.

Úloha 4:

Otvorte si stránku **NCBI**. Vyhľadajte známe nukleotidové sekvencie vírusu osýpok. Postupujte takto:

Zvoľte si v rozbaľovacom políčku databázu nukleotidov **Nucleotide**. Do vyhľadávacieho políčka napíšte názov ochorenia (morbilli). Zobraza sa záznamy o sekvenciách známych génov. Kliknutím na link [FASTA](#) sa zobrazí prislúchajúca sekvencia cRNA.

Analogicky, ak si v rozbaľovacom políčku zvolíte databázu **Protein**, viete si pomocou linku FASTA zobrazíť zodpovedajúcu sekvenciu aminokyselín.

Ak chcete nájsť sekvencie cRNA génov alebo aminokyselín proteínov ľudských osýpok, zadajte do vyhľadávacieho políčka **homo sapiens morbillivirus**.