

Fakulta agrobiológie a potravinových zdrojov

Ústav rastlinných a environmentálnych vied

prof. Ing. Katarína Ražná, PhD. – prof. RNDr. Milan Bežo, CSc.  
prof. Ing. PaedDr. Jana Žiarovská, PhD.

**NÁVODY NA CVIČENIA Z GENOMIKY  
A BIOINFORMATIKY – BIOLOGICKÉ DATABÁZY  
A DATABÁZOVÉ ZÁZNAMY**

Tretie nezmenené vydanie

Nitra 2023

Vydala Slovenská poľnohospodárska univerzita v Nitre  
vo Vydavateľstve SPU

Autori: prof. Ing. Katarína Ražná, PhD. (3,53 AH)  
Ústav rastlinných a environmentálnych vied  
FAPZ, SPU v Nitre

prof. RNDr. Milan Bežo, CSc. (1 AH)

prof. Ing. PaedDr. Jana Žiarovská, PhD. (1 AH)  
Ústav rastlinných a environmentálnych vied  
FAPZ, SPU v Nitre

Recenzenti: doc. Ing. Juraj Candrák, PhD.  
Ústav výživy a genomiky  
FAPZ, SPU v Nitre

doc. Mgr. Ildikó Matušíková, PhD.  
Fakulta prírodných vied  
Univerzita sv. Cyrila a Metoda v Trnave

Schválila rektorka Slovenskej poľnohospodárskej univerzity v Nitre dňa 7. 7. 2023  
ako návody na cvičenia pre študentov SPU.

Táto publikácia bola vytlačená na ekologickom papieri.



© K. Ražná, M. Bežo, J. Žiarovská, Nitra 2023

ISBN 978-80-552-2620-0

## OBSAH

Predhovor  
PodĎakovanie

|                                                                          |           |
|--------------------------------------------------------------------------|-----------|
| <b>1. Úvod</b>                                                           | <b>5</b>  |
| <b>2. Databázy biologických dát</b>                                      | <b>6</b>  |
| 2.1 Rozdelenie a zameranie biologických databáz                          | 8         |
| 2.2 Primárne databázy                                                    | 11        |
| 2.3 Sekundárne databázy                                                  | 16        |
| 2.4 Špecializované databázy                                              | 18        |
| 2.5 Integrované databázové systémy a biologické portály                  | 21        |
| <b>3. Formáty databázových záznamov</b>                                  | <b>24</b> |
| 3.1 Formát <i>flatfile</i>                                               | 26        |
| 3.2 Formát FASTA                                                         | 29        |
| 3.3 Formát PIR/NBRF                                                      | 31        |
| 3.4 Ďalšie typy formátov                                                 | 31        |
| <b>4. Štruktúra databázového formátu <i>flatfile</i></b>                 | <b>34</b> |
| 4.1 Štruktúra sekvenčného záznamu ( <i>flatfile</i> )                    | 34        |
| 4.2 Anotácia sekvencií treťou stranou, TPA (Third Party Annotation)      | 44        |
| 4.3 Referenčné sekvenčné záznamy, RefSeq (Reference Sequences)           | 45        |
| 4.4 EMBL Genome reviews                                                  | 46        |
| 4.5 Zasielanie sekvencií do databázy                                     | 46        |
| <b>5. Vyhľadávanie a získavanie informácií v biologických databázach</b> | <b>48</b> |
| 5.1 Štruktúra a možnosti vyhľadávania v rozhraní Entrez                  | 49        |
| 5.2 Vyhľadávanie sekvenčných záznamov v databáze GenBank (NCBI)          | 55        |
| 5.3 Vyhľadávanie publikačných záznamov PubMed/MEDLINE                    | 60        |
| <b>6. Úlohy</b>                                                          | <b>67</b> |
| <b>7. Terminológia</b>                                                   | <b>69</b> |
| <b>8. Zoznam použitej literatúry</b>                                     | <b>72</b> |

## **Predhovor**

Bioinformatika sa stáva neoddeliteľnou súčasťou výskumu a bádania. Je to vedná oblasť interdisciplinárneho charakteru s významným dosahom pre vedu a prax. Teoretické znalosti a praktické skúsenosti v danej oblasti tvoria platformu vedeckej práce odborníkov pôsobiacich vo výskumných, univerzitných pracoviskách a v súkromnom sektore biologického zamerania.

Návody na cvičenia z enomiky a bioinformatiky – biologické databázy a databázové záznamy sú študijnou literatúrou určenou pre študentov predmetu Genomika a bioinformatika a študentov študijného programu Genetické technológie rastlín. Rovnako nájdú uplatnenie v širokom spektre biologicky zameraných odborov a programov univerzitného smeru. Publikácia poskytuje základné informácie o databázach biologických údajov, o formátoch databázových záznamov a ich obsahovej štruktúre, ako aj o spôsoboch vyhľadávania biologických a bibliografických informácií v databázach. Súčasťou skrípt je pomerne rozsiahla obrázková príloha uľahčujúca čitateľovi priamu aplikáciu získaných poznatkov vo vyhľadávaní informácií v databázach, otázky, úlohy a terminológiu z danej oblasti.

Zámerom autorov je pripraviť nadstavbu učebnej pomôcky, ktorá bude obsahovo venovaná pokročilým postupom bioinformatiky, zahrňujúcim analýzy na úrovni sekvencií nukleotidov, posudzovanie podobnosti sekvencií a fylogenetické analýzy, ako aj oboznámenie sa so špecializovanými databázami a špecifickými algoritmami vyhodnocovania sekvenčných zápisov.

Budeme radi, ak daná študijná literatúra, okrem svojho základného poslania, prispeje aj k získaniu náklonnosti čitateľa o daný vedný odbor a najmä podporí využívanie bioinformatiky v štúdiu, príprave záverečných prác a vo výskume.

Autori

## **PodĎakovanie**

Autori vyjadrujú poďakovanie recenzentom Mgr. Ildikó Matušíkovej, PhD a doc. Ing. Jurajovi Candrákovi, PhD. za zodpovedné zhodnotenie rukopisu, cenné pripomienky a návrhy pre skvalitnenie obsahovej stránky publikácie. Ďakujeme svojim najbližším za zázemie, ktoré nám vytvárajú pre našu vedeckú a pedagogickú realizáciu.

## 1. Úvod

Bioinformatika je v súčasnosti jednou z najintenzívnejšie sa rozvíjajúcich vedných oblastí. Svojim interdisciplinárnym obsahom a aplikáciou, významne ovplyvňuje viaceré vedné odbory (genetiku, biológiu, biochémiu, informatiku, matematiku, medicínu, farmakológiu, poľnohospodárstvo a potravinárstvo). Metódy bioinformatiky umožňujú analyzovať genómy a molekulárno-fenotypové interakcie organizmov z doteraz nepoznanej perspektívy. Vedecká hodnota informácií a znalostí z tohto odboru má priamu aplikáciu pre efektívnu poľnohospodársku produkciu a zabezpečenie kvalitných potravinových produktov. Aplikčný význam a poznatkový potenciál bioinformatiky pre poľnohospodárske vedy podnietil vznik samostatného odvetvia tzv. omických vied vo forme agroinformatiky, respektíve agrobioinformatiky.

V oblasti poľnohospodárstva a potravinárstva vyúsťuje aplikácia bioinformatiky do viacerých výstupov:

- identifikácie génov podmieňujúcich produkčne a výživovo cenné vlastnosti (genomika),
- štúdiá adaptability genómu na abiotické a biotické faktory stresu (transkriptomika, proteomika, metabolomika),
- identifikácie faktorov podmieňujúcich efektívne využívanie živín z pôdy (metagenomika),
- sledovanie interakcií medzi výživou zvierat, resp. človeka a expresiou génov (nutrigenomika).

Teoretické znalosti a praktické skúsenosti z oblasti bioinformatiky sa stávajú nevyhnutnou súčasťou vedy a výskumu. Nastolený trend napredovania v danej oblasti, ktorý je prezentovaný vývojom špecializovaných bioinformačných databáz, nových algoritmových nástrojov *in silico* analýz, zákonite podmieňuje hľadanie ďalších možností archivácie, katalogizácie, spravovania dát a ich vyhodnocovania. Uvedený vývoj je nepochybne podienený aj zvyšujúcou sa úrovňou technickej a prístrojovej infraštruktúry, čo však zároveň so sebou prináša nepriaznivý fenomén vo forme limitovanej kvality dát a množstva opakujúcich sa údajov.

**Autori:**

prof. Ing. Katarína Ražná, PhD. – prof. RNDr. Milan Bežo, CSc.  
prof. Ing. PaedDr. Jana Žiarovská, PhD.

**Názov:**

NÁVODY NA CVIČENIA Z GENOMIKY A BIOINFORMATIKY  
– BIOLOGICKÉ DATABÁZY A DATABÁZOVÉ ZÁZNAMY

Vydavateľ: Slovenská poľnohospodárska univerzita v Nitre

Vydanie: tretie nezmenené

Počet strán: 74

Rok vydania: 2023

AH – VH: 5,53 – 5,67

Neprešlo redakčnou úpravou vo Vydavateľstve SPU.

ISBN 978-80-552-2620-0