

## Praktická úloha č. 2

### Téma: Genómy

Pojem genóm označuje súhrn celkovej genetickej informácie v bunke. Technológie sekvenovania DNA nám dnes umožňujú skúmať nielen jednotlivé gény v organizmoch, ale aj ich celé genómy, čo bude téma tejto teoreticko-praktickej úlohy. Na jej vyriešenie máte 60 minút a maximálny počet bodov je 40. Veľa šťastia!

### Úloha 1

Za priekopnícku prácu na vývoji technológie sekvenovania DNA bola v roku 1980 udelená Nobelova cena za chémiu Frederickovi Sangerovi, Walterovi Gilbertovi a Paulovi Bergovi. Podľa prvého menovaného bola pomenovaná tzv. Sangerova metóda, ktorá umožnila osekvenovať celý genóm organizmov a s určitými vylepšeniami bola využitá aj pri sekvenovaní ľudského genómu začiatkom tohto tisícročia. Napriek tomu, že ide o veľmi precíznu metódu, je tiež veľmi pomalá a neumožňuje spracovať veľmi dlhé sekvencie.

**1A:** Predpokladajte, že chybovosť Sangerovej metódy, teda pravdepodobnosť, že nesprávne určí konkrétny nukleotid v sekvencii je 0.001%. Vypočítajte, minimálne koľko básových párov musí mať sekvenovaná molekula DNA, aby bola pravdepodobnosť, že sú všetky nukleotidy určené správne nižšia než 50%.

(pomôcka: riešením rovnice  $a^x = b$  je  $x = \log_a b = \log b / \log a$ ).

**3b**

**1B:** Napriek výsledku v predchádzajúcej časti, maximálna dĺžka DNA, ktorú je schopná Sangerova metóda zvládnuť je asi 1 000 básových párov (bp; ak myslíme na dvojvláknovú DNA). Jeden inštrument dokáže za približne 2 hodiny spracovať a osekvenovať naraz paralelne maximálne 384 vzoriek. Ak máte 100 takýchto



zariadení, vypočítajte, za aký najkratší čas (predpokladajúc, že už máte všetky vzorky pripravené!) dokážete osekvenovať celý jadrový obsah DNA malej kvitnúcej rastliny *Paris japonica*, organizmu s najväčším známym genómom, ak jej haploidný genóm má veľkosť 150 Gbp (teda asi 50-krát viac než človek) a jej bunky sú oktaploidné. Výsledok uveďte zaokrúhlený v dňoch. (Gbp = giga bázových párov).

**3b**

## Úloha 2

**2A:** Rozmanitosť veľkosti a štruktúry genómov je skutočne obrovská, ak sa pozeráme naprieč všetkými organizmami, ale aj v rámci určitých skupín. Veľkosť genómu môže byť popísaná viacerými údajmi, či už hmotnosťou DNA v jadre somatických buniek, C-hodnotou (hmotnosťou haploidného genómu), stupňom ploidie somatických buniek alebo množstvom bázových párov (bp).

Vašou úlohou je správne doplniť údaje v nasledujúcej tabuľke, ktorá obsahuje rôznych zástupcov vyšších rastlín ak viete, že 1 pikogram DNA obsahuje približne 978 miliónov bázových párov (1pg = 978 Mbp). Všetky výsledky uveďte zaokrúhlené na 1 desatinné miesto:

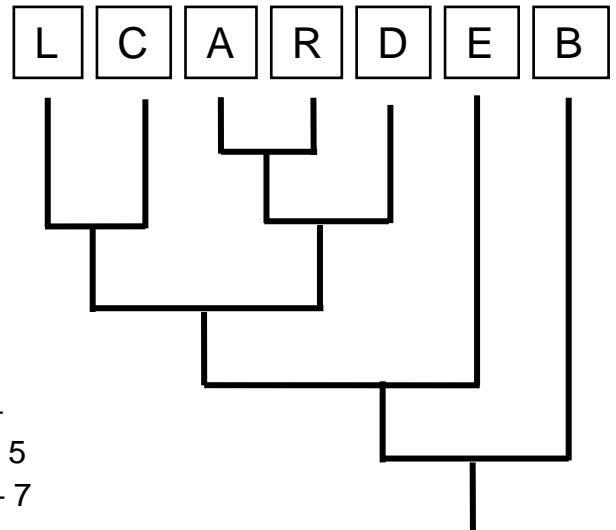
**5b**

| druh | hmotnosť DNA v jadre (pg) | C-hodnota (pg) | ploidia   | veľkosť haploidného genómu (Mbp) |
|------|---------------------------|----------------|-----------|----------------------------------|
| A    |                           | 17,3           | hexaploid |                                  |
| B    | 29,3                      |                |           | 14 328                           |
| C    | 0,72                      | 0,18           |           |                                  |
| D    |                           | 2,8            | diploid   |                                  |
| E    |                           |                | diploid   | 32 206                           |

**2B:** Uvedený dendrogram zobrazuje fylogenetické vzťahy medzi druhmi uvedenými v tabuľke vyššie a dvoma ďalšími druhmi. Na základe vašich znalostí o rastlinnej ríši sa pokúste doplniť priradenie názvu druhu a jeho pozície v dendrograme a označte, či je tvrdenie nižšie pravdivé (P) alebo nepravdivé (N) zakrúžkovaním príslušnej možnosti.

**6b**

1. *Genlisea pygmaea* (z radu hluchavkotvaré)
2. praslička roľná (*Equisetum arvense*)
3. pšenica letná (*Triticum aestivum*)
4. borovica jedlá (*Pinus edulis*)
5. levanduľa úzkolistá (*Levandula augustifolia*)
6. kokosovník obyčajný (*Cocos nucifera*)
7. ryža siata (*Oryza sativa*)



A –  
B –  
C –  
D –

E –  
L – 5  
R – 7

Rastlina z tabuľky s najväčším haploidným genómom má semená vyživované triploidným endospermom: **N P**

Rastlina so všeobecne najväčším haploidným genómom má semená vyživované triploidným endospermom: **N P**

### Úloha 3

**3A:** Rozmanitosť nie je len vo veľkosti genómov ale aj v tom, čo sa v nich nachádza. Vašou úlohou bude s použitím uvedenej tabuľky zostrojiť graf závislosti počtu génov (os y) od veľkosti genómu (os x) pre vybrané živočíšne druhy. Pokúste sa tiež zostrojiť priamku, ktorá túto závislosť čo najlepšie popisuje.

**5b**

| Slovenský názov  | Latinský názov                  | Veľkosť genómu (Mbp) | Počet génov |
|------------------|---------------------------------|----------------------|-------------|
| včela domáca     | <i>Apis mellifera</i>           | 250                  | 10 157      |
| mihuľa morská    | <i>Petromyzon marinus</i>       | 886                  | 26 046      |
| tučniak cisársky | <i>Aptenodytes forsteri</i>     | 1 254                | 13 929      |
| vtákopysek divný | <i>Ornithorhynchus anatinus</i> | 1 996                | 18 527      |
| pes domáci       | <i>Canis lupus familiaris</i>   | 2 411                | 19 300      |



**3B:** Na základe dát z tabuľky a nižšie uvedeného vzorca pre jednoduchú lineárnu regresiu vypočítajte parametre priamky  $Y = aX + b$  z úlohy 3A. Na základe tohto vzťahu, odhadnite počet ľudských génov, ak viete, že človek má genóm s veľkosťou asi 3 Gbp. **3b**

$$a = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$
$$b = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$n$  je počet bodov nanesených do grafu

$\sum y$  označuje súčet  $y$ -ových súradníc všetkých bodov v grafe

$\sum x$  označuje súčet  $x$ -ových súradníc všetkých bodov v grafe

$\sum xy$  označuje súčet súčinov  $x$ -ových a  $y$ -ových súradníc pre všetky body v grafe

$\sum x^2$  označuje súčet druhých mocnín  $x$ -ových súradníc všetkých bodov v grafe

Miesto na pomocný výpočet:

**$a$  (s presnosťou na desatiny):**

**$b$  (s presnosťou na jednotky):**

**Odhad počtu ľudských génov:**

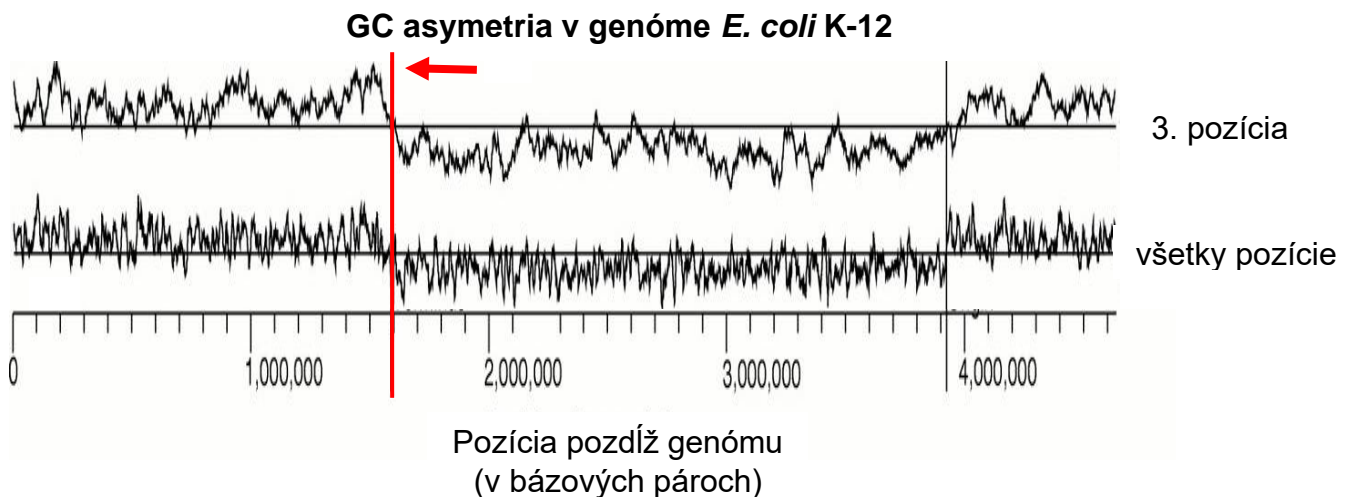
**3C:** Na základe vašich znalostí o živočíšnej ríši sa pokúste zostrojiť dendrogram podobný tomu v úlohe 2B, ktorý by reprezentoval fylogenetické vzťahy medzi druhmi z tabuľky v úlohe 3A. **2b**

**3D:** Genómy živočíchov (ale aj rastlín a iných eukaryotických organizmov) veľmi často obsahujú okrem exónov (kódujúcich oblastí a 3', 5' neprekladaných oblastí génov) aj ďalšie elementy, ktoré prispievajú k veľkosti genómu. Z uvedených pojmov zakrúžkujte dva, ktoré označujú elementy, ktoré najvýraznejšie prispievajú k veľkosti živočíšnych genómov: **2b**

transpozóny      STOP kodóny      intróny      Shine-Dalgarnova sekvencia

#### Úloha 4

Ďalšou zaujímavou vlastnosťou genómov je ich chemické zloženie, teda relatívne zastúpenie jednotlivých nukleotidov. Tentokrát sa pozrieme na prokaryotický genóm baktérie *Escherichia coli*, kmeň K-12. Blattner et al. (1997) vzal sekvenciu len na jednom vlákne DNA, rozdelil ju na regióny s dĺžkou 10 kbp a pre každý z nich spočítal tzv. GC asymetriu podľa nižšie uvedeného vzorca (G a C označuje počet jednotlivých nukleotidov). Výsledky analýzy vidíte v nasledujúcom grafe, ktorý zobrazuje GC asymetriu na 3. pozícii v kodónoch a na všetkých pozíciách v kódujúcich oblastiach. Horizontálne čiary predstavujú nulovú asymetriu a šípka označuje pozíciu počiatku replikácie:



$$\text{GC asymetria} = \frac{G-C}{G+C}$$

Z nasledujúcich možností zakrúžkujte 6 tých, ktoré sú pravdivé (za každú zakrúžkovanú nepravdivú odpoveď sa vám 2 body odpočítajú): **6b**

- A) GC asymetria pre sekvenciu AGCGTGTGATAATGGA je kladná, čo znamená, že je obohatená o cytozín.
- B) Druhé vlákno genómu *E. coli* K12 (komplementárne k tomu v grafe) bude mať vzor GC asymetrie zrkadlovo obrátený okolo x-ovej osi.
- C) Z grafu vyplýva, že v oblasti medzi približne 1,60 a 3,90 Mbp má zobrazené vlákno nadbytok cytozínu v kódujúcich oblastiach.
- D) Zobrazený vzor v asymetrii rozloženia G a C nukleotidov by mohlo súvisieť s mechanizmom DNA replikácie, keďže sa s vláknami zaobchádza rozdielne - jedno vlákno je vedúce a druhé zaostávajúce.
- E) Na rozdiel od človeka, chromozóm *E. coli* neobsahuje teloméry.
- F) Nie všetky baktérie majú kruhový chromozóm, napr. *Borrelia burgdorferi*, pôvodca Lymfatickej choroby, má lineárny chromozóm.
- G) Replikácia DNA prebieha v *E. coli* len jedným smerom, takže počiatok replikácie je hneď veľa miesta ukončenia replikácie (v opačnom smere).
- H) Zmena vo vzore GC asymetrie, ako nastáva napríklad v *E. coli* K12 na pozícii ~ 1,6Mbp, by mohla identifikovať aj pozíciu počiatku aj pozíciu terminácie replikácie.
- I) Zo znalosti GC asymetrie pre dané vlákno DNA možno odvodiť aj AT asymetriu (definovanú úplne analogicky ako GC asymetria).

## Úloha 5

Vráťme sa opäť k eukaryotickým genómom. Predstavte si, že chcete osekvenovať celý ľudský genóm, všetku DNA ktorú zdravé ľudské bunky obsahujú. Na to, aby ste zvýšili presnosť svojej analýzy ho v skutočnosti chcete osekvenovať niekoľkokrát. Parameter, ktorý popisuje koľkokrát ste osekvenovali nukleotid na určitej pozícii v genóme sa nazýva pokrytie (coverage). Ak by ste to urobili len raz alebo dvakrát, a našli by ste nečakanú substitúciu, nevedeli by ste povedať, či ide o skutočnú mutáciu v genóme alebo len chybu pri sekvenovaní. Chcete ho teda osekvenovať do pokrytia aspoň 20 (tiež nazývaného hĺbka sekvenovania).

