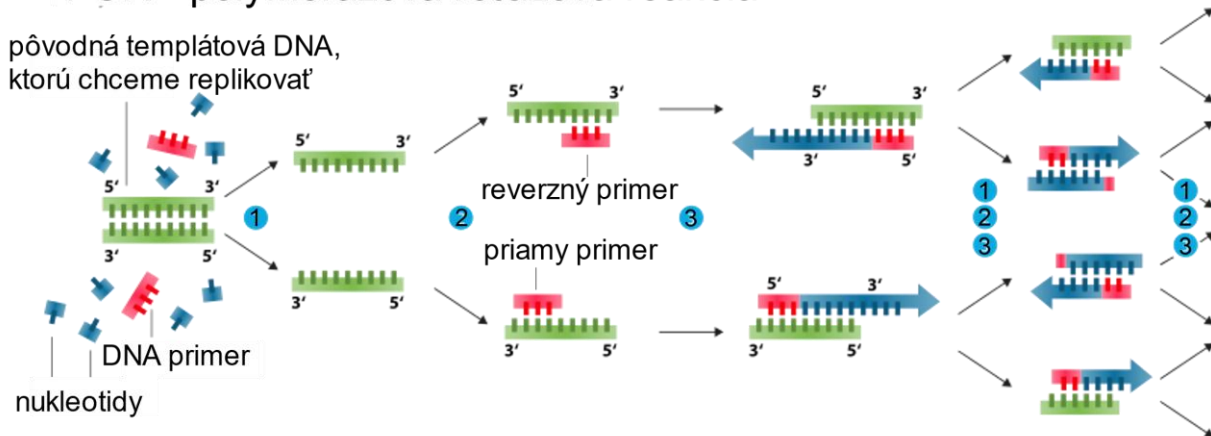


**Téma: Bioinformatika a molekulárna systematika****Úvod:**

Správne určenie príbuzenských vzťahov medzi organizmami a vymedzenie druhových hraníc je nevyhnutné nie len pre poznanie biodiverzity, ale často aj pre ďalší aplikovaný výskum, ochranu prírody, či boj proti fytopatogénom. Veľa nových zistení pritom priniesol v posledných desaťročiach rozvoj molekulárnej biológie a využitie sekvencií DNA. Jedným z odborov, kde molekulárne dáta výrazne posunuli naše poznanie, je aj mykológia. Obzvlášť v prípade mikroskopických húb sme často schopní sledovať len výrazne obmedzené množstvo morfológických znakov.

Pokiaľ chcete získať sekvenciu určitého génu, obvykle musíte najskôr z vášho materiálu izolovať DNA. Následne potrebujete úsek DNA, ktorý vás zaujíma, namnožiť, k čomu slúži polymerázová reťazová reakcia – PCR. PCR je založená na cyklickom striedaní rôznych teplôt, pri ktorých postupne dochádza k denaturácii DNA, nasadnutiu primerov a k predlžovaniu ich 3' konca polymerázou podľa templátovej DNA. Polymeráza totiž nedokáže syntetizovať nové vlákno od začiatku a potrebuje krátky oligonukleotid – primer, na ktorý nasadne a ktorý predlžuje. Primer sa pritom na DNA viaže špecificky iba na miesto, ktoré je komplementárne s jeho sekvenciou. Pokiaľ teda do reakcie pridáte dva oligonukleotidy – priamy a reverzný primer, ktorými si váš vybraný úsek z oboch strán ohraničíte, polymeráza bude množiť špecificky práve tento úsek. Váš vybraný úsek DNA sa tak po niekoľkých cykloch začne vo vzorke hromadiť a po 20 – 40 cykloch už vo vzorke prevažuje. Zjednodušene je priebeh prvých cyklov PCR zobrazený aj na obrázku nižšie.

## PCR - polymerázová reťazová reakcia



- 1 **Denaturácia** pri 94-96 °C
- 2 **Nasadenie primerov** pri 45-65 °C
- 3 **Polymerizácia** pri 72 °C

Potom ešte PCR produkt skontrolujete na gélovej elektroforéze a úspešné reakcie prečistíte. Samotná sekvenácia prebieha obvykle v špecializovanom laboratóriu so sekvenátorom, kam svoje vzorky zašlete. Existuje niekoľko rôznych metód sekvenovania, no pre naše účely sa používa najčastejšie modifikácie Sangerovej metódy. Tá je podobne ako PCR založená na replikácii DNA pomocou polymerázy. Na rozdiel od PCR sa však používa len jeden primer a okrem klasických nukleotidov sa do reakcie pridajú aj fluorescenčne značené dideoxynukleotidy (ddNTP). Po naviazaní ddNTP dôjde k zastaveniu elongácie reťazca, a keďže sú jednotlivé ddNTP rôzne fluorescenčne značené, je možné určiť, ktorý ddNTP sa naviazal a ktorý nukleotid bol v pôvodnom reťazci. Vďaka tomu, že to či, a kedy sa do nového rastúceho reťazca v priebehu reakcie začlení ddNTP, je náhodné, vznikne zmes rôzne dlhých vlákien DNA, ktoré majú špecifické fluorescenčné označenie zodpovedajúce ddNTP, ktorý sa naviazal na ich 3' konci. Získané produkty sa potom rozdelia elektroforézou na základe ich dĺžky a vďaka rôznemu fluorescenčnému označeniu jednotlivých ddNTP je možné zrekonštruovať sekvenciu vášho génu. Práve záznam fluorescencie rôzne dlhých reťazcov v podobe chromatogramu potom po pár dňoch obdržíte ako výsledok sekvenovania.

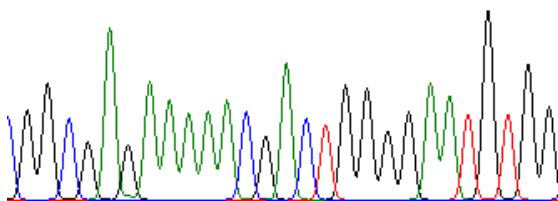
Tu začína vaša úloha, predstavte si hypotetickú situáciu, že skúmate druh parazitickej vreckatej huby – lesklospórku (*Lamprospora maireana*). Tá má podľa súčasnej literatúry široké hostiteľské spektrum a vy chcete overiť, či sa skutočne jedná o jeden druh alebo o niekoľko morfológicky podobných, no hostiteľsky špecializovaných druhov. Po niekoľkých dňoch v teréne sa vám podarí nájsť 8 vzoriek lesklospórky (zberov) na 3 rôznych hostiteľských druhoch. Informácie o jednotlivých zberoch sú uvedené v nasledujúcej tabuľke.

Číslo zberu	Lokalita	Hostiteľ
1	Komárno, breh Dunaja, 110 m n. m.	pečeňovka <i>Fossombronia sp.</i>
2	Trenčín, breh Váhu, 200 m n. m.	pečeňovka <i>Fossombronia sp.</i>
3	Štúrovo, Kamenínske slanisko, 150 m n. m.	pečeňovka <i>Fossombronia sp.</i>
4	Nitra, Zobor, 500 m n. m.	mach <i>Trichostomum sp.</i>
5	Bratislava, Devínska Kobyla, 150 m n. m.	mach <i>Trichostomum sp.</i>
6	Záhorská nížina, Šaštín-Stráže, 250 m n. m.	mach <i>Weissia controversa</i>
7	Terchová, Jánošíkove diery, 600 m n. m.	mach <i>Weissia controversa</i>
8	Belianske Tatry, Široké sedlo, 1800 m n. m.	mach <i>Weissia controversa</i>

Vaše zbery ste spracovali a zo sekvenovacieho laboratória vám prišli sekvencie génu 28S rRNA. Na začiatku musíte skontrolovať ich kvalitu – na výslednom chromatograme by sa vrcholy zodpovedajúce jednotlivým bázam nemali prekryvať a malo by byť jednoznačné, aká báza sa na danej pozícii vyskytuje (Obrázok A). Pokiaľ PCR reakcia neprebehla úspešne, napr. kvôli chybe pri pipetovaní alebo kontaminácií vášho materiálu, môžu byť vrcholy na chromatograme nerozlíšiteľné alebo sa prekryvať (Obrázok B). Aj v prípade bezproblémovo pripravených vzoriek sú však typicky výsledky na začiatku a na konci sekvenovacej reakcie menej kvalitné.

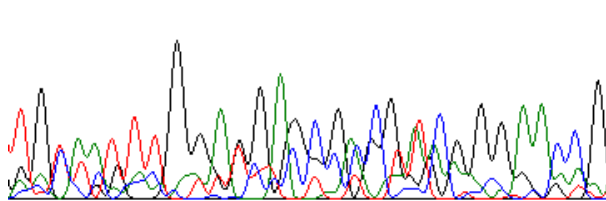
**A**

480 490 500  
 >GG CG AG AAAA AC GACT G GGG AAT GTGG



**B**

90 100 110  
 T G CAAT TT G G AG G AGC GCCTT CG G GAACC G.



**Postupujte podľa zadania a pozor, súbory ktoré vytvoríte, uložte na plochu vášho počítača, budete hodnotení nie len za vaše odpovede, ale práve aj za tieto súbory. V prípade že vám niečo nebude fungovať alebo sa bude diať niečo neočakávané (napr. sa po dvojkliku nespustí očakávaný program), skúste si najskôr zadanie ešte raz dôkladne prečítať. Pokiaľ by problém pretrvával, prihláste sa a my vám prídeme na pomoc. Veľa šťastia!**

1. Vašou prvou úlohou bude skontrolovať kvalitu sekvencií, ktoré nájdete v priečinku „Sekvencie“ na pracovnej ploche vášho počítača. Po tom, čo na každú sekvenciu dvakrát kliknete, mala by sa otvoriť v programe „BioEdit“, kde už vidíte aj chromatogram. Prezrite si postupne všetky sekvencie, ďalej totiž môžete pracovať len so sekvenciami, ktoré sú kvalitné a dobre čitateľné. Pri prezeraní sekvencií si môžete výšku vrcholov zvýšiť pomocou posuvníku vľavo od chromatogramu.

Sekvencie ktorých zberov by podľa vás nemali byť použité v ďalšej analýze?

2. Z čitateľných sekvencií potrebujete odstrániť menej kvalitné konce a previesť ich do textovej podoby, v akej s nimi môžete ďalej pracovať. Pri sekvenciách označte čitateľný úsek (ideálne čo najdlhší) bez nekvalitne osekvenovaných koncových úsekov, skopírujte ho a vložte do nového dokumentu v poznámkovom bloku (program „Poznámkový blok“ máte na ploche). V riadku nad sekvenciou potom ešte musíte vytvoriť hlavičku, ktorú tvorí znak „>“ a názov príslušného zberu. Do jedného súboru takto pod seba postupne vložte všetky kvalitné sekvencie. To, ako by to mohlo vyzeráť pre prvé dva zbery, vidíte na nasledujúcom príklade. Výsledný súbor uložte na ploche pod názvom „sekvencie-vaše súťažné číslo.txt“.

>Zber1

CTA.....

>Zber2

AGA.....

3. K tomu, aby ste mali lepšiu predstavu o fylogenetickej pozícií vzoriek lesklospórky odobratých v rámci tejto štúdie, je potrebné do analýzy zahrnúť aj sekvencie od ďalších príbuzných druhov. Pokiaľ ich sami nechcete alebo nemáte možnosť sekvenovať, môžete si sekvencie niektorých z nich stiahnuť aj z online databáz, kde sú uložené sekvencie z publikovaných prác iných autorov. Niekoľko takýchto sekvencií nájdete v priečinku „Alignment“. Otvorte si v poznámkovom bloku súbor „Lamprospora.txt“ a pod sekvencie, ktoré v ňom sú, prekopírujte vaše sekvencie zo súboru v úlohe 2. Súbor potom uložte na plochu vo formáte fasta ako „Lamprospora.fas“. Spustite si program „MEGA7“, otvorte súbor „Lamprospora.fas“ („File“ > „Open a File/Session“) a zvolte možnosť „align“. Vašou ďalšou

úlohou bude sekvencie zarovnať a vytvoriť tzv. zarovnanie (alignment). Pokiaľ totiž chcete rekonštruovať fylogenetický strom, musíte sekvencie medzi sebou zarovnať tak, aby boli zodpovedajúce si pozície pod sebou, sekvencie neboli medzi sebou rôzne posunuté a dali sa porovnať. Označte všetky sekvencie, kliknite na „Alignment“ a vyberte možnosť „Align by ClustalW“. Nechajte prednastavené parametre a spustíte analýzu. Po pár sekundách by vaše sekvencie mali byť zarovnané. Začiatky a konce sekvencií orežte podľa najkratšej tak, aby žiadne zo sekvencií na začiatku alebo konci nepresahovali (označte si stĺpce, ktoré chcete zmazať a stlačte „delete“). Výsledné zarovnanie uložte na plochu ako „alignment-vaše súťažné číslo.fas“ tak, že kliknete na „Export“ a zvolíte možnosť „FASTA format“.

4. Teraz môžete začať so zobrazením príbuzenských vzťahov vašich vzoriek. Existujú viaceré metódy tvorby fylogenetických stromov, vy si dnes vyskúšate výpočetne menej náročnú a rýchlejšiu alternatívu – metódu najbližšieho suseda (Neighbour joining), ktorá je založená na postupnom zhlukovaní sekvencií podľa ich podobnosti – genetickej vzdialenosti. V programe „MEGA7“ by ste stále mali mať otvorené vaše zarovnanie, kliknite na „Data“ a vyberte možnosť „Phylogenetic Analysis“. Následne by sa vám malo otvoriť okno, v ktorom sa vás program pýta, či sa jedná o proteín-kódujúce sekvencie, vy kliknite na „No“. Prekliknite na druhé okno v programe „MEGA7“ (to bez farebného zarovnanie) a kliknite na „Phylogeny“ a „Construct/Test Neighbor-Joining Tree“, program sa vás ešte spýta, či chcete použiť aktuálne aktívne dáta a vy len potvrdíte „Yes“ a spustíte analýzu stlačením „Compute“. Po pár sekundách by sa malo objaviť nové okno s fylogenetickým stromom. Tento strom je však zatiaľ nezakorenený, a aby ste mohli správne interpretovať príbuzenské vzťahy medzi jednotlivými druhmi a ich evolúciu, musíte ho zakoreniť pomocou sekvencie príbuzného organizmu, ktorý nepatrí do nami študovanej skupiny organizmov (Outgroup). V našom prípade sa jedná o druh *Rhodotarzetta rosea*, kliknite teda pravým tlačidlom myši na vetvu vedúcu k tomuto druhu a vyberte možnosť „Place Root“. Výsledný strom už iba uložte („Image“ > „Save Enhanced Meta File (EMF)“) ako „Strom-vaše súťažné číslo.emf“.

5. Prehliadnite si fylogenetický strom, ktorý ste vytvorili, a vypíšte mená všetkých taxónov (druhov, rodov), ktoré nie sú monofyletické – nachádzajú sa na viacerých miestach v strome alebo aspoň nezahŕňajú spoločného predka a všetkých jeho potomkov.

6. Čo sa dá o súčasnom druhovom koncepte *Lamprospora maireana* usudzovať na základe vášho fylogenetického stromu? Jedná sa o jeden druh so širším hostiteľským spektrom alebo niekoľko hostiteľsky špecializovaných druhov? Vysvetlite svoju odpoveď.

Sekvencie, ktoré vám prídu zo sekvenátora, nemusia vždy patriť iba organizmu ktorý sledujete a ktorý ste sa pokúšali osekvenovať. Mohlo sa totiž stať, že sa pri PCR namiesto jeho sekvencie amplifikovala sekvencia iného druhu, ktorým bola vaša vzorka/chemikálie kontaminované. K približnému zaradeniu vašej sekvencie a odhaleniu možnej kontaminácie slúži napr. program „BLAST“, ktorý vašu sekvenciu porovná so sekvenciami z publikovaných prác iných autorov a ukáže vám tabuľku tých najpodobnejších sekvencií. Občas sa však kontaminácie zbaviť nedokážete a aj po opakovanej izolácii DNA získate len jej sekvenciu. Riešením tohto problému môže byť napr. návrh nových primerov, ktoré budú špecifické len pre skupinu organizmov, ktorú skúmate. Ideálne primery by pritom mali spĺňať niekoľko podmienok:

I. dĺžka 16 – 25 nukleotidov

II. oba primery majú približne rovnakú teplotu topenia

III. sekvencia primeru na 3'-konci je úplne komplementárna k templátu

IV. obsah G a C je spolu v primere v rozmedzí 40 – 60 %

V. primer netvorí sekundárne štruktúry, nie je komplementárny voči sebe

VI. priamy a reverzný primer nie sú ani navzájom komplementárne (netvorí diméry)

7. Na nasadnutie polymerázy a priebeh polymerizačnej reakcie stačia aj podstatne kratšie primery (napr. už 6 nukleotidov). Prečo sa teda obvykle pri PCR používajú výrazne dlhšie primery?

8. Prezrite si nasledujúci primer a vysvetlite, ktoré z podmienok (I. – V.) nespĺňa a prečo? O ktorých podmienkach nemôžete rozhodnúť, vysvetlite? 5' GGTATCGCAATTGCGATACC 3'

9. Predstavte si hypotetickú situáciu, opakovane sa vám pri PCR s primermi zaciľujúcimi gén 28S rRNA amplifikovala sekvencia kontaminujúcej DNA. Zo sekvencií väčšiny druhov rodu *Lamprospora* a *Octospora*, ktoré študujete a ich častých kontaminantov ste si vytvorili zarovnanie. To môžete nájsť uložené na ploche pod názvom „Alignment\_kontaminacia.fas”. Otvorte si ho v programe „MEGA7“ a na základe tohto zarovnaní sa pokúste navrhnúť primery s dĺžkou 20 nukleotidov, ktoré sú špecifické len pre zástupcov rodov *Lamprospora* a *Octospora* a umožnili by amplifikovať čo možno najdlhší úsek z pôvodného génu (ideálne by sa na 3' konci mali líšiť od kontaminácie v aspoň dvoch nukleotidoch). Sekvencie v zarovnaní sú zapísané v smere 5' -> 3', sekvencia priameho primeru by mala byť identická so sekvenciou príslušného úseku sekvencie v zarovnaní, zatiaľ čo sekvencia reverzného primeru by mala byť reverzným komplementom vybraného úseku sekvencie zo zarovnaní (ak by mal napríklad príslušný úsek v zarovnaní sekvenciu ATTGC, tak reverzný komplement k tejto sekvencii by bol GCAAT). Sekvencie oboch navrhnutých primerov zapíšte v smere 5' -> 3'.

**10.** Teplota topenia primeru  $T_m$  závisí na dĺžke primeru a zastúpení jednotlivých nukleotidov a zjednodušene sa dá spočítať podľa vzorca  $T_m = (A+T) * 2 + (G+C) * 4$  (za A, T, G, C dosadíte počet príslušných nukleotidov). Vypočítajte  $T_m$  vašich dvoch primerov a rozhodnite, či sa od seba výrazne nelíšia (rozdiel by mal byť menší ako 5 °C).

**11.** Gén, ktorý ste sekvenovali a dnes používali na rekonštrukciu fylogeniezy, má názov 28S rRNA a podieľa sa na tvorbe veľkej ribozomálnej podjednotky. Keby ste sa pritom bližšie pozreli na jeho sekvenciu, našli by ste v nej niekoľko stop kodónov vo všetkých troch čítacích rámcoch. Napriek tomu sa ale jedná o normálne funkčný gén, pokúste sa vysvetliť, ako je to možné.

Autor: Bc. Lukáš Janošík

Recenzia: Mgr. Katarína Juríková