

Praktická úloha č. 1

Téma: Molekulárna fylogenetika

Úvod:

Správne určenie príbuzenských vzťahov medzi organizmami a vyčlenenie druhových hraníc je často nevyhnutné pre ďalší aplikovaný výskum, ochranu prírody a poznanie biodiverzity. Po dlhú dobu v biológii prevládal hierarchický systém klasifikácie, v ktorom sa príbuzné druhy združujú do spoločných vyšších taxonomických jednotiek, ako sú rody, čeľade, triedy a pod.. Od polovice 20. storočia sa však do popredia dostáva aj moderná fylogenetická systematika využívajúca kladistickú analýzu. Vzťahy medzi organizmami sú potom vyjadrené vo fylogenetickom strome, tzv. kladograme, ktorý pozostáva zo série dichotomických vetvení, kde každý vetviaci bod zodpovedá rozchodu dvoch druhov zo spoločného predka. Veľa nových zistení priniesol v posledných desaťročiach rozvoj molekulárnej biológie a využitie sekvencií DNA na rekonštrukciu kladogramov. Tie totiž poskytujú veľké množstvo znakov s podobnou pravdepodobnosťou zmeny a umožňujú určiť príbuzenské vzťahy aj medzi organizmami s veľmi jednoduchou stavbou tela, z ktorých by sme inak nedokázali získať dostatok morfológických dát.

Jednou z metód, ako sa dá takýto kladogram zo sekvencií získať je UPGMA (Unweighted Pair Group Method with Arithmetic Mean) a práve to ako táto metóda funguje si vyskúšate v nasledujúcej úlohe.

Pomôcky: písacie potreby, pravítko

Postup:

Základným princípom tejto metódy je, že sa dvojica druhov (klastrov) s najnižšou hodnotou vzdialenosti (vyjadrenou napríklad jednoducho ako počet substitúcií medzi homologickými sekvenciami dvoch druhov) v matici spojí do nového klastra vyššieho rádu. To sa opakuje v niekoľkých iteráciách, až kým nezískame jeden výsledný spoločný klastor. Nižšie je na ukážku detailne popísaný postup na príklade štyroch hypotetických druhov A–D. Na základe toho potom budete následne spracovávať váš súbor dát.

	A			
A	0			
B	6	0		
C	7	8	0	
D	8	9	4	0

Iterácia 1: Dvojica s najnižšou hodnotou vzdialenosti (počtom substitúcií) je v tomto prípade C a D, ktoré teda spojíme do prvého klastru (C, D). Relatívny vek nového klastru vypočítame tak, že

hodnotu vzdialenosti medzi dvoma sekvenciami (resp. ich klastrami) jednoducho vydělíme dvomi. V tomto prípade je teda relatívny vek nového klastru $4/2 = 2$. Potom vytvoríme novú maticu vzdialeností, v ktorej už nevystupujú druhy C a D samostatne ale ako klastery (C, D). Vzdialenosť nového klastru (C, D) od druhu A vypočítame ako priemer pôvodnej vzdialenosti druhu C od A a druhu D od A. Výsledná vzdialenosť klastru (C, D) od druhu A je teda $(7+8)/2 = 7,5$. Podobne potom vypočítame aj vzdialenosť klastru (C, D) od druhu B. Teda $(8+9)/2 = 8,5$. Hodnoty zapíšeme do novej matice vzdialeností.

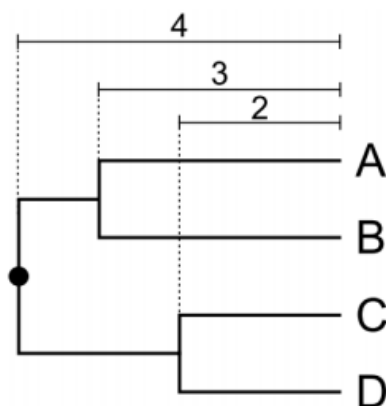
	A		
A	0	B	
B	6	0	(C,D)
(C,D)	7.5	8.5	0

Iterácia 2: Dvojicu s najnižšou vzdialenosťou teraz predstavujú druhy A a B, ktoré zlúčime do klastru (A, B) s relatívnym vekom $6/2 = 3$. Vzdialenosti medzi klastrami (A, B) a (C, D) vypočítame ako priemer vzdialenosti A od (C,D) a vzdialenosti B od (C,D). V tomto prípade je to $(7,5+8,5)/2 = 8$.

	(A,B)	
(A,B)	0	(C,D)
(C,D)	8	0

Iterácia 3: V poslednej iterácii dva zostávajúce taxóny skombinujeme do nového klastru ((A, B),(C, D)), ktorého relatívny vek je $8/2 = 4$.

Na základe výsledného klastru potom môžeme zostaviť kladogram, dĺžka vetiev pritom zodpovedá relatívnemu veku vypočítanému pre daný klastery.



Sekvence druhov, ktoré budete v úlohe analyzovať:

Druh A: **GTCTTTTTTGGTGCTCGAGTTGTAATTTGGAGGGGCACTTT**

Druh B: **GTCTTTTTCAGGCGTCCGAATTGTAATTTGGAGGAGCGTCTTC**

Druh C: **GTTTTTTTCAGGCGTCAGAGTTGTAATTTGGAGGAGTATCTTC**

Druh D: **GTTTTTTTCAGGCGTCCGAGTTGTAATTTGGAGGAGTATCTTC**

Druh E: **GTTTTTTTAGGCGTCCGAGTTGTAATTTGGAGGAGCGTCTTC**

Druh F: **GTTTTTGGTAGGCGTCCGAGTTGTAATTTGGAGGAGCGTCTTC**

Vypracovanie:

1) Pred tým než začnete samotnú analýzu, spočítajte jednoducho počet substitúcií medzi jednotlivými sekvenciami a doplňte chýbajúce hodnoty v nasledujúcej tabuľke:

Tabuľka 1: Počet substitúcií medzi sekvenciami.

	A					
A	-	B				
B		-	C			
C	11		-	D		
D		4		-	E	
E	9		4		-	F
F		5		5		-

2) Do nasledujúcej tabuľky vyplňte hodnoty vzdialeností po prvej iterácii.

Tabuľka 2: Počet substitúcií medzi sekvenciami, resp. ich klastrami po prvej iterácii.

	-				
		-			
			-		
				-	
					-

Relatívny vek nového klastru: _____

3) Do nasledujúcej tabuľky vyplňte hodnoty vzdialeností po druhej iterácii.

Tabuľka 3: Počet substitúcií medzi sekvenciami, resp. ich klastrami po druhej iterácii.

	-			
		-		
			-	
				-

Relatívny vek nového klastru: _____

4) Do nasledujúcej tabuľky vyplňte hodnoty vzdialeností po tretej iterácii.

Tabuľka 4: Počet substitúcií medzi sekvenciami, resp. ich klastrami po tretej iterácii.

	-		
		-	
			-

Relatívny vek nového klastru: _____

5) Do nasledujúcej tabuľky vyplňte hodnoty vzdialeností po štvrtej iterácii.

Tabuľka 5: Počet substitúcií medzi sekvenciami, resp. ich klastrami po štvrtej iterácii.

	-	
		-

Relatívny vek nového klastru: _____

6) Napíšte výsledný klastre po piatej iterácii: _____

Relatívny vek výsledného klastru: _____

7) Na základe výsledného klastru zostavte kladogram, dĺžka vetiev klastrov by mala zodpovedať veku klastru, použite mierku: 1:2, t.j. relatívny vek 1 = 2 cm.

Doplňujúce otázky:

1) Na rekonštrukciu fylogény sa môžu použiť sekvencie rôznych úsekov DNA. Tie sa však od seba môžu výrazne líšiť v tom ako rýchlo sa v nich hromadia mutácie a nie všetky gény sú preto dobre použiteľné vo všetkých štúdiách.

I. Vyberte z nasledujúcich dvojíc vždy možnosť, u ktorej by ste predpokladali rýchlejšie hromadenie mutácií.

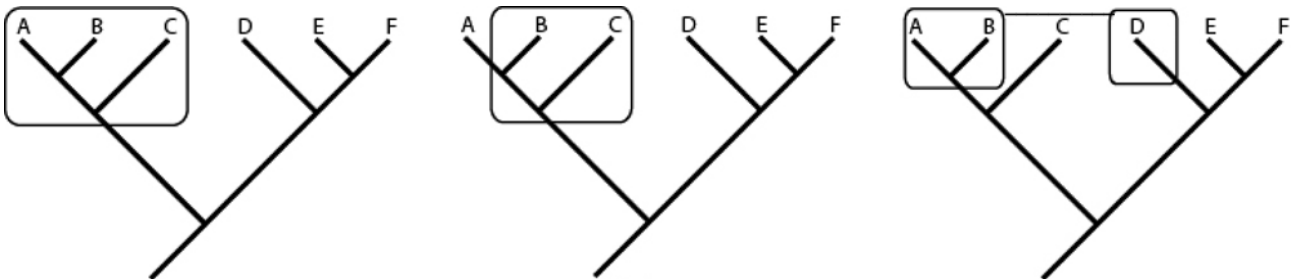
- a) exón / intrón
- b) gén kódujúci ribozomálnu RNA / pseudogén
- c) vystrihovaná nekódujúca oblasť génu / regulačná oblasť génu
- d) mitochondriálny gén / jadrový gén

II. Vyberte z nasledujúcich génov ten, ktorý by bol podľa vás najlepšie použiteľný v štúdiu zaoberajúcej sa príbuzenskými vzťahmi medzi rôznymi živočíšnymi kmeňmi.

- a) intrón v géne kódujúcom tubulín
- b) konzervatívny chloroplastový gén
- c) gén pre ribozomálnu RNA
- d) pseudogén v oblasti teloméry

2) Na ktorej pozícií v rámci tripletu by ste podľa vás pri proteín-kódujúcich génoch našli pri dvoch príbuzných druhoch najviac substitúcií a prečo?

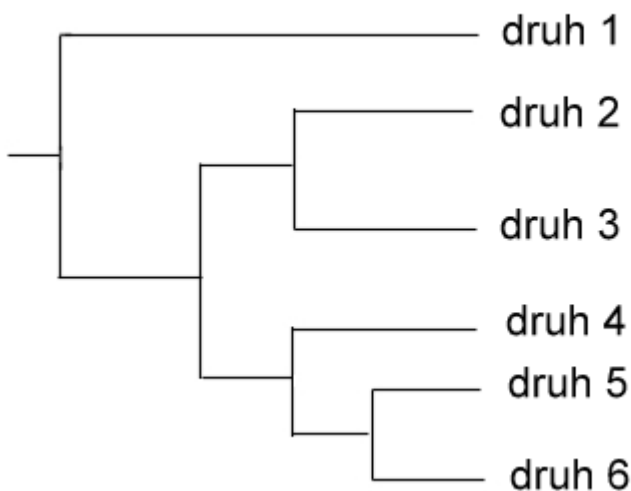
3) Z hľadiska toho, či taxón zahŕňa spoločných predkov a všetkých jeho potomkov rozlišujeme taxóny na polyfyletické, monofyletické a parafyletické. Nasledujúci obrázok znázorňuje tri rôzne taxóny. Priradiť k obrázkom o aký typ sa v jednotlivých prípadoch jedná.



a) _____ b) _____ c) _____

4) Na popis kladogramu slúžia rôzne termíny a ich porozumenie je často nevyhnutné pre pochopenie a správnu interpretáciu výsledkov štúdie. Na nasledujúcom kladograme vyznačte:

- a) koreň
- b) sesterský taxón druhu 5
- c) vonkajšiu skupinu (outgroup)
- d) posledného spoločného predka druhov 4 a 6



- 5) Vyznačte, ktoré z nasledujúcich tvrdení o fylogenetike je/sú pravdivé (P) a ktoré nepravdivé (N).
- a) V prípade využitia morfológických dát by sme mali do analýzy zahrnúť iba znaky, ktoré vznikli viackrát nezávisle.
 - b) Pri taxónoch s medzidruhovou hybridizáciou nie sme schopní pomocou jednoduchého dichotomicky sa vetviaceho kladogramu popísať všetky príbuzenské vzťahy.
 - c) Výhodou sekvencií DNA je, že sa dajú do analýzy zahrnúť aj fosílné druhy, z ktorých nemáme dostatok morfológických údajov.
 - d) Sekvencie DNA obvykle poskytujú väčšie množstvo údajov využiteľných na analýzu v porovnaní s morfológickými údajmi.
 - e) Bez ohľadu na to aké druhy sledujeme môžeme na určenie blízkych príbuzenských vzťahov vždy použiť chloroplastovú DNA, ktorá má vyššiu mutačnú rýchlosť.

Autor: Lukáš Janošík

Recenzia: prof. RNDr. Peter Fedor, PhD.