

MOLEKULÁRNÍ FYLOGENETIKA A TAXONOMIE 2019 - 1

Taxonomie neboli systematika se snaží katalogizovat diverzitu organismů a uspořádat je do hierarchicky uspořádaných skupin – příbuzné druhy do rodů, příbuzné rody do čeledí atd. Každá taxonomická skupina může patřit jen do jedné vyšší skupiny. Každá taxonomická skupina je v ideálním případě definována znaky nebo kombinacemi znaků, které její členové sdílí a zároveň se jimi odlišují od jiných organismů. Postupem času převládá v taxonomii názor, že nejlepším hlediskem, podle kterého organismy třídit, je jejich příbuzenský vztah, tedy fylogeneze. Toto hledisko je objektivní a také obecně použitelné. Předpokládáme totiž, že všechny organismy prošly jednou určitou fylogenezí¹ a naším cílem je ji zrekonstruovat. Rekonstrukcí fylogeneze se zabývá **fylogenetika**. Ta se snaží vystopovat pořadí větvení taxonů (**kladogenezi**), ale všímá si také vývoje vlastností v rámci linií (**anageneze**). Molekulární fylogenetiku, tedy fylogenetiku založenou na molekulárních znacích, můžeme považovat jednak za pomocnou disciplínu molekulární taxonomie, ale je to samozřejmě také samostatný rozvinutý obor, jehož výsledky nemusí být vždy používány pro taxonomické účely. Je třeba zdůraznit, že na výstupy fylogenetických metod je však třeba pohlížet jako na hypotézy, které jsou předmětem testování a mohou být zavrhovány. To však neotřásá naší představou o existenci jedné určité fylogeneze organismů.

Molekulární fylogenetika si při rekonstrukci fylogeneze pomáhá **molekulárními znaky**, pod kterými rozumíme znaky uloženými v **sekvencích informačních makromolekul** – DNA, RNA a proteinů. Nebudeme si tedy vůbec všimnout sacharidů, i když sacharidy jsou používány jako diagnostické znaky (typ buněčné stěny bakterií, druh zásobních sacharidů eukaryotických skupin), a vlastně ani vyšší než primární struktury informačních makromolekul (kličky na rRNA a podobně). Ty sice mohou být používány pro definici taxonů, ale jedná se v podstatě o morfologický znak. Molekulárnímu fylogenetikovi mohou pomoci třeba k tvorbě přesnějšího alignmentu (co to je alignment se dozvíte hned v druhé přednášce), ale to, co molekulární fylogenetika zajímá především, jsou primární sekvence informačních makromolekul. To proto, že primární struktura DNA je právě tou úrovní, na které vznikají evoluční novinky ve formě mutací a navíc je to také zrovna DNA, která je replikátorem a předává se fyzicky do dalších generací podle jasných pravidel. Pro morfologické znaky, ale ani proteiny, to neplatí. Ty se dědí nepřímo skrze DNA. Naštěstí pro molekulární fylogenetiku, je DNA replikátor nepřesný, tu tam se přepíše špatně. V rozšíření těchto odchylek mezi studovanými organismy se skrývá **informace o historii daného lokusu** DNA nebo proteinu. V závislosti na tom, o jaký lokus DNA nebo o jaký protein se jedná, nám ve srovnání s odpovídajícími lokusy jiných organismů umožňují určit identitu jedince a jeho příbuznost s ostatními jedinci v populaci, druhovou příslušnost nebo dokonce příbuzenský vztah tohoto druhu nebo příslušné taxonomické skupiny s jinými.

Jak jsem již uvedl výše, je snaha, aby klasifikace organismů byla v souladu s jejich fylogenezí. Systematické skupiny organismů jsou tedy skupiny „historické“ – spojuje je společný původ, který je jediným nutným předpokladem pro jejich vytváření. Na rozdíl od ahistorických skupin (např. systém chemických prvků), které jsou vytvářeny na základě definujících znaků (podle počtu protonů a neutronů v jádře) a vznikly jednotlivých atomů určitého prvku jsou zcela nezávislé. U skupin organismů, na rozdíl společného původu, striktně nevyžadujeme přítomnost definujících znaků u všech členů skupiny. Vůbec nás nezaráží, že hadi jsou řazeni

¹ I když u prokaryot je to možná trochu komplikovanější díky velmi časté výměně genetické informace mezi nepříbuznými skupinami – horizontální genový přenos

mezi tetrapody aniž by měli končetiny. Věříme totiž v jejich společný původ s ostatními tetrapody, protože tato fylogenetická hypotéza se nám jeví jako velmi robustní, je podpořená mnoha jinými znaky, a umíme si představit, že hadi končetiny prostě ztratili.

Znalost fylogeneze skupiny nám jednoznačně ukazuje, které taxony nesmíme vytvářet. Základním požadavkem, který by měl taxon splňovat je monofyletyčnost. **Monofyletický** taxon neboli **klád**, je takový taxon, jehož členové si jsou vzájemně příbuzní více, než je kdokoli z nich příbuzný druhu mimo. Obsahuje tedy předka a všechny jeho potomky. Kladisti jsou přesvědčeni, že takové taxony jsou existující objekty, které vznikly se společným předkem a budou existovat, dokud nevymře jeho poslední potomek. Cílem taxonomie je tyto objevovat a označovat jmény. Evoluční taxonomové (na rozdíl od kladistů) uznávají také **parafyletické taxony**. Parafyletický je takový taxon, který zahrnuje společného předka a některé (ne však všechny) jeho potomky. Takovým taxonem jsou například plazi. Společným předkem všech plazů byl plaz, ale z některých potomků tohoto plaza se vyvinuli ptáci a savci, které do plazů neřadíme. Parafyletický taxon někdy nazýváme vývojovým stupněm. Všichni taxonomové se shodují na tom, že je nepřípustné vytvářet polyfyletické taxony. **Polyfyletický** je takový taxon, který nejen že nezahrnuje všechny potomky společného předka, ale nezahrnuje ani tohoto předka, protože ten neměl znaky, které tento taxon definují. Příkladem polyfyletického taxonu jsou třeba slunivky – amébovití prvoci s kulovitou buňkou radiálně rozprostřenými axopodiemi. Ukázalo se, že různé slunivky spadají do několika zcela nepříbuzných skupin eukaryot. Společný předek slunivek byl zároveň také předkem např. chaluh, výtrusovců, skrytěnek, nálevníků nebo dírkonošců a rozhodně nevypadal jako slunivka. Dalším polyfyletickým taxonem jsou například řasy.

Znalost fylogeneze nám tedy napovídá, které taxony tvořit nesmíme, ale nepřikazuje nám, které taxony tvořit máme nebo musíme. Toto rozhodnutí je již na každém z nás a je potřeba přihlížet zejména ke stupni anagenese jednotlivých větví, tj. jak významné novinky se u nich vyvinuly a zda jsou natolik významné, že tuto větev oddělíme jako samostatný taxon.

Je vhodné rozlišovat pojmy **znak** a **forma znaku**. Znakem rozumíme část těla, rys či vlastnost, která je kvazi-nezávislá (přistupujeme k ní, jako by byla nezávislá, i když mnohdy to tak nebude) na ostatních znacích. Formou znaku rozumíme popis znaku konkrétního organismu používaný pro srovnání s jiným. Pod znakem si můžeme představit například horní končetinu obratlovců a tento znak může nabývat forem – ruka, přední noha, křídlo, ploutev nebo nepřítomnost. V případě DNA je znakem nukleotidová pozice X v genu Y, která může nabývat forem A, C, T, G nebo nepřítomnost.

V závislosti na tom, kdy v kladogenezi daná forma znaku vznikla, můžeme rozlišit několik kategorií, které jsou různě významné pro rekonstrukci fylogeneze. U společného předka taxonů A-F nabývaly znaky \underline{x} , \underline{y} a \underline{z} forem $\underline{x1}$, $\underline{y1}$ a $\underline{z1}$. Jsou to původní formy znaků celé skupiny a nazýváme je **plesiomorfie**. Přítomnost původních forem znaků u taxonů A a F (tzv. **synplesiomorfie**) nemá význam pro rekonstrukci fylogeneze a neznamena, že by taxony A a F měly svého exkluzivního společného předka, který nebyl předkem ostatních. Jedná se totiž o původní formu znaku, která se na větvích vedoucích k taxonům B-E se změnila. Větve, které si tyto znaky uchovaly v původní podobě, si nemusí být vzájemně příbuznější než ostatní, tak jako ryby a obojživelníci si nejsou příbuzní více než každá s těchto skupin s ostatními tetrapody, přestože sdílí podobné uspořádání vaječných obalů. Toto původní uspořádání vaječných obalů je plesiomorfni znak. Stejně tak některá jednobuněčná

eukaryota jsou příbuznější mnohobuněčným skupinám než jiným jednobuněčným eukaryotů. Jednobuněčnost u eukaryot je opět synplesiomorfie. Tvorba taxonů na základě synplesiomorfii vede k vytváření parafyletických taxonů.

Znak x2 je **apomorfie** vzniklá u společného předka BCD, nazýváme jej proto **synapomorphií** skupiny BCD. Tento znak je významný pro rekonstrukci fylogeneze, protože ukazuje, že BCD měly společného předka, u kterého tento znak vznikl. Příkladem mohou být naopak vaječné obaly amniot, které jsou synapomorphií plazů, ptáků a savců a vznikly u jejich společného předka. Synapomorfie nám definují monofyletické taxony. Znak y2 vznikl jen u taxonu B, nazýváme jej **autapomorfii** a nemá význam z hlediska kladogeneze (pořadí větvení), má však význam z hlediska anageneze tak, jako peří ptáků nemá význam pro rekonstrukci vztahů mezi třídami obratlovců. Znak z2 se vyvinul nezávisle na sobě u taxonů D a E. Tento znak nazýváme **homoplázie** neboli **konvergence**. Z hlediska rekonstrukce fylogeneze je nejen nepoužitelný, ale dokonce zavádějící, protože navozuje dojem, že taxony D a E měly společného předka, u kterého by tento znak vznikl jen jednou. Nezávislý opakovaný vznik téhož znaku je totiž obvykle považován za málo pravděpodobný. Příkladem může být třeba endotermie savců a ptáků, která u těchto skupin vznikla nezávisle. Jejich společný předek nebyl endotermní. Používání homoplázií v taxonomii je nebezpečné, protože vede k tvorbě polyfyletických taxonů.

Existují dva základní přístupy (názory), jak používat molekulární (ale i jiné) znaky pro rekonstrukci fylogeneze. **Fenetický přístup** se nezajímá o to, do jaké z výše uvedených kategorií znaky spadají. Při porovnávání organismů používá všechny znaky. Proto se říká, že fenetici analyzují podobnost mezi organismy. Na fenetickém přístupu je založena numerická taxonomie, která se stala populární v 60 letech minulého století. Jejimi zakladateli byli především Sneath a Sokal. Numerická taxonomie byl první pokus o objektivizaci taxonomie. Numeričtí taxonomové kladli důraz na použití velkého množství dat a vyvinuli matematické postupy, jak z nich vypočítat celkovou podobnost (nebo naopak odlišnost - distanci) mezi taxony.

Druhým přístupem je **kladistický**, který pro rekonstrukci fylogeneze používá výhradně synapomorfie. Tento přístup je v principu správnější, ale v praxi naráží na podobné problémy jako fenetický přístup, protože synapomorfie nelze velmi často odlišit od ostatní typů znaků. Metody rekonstrukce fylogeneze označované jako fenetické (založené na distancích) byly nebo jsou kladisty neprávem zavrhovány. Kladistické metody (maximální parsimonie) se ovšem v praxi dostávají do podobných obtíží, nemají vodítko, jak rozeznat homoplázie a konflikty mezi znaky řeší nakonec podobně jako „fenetické“ metody.

Důležitým pojmem v taxonomii je pojem **homologie**. Homologické jsou takové struktury, vzorce chování nebo geny, které mají jednotný původ, tj. jsou odvozeny od téže struktury společného předka. V případě genů nebo úseků DNA je situace jednoduchá a homologické jsou takové geny či nukleotidy, které vznikly přímým kopírováním téhož genu (nukleotidu) společného předka. U morfologických struktur nelze mluvit i kopiích, protože např. přední končetiny obratlovců se nekopírují do dalších generací. Kopíruje se opět genetická informace, na základě které se potom tyto struktury staví. Obecně lze tedy homologii definovat následovně:

„Homologie jsou podobnosti mezi komplexními strukturami nebo vzory, které jsou způsobeny kontinuitou biologické informace. (Riedl a Hazsprunar)

Synapomorfní i symplesiomorfní formy znaků jsou zároveň také homologické. Naopak homoplázie (konvergence) homologické nejsou a hovoříme o analogiích. Často situace není tak jednoduchá a je třeba si upřesnit kontext. Příkladem mohou být křídla ptáků, netopýrů a pterodaktylů. Jsou to homologické struktury nebo ne? Zde je opět potřeba rozlišovat mezi znakem a jeho formou. Jako *znak* (přední končetina obratlovců) jsou beze sporu homologické, kosti, které je tvoří, jsou také vzájemně homologické, ale *forma tohoto znaku* (přední končetina adaptovaná k letu) homologická není, protože se adaptovala nezávisle a jedná se o homoplázie. Společný předek měl sice přední končetinu nikoli však ve formě křídla. Rozpoznat homologické znaky je pro taxonoma i fylogenetika (molekulárního i klasického) velmi důležité, protože z hlediska fylogenetiky a taxonomie je smysluplné porovnávat formy pouze homologických znaků.

Homologický vztah znaků lze rozpoznat na základě odpovídající polohy a vzájemné podobnosti ve stavbě. Vřetení kosti v ploutvi velryby, horní končetině člověka a křídle ptáka jsou tvarově velmi odlišné, a kdybychom je samotné položili vedle sebe, těžko bychom rozpoznali, že se jedná o homologní kosti. Díky jejich poloze v končetině a návaznosti na ostatní kosti však o jejich homologii nepochybuje. V případě sekvencí genů, potřebujeme pro rozpoznání homologických pozic provést „**alignment**“ sekvencí. To znamená položit si nad sebe sekvence tak, aby si nukleotidy ve sloupcích byly co nejpodobnější. Až v kontextu takového alignmentu, jehož správnost naznačuje velké množství pozic distribuovaných po celé jeho délce, můžeme doufat, že nukleotidy ve sloupcích jsou homologní. Jinak bychom na to nebyli schopni přijít, protože nukleotidy jsou příliš jednoduché znaky vyskytující se jen ve čtyřech formách, a proto je u nich velké riziko konvergence (nehomologní znaky vypadají stejně). I tak si někdy musíme pomoci porovnáním sekundárních struktur RNA nebo proteinů. Motivy na těchto strukturách, pokud jsou důležité pro funkci, jsou v průběhu evoluce neměnné, nebo se mění velmi pomalu, a lze předpokládat, že nukleotidy nebo aminokyseliny, které je tvoří, si jsou homologické. Nukleotidy v homologické pozici jsou *homologické znaky*, všechny přece vznikly kopírováním, byť nepřesným, nukleotidu společného předka. I homologické morfologické znaky se mohou měnit a nabývat různých forem, někdy hovoříme o tom, že vytvářejí transformační série. Jen homologické znaky je smysluplné porovnávat, a proto je alignment prvním a zásadním krokem molekulárně fylogenetických analýz.

Homologický vztah mezi znaky je možné vyvrátit nebo minimálně zkomplikovat, pokud nalezneme organizmus, který nese více forem jednoho znaku. Homologický vztah mezi křídlem ptáků a horní končetinou člověka by byl zkomplikován, pokud by existovali andělé s rukama a křídly. Pak bychom se mohli domnívat, že předek čtvernožců měl vlastně tři páry končetin (dva páry kráčivých a jedna křídla). Křídla ptáků by pak byla homologická křídlym andělů, lidské ruce zase jejich rukám, u ptáků bychom přední kráčivé končetiny považovali za ztracené, u nelétavých savců by byla zase hypoteticky ztracená křídla. Ukázalo by se, že přítomnost těchto struktur nekopíruje věrně fylogenezi organismů a jejich použitelnost pro stanovování příbuzenských vztahů mezi organizmy by tím byla omezena. Přesto by nebylo zcela správné tvrdit, že si taková křídla a kráčivé končetiny nejsou homologické. Samozřejmě že by byly, vždyť i přední a zadní končetina jsou homologické struktury. Tělní plán živočichů

(zejména článkovců) se sestaven z vzájemně homologických modulů. Tento typ homologie ovšem nekopíruje fylogenezi, pro fylogenetiku je tedy nepoužitelný a někdy se označuje termínem homonomie.

Molekulární fylogenetici se setkávají s podobnými případy také velmi často, jedná se o jev, který nazýváme paralogie (více o termínu paralog a ortolog se dozvíte v některé z následujících přednášek). Všechna eukaryota používají při elongaci translace určitý GTP štěpící enzym. U některých se tento enzym nazývá elongační faktor-1 α . Jiná eukaryota používají sekvenčně podobný enzym zvaný elongation factor like (EFL). Existují eukaryota, která mají oba, a proto je jasné, že se jedná o dva paralogy – geny, které vznikly genovou duplikací jedné původní GTPázy v dávné minulosti. Směs takových paralogů nelze používat na rekonstrukci fylogeneze organismů.

Dosud jsem se zabýval jen homologii znaků. V okamžiku, kdy jsme si ujasnili, které znaky jsou homologické, tak přistoupíme ke srovnávání jejich forem. V molekulární fylogenetice to znamená, že začneme porovnávat nukleotidy či aminokyseliny v jedné pozici. **Nalezneme-li u víc organismů v jedné pozici ten samý nukleotid (např. adenin) budeme stát před otázkou, zda je tato forma znaků homologická** (předek měl adenin) nebo zda některé z nich vznikly konvergencí (předek neměl adenin). Jak rozhodnout? Odpověď na tuto otázku je nesmírně těžká a statistické metody molekulární fylogenetiky se snaží s různým úspěchem odhadnout pravděpodobnost obou scénářů. Kladistika nám v této situaci nabízí na pomoc tzv. **auxiliární princip**, který říká, že stejné (nebo velmi podobné v případě morfologie) znaky máme považovat za homologické, pokud není důvod se domnívat opak. Homologie je v takovém případě nulová hypotéza. Důvodem pro opačnou domněnku je nejčastěji skutečnost, že námi zkoumaný znak svou distribucí forem mezi taxony odporuje velkému množství jiných nezávislých znaků, tj. je přehlasován.

Přítomnost velkého množství polymorfismu na úrovni DNA (dnes víme, že dva lidé se liší zhruba ve 3 miliónech bazí) nejen v nekódujících, ale i v kódujících oblastech genomu, bylo z počátku velmi překvapivé. Vždyť kdyby tyto různé alely genů zvyšovaly nebo snižovaly fitness, byly by přírodním výběrem rychle eliminovány nebo naopak fixovány a žádný polymorfismus bychom v daném místě nepozorovali. Existenci velkého množství polymorfismů vysvětluje **teorie neutrální evoluce** (Kimura 1991). Podle této teorie drtivá většina znaků neovlivňuje fenotyp, a je proto “neviditelná” pro přírodní výběr – jsou selekčně neutrální. Frekvence takových znaků (alel) v populaci je ovlivňována výhradně náhodou – genetickým driftem. Genetický drift je způsoben tím, že do následující generace se dostane konečný počet náhodně vybraných alel z rodičovské populace. Frekvence alel během generací tak náhodně fluktuuje, dokud se jednou nesníží na 0, nebo nezvýší na 100. V takových případech mluvíme o vymizení alely nebo o její fixaci. Genetický drift funguje rychleji v malých populacích. V malých populacích frekvence alel rychleji dosáhne jednoho z extrémů (0% nebo 100%). V takových populacích proto vliv driftu převáží i nad působením přírodního výběru v případě mírně selekčně výhodných nebo mírně selekčně nevýhodných mutací (**nearly neutral theory**). Protože drift je u velkých populací pomalejší než selekce, přetrvávají obě varianty v populaci po dlouho dobu a my je detekujeme jako polymorfismus.

Aby bylo jasno, neutrální teorie netvrdí, že většina genů je postradatelná a mohou být zaplaveny mutacemi, ale tvrdí, že mutace v genech se v populaci udržují a někdy i fixují

proto, že většina forem (alel) téhož genu je funkčně stejně dobrá. Neutrální teorie netvrdí, že neexistují mutace se škodlivým efektem, které jsou eliminovány přírodním výběrem. Netvrdí ani, že neexistují místa v DNA, kde téměř každá mutace je škodlivá, takže je eliminována přírodním výběrem (říkáme tomu čistící selekce) a sekvence v takovém místě se nemění - je **sekvenčně konzervované** (např. aktivní místo enzymu). Tvrdí však, že škodlivých mutací a sekvenčně konzervovaných míst je menšina. Neutrální teorie samozřejmě nezavrhuje darwinistickou adaptivní evoluci poháněnou přírodním výběrem, ale tvrdí, že většina mutací je pro přírodní výběr „neviditelná“ a k adaptivní evoluci nepřispívá.

Aby to nebylo tak jednoduché, může být osud alely ovlivňován také alelami v sousedství. Pokud na určitém lokusu vznikne selekčně výhodná alela, která se v populaci rychle zafixuje. Zafixují se s ní i neutrální alely v sousedství, které jsou s ní v silné vazbě. V takovém případě mluvíme o genetickém svezení se (genetic draft). Snížení polymorfizmu v lokusech kolem výhodné mutace se nazývá vymetení (selective sweep).

Molekulární znaky mají oproti jiným znakům mnoho podstatných výhod.

1. Jsou to znaky právě z té úrovně, kde vznikají evoluční novinky (mutace).
2. Obvykle víme docela dobře, jak se dědí.
3. Mnoho z nich nezávisí na prostředí - jsou stejné, ať rostlina roste na chudém substrátu a je tedy celkově menší nebo na bohatém substrátu a je celkově větší.
4. Jsou velmi často selekčně neutrální, nejsou ovlivňovány přírodním výběrem (viz. neutrální evoluce). Navíc dokážeme dost dobře odhadnout, které lokusy DNA jsou selekčně neutrální a které nikoli. U selekčně neutrálních znaků je nižší riziko konvergence jejich forem (analogie). Přesto i zde se konvergence musíme obávat. Konvergence je zde způsobena jen náhodou – s určitou pravděpodobností mohou dvě nepříbuzné sekvence nebo pozice v sekvencích mutovat do stejné podoby. V případě znaků ovlivňovaných přírodním výběrem může kromě náhody, konvergenci způsobit právě přírodní výběr a přizpůsobování se stejným podmínkám, či funkcím.
5. Je jich obrovské množství. Velikost genomů se pohybuje od $0,5 \cdot 10^6$ – $600 \cdot 10^9$. Lidský genom obsahuje přes 3 miliardy párů bazí. Odhaduje se, že lidé se mezi sebou liší v 0,1% tj. 3 miliónech bazí. Např. genom *C. Ventera* publikovaný v roce 2007 se od prvního publikovaného lidského genomu liší ve 3,2 miliónech jednonukleotidových záměnách a asi 900 000 strukturálních variací (drobných přestavbách). Genom J. D. Watsona se od referenčního genomu lišil v 3,3 miliónech jednonukleotidových substitucí, strukturálních variací bylo nalezeno méně.
6. Jsou použitelné na všech úrovních taxonomie. Od porovnávání jedinců v rámci populace až po rekonstrukci velmi hluboké fylogeneze. Záleží jen na tom, jaký lokus si zvolíme.
7. Dají se jednoznačně popsat, protože nabývají několika diskrétních stavů (4 nukleotidy, 20 aminokyselin). Jsou v podstatě digitální.
8. Jednotlivé znaky jsou na sobě obvykle nezávislé. To u morfologických neplatí. Velikost orgánů je závislá na velikosti těla, duté kosti, křídla a mohutné prsní svaly vznikly společně jako přizpůsobení k letu a podobně.
9. Jsou lépe vážitelné, protože neuděláme velkou chybu, když jim přisoudíme stejnou váhu. Váha morfologických znaků může být dost rozdílná (znaky na kostře vs. zbarvení srsti).

10. Lépe se kvantifikuje stupeň nejistoty, protože můžeme statistiku založit na velkém množství selekčně neutrálních znaků se stejnou váhou.
11. S jistým stupněm opatrnosti můžeme vybrané molekulární znaky používat jako molekulární hodiny, které nám po důkladné kalibraci mohou poskytnout odhady stáří.
12. Některé znaky na molekulách, a tentokrát se nejedná o jednonukleotidové záměny ale spíše o větší mutace, vložení transpozonů či retropozónu, přestavba chromozomu, fúze genů nebo jejich duplikace, se velmi pravděpodobně odehrávají jen jedním směrem nebo lze s poměrně velkou jistotou poznat, jaký stav je původní – lokus s nevloženým transpozonem, nefúzovaný gen, neduplikovaný gen, chromozom s obvyklou strukturou jakou vidíme u příbuzných skupin. Takové znaky nám umožní fylogenezi polarizovat – organizmy s odvozenou formou znaku měly společného předka (tvoří monofylum, klád), organizmy s původní formou znaku se odvětvily před tímto společným předkem a leží tedy blíže kořenu stromu.
13. Srovnáváním molekulárních polymorfizmů různých jedinců v populaci můžeme usuzovat na vlastnosti této populace. Například jaká je její efektivní velikost (N_e), jestli je panmiktická nebo složená ze subpopulací, které mezi sebou vzájemně geneticky nekomunikují, jestli je inbrední nebo jestli v minulosti prošla obdobím, kdy se její velikost výrazně zmenšila – hrdlo láhve.

Samozřejmě, že molekulární znaky mají také nevýhody

1. Neposkytují informaci o anagenezi. Z výše uvedeného vyplývá, že většina molekulárních znaků se vůbec neprojeví na fenotypu, takže nijak nesouvisí s anagenezí organismu. I když by se nejednalo o selekčně neutrální znak, téměř nikdy nevíme, jak konkrétně se daný molekulární znak na fenotypu projeví. Z molekulárních znaků samotných nelze poznat, zda je jeho nositel ještě plaz nebo už pták či savec.
2. I když se náklady rok od roku snižují, získávání molekulárních znaků je v průměru stále ještě dražší než získávání znaků morfologických.
3. Při získávání molekulárních znaků je někdy nutné organismus nebo jeho část nenávratně zničit.

SEKVENACE DNA

Protože znaky, které molekulární taxonomie a fylogenetika používá, vznikají mutacemi DNA, je sekvenace DNA tou základní úrovní, kterou molekulární taxonomové studují. Jiné metody pro detekci, které si představíme ve 4. Přednášce (isozymy, restriční analýzy, mikrosatelity) představují jen způsoby jak odhalit některé polymorfismy mezi organizmy rychleji a levněji. Pokud bychom však u daných organismů znali sekvenace jejich genomů, všechny tyto další metody by byly zbytečné. Hluběji než k sekvenci DNA v molekulární taxonomii jít nelze.

Klasickou metodou sekvenace DNA, která opanovala toto pole od 70. let do začátku nového tisíciletí je **Sangerova dideoxy metoda**. I dnes je nepostradatelná, pokud potřebujeme osekvenovat jeden určitý úsek DNA. Pro Sangerovu metodu je nejprve potřeba připravit úsek DNA, který chceme sekvenovat, v mnoha identických kopiích – amplifikovat jej. K tomu se používá metoda **PCR** (polymerase chain reaction).

Princip PCR najdete na snímku 27. Templátem je obvykle genomová DNA a pro ohraničení úseku, který hodláme neamplifikovat, se používají 15-30 nt dlouhé oligonukleotidy (), tzv. primery. Jeden primer je komplementární s + řetězcem, druhý s - řetězcem templátu a oba mají 3' konce nasměrovány dovnitř úseku, který ohraničují. PCR probíhá v teplotních cyklech (30-40 opakování), které zahrnují denaturaci templátu (95°C), nasednutí primerů (40-70°C), a polymeraci (60-72°C podle použité polymerázy). Aby DNA polymeráza vydržela mnoho cyklů v těchto krutých podmínkách, musí být termostabilní (např. Taq polymeráza).

1. V 1. cyklu dojde k denaturaci templátu (gDNA), nasednutí primerů a polymeraci. Vznikají tak molekuly, u nichž je jeden řetězec tvořený templářem gDNA a druhý nově nasyntetizovanou DNA, která je ohraničena primerem na jednom konci a je tak dlouhá, kolik polymeráza zvládla nepolymerovat.
2. V druhém cyklu dojde opět k denaturaci všech templátů, nasednutí primerů a polymeraci. Vzniknou jednak tytéž molekuly jako v prvním cyklu (ty budou vznikat v každém dalším cyklu). Kromě toho však vzniknou také molekuly ohraničené primery na obou stranách.
3. Ve třetím a dalších cyklech se děje stále to samé. Vždy se z původního templátu gDNA nepolymerují řetězce ohraničené z jedné strany. Z jednostranně ohraničených a oboustranně ohraničených řetězců se napolymerují nové oboustranně ohraničené řetězce. Protože jednostranně ohraničené řetězce vznikají pouze z původního templátu DNA vznikne jich v každém cyklu množství omezené počtem kopií daného úseku v gDNA. V případě 1000 kopií jich tedy za 30 cyklů vznikne 30 000. Oboustranně ohraničené řetězce vznikají z jednostranně ohraničených, ale hlavně jsou templátem i samy pro sebe a proto se jejich počet v každém cyklu více než zdvojnásobí. Pokud jich v 2. cyklu vzniklo 1000, bude jich na konci třicátého cyklu $2,6 \cdot 10^{11}$. Je zřejmé, že tyto ohraničené řetězce zcela převáží nad ostatní DNA.

Pokud se v našem vzorku namnožil jeden typ oboustranně ohraničených řetězců, je vzorek rovnou použitelný pro sekvenaci. Pokud se ve vzorku namnožilo více různých úseků je potřeba je nejprve klonovat a namnožit si tyto úseky samostatně. To se dělá pomocí bakterií. Nejprve zaklonujeme směs úseků do plasmidů. Těmi natransformujeme bakterie (do jedné bakterie se dostane jen jeden plasmid). Poté, co vyselektujeme a namnožíme jeden bakteriální klon (kolonii), získáme z ní klon plasmidu obsahující klon našeho úseku. Plasmid můžeme použít na sekvenaci.

Při sekvenci Sangerovou metodou dochází opět k cyklické polymeraci, jako v případě PCR. Templátem je náš namnožený úsek (šedá molekula nahoře). Na rozdíl od PCR je polymerace zahajována jen od jednoho primeru a nedochází tedy k namnožování úseků geometrickou řadou, řetězce se syntetizují vždy jen na základě námi přidaného původního templátu. Vtip této metody spočívá v tom, že do směsi dNTP, které polymeráza potřebuje k syntéze, přidáme určité množství ddNTP. ddNTP jsou nukleotidy, které nemají 3'-OH skupinu a není možné na ně navázat další nukleotid. Zařazením ddNTP polymerace končí. Na obrázku je ukázáno, co se stane, pokud do reakce přidáme kromě dATP, dCTP, dTTP a dGTP určité množství ddGTP. Tu a tam se na konec prodlužujícího řetězce zařadí ddGTP a tím se polymerace zastaví. Získáme směs řetězců různých délek, které jsou na jedné straně ohraničeny primerem a na druhé končí ddGTP. Kdybychom si je seřadili podle délky a určili jejich délky, zjistili bychom, ve kterých pozicích, počítáno od primeru, je v templátové DNA

nukleotid C. Pokud do reakční směsi přidáme určité množství dd formy od každého nukleotidu a každou z nich označíme jinou fluorescenční barvičkou, získáme směs řetězců různých délek a různých barev, které ukazují, jaký je jejich terminální nukleotid. Pokud si vniklé řetězce seřadíme podle délek, můžeme z barev odečíst pořadí nukleotidů. Řazení podle délek se provádí kapilární elektroforézou. Tenká kapilára je naplněna gelem a připojena k elektrickému poli. Řetězce DNA migrují ke kladnému pólu, čím je řetězec kratší, tím migruje rychleji. V určitém místě je umístěn laser excitující fluorescenci barviček a detektor zaznamenávající barvu fluorescence. Na chromatogramu sledujeme, jak tímto místem postupně prochází barevné signály od nejkratších molekul až po ty nejdelší a čteme sekvenci.