

## MOLEKULÁRNÍ TAXONOMIE - 4

V této přednášce si představíme metody, které získávají molekulární znaky bez použití sekvenace.

**Všechny tyto metody je teoreticky možné sekvenací nahradit.** Oproti sekvenaci celých genomů totiž neposkytnou žádnou informaci navíc. Jejich **výhodou je však rychlost a láce.** Výstupy těchto metod lze navíc okamžitě dále zpracovávat relativně jednoduchými zavedenými postupy, kdežto analýza celogenomové sekvenace je výpočetně i mentálně dost náročná.

### DNA - DNA hybridizace

Tato metoda pracuje s předpokladem, že komplementární řetězce DNA spolu drží lépe (denaturují za vyšších teplot) než řetězce, které si nejsou přesně komplementární. Změříme-li střední teplotu tání ( $T_m$ ), tj. teplotu, při které polovina dvojřetězců ještě drží a polovina je denaturovaná, u zcela komplementárních dvojřetězců (tzv. homoduplexů) a  $T_m$  dvojřetězců tvořených jedním řetězcem ze vzorku A a druhým řetězcem ze vzorku B (tzv. heteroduplexů) můžeme z rozdílu mezi  $T_m$  a  $T_{ms}$  usoudit v jakém procentu nukleotidů se vzorky A a B liší.

Technicky se  $T_{ms}$  stanovuje například tak, že se smíchá malé množství značené DNA (radioaktivně či jinak) jednoho vzorku, této DNA říkáme *tracer*, s velkým množstvím neznačené DNA druhého vzorku, tzv. *driver*. Směs se nejprve denaturuje, aby se řetězce oddělily, a pak se nechá zrenaturovat. Uvedené uspořádání pokusu nám zaručuje, že značená DNA, kterou můžeme specificky detekovat, renaturuje výhradně s DNA z druhého vzorku, která je ve velkém nadbytku, a ne sama se sebou. Po renaturaci se vzorek DNA nechá protéct hydroxyapatitovou kolonou, kde se zachytí jen dvojřetězcovou DNA, kdežto nezrenaturované jednořetězce protečou. Poté se postupně zvyšuje teplota a měří se množství značené DNA, které se uvolňuje z kolony.  $T_{ms}$  je teplota, při které se uvolní polovina značené DNA.

Z naměřených hodnot můžeme vypočítat  $p$  (podíl rozdílných nukleotidů) mezi vzorky A a B následujícím způsobem:

1) Stanovíme průměr střední teploty tání homoduplexů vzorků A a B

$$T_m = (T_{mA} + T_{mB})/2$$

2) Stanovíme střední teplotu tání heteroduplexu ( $T_{ms}$ ) postupem uvedeným výše.

3) Vypočítáme rozdíl mezi  $T_m$  a  $T_{ms}$ .

$$\Delta T_m = T_m - T_{ms}$$

4) Z rozdílu vypočítáme  $p$  (procento rozdílných nukleotidů)

$$p = \Delta T_m * 0,01 \text{ (případně } 0,015)$$

Metodu DNA-DNA hybridizace proslavily zejména práce Charlese Sibleyho a Jona Ahlquista, kteří ji v 70. a 80. letech použili na rekonstrukci fylogeneze ptáků a posléze primátů. Výhodou této metody je její totilokusový charakter (porovnává celé genomy) a to že je použitelná od blízkých po poměrně vzdálené taxony, na velmi vzdálené však vhodná není. Jejím nevýhodou je, že poskytuje pouze genetické distance (procento rozdílných nukleotidů mezi vzorky) a ne konkrétní znaky. Další nevýhodou je značná experimentální náročnost. Ta spočívá nejen ve vlastním měření, ale také ve skutečnosti, že je nutné porovnávat všechny dvojice analyzovaného souboru dat. Pokud máme v souboru  $N$  vzorků, musíme provést  $N*(N-1)$  experimentů. Velkou nevýhodou je, že výsledek této

metody může být velmi ovlivněn přítomností repetitivních sekvencí v genomech. To je problém, zejména u eukaryotických genomů, které jsou obvykle plné nekódující DNA, z níž je velké procento repetitivních. Podobnost v těchto úsecích často neodráží příbuznost organismů a je proto zavádějící. Existují sice postupy, jak repetitivní sekvence ze vzorku odstranit, nefungují však stoprocentně.

DNA-DNA hybridizace se stále používá v bakteriální taxonomii, kde je dosud považována za zlatý standard. Za druhovou hranici bývá považovaný pokles  $T_m$  o 30%, který odpovídá zhruba 5% rozdílu nukleotidů. Vzhledem k technické náročnosti metody se ovšem používá až jako druhý krok následující po sekvenaci 16 rRNA. Pokud se 16S rRNA gen studované bakterie liší o více než 3 % bazí od známých sekvencí, je považován za nový druh. Pokud se liší o 3 % a méně, je potřeba přistoupit k DNA-DNA hybridizaci. V době, kdy se sekvenace bakteriálního genomu stala téměř rutinní a levnou záležitostí, je pochopitelná snaha nahradit pracnou DNA-DNA hybridizaci, nějakým indexem podobnosti mezi dvojicí genomů vypočtených z celogenomového srovnání. Tento přístup bezesporu brzy nahradí DNA-DNA hybridizaci. V současnosti bylo navrženo několik metod poskytujících míru odlišnosti genomů. K výpočtu některých z nich se vrátíme na následující přednášce.

- **ANI** (average nucleotide identity) – průměrná identita všech homologních úseků rozpoznaných pomocí BLASTN. Postupuje se tak, že se jeden genom rozseká na úseky dlouhé 1020 pb a tyto se BLASTují proti druhému genomu. U každého úseku, který najde odpovídající oblast se spočítá identita s nalezenou oblastí. Identity se zprůměrují. Hranice bakteriálního druhu leží mezi 95-96% (Goris a kol. 2007).  
<https://www.ezbiocloud.net/tools/ani>
- Rozdíl ve **frekvenci k-merů (X nukleotidových „slov“)** mezi porovnávanými genomy (Richter a Rosseló-Móra 2009).
- **GBDP** – průměrná genetická vzdálenost vypočtená z celkového alignmentu dvou genomů. Pod celkovým alignmentem se rozumí souhrn alignovatelných oblastí (např. HSPs, „high-scoring segment pairs“, nalezených BLASTem), nikoliv alignment genomů od začátku do konce. Hranice bakteriálního druhu je stanovena přibližně na hodnotě 0,26. Další čtení: Meier-Kolthoff a kol. 2013, Auch a kol. 2010. <http://ggdc.dsmz.de/>
- **MUMi** – podíl oblastí s dokonalou shodou na celkové délce porovnávaných genomů.

Dále lze DNA-DNA hybridizaci použít pro odhad biodiverzity vzorku. Vzorek bakteriální DNA izolované z prostředí, který je tvořen jedním dominantním druhem, bude vykazovat vyšší  $T_m$  než vzorek tvořený směsí mnoha nepříbuzných druhů. I v tomto případě sekvenace metagenomu (směsného) tohoto směsného vzorku poskytne přesnější a úplnější informace a DNA-DNA hybridizaci plně nahradí.

### Izoenzymová analýza

Tato metoda se v dnešní době příliš nepoužívá pro účely molekulární taxonomie, protože byla nahrazena lepšími metodami (mikrosatelity, SNP). V minulosti však byla hojně využívána a jednalo se o jednu z prvních metod pro studium vnitrodruhového polymorfismu. Pro úplnost zde uvádím stručný popis této metody. Na úvod několik pojmů

Isozymy – enzymy vykazující stejnou nebo podobnou aktivitu.

Alozymy – podmnožina izozymů, alelické formy téhož enzymu (leží v jednom lokusu)

Zymodem -taxonomická jednotka vymezená na podkladě izozymového vzoru

Zymogram - rozložení zón a tedy i jednotlivých enzymů na elektroforetogramu

Princip spočívá v tom, že různé izozymy vykazují různou elektromobilitu, protože se liší sekvencí aminokyselin. Pokud rozdělíme tyto izozymy na elektroforéze, získáme elektroforetický vzor pruhů. Čím podobnější vzory nám vzorky poskytnou, tím jsou si geneticky podobnější. Statistickému zpracování těchto obrazců se budeme věnovat v další přednášce. Několik dalších metod, o kterých pojednává tato přednáška, se vyznačuje tvorbou elektroforetických vzorů. Těmto vzorům se někdy říká fingerprinty (otisky prstů) a metodám molekulární fingerprinting.

Jednoduchost isoenzymové analýzy spočívá v tom, že není potřeba ze vzorku separovat isoenzymy, které hodláme studovat. Na nativní elektroforézu se nanáší celý vzorek proteinů, které se společně rozdělí, ale pak detekujeme jen isoenzymy našeho zájmu. K detekci se využívá jejich specifické enzymatické aktivity, kterou spřáhneme s reakcí poskytující barevný produkt. Tak nám na elektroforetogramu vzniknou barevné pruhy v místech, kam naše isoenzymy doputovaly. Za dobu používání této analýzy byly vyvinuty detekční systémy asi na 40 enzymů.

### **Restrikční analýza (RFLP)**

Restrikční analýza je dalším typem fingerprintingu. Je založena na použití restrikčních endonukleáz - enzymů štěpících DNA specificky v sekvenci 4 nebo 6 nukleotidových DNA palindromů. Například první objevená restrikční endonukleáza EcoRI štípe vždy sekvenci GAATTC za nukleotidem G v tomto řetězci a za G komplementární s C na druhém řetězci. Při restrikční analýze se vzorek DNA naštípe zvolenou endonukleázou, vzniklé fragmenty se rozdělí na elektroforéze podle velikosti, takže vznikne vzor pruhů charakteristický pro daný vzorek. Název této metody se zkracuje RFLP (Restriction Fragment Length Polymorphism).

Problém je, že restrikční analýza velmi komplexního vzorku, např. eukaryotického genomu, který obsahuje velké množství unikátních sekvencí, poskytne velmi komplexní elektroforetický vzor, velké množství pruhů, které nakonec splynou v jednu šmouhu. Abychom dospěli ke vzorům, které je možné dále analyzovat, je potřeba používat pro restrikční analýzy méně komplexní vzorky (bakteriální nebo organelové genomy, PCR produkty) a dobře fragmenty rozdělít. Druhou možností je použít komplexní vzorek, ale na elektroforéze detekovat pouze některé fragmenty. Například fragmenty pocházející z mitochondriální DNA. V takovém případě se používá metoda Southern blotting. DNA rozdělená na elektroforéze se přeblotuje (přesaje) na membránu. Cílová DNA na membráně se poté zviditelní značenou sondou (próbou). V prezentaci je příklad, ve kterém byla přebloťovaná DNA detekovaná radioaktivně značenou sondou připravenou z mitochondriální DNA.

Asi nejoblíbenějším využitím RFLP je analýza počtu kopií tandemových repetitivních minisatelitů - VNTR (Variable Number of Tandem Repeats). Minisatelity jsou tandemové repetice s jednotkou opakování dlouhou 10-60 nukleotidů. V lidském genomu nalezneme takové repetice asi v 1500 lokusech.

Minisatelitové lokusy jsou velmi variabilní, co se týče počtu repetic v tandemu. Toho se využívá při restričních analýzách zaměřených obvykle na určování identity jedince nebo rodičovství.

Postupuje se tak, že se celková DNA naštěpí restriční endonukleázou, která neštěpí uvnitř minisatelitu, takže dostaneme velké množství fragmentů, z nichž některé budou obsahovat minisatelity. Naštípanou DNA přebloujeme na membránu a hybridizujeme se značenou próbou. Próba může být namířena čistě proti sekvenci minisatelitu a potom zviditelníme všechny lokusy toho konkrétního minisatelitu (v prezentaci schéma vpravo). Pokud bychom chtěli zviditelnit pouze jeden minisatelitový lokus, použijeme jako próbu sekvenci minisatelitu a části unikátní sekvence, která na ni navazuje (v prezentaci, dolní obrázek). V obou případech je podoba elektroforetického vzoru určována jednak délkou minisatelitových lokusů, ale také polohou nejbližších restričních míst.

Výhodami restričních analýz obecně je experimentální nenáročnost a láce. Výsledky lze dobře reprodukovat. Znaky se jeví jako kodominantní, tj. vidíme obě alely nejen tu dominantní. Znaky obvykle pocházejí z mnoha lokusů a jedná se obvykle o selekčně neutrální znaky. Nevýhody spočívají v potřebě relativně velkého množství DNA, která musí být navíc poměrně kvalitní (nedegradovaná) a čistá. Elektroforetogramy mohou být špatně čitelné. Znaky nemusí být nezávislé – ztráta jednoho restričního místa ovlivní více fragmentů (dva se spojí v jeden s jinou délkou). Přestože tyto metody poskytují znaky (fragменты o určitých délkách), obvykle se s nimi, jako se znaky nepracuje, ale převádějí se nejprve na genetickou vzdálenost (další přednáška) výjimku tvoří VNTR.

## **Mikrosatelity**

Mikrosatelity nazývané také STR (Short Tandem Repeat) jsou krátké sekvenční motivy (dinukleotidy až hexanukleotidy) vyskytující se na některých místech genomu v mnoha tandemově uspořádaných kopiích, např. "CACACACACA", o celkové délce až 150 opakování. Mikrosatelity velmi rychle mutují. Ke změnám v počtu opakování repetice dochází v daném lokusu průměrně 1x za 1000 generací, což je o několik řádů častěji, než vznikají jednonukleotidové mutace v jiných částech genomu. Proto se každý jednotlivý lokus mikrosatelitu obvykle vyznačuje velkým vnitropopulačním polymorfismem v délce. K prodlužování nebo zkracování počtu kopií dochází při replikaci DNA mechanismem sklouznutí templátu – polymeráza při tvorbě nového řetězce sklouzne po templátu vpřed nebo vzad, a tak nasyntetizuje o jednu nebo více jednotek repetice méně či více. Čím větší počet repetic daný lokus má, tím snadněji dojde ke sklouznutí. Delší lokusy jsou tedy polymorfnější – mají větší počet různých alel. Polymorfismus v délce mikrosatelitu v daném lokusu můžeme snadno studovat pomocí PCR amplifikace s primery specifickými na unikátní oblast v okolí mikrosatelitu. PCR produkty rozdělíme na elektroforéze a porovnáme jejich délky (viz obrázek v prezentaci). Elektroforéza musí být dostatečně citlivá, aby bylo možné rozlišit fragmenty lišící se jednou jednotkou repetice. V úvahu tedy připadá akrylamidová elektroforéza, ale v dnešní době se nejčastěji používá kapilární elektroforéza prováděná na sekvenátorech používaných k Sangerově sekvenaci. V takovém případě můžeme při jedné reakci skrínovat až 4 různé lokusy, pokud si příslušné primery označíme 4 různými fluorescenčními barvičkami (obrázek v prezentaci).

Pro mnoho organismů jsou již popsány lokusy s mikrosatelity včetně primerů na amplifikaci a v takovém případě je situace jednoduchá. Pokud bychom chtěli analyzovat mikrosatelity u nového organismu, musíme si nejprve takové lokusy najít a primery navrhnout. Při troše štěstí budou

fungovat primery od blízce příbuzného organismu, ale použitelnost mikrosatelitových primerů se vzrůstající genetickou vzdáleností rychle klesá. Hledání vhodných mikrosatelitních lokusů u “nového” organismu představuje obvykle nejobtížnější a nejriskantnější část projektu. Klasický postup při hledání mikrosatelitů byl přes přípravu genomové knihovny a skrínování této knihovny pomocí sondy komplementární s mikrosatelity. Ze sekvence klonu obsahujícího mikrosatelit obvykle získáme také sekvence unikátních oblastí, které jej ohraničují. Při tomto postupu je vhodné si nejprve vzorek DNA obohatit o sekvence mikrosatelitů hybridizací s mikrosatelitovou sondou. Dnes je možné také využít služeb komerčních laboratoří, které pomocí metod next generation sequencing nagenarují množství genomických dat pro daný organismus, ve kterých hledají vhodné mikrosatelity. Na internetu jsou také dostupné různé nástroje pro vyhledávání vhodného lokusu v genomových datech.

Mikrosatelity stále představují oblíbenou metodu pro studie vnitrodruhové variability. Jedná se v podstatě o přesnější obdobu alozymové analýzy. Jejich výhodou je vysoká reprodukovatelnost, kodominance znaků (vidíme obě alely), velké množství alel v lokusu, Mendelovská dědičnost těchto znaků. Navíc jsou to znaky, u kterých máme určitou představu, jak mutují. Při sklouznutí řetězce dojde nejčastěji k prodloužení nebo zkrácení o jednu repetici. Proto lze počítat s tím, že alely v jednom lokusu, které si jsou podobnější délkou, si jsou také příbuznější.

### **Random Amplified Polymorphic DNA (RAPD)**

Další metoda patří mezi DNA fingerprinting. Vzorek DNA je podroben PCR reakci s jedním nebo i více krátkými primery (8-12 nukleotidů), u kterých je předpoklad, že budou komplementární k mnoha místům v genomu. Lokusy, které obsahují komplementární místa pro primery dostatečně blízko sebe, budou naamplifikovány. PCR produkty rozdělíme na elektroforéze a výsledné vzory porovnáme. Jejich podobnost by měla odrážet genetickou podobnost vzorku. Obvykle nás nezajímá odkud získané PCR produkty pocházejí. Pro RAPD analýzu je potřeba nashromáždit větší množství znaků, tj. provést PCR z několika desítkami primerů. Pokud PCR neprodukuje dostatek pruhů, můžeme zvýšit počet znaků tak, že budeme kombinovat více primerů, že budeme výsledné produkty ještě navíc štěpit restrikcími endonukleázami nebo, že použijeme primery komplementární s repetitivními elementy (např. transpozony), o kterých víme, že je jich v genomu velké množství. Primery ovšem musí směřovat ven z repetitivního elementu, aby amplifikovaly unikátní oblasti v jejich sousedství. Varianty využívající retrotranspozony se označují různými zkratkami: IRAP – oba primery nasedají na LTR, REMAP – druhý primer nasedá na mikrosatelity a R-RAP – druhý primer je náhodný jako u RAPD. Důležité je dbát na vysokou kvalitu elektroforetogramů, aby bylo možné odlišit produkty o nepatrně odlišných délkách.

Výhodou RAPD metody je nízká cena a rychlost, univerzální použitelnost, potřeba malého množství DNA, která nemusí být příliš kvalitní (může být degradovaná), ale nesmí obsahovat kontaminaci cizorodé DNA. Metoda umožňuje získat velké množství dat, i když tato výhoda s klesající cenou strmě rostoucí výkonností sekvenčních technologií rychle bledne.

Nevýhodami RAPD je nízká reprodukovatelnost - PCR provedené za nepatrně odlišných podmínek, třeba na různých přístrojích a v různých laboratořích mohou poskytnout různé vzory. Proto je možné porovnávat jen vzory z vzniklé v tom samém pokusu na jedné určité elektroforéze.

Reprodukovatelnost je lepší u variant, které využívají retrotranspozony. Znaky, které RAPD poskytuje

jsou poměrně nevěrohodné - stejně dlouhé produkty mohou ve skutečnosti pocházet z různých míst. Z toho plyne vysoké riziko homoplázií. Podobné je to u restričních analýz. Tato data proto není možné analyzovat jako znaky, ale vzory se nejprve převedou na genetickou vzdálenost a dále se pracuje s touto vzdáleností. Tím se vliv homoplázií sníží. RAPD je použitelná jen pro příbuzné druhy, se zvyšující se genetickou vzdáleností se prudce snižuje počet společných znaků a roste riziko homoplázií. Znaky nejsou kodominantní, tj. z jednoho lokusu se obvykle amplifikuje jen jedna alela - ta která obsahuje komplementární místa nebo ta která se snáze amplifikuje snáze, protože je například kratší.

### **Amplified Fragment Length Polymorphism - AFLP**

Další z fingerprintigových metod, která je v podstatě kombinací RFLP a RAPD. Kombinuje výhody obou metod a minimalizuje jejich nevýhody. Je dobře reprodukovatelná s nízkým rizikem konvergence, potřebuje malé množství vzorku, produkuje velké množství znaků, ale vzory nejsou tak komplexní jako u RFLP. Nevýhodou je o něco větší experimentální náročnost, i když jsou k dispozici i kity na AFLP, které jsou ovšem zase dražší (15-30 tis.) než jen obyčejné primery nebo restriční enzymy.

Experimentální procedura začíná restrikcí vzorku DNA pomocí dvojice restričních endonukleáz, na obrázku v prezentaci jsou to endonukleázy EcoRI a MseI. Na fragmenty vzniklé restrikcí jsou v následujícím kroku naligovány krátké adaptéry. Fragmenty s adaptéry jsou posléze amplifikovány pomocí PCR s primery komplementární k adaptérům, ale přesahujícím dovnitř fragmentu o 1-3 nukleotidy. Amplifikace může probíhat ve více kolech. Tímto způsobem dojde k amplifikaci jen určité podmnožiny původních fragmentů, proto vzor není tak komplexní jako u RFLP. Poté jsou produkty velmi jemně rozděleny podle délek na akrylamidové elektroforéze nebo v kapilární elektroforéze. Je možné produkty značit fluorescenčními barvičkami navázanými na primery.

Díky vysoké reprodukovatelnosti jsou AFLP znaky použitelné nejen na analýzu příbuzenských vztahů mezi blízkými příbuznými druhy či jedinci v populaci, ale také, podobně jako mikrosatelity, pro určování identity či rodičovství. Na rozdíl od mikrosatelitů nepotřebujeme žádnou předchozí znalost o vzorku - nemusíme hledat lokusy a navrhnout primery. Znaky jsou opět kodominantní (vidíme obě alely).

### **Protein a DNA mass fingerprinting**

Tyto metody využívají velmi rozvinutou metodiku hmotnostní spektrometrie. Hmotnostní spektrometrie detekuje a rozlišuje různé molekuly na základě jejich pohybu v elektrickém poli. Měří například dobu letu (time of flight - TOF) ionizovaných proteinů nebo jejich štěpů. Protože tato doba letu závisí na hmotnosti a náboji molekuly, je pro identické molekuly stejná a jiné molekuly se v ní obvykle liší. Lze ji proto použít na jejich identifikaci. Hmotnostní spektrofotometr je nákladné zařízení (10 mil Kč), ale pokud k němu má taxonom přístup, může jej využít i na "mass fingerprintigové" metody, protože zpracování jednoho vzorku je snadné a levné. Princip je stejný jako u ostatních fingerprintingů, extrakt všech proteinů z organismu (tedy nejen isozymy) nebo jejich štěpů je rozdělen na hmotnostním spektrometru a vznikne obrazec – spektrum – ve kterém jsou vidět vrcholy nejvíce zastoupených molekul. Spektrum je specifické pro jedince či druh a lze jej použít pro identifikaci či pro zjišťování příbuznosti. Ukazuje se, že reprodukovatelnost výsledků je vysoká, ale pro jistotu je dobré provést více replikátů pro jeden vzorek. Podobnosti vzorů se analyzují metodami

shlukující analýzy (např. UPGMA v příští přednášce) nebo metodami jako je principal component analysis (PCA). Nevýhodou protein fingerprintingu je skutečnost, že fingerprinty jsou ovlivněny aktuálním fyziologickým stavem organismu, který se projevuje do skladby proteinů. Např. dva stejné kmeny bakterií vykazují různé fingerprinty pokud jsou kultivovány na různých médiích. Mass fingerprinting lze využít i pro analýzu vzorků nukleových kyselin - např. PCR produktů nebo jejich štěpů. Protože RNA se lépe ionizuje, přepisuje se při přípravě vzorků DNA na RNA. Porovnáním s referenční databází (genomem) můžeme identifikovat o jaký organismus se jedná, nebo porovnáváním vzorů odhadovat příbuznost vzorků.

### **Short Interspersed repetitive Elements (SINE)**

SINE jsou retroposony derivované z tRNA, případně 7 SL RNA (Alu element u člověka). Jejich velikost je 70-500 bazí a vytvářejí značnou část eukaryotického genomu, často až 10 000 kopií na genom. Jejich výhoda pro molekulární fylogenetiku spočívá v tom, že je jen malá pravděpodobnost, že by se vložily do stejného místa dvakrát nezávisle na sobě a je téměř vyloučeno, že by se beze stopy z daného místa vystříhly. Po jejich vystřížení zůstane v místě tandemová repetice. Přítomnost SINE v určitém lokusu u více organismů, lze tedy s vysokou pravděpodobností považovat za synapomorfii.

V analýze pomocí SINE postupujeme následovně. Pokud jsou pro naši skupinu organismů známy lokusy se SINE, využijeme je. Pokud potřebujeme nalézt nové lokusy můžeme pátrat po SINE v genomových sekvencích organismů. Pokud nejsou dostupné genomové sekvence na internetu, můžeme si náhodný vzorek genomové sekvence vygenerovat třeba metodami "next generation sequencing". Vzhledem k velkému počtu SINE lokusů v genomu (známo u savců), mělo by stačit získat sekvence několika desítek kilobází. Alternativně můžeme připravit genomovou knihovnu a skrínovat klony pomocí značené sondy proti sekvenci známého SINE. U nově objevených lokusů je dobré vyzkoušet na menším souboru taxonů, zda jsou na naší taxonomické úrovni dostatečně polymorfni. Jakmile dospějeme k sadě použitelných SINE lokusů, provedeme u každého z nich PCR amplifikaci pomocí primerů komplementárních s unikátní sekvencí, které obklopují SINE, a rozdělíme produkty na elektroforéze. Delší fragmenty zřejmě představují lokusy obsahující SINE element. Pro ověření, že PCR naamplifikovala námi zvolený lokus a že delší produkty skutečně obsahují SINE elementy, musíme PCR produkty osekvenovat nebo hybridizovat s próbami namířenými jednak proti unikátní sekvenci lokusu a jednak proti sekvenci SINE. Sekvenace je výhodnější v tom, že nám může odhalit případy, kdy došlo k vystřížení SINE a vzniku tandemové repetice v místě, odkud byl vystřížen.

Informaci o přítomnosti a nepřítomnosti SINE u taxonů v několika zkoumaných lokusech zaznamenáme do tabulky. Přítomnost SINE je vždy synapomorfie - SINE se do stejného místa nevloží 2x nezávisle na sobě a zároveň víme, že jeho nepřítomnost je původní stav. Proto se tato data velmi snadno interpretují a k sestavení stromu příbuzenských vztahů nám často postačí papír, tužka a metoda maximální parsimonie (o 3 přednášky dále).

### **Single nucleotide polymorphism - SNP**

Polymorfismus DNA, kdy se jedinci nebo druhy liší v jedné nukleotidové záměně

AAGCCTA

AAGCTTA

V tomto případě mluvíme o alelách C a T. Téměř všechny SNPy mají jen 2 alely, protože je málo pravděpodobné, že by v populaci konkrétní nukleotid zmutoval hned dvakrát. Genom dvou lidí se liší zhruba ve 3 mil. bází, ale ne všechno jsou SNP. Databáze SNP v rámci NCBI (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) k říjnu 2018 eviduje přes 6 miliard miliónů různých SNP u člověka.

SNP se využívají pro mapování genomu a hledání genů souvisejících s fenotypy, které nás zajímají (například choroby). Pokud přítomnost SNP v populaci přesně odpovídá výskytu choroby nebo s ní signifikantně koreluje, je zřejmé, že tento SNP je nějak svázan s genem, jehož porucha tuto chorobu vyvolává. Buď je tento SNP uvnitř tohoto genu a jedna jeho alela přímo způsobuje tuto chorobu nebo, a to je častější případ, SNP leží poblíž genu a je s ním v genetické vazbě. Pátrání po genech v okolí takového SNP může vést k odhalení problematického genu. V molekulární taxonomii mají SNP podobné využití jako mikrosatelity. Stejně jako mikrosatelity se Mendelovsky dědí, kombinace SNP je specifická pro jedince a ukazuje na jeho příbuznost s jinými jedinci v populaci. Proto můžeme SNP využít v identifikaci jedinců, určování rodičovství nebo v populačních studiích.

Pokud chceme nalézt SNP v genomech organismů, pro které zatím SNP známy nejsou, je nejjednodušší způsob osekvenovat genomy od více jedinců a porovnat je vzájemně. Problém je, že velké množství zdánlivých SNP budou ve skutečnosti chyby. Z tohoto důvodu je s výhodou používat metody sekvenace, které jsou méně citlivé na chyby typu substituce. Takovou metodou je SOLiD. Pokud nemůžeme investovat tolik prostředků do celogenomového sekvenování, můžeme postupovat tak, že si připravíme směsný vzorek DNA dvou jedinců a budeme sekvenovat náhodné části smíchaných genomů. V některých místech uvidíme dvojitý signál. Některé dvojitě signály budou opět chyby nebo nepřesnosti sekvenace. Některé mohou představovat SNP - místa, ve kterých mají dva použité vzorky DNA jiný nukleotid.

Pokud však hodláme genotypizovat na SNP jedince modelového organismu, pak máme k dispozici nepřehledné množství metod SNP genotypizace, které se neustále rozvíjejí a mění. Některé z nich si představíme. Výhodou mnoha z nich je, že nám v jenom kroku otypují obrovské množství SNP. Tyto metody jsou založeny na hybridizačních technikách, enzymatických reakcích nebo na jiných principech.

Hybridizační technikou je například Molecular beacon. U této metody se používají fluorescenčně značené próby, které mají uprostřed své molekuly místo přesně komplementární s lokusem obsahujícím konkrétní alelu SNP a na krajích mají sekvence komplementární navzájem. Na jednom konci próby je navázána fluorescenční barva na druhém konci tzv. quencher, který inhibuje fluorescenci. V "klidovém" stavu próba vytváří formu vlásenky s kličkou. U této formy jsou fluorochrom a quencher blízko sebe a sonda nesvítí. Pokud se dostane do kontaktu s komplementární DNA, tedy se "svou" alelou, její struktura se otevře a my můžeme detekovat fluorescenci. Pokud necháme vzorek DNA inkubovat s próbami na všechny známé alely SNP obarvené různými barvičkami, můžeme podle barevného signálu odečíst, která alela/alely se u jedince vyskytuje.

Hybridizační metody založené na mikroarray čípech dovolují skrínovat velké množství SNP najednou. Oligonukleotidy komplementární s lokusem, kde je známý SNP jsou imobilizovány na sklíčku těsně vedle a známe jejich polohu. Oligonukleotidy představující čtyři možné formy daného SNP, tj. lišící se v jednom nukleotidu představujícím SNP, jsou obvykle vedle sebe. Vzorek DNA je naštipán na krátké



fragmenty, fluorescenčně označen, hybridizován na čip a pak je odečten světelný signál. Čipy firmy Affimetrix umožňují naráz oskrínovat 906 tis. známých lidských SNP. Problém hybridizačních metod spočívá především v jemném vyladění podmínek hybridizace tak, aby próby na alely SNP vzájemně nekrosreagovaly. U hromadného skrínování je tento problém ještě větší, protože různé oligonukleotidy vyžadují různé optimální podmínky hybridizace. Proto je na mikroarray čipech každý SNP analyzován víckrát na různých místech čipu v rámci různých oligonukleotidů z jeho lokusu.

Z enzymatických metod si představíme metodu Infinium od firmy (Illumina). Na sklíčku jsou těsně vedle sebe do známých míst připevněny kuličky, na kterých jsou navázány oligonukleotidy. Každá kulička nese oligonukleotidy jednoho typu, které jsou komplementární se sekvencí známého SNP lokusu a jejich volný 3' konec končí o jeden nukleotid před SNP. Vzorek DNA se naštípe na náhodné fragmenty určitých délek, které se denaturují a hybridizují na sklíčko, takže fragment se SNP lokusem se zachytí na oligonukleotidu čouhající z kuličky. Následně dojde k polymeraci a DNA polymeráza prodlouží oligonukleotid navázaný na kuličce o jeden nukleotid podle templátu z DNA vzorku - jedná se právě o polymorfni SNP nukleotid. Pomocí fluorescenčně značených protilátek proti čtyřem možným nukleotidům se detekuje, jaký nukleotid, v případě heterozygota jaké dva nukleotidy, byly připolimerovány. Ten představuje SNP genotyp.

Starou a jednoduchou, avšak stále používanou metodou analýzy SNP polymorfismu v jednom lokusu je SSCP (Single Strand Conformation Polymorphism). SNP lokus amplifikujeme pomocí PCR. Produkty pak denaturujeme teplotou, aby se rozdělily na jednotlivé řetězce, a necháme je renaturovat. Odstraníme dsDNA a zbudou nám ssDNA řetězce, které renaturovaly samy se sebou, přičemž vytvořily komplexní 3D struktury. Ty rozdělíme na elektroforéze. Rychlost jejich migrace je dána ani ne tak délkou, jako tvarem, který zaujmou. Na tvaru těchto 3D strukturách, a tedy i na jejich elektromobilitě, se projeví i substituce v jednom nukleotidu, která by elektromobilitu dsDNA neovlivnila. U homozygotů nalezneme dva pruhy, každý odpovídá jednomu vlákně DNA. U heterozygotů 4 pruhy.