

Titulna strana 1
TECHNICKÁ UNIVERZITA VO ZVOLENE
LESNÍCKA FAKULTA

Dušan Gömöry
Diana Krajmerová

GENETIKA A ANALÝZA DNA

2021

Titulna strana 2

Recenzenti: prof. Ing. Jaroslav Kmeť, PhD., LF TU vo Zvolene
Ing. Roman Longauer, CSc., Národné lesnícke centrum Zvolen

Schválené: Rektorom Technickej univerzity vo Zvolene dňa xx.2.2021, číslo EP 1/2021
ako skriptum pre FEE – študijný program Forezná a kriminalistická environmentalistika

© Technická univerzita vo Zvolene

© prof. Ing. Dušan Gömöry, DrSc., Ing. Diana Krajmerová, PhD.

Lesnícka fakulta, KF, T.G. Masaryka 24, 96001 Zvolen

ISBN

Všetky práva vyhradené. Nijaká časť testu ani ilustrácie nemôžu byť použité na ďalšie šírenie akoukoľvek formou bez predchádzajúceho súhlasu autora alebo vydavateľa.

Obsah

Úvod	5
VŠEOBECNÁ ČASŤ	7
Molekulárne základy dedičnosti	7
<i>Biologicky aktívne molekuly</i>	7
<i>Nukleové kyseliny</i>	7
<i>Proteíny</i>	11
<i>Replikácia DNA</i>	14
<i>Expresia génu</i>	15
<i>Genóm</i>	20
Štruktúra bunky a bunkový cyklus	22
<i>Prokaryotická bunka</i>	22
<i>Eukaryotická bunka</i>	23
<i>Štruktúra bunky</i>	23
<i>Chromozóm</i>	23
<i>Bunkový cyklus a mitóza</i>	25
<i>Meióza, väzba génov</i>	27
<i>Gonozómy</i>	29
Dedičnosť fenotypových znakov	30
<i>Autozomálna dedičnosť kvalitatívnych znakov</i>	30
<i>Gonozomálna dedičnosť kvalitatívnych znakov</i>	32
<i>Dedičnosť kvantitatívnych znakov</i>	33
Mutácie ako zdroj dedičnej premenlivosti	33
<i>Génové mutácie</i>	33
<i>Chromozómové mutácie</i>	34
<i>Genómové mutácie</i>	35
Genetika populácií	35
<i>Populácia a jej štruktúra</i>	35
<i>Príbuzenstvo a inbreeding</i>	39
<i>Panmiktická rovnováha (Hardy-Weinbergov zákon)</i>	42
<i>Väzbová rovnováha</i>	47
ŠPECIÁLNA ČASŤ	48
Biologická stopa	48
Nástroje forenznej genetickej analýzy	50
<i>Genetické a génové markéry</i>	50
<i>Proteínové markéry</i>	50
<i>DNA markéry</i>	51
<i>Princípy analýzy DNA</i>	51
<i>DNA markéry založené na fragmentačnej analýze</i>	53
<i>Sekvenovanie DNA</i>	55
<i>SNP markéry</i>	56
Identifikácia druhovej príslušnosti	58
<i>Princíp a používané markéry</i>	58
<i>Bioinformatické spracovanie</i>	59
Individuálna identifikácia	59
<i>Princíp a používané markéry</i>	59
<i>Štatistické hodnotenie</i>	62
<i>Pravdepodobnosť náhodnej zhody</i>	63

<i>Vierohodnosť vzájomne sa vylučujúcich hypotéz</i>	66
<i>Dôkaz založený na Bayesovskej logike</i>	66
Hodnotenie príbuznosti	69
<i>Princíp a používané markéry</i>	69
<i>Štatistické hodnotenie</i>	69
<i>Testovanie paternity</i>	69
<i>Odhad koeficientu príbuznosti</i>	72
Určovanie geografického pôvodu	73
<i>Princíp a používané markéry</i>	73
<i>Štatistické hodnotenie</i>	75
Literatúra	82
Literatúra odporúčaná pre štúdium	82

Úvod

Predstavme si nasledujúcu situáciu: polícia dostane hlásenie o ilegálne ulovenom rysovi u preparátora. Rys sa skutočne nájde, ale preparátor nepozná zákazníka (prinajmenšom tvrdí, že ho nepozná). Na základe analýzy kamerových záznamov z ulice polícia identifikuje podozrivý pickup. Na ložnej ploche vozidla sa nájdu stopy krvi, ale majiteľ tvrdí, že v ňom viezol polku prasaťa zo zabíjačky. Takže polícia musí nájsť odpovede na dve otázky: 1. je krv vo vozidle z prasaťa alebo z rysa? 2. ak je z rysa, ide o to isté zviera, ktoré sa našlo u preparátora? Bez odpovede na tieto dve otázky je zrejme možné vzniesť obžalobu voči preparátorovi, ale ten sa dopustil menej významného prečinu. Voči strelcovi, ktorý prírodu pripravil o vzácného živočícha, nie.

Iná situácia: na internete sa objaví inzerát na predaj mláďat chránenej korytnačky. Predajca tvrdí, že ich neodchytil v prírode, ale sú produktom jeho legálneho chovu. Rovnako vzniká otázka: je možné vylúčiť, že mláďatá sú potomstvom chovaných jedincov?

Posledný príklad: v kvetinárstve sa nájde veľké množstvo papradí *Ruscus hypoglossum*. Predajca síce netvrdí, že ich vypestoval na svojej záhradke, pretože žiadnu nemá, ale údajne ich doviezol z Chorvátska, kde tento druh chránený nie je. Otázka: je možné dokázať, že sa predajca nedopustil colného priestupku, ale prečinu proti zákonu 543/2002 Z.z. o ochrane prírody a krajiny?

Na všetky tieto otázky umožňuje odpovedať genetická analýza. Otázky sú však rozdielne, a preto aj odpovede na ne sú rozdielne. Prvá otázka sa týka identifikácie druhu, odpoveďou je „genetický čiarový kód“ (*genetic barcode*). Spravidla sa pre tento typ forenznej analýzy využíva gén s pomerne malou vnútrodruhovou premenlivosťou, ktorý ale vykazuje rozdiely medzi druhmi (samozrejme rozlíšenie rysa a prasaťa nie je problémom, ale rozlíšenie napr. dvoch blízko príbuzných druhov sýkoriiek, z ktorých jedna je chránená a druhá nie, problémom byť môže). V prípade rysa by sa použilo sekvenovanie niektorého konzervatívneho mitochondriálneho génu, najčastejšie génu pre 1. podjednotku cytochróm *c* oxidázy (*cox1*). Druhá otázka sa týka identifikácie jedinca, a odpoveďou je „genetický odtlačok prsta“ (*genetic fingerprint*). Pre tento účel je naopak potrebné použiť také genetické znaky (markéry), ktoré v rámci druhu vykazujú čo najväčšiu variabilitu, aby ich kombinácia bola pre uloveného jedinca v rámci populácie rysov na Slovensku jedinečná. Najbežnejšou voľbou tu budú nekódujúce repetitívne sekvencie jadrovej DNA (tzv. minisatelity alebo mikrosatelity). Podobný typ genetických markérov je potrebný aj pre odpoveď na tretiu otázku, ktorá sa týka určenia príbuzenstva. Kedysi sa pre tento účel u človeka (určenie otcovstva v rodinnom práve) používali krvné skupiny, preto sa takáto analýza dodnes označuje ako „krvná skúška“. Dnes sú najbežnejšími nástrojmi opäť jadrové minisatelity alebo mikrosatelity. Posledná otázka sa týka identifikácie populácie, z ktorej jedinec alebo jedince pochádzajú. Odpoveď na ňu môže byť jednoduchá aj zložitá. Ak máme šťastie, populácie daného druhu v Chorvátsku pochádzajú z iného glaciálneho refúgia, než slovenské, a preto predstavujú odlišnú genetickú líniu. V takom prípade je pre identifikáciu zdrojovej populácie možné použiť nejaký vhodný maternálne dedený markér (mitochondriálny, v prípade rastlín aj chloroplastový); takúto analýzu je možné vykonať za pár eur. Ak šťastie nemáme, populácie sú len do väčšej či menšej miery geneticky diferencované, takže záver o pôvode je možné spraviť len na základe rozdielu genetickej štruktúry podozrivého súboru jedincov a udávanej zdrojovej populácie.

Jedným problémom je výber vhodného nástroja, druhou jeho spoľahlivosť či skôr výpovedná hodnota. Pri genetickom fingerprinte platí to isté, čo pri fyzickom odtlačku prstov – môžu mať dvaja ľudia rovnaký odtlačok prsta? Väčšinu laikov napadne v prípade fyzického odtlačku automatická odpoveď – nemôže. Kriminalisti však často nenájdu pekný odtlačok celého bruška palca, ale väčší či menší fragment. Takže otázku by bolo správne skôr modifikovať – stačí na identifikáciu fragment odtlačku veľkosti 100 mm²? stačí 10 mm²?

stačí 1 mm²? Odpoveď je jednoduchá – stačiť môže možno aj 1 mm², ak sa medzi nájdenou stopou a odtlačkom podozrivého **nezhoduje**. V takom prípade možno podozrivého vylúčiť jednoznačne. A stačiť nemusí ani 100 mm², ak je zhodný. V tom prípade sa súd logicky bude kriminalistu pýtať, aká je pravdepodobnosť, že sa v populácii vyskytujú dvaja alebo viacerí ľudia, u ktorých sa fragment odtlačku s danou veľkosťou môže zhodovať. Ak odpoveď bude 10%, súd bude posudzovať takýto dôkaz iste s inou váhou, než keď odpoveď bude 0,0001%. Pre genetické aj iné nástroje platí, že zhodu medzi stopou a podozrivým objektom (či už ide o človeka, rysa alebo rastlinu) je možné vylúčiť s úplnou jednoznačnosťou, ale potvrdiť len s určitou pravdepodobnosťou. Aj relatívne primitívny nástroj môže zhodu jednoznačne vylúčiť. Ak by sme zobrali ako príklad takéhoto „primitívneho“ nástroja krvné skupiny (čo bol úplne prvý nástroj genetickej analýzy pre účely kriminalistiky): ak biologickú stopu na mieste činu predstavuje krv skupiny A a podozrivý na krvnú skupinu B (alebo AB alebo 0), je ako páchatel' vylúčený (prinajmenšom sa táto stopa ako dôkaz obžaloby vobec nemôže použiť resp. ide o dôkaz v prospech obvineného). Ak má aj podozrivý krvnú skupinu A, tak dôkaz použiteľný je. Ovšem frekvencia krvnej skupiny A (teda pravdepodobnosť, že náhodne vybraný človek bude mať krv skupiny A) na Slovensku je cca 30%, takže sila takéhoto dôkazu je minimálna, a ak by obžaloba žiadny iný dôkaz nemala, súdu by iste nestačil. Z tejto ilustrácie je zrejmé, že pre posúdenie dôkaznej sily akejkoľvek genetickej forenzej analýzy je potrebná znalosť nielen genetického profilu biologickej stopy a podozrivého, ale aj kontextu populácie, z ktorej potenciálny páchatel' či obeť pochádzajú.

Z uvedeného vyplýva, ktoré vedomosti budú potrebné pre úspešné zvládnutie tohoto predmetu. Jeho náplňou budú informácie o samotných technických nástrojoch, používaných pre genetickú forenznú analýzu, teda nukleových kyselinách (DNA a RNA) a proteínoch (krvné skupiny, izoenzyémy a pod.), čo vyžaduje aspoň elementárne znalosti organickej chémie a biochémie. Zároveň je potrebné chápať, ako tieto molekuly v organizme fungujú, čím sú jedinečné, kde v bunke sa nachádzajú, ako sa vytvárajú, ako sa dostávajú do každej bunky pri raste organizmu a rozmnožovaní. Druhou stránkou je kvantifikácia výpovednej hodnoty forenzej analýzy, ktorá si vyžaduje aspoň základy počtu pravdepodobnosti a matematickej štatistiky. Vzhľadom na skutočnosť, že študijný program *Forenzná a kriminalistická environmentalistika* nezahŕňa v bakalárskom štúdiu žiadny predmet, obsahujúci základy genetiky či molekulárnej biológie, je nutné značnú časť týchto učebných textov venovať práve tejto problematike, teda štruktúre a vlastnostiam organických makromolekúl, zúčastňujúcich sa na procesoch uchovávaní, kopírovania a expresie genetickej informácie, ich súvisu s bunkovými štruktúrami a ich správaním v priebehu životného cyklu bunky, ich dopadu na fenotypové znaky, a rovnako aj základom populačnej genetiky. Pre pripomenutie sa učebné texty vracajú aj k otázkam pravdepodobnosti a testovania štatistických hypotéz, ktoré su nutné pre pochopenie základov forenzných genetických analýz.

VŠEOBECNÁ ČASŤ

Molekulárne základy dedičnosti

Biologicky aktívne molekuly

Pod dedičnou informáciou rozumieme informáciu, ktorú pri rozmnožovaní odovzdávajú rodičia potomstvu a ktorá organizmu potomstva umožňuje vyvinúť všetky druho- špecifické aj individuálne znaky. Jej fyzickým nositeľom sú u všetkých známych živých organizmov molekuly nukleových kyselín. U všetkých prokaryotických organizmov (organizmy bez diferencovaného bunkového jadra; baktérie a archej) rovnako ako u všetkých eukaryotov (organizmy s bunkovým jadrom ako samostatnou bunkovou štruktúrou, oddelenou od cytoplazmy jadrovou membránou; rastliny, huby, živočíchy, rôzne typy jednobunkových organizmov) je to kyselina deoxyribonukleová (DNA). V prípade vírusov to neplatí (aj keď je potrebné pripomenúť, že vírusy nevykazujú všetky znaky života, biológovia ich spravidla nepovažujú za živé organizmy *sensu stricto*); ich veľkú časť predstavujú retrovírusy resp. RNA-vírusy (napr. vírus HIV, chrípky či koronavírus), u ktorých je nositeľom dedičnosti molekula kyseliny ribonukleovej.

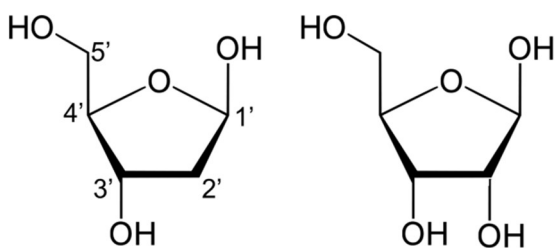
DNA v bunke neplní v princípe žiadnu ďalšiu úlohu než uskladňovanie genetickej informácie. Je istou formou stavebného plánu, ktorý umožňuje bunke vytvoriť si všetky látky, ktoré potrebuje pre svoju existenciu. Pochopiteľne neuskladňuje informácie o všetkých látkach, ich štruktúre a procesoch ich syntézy. Všetky biochemické reakcie aj prenos látok v bunke sú riadené, pričom obrovskú väčšinu týchto procesov riadia dva typy látok: kyselina ribonukleová (RNA) a bielkoviny (proteíny). RNA sa podieľa predovšetkým na riadení procesov, ktoré bezprostredne súvisia s prenosom dedičnej informácie a jej vyjadrením (expresiou) vo fenotypových znakoch. Existuje aj malá časť bežných biochemických procesov, ktoré riada RNA molekuly, tie sa v tomto prípade označujú ako ribozýmy a predstavujú pravdepodobne rudiment dávnej evolučnej minulosti, keď funkcie RNA v metabolizme primitívnych organizmov boli oveľa širšie. V súčasnosti väčšinu týchto funkcií prevzali proteíny, ktoré ich dokážu vykonávať efektívnejšie a s menšími nárokmi na energiu. Fungujú ako biologické katalyzátory (enzýmy) pre väčšinu reakcií v bunke, ako prenosové kanály sprostredkujúce riadený prenos látok cez vonkajšiu či vnútorné bunkové membrány a pod. RNA a proteíny sú schopné fungovať autonómne, nepotrebujú ďalšie riadenie. Bunke teda nepotrebuje ukladať informáciu o všetkých svojich stavebných a funkčných súčiastkach. Stačí jej uskladňovať informáciu, potrebnú pre vytvorenie RNA a proteínov, a tie už sú schopné cez riadenie metabolizmu zabezpečiť syntézu všetkého ostatného. Navyše, ako bude uvedené ďalej, DNA túto informáciu uskladňuje veľmi úsporným spôsobom.

Bunky musia byť schopné sa rozmnožovať, či už pre zabezpečenie rastu organizmu (platí pri mnohobunkových organizmoch), alebo pre zabezpečenie rastu populácie (teda zväčšovania počtu jedincov), ktorý je predpokladom prežitia druhu. Svoj stavebný plán potrebuje každá bunke, teda pred rozmnožením (delením) sa dedičná informácia materskej bunky musí nakopírovať, aby ju mohli prevziať obe dcérske bunky. DNA ako nositeľ stavebného plánu bunky teda musí byť schopná autonómneho kopírovania (replikácie), pri ktorom z jednej molekuly DNA vzniknú dve nové, identické s pôvodnou.

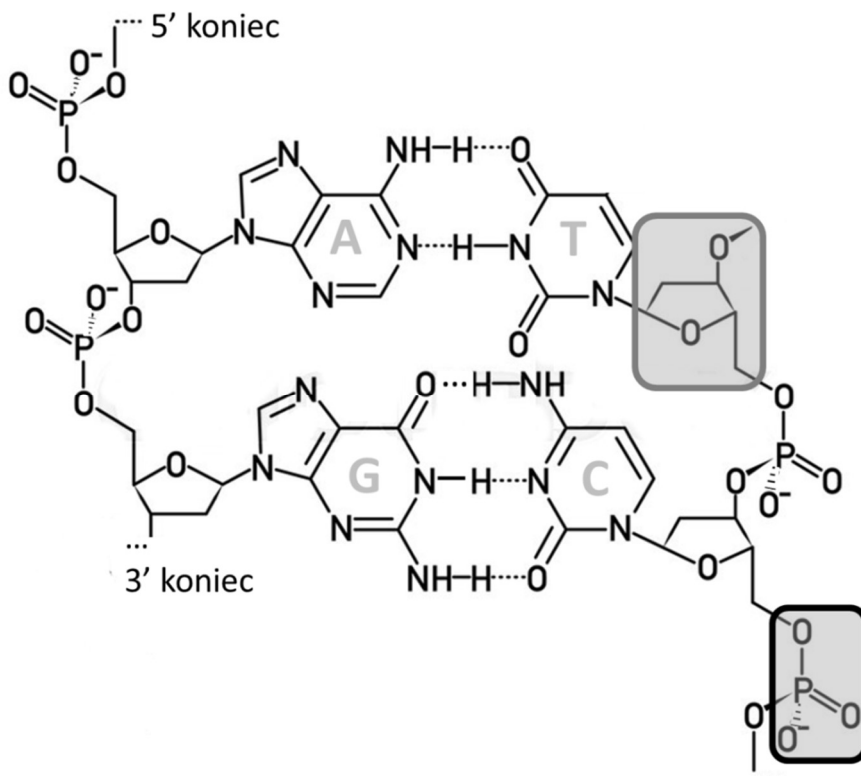
Nukleové kyseliny

Nukleové kyseliny sú polymérne organické makromolekuly, teda molekuly zložené z veľkého množstva menších stavebných kameňov. Ich základné stavebné zložky sa označujú ako nukleotidy (v prípade DNA je terminologicky správny názov 2-deoxynukleotid). Tieto stavebné kamene sú v molekule uložené za sebou, molekula DNA ani RNA sa nerozvetvuje a predstavuje dlhé súvislé vlákno. Každý nukleotid sa skladá z troch zložiek. Jednou z nich je monosacharid s piatimi atómami uhlíka (pentóza), ktorou je pri DNA 2-deoxyribóza (obr. 1).

Ďalšiu predstavuje zvyšok kyseliny fosforečnej, ktorý je naviazaný na 5. uhlíku pentózy. Tieto dve zložky sú v každom nukleotide rovnaké, a teda nie sú nositeľmi žiadnej informácie. Poslednou zložkou je jedna zo štyroch heterocyklických organických báz, z ktorých dve sú derivátmi purínu (dvojitý cyklus: adenín, guanín) a druhé dve derivátmi pyrimidínu (jednoduchý cyklus: cytozín, tymín) (obr. 2). V molekule pentózy je báza naviazaná na atóm uhlíka v pozícii 1', fosfátový zvyšok na uhlíku v pozícii 5' sa naväzuje na uhlík v pozícii 3' predchádzajúceho nukleotidu. Kostru vlákna DNA teda vytvárajú molekuly 2-deoxyribózy s naviazanými fosfátovými zvyškami. Nositeľom dedičnej informácie sú heterocyklické bázy, ich poradie (sekvencia) na molekule DNA predstavuje zápis genetickej informácie podobne ako poradie písmen určuje zápis informácie v napísanom texte či poradie 1 a 0 (resp. ich technických ekvivalentov) určuje zápis informácie v digitálnej forme.



Obr. 1 Štruktúrny vzorec cyklickej formy 2-deoxyribózy (vľavo) a ribózy (vpravo)



Obr. 2 Znáozornenie chemickej štruktúry úseku molekuly DNA. Sivý rámik: 2-deoxyribóza; čierny rámik: fosfátový zvyšok. Sivé písmená: A – adenín, T – tymín, G – guanín, C – cytozín. Plné čiary – kovalentné väzby, prerušované čiary – vodíkové mostíky.

Označovanie nukleotidov v sekvenciách je uvedené v tab. 1. Vo všeobecnosti sa bázy v zápise sekvencie označujú prvými písmenami ich názvu (A, G, C, T). V rámci konkrétnej definovanej skupiny jedincov (napr. populácie, druhu a pod.) sa však v sekvencii môže

vyskytovať premenlivosť. Označenie A, C, G, T sa potom vzťahuje na tzv. konzervované pozície, teda také, na ktorých sa v danom súbore vyskytuje vždy konkrétny nukleotid. Variabilné pozície sú v zápise označované inými písmenami (tab. 1). Nukleotid chýbajúci v sekvencii v dôsledku mutácie (delécie) sa označuje pomlčkou. V zápise sa pre bázy, ktoré sú derivátmi purínu (A, G) môže použiť aj označenie Pu, pre deriváty pyrimidínu (C, T/ U) označenie Py.

Tab. 1 označovanie nukleotidov v zápise sekvencií nukleových kyselín

ozn.	Nukleotid	ozn.	nukleotid	ozn.	nukleotid
A	adenín	R	A/G	B	C/G/T
G	guanín	Y	C/T	D	A/G/T
C	cytozín	K	G/T	H	A/C/T
T	tymín	M	A/C	V	A/C/G
		S	G/C	N	A/G/C/T
U	uracil*	W	A/T	-	delícia

* uracil je v RNA ekvivalentom tymínu

Keďže molekuly 2-deoxyribózy aj ribózy nie sú symetrické, sú molekuly nukleových kyselín priestorovo orientované. Syntéza reťazca prebieha vždy v smere 5'→3' (vid' obr. 1 a 2), t.j. voľné nukleotidy sa vždy naväzujú na hydroxylovú skupinu na 3' uhlíku na konci reťazca nukleovej kyseliny. Sekvencia akéhokoľvek úseku DNA (génu či nekódujúceho úseku) sa vždy zapisuje v rovnakom smere, teda aj ak je vynechané označenie koncov úseku 5' a 3', automaticky sa predpokladá, že zápis je vykonaný v tomto smere.

Pokiaľ sa zápis týka skupiny organizmov (lokálna či regionálna populácia, taxón a pod.), môžu byť niektoré pozície v ňom v dôsledku bodových mutácií variabilné. Zápis sekvencie s celkovou dĺžkou 18 nukleotidov v tvare 5'AGRACTTAGBTGNCCCTA3' znamená, že väčšina pozícií v tomto úseku je konzervovaných (teda u všetkých jedincov v skupine sa v danom mieste vyskytuje rovnaký nukleotid), na 3. pozícii sa vyskytuje jeden z purínových nukleotidov (R=A/G), na 10. pozícii sa vyskytuje cytozín, guanín alebo tymín (B=C/G/T), a na 13. pozícii sa môže vyskytovať ktorýkoľvek nukleotid (alebo nukleotid nemožno určiť; N=A/G/C/T). Zápis v tvare 5'AGAACTTAGCTGAC-CTA3' pre konkrétneho jedinca z tejto skupiny znamená, že v porovnaní s ostatnými na 15. pozícii jeden nukleotid chýba (delícia). Dĺžka sekvencií DNA sa udáva v básových pároch (bp), teda v počte dvojíc nukleotidov, resp. v ich násobkoch (kbp, Mbp atď.). Sekvencie, ktoré majú spoločný evolučný pôvod, sa označujú ako homologické. Môže sa jednať o úseky DNA v dvoch rozdielnych jedincoch (teda umiestnené vo fyzicky rozdielnych molekulách DNA), zdedené od spoločného predka (ortologické sekvencie), alebo o sekvencie pochádzajúce z duplikácie úseku v rámci tej istej molekuly DNA (paralogické sekvencie). Ortology ani paralogy nemusia byť úplne identické, môžu byť pozmenené následnými mutáciami (zmenou genetickej informácie, vid' kap. Mutácie). Podobnú štruktúru môžu však mať aj úseky, ktoré vznikli v evolúcii u rôznych organizmov nezávisle na sebe ale ich produkty plnia rovnakú alebo podobnú funkciu, a preto sa tlakom prírodného výberu v nich vyvinula rovnaká alebo úpodobná sekvencia nukleotidov (konvergentná evolúcia). V prípade kratších úsekov môže rovnaká sekvencia vzniknúť aj náhodne. Takéto sekvencie sa označujú ako analogické.

Molekula DNA je v skutočnosti dvojité, a jej priestorová štruktúra je špirálovitá (obr. 3). Je tvorená dvomi antiparalelne prebiehajúcimi reťazcami (protismerne; orientácia reťazcov v smere 5'→3' je opačná), ktoré sú vzájomne prepojené vodíkovými väzbami medzi opačne elektricky nabitými miestami na heterocyklických bázach (obr. 2). Vodíková väzba (vodíkový mostík) je jeden z najjednoduchších a súčasne najslabších typov chemickej väzby. Na heterocyklických bázach sa nachádzajú miesta s opačnými nábojmi. Atómy kyslíka či dusíka

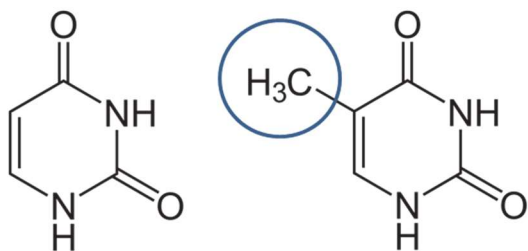
majú vo valenčnej vrstve voľné elektrónové páry (dusík jeden, kyslík dva), teda predstavujú záporne nabitú body. Naopak vo väzbe medzi dusíkom a vodíkom je elektrónový oblak koncentrovaný okolo atómu dusíka s vyššou elektronegativitou, teda atóm vodíka je nabitý kladne. Tieto opačne nabitú body na protiľahlých bázach sa môžu navzájom elektrostaticky priťahovať. Vzhľadom na priestorovú štruktúru báz a rozmiestnenie elektrického náboja v nich sa vodíkové väzby vytvárajú vždy medzi konkrétnymi dvojicami báz, jednu purínovou a jednu pyrimidínovou (obr. 2): adenín sa viaže s tymínom dvomi vodíkovými mostíkmi, guanín s cytozínom tromi ($A=T$, $G=C$). Oba reťazce DNA predstavujú navzájom svoj „zrkadlový obraz“; sú prísne komplementárne, identita nukleotidu na konkrétnej pozícii v jednom z nich jednoznačne určuje identitu protiľahlého nukleotidu v druhom reťazci. Žiadny iný než „zrkadlový“ nukleotid sa naň nedokáže naviazať vodíkovým mostíkom buď z priestorových dôvodov (dve purínové bázy by boli príliš veľké, dva pyrimidíny naopak príliš malé) alebo kvôli inkompatibilite elektrických nábojov. Pre úplnosť je treba uviesť, že v bunke sa môžu nachádzať aj fragmenty jednoreťazcovej DNA, ale podstatná časť biologicky aktívnych DNA molekúl sú dvojreťazcové molekuly.



Obr. 3 Priestorová štruktúra DNA (B-konformácia)

Štruktúra molekuly RNA je analogická. Stavebnými kameňmi sú takisto nukleotidy, rovnako tvorené molekulou pentózy (v prípade RNA je to ribóza; obr. 1), fosfátovým zvyškom, a jednou zo štyroch organických báz, opäť je to adenín, guanín a cytozín, posledná báza je v tomto prípade odlišná, ide o uracil, ktorý má ale veľmi podobnú chemickú štruktúru ako tymín, rozdiel je len v prítomnosti jednej metylovej skupiny (obr. 4), a má rovnaké chemické vlastnosti aj rovnakú väzobnú afinitu s adenínom.

Molekula RNA je spravidla tvorená len jedným reťazcom (existujú aj dvojreťazcové fragmenty s analogickou štruktúrou ako DNA). Ak sa tento reťazec ohne, nukleotidy na ňom sa tiež dostávajú do protismerného postavenia, ktoré umožňuje vytvorenie vodíkových väzieb medzi komplementárnymi („zrkadlovými“) bázami ($A=U$, $C=G$). V tomto prípade však ide o nukleotidy nachádzajúce sa na tom istom reťazci. Tieto väzby stabilizujú molekulu RNA v trojrozmernom tvare, ktorý je určujúci pre jej funkčnosť.



Obr. 4 Chemická štruktúra uracilu (vľavo) a tymínu (vpravo)

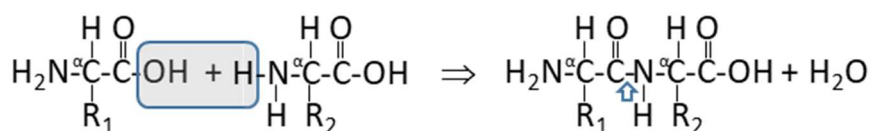
Jednotlivé súčasti nukleotidov (fosforylovaná pentóza, báza) aj jednotlivé nukleotidy sú navzájom prepojené kovalentnými väzbami, ktoré majú vysokú energiu a na ich vytvorenie aj rozštiepenie je potrebný katalyzátor (enzým), ktorý reakciu sprostredkuje. Enzýmy štiepiace nukleové kyseliny sa označujú ako nukleázy (DNáza, RNáza). Vodíkové väzby medzi komplementárnymi nukleotidmi sa vytvárajú spontánne, na ich tvorbu nie je nutná enzymatická katalýza ani iný reakčný mechanizmus. Všetky jednoreťazcové molekuly nukleových kyselín (RNA, jednoreťazcová DNA) majú preto tendenciu k párovaniu komplementárnych častí reťazcov. Po denaturácii dvojreťazcovej DNA (vodíkové mostíky sa narušia napr. chemicky alebo vysokou teplotou, a reťazce sa oddelia), postačuje obnovenie normálnych podmienok na obnovenie pôvodnej štruktúry (renaturáciu). Pri absencii komplementárneho reťazca vykazuje jednoreťazcová molekula resp. jednoreťazcový úsek spravidla silnú tendenciu sa prehnúť a vytvoriť vodíkové väzby medzi komplementárnymi (teda vzájomne sa doplnujúcimi) miestami v rámci rovnakého reťazca. Takáto štruktúra sa označuje ako „vlásenka“ (hairpin). Rovnako ak sa v reťazci DNA tandemovo opakuje viackrát rovnaký sekvenčný motív (minisatelity a mikrosatelity, tak pri renaturácii môže dôjsť posunu komplementárneho reťazca pri párovaní reťazcov, teda k spárovaniu úsekov, ktoré pôvodne neboli protiľahlé.

Na schopnosti vzájomne komplementárnych jednoreťazcových fragmentov nukleových kyselín vyhľadávať a párovať sú založené mnohé metódy molekulárnej biológie a genetických manipulácií. Viaceré z technických nástrojov využívaných vo forenzných analýzach sú založené práve na tejto vlastnosti nukleových kyselín.

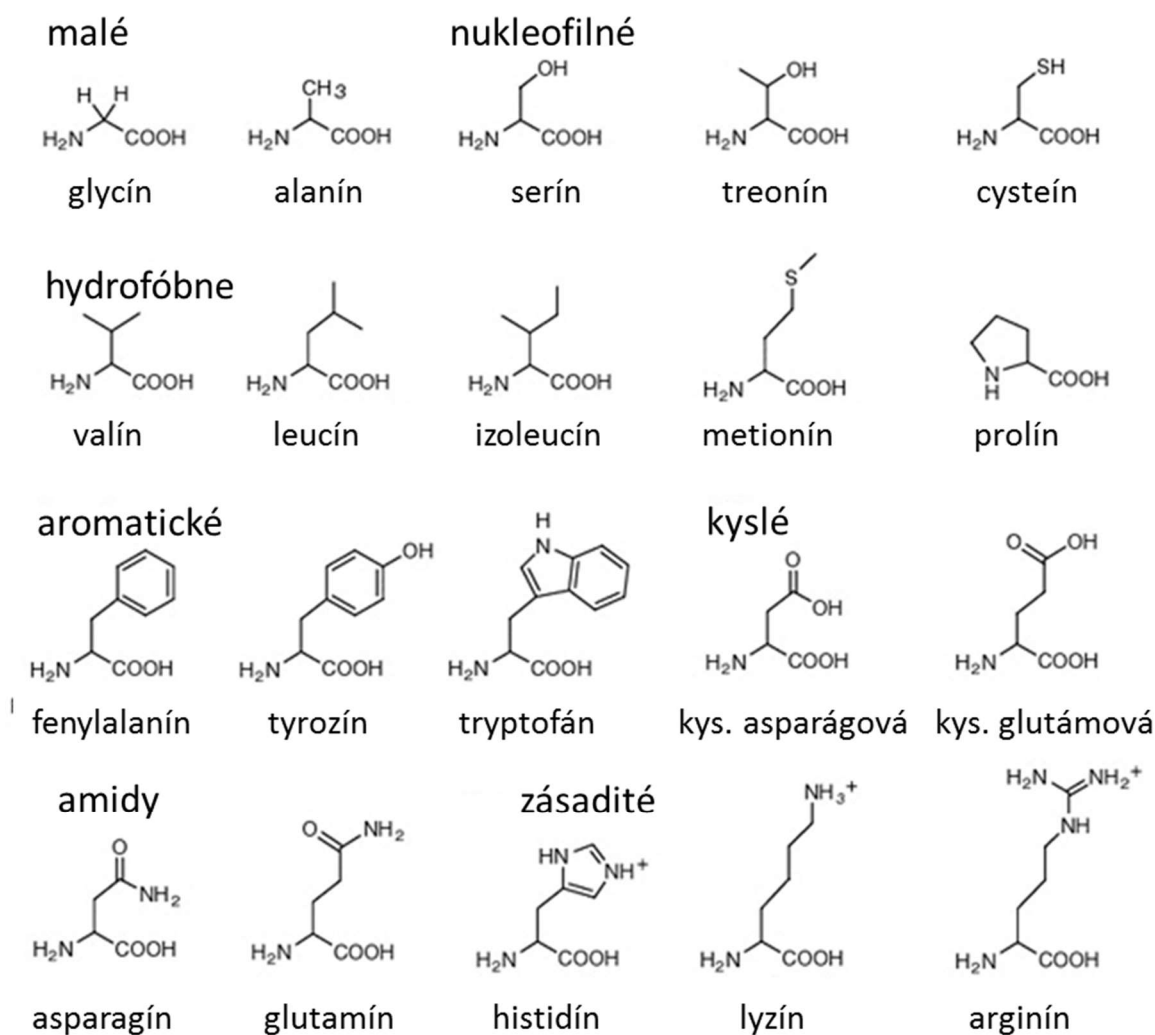
Proteíny

Proteíny (bielkoviny) sú síce v porovnaní s DNA malé molekuly, ale v porovnaní s väčšinou iných látok v bunke majú ešte stále veľkú molekulovú hmotnosť. Ich základnou stavebnou zložkou sú polypeptidové reťazce. Mnohé proteíny sú tvorené jediným polypeptidom, iné sú oligomérne (skladajú sa z dvoch alebo viacerých rovnakých alebo rozdielnych polypeptidov), niektoré obsahujú aj ďalšiu zložku (lipid – lipoproteíny, sacharid – glykoproteíny a pod.).

Polypeptid je polymér tvorený lineárnym (nerozvetveným) reťazcom stavebných kameňov označovaných ako aminokyseliny (presnejšie α -aminokyseliny resp. 2-aminokyseliny). Štruktúra aminokyseliny je na obr. 6: na 2. uhlíku (α -uhlíku) je naviazaná jedna aminoskupina ($-\text{NH}_2$), jedna karboxylová skupina ($-\text{COOH}$) a jeden atóm vodíka; tieto tri zložky sú vo všetkých aminokyselinách rovnaké. Odlišná je štvrtá zložka, tzv. bočný reťazec (*side chain*; na obr. 5 označená ako R). Pri spájaní aminokyselín do peptidového reťazca sa z karboxylovej skupiny jednej aminokyseliny odštiepi hydroxyl ($-\text{OH}$), z aminoskupiny nasledujúcej aminokyseliny sa odštiepi atóm vodíka (spolu vytvorí molekulu vody), a atómy uhlíka jednej a dusíka nasledujúcej aminokyseliny sa spoja kovalentnou väzbou, ktorá sa označuje ako peptidická väzba. Tvorba peptidov ovšem neprebíha spontánne, aminokyseliny sú za sebou radené zákonitým spôsobom, ktorý určuje genetická informácia, a celý proces je katalyzovaný spôsobom, popísaným v ďalšom texte. Rovnako ako DNA je aj molekula proteínu priestorovo orientovaná, na jednej strane je ukončená $-\text{NH}_2$ skupinou (N-koniec), na druhej $-\text{COOH}$ skupinou (C-koniec).



Obr. 5 Schéma štruktúry aminokyselín a syntézy peptidu. Prázdnu šípkou je označená peptidická väzba, symbolom α je označený α -atóm uhlíka.

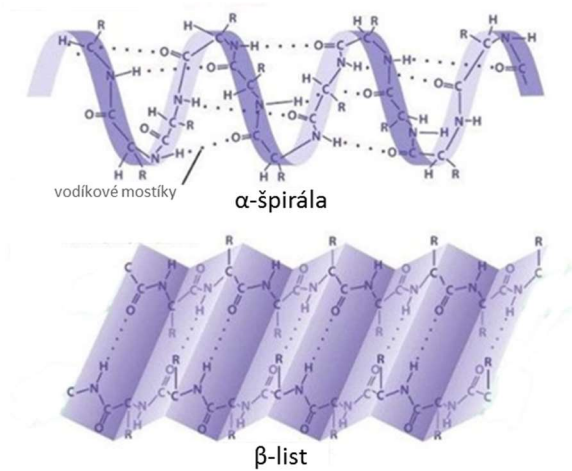


Obr. 6 Štruktúrne vzorce esenciálnych aminokyselín (ccrma.stanford.edu, upravené)

V bielkovinách sa vyskytuje 20 rôznych druhov aminokyselín, označovaných ako esenciálne aminokyseliny. Ich chemické vlastnosti závisia vždy na zvyšku (R; na obr. 6 predstavuje časť molekuly nad väzbou $\text{H}_2\text{N}-\text{COOH}$). Chemické vlastnosti a teda aj funkčnosť celého proteínu určuje typ aminokyselín, ktoré sú v ňom obsiahnuté, a ich usporiadanie.

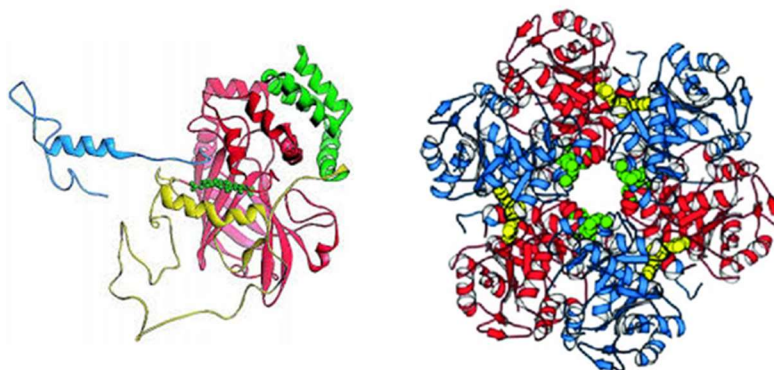
Genetická informácia určuje, v akom poradí sú aminokyseliny za sebou zoradené v reťazci polypeptidu (spôsobom, ktorý bude popísaný ďalej). Toto poradie sa označuje ako primárna štruktúra proteínu. Jednotlivé úseky reťazca polypeptidu sa však spravidla usporiadajú do priestorovej štruktúry, ktorá je stabilizovaná opäť vodíkovými mostíkmi medzi rôznymi aminokyselinami v rámci kostry rovnakého reťazca, teda tými časťami aminokyselín, ktoré

sú prepojené peptidickými väzbami (atómami kyslíka karboxylovej skupiny jedne aminokyseliny a vodíkom v rámci NH väzby inej aminokyseliny). Toto usporiadanie sa označuje ako sekundárna štruktúra, najbežnejšie typy sekundárnej štruktúry su α -špirála a β -list (obr. 7).



Obr. 7 Sekundárna štruktúra bielkovín (www.quora.com, upravené)

Následne sa molekula proteínu usporiada do definitívneho trojrozmerného tvaru, ktorý je rozhodujúci z hľadiska jeho funkcie. Tento trojrozmerný tvar sa označuje ako terciárna štruktúra, a v jeho rámci sa spravidla striedajú úseky so špirálovitou či plochou sekundárnou štruktúrou a voľné úseky.



Obr. 8 Terciárna štruktúra monomérneho proteínu (vľavo) a kvartérna štruktúra hexaméru (vpravo)

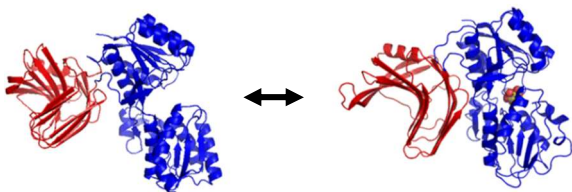
Ako bolo uvedené, niektoré proteíny sú tvorené viacerými polypeptidmi, buď rovnakými (homooligoméry) alebo rozdielnymi (heterooligoméry). V niektorých prípadoch môžu obsahovať aj ďalšiu, nepeptidovú zložku. Táto štruktúra sa označuje ako kvartérna (obr. 8). Typickým príkladom je tetramérna molekula hemoglobínu – je tvorená štyrmi molekulami globínu, ktoré obsahujú hémové jadrá s atómami železa.

Nadobudnutie definitívneho trojrozmerného tvaru (skladanie resp. zbaľovanie bielkovín; *foldng*) prebieha u časti proteínov spontánne, ale väčšinou býva podmienené interakciou s inými bielkovinami, ktoré proteínu pomáhajú zbaľiť sa do správneho tvaru, označovanými ako šaperóny (*chaperons*). Nesprávne poskladanie môže viesť k závažnej dysfunkcii proteínu: príkladom môžu byť prióny, čo sú bielkovinné molekuly spôsobujúce nesprávne zbalenie proteínov nervových tkanív a spôsobujúce spongiformnú encefalopatiu („choroba šialených kráv“, prenosná aj na človeka ako Creutzfeldt-Jacobova choroba).

Terciárna aj kvartérna štruktúra sú stabilizované interakciami medzi zvyškami (bočnými reťazcami) aminokyselín, závisí teda od ich chemických vlastností. Príkladom môže byť tzv. disulfidický mostík ($-S=S-$), ktorý sa vytvára medzi dvoma zvyškami cysteínu a predstavuje

veľmi pevnú väzbu. Väzby v kostre bielkovinného reťazca (t.j. medzi dusíkom aminoskupiny, α -uhlíkom, uhlíkom karboxylovej skupiny + peptidická väzba) sú jednoduché, takže reťazec okolo nich môže voľne rotovať. Výnimku predstavuje prolín, kde α -C je súčasťou cyklickej molekuly; prítomnosť prolínu teda vedie k stabilnému priestorovému lomu proteínového reťazca. Od primárnej štruktúry (poradie aminokyselín) teda závisí aj terciárna štruktúra a teda funkčnosť proteínu. Zmena genetickej informácie (mutácie) vedúca k zámene aminokyseliny v reťazci teda môže spôsobiť zmenu tvaru a tým funkciu proteínu zmeniť alebo (častejšie) proteín znefunkčnit'.

Niektoré proteíny sú schopné nadobúdať viac stabilných priestorových konformácií, často sú v jednej z nich stabilizované interakciou s inou, nebielkovinnou zložkou. Označujú sa ako alosterické proteíny (obr. 9).



Obr. 9 Dve možné priestorové usporiadania enzýmu xylanáza. V konformácii vpravo je enzým funkčný, je v nej stabilizovaný naviazaním molekuly monosacharidu xylóza.

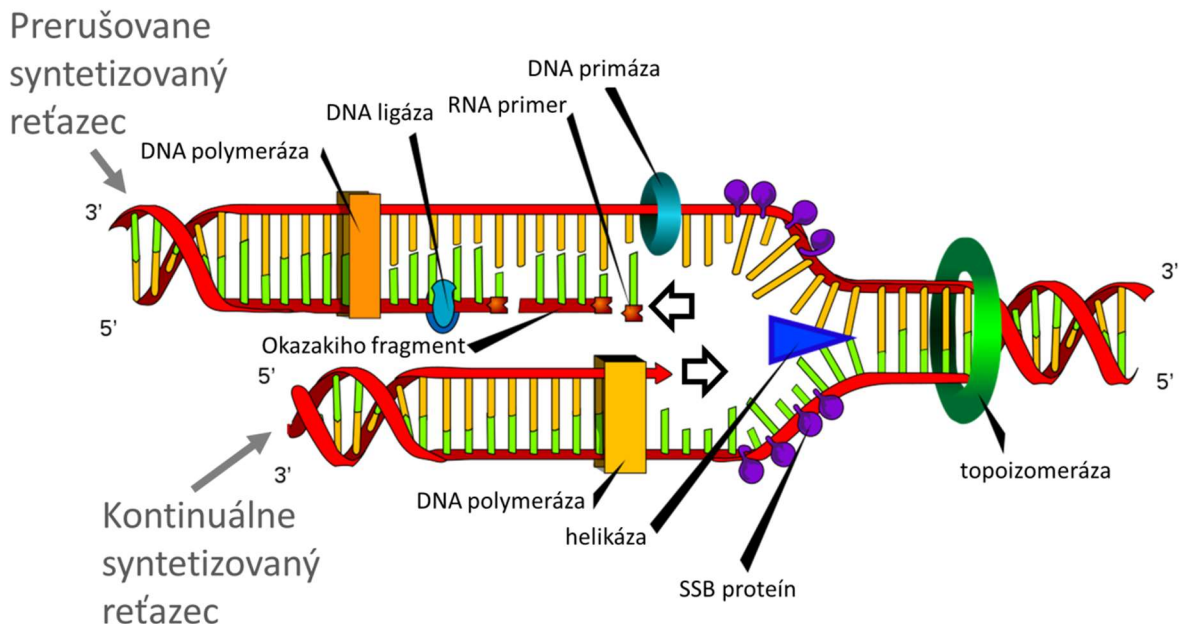
Replikácia DNA

Ako už bolo uvedené, DNA nesie informáciu pre všetky bunečné funkcie a fenotyp organizmu, preto všetky bunky musia obsahovať identickú verziu molekuly (alebo molekúl) DNA. Všetky sa organizmy prostredníctvom delenia buniek množia, a mnohobunkové organizmy touto cestou aj rastú (aj keď časť rastu s uskutočňuje aj zväčšovaním objemu buniek), takže pred delením sa všetky molekuly DNA v materskej bunke musia identicky nakopírovať tak, aby dcérske bunky mohli dostať po jednej kópii. Tento proces sa označuje termínom replikácia. Pre pripomenutie – molekula DNA je tvorená dvoma reťazcami, ktoré sú orientované protismerne a sú navzájom komplementárne, teda konkrétnej báze v jednom reťazci zodpovedá jediná možná konkrétne báza v druhom reťazci. Táto vlastnosť je rozhodujúca pre mechanizmus replikácie: jej výsledkom sú dve molekuly, z ktorých každá je tvorená jedným originálnym a jedným novo syntetizovaným reťazcom. Tento mechanizmus sa označuje ako semikonzervatívna replikácia. Vzhľadom na veľkosť molekúl DNA (u eukaryotov môže mať jedna molekula dĺžku až niekoľko desiatok či stoviek miliónov bázových párov) sa proces replikácie spravidla začína na niekoľkých miestach naraz. Proces replikácie riadi komplex enzýmov, z ktorých každý zabezpečuje inú úlohu. Tento komplex sa posúva od iniciálneho miesta a postupne syntetizuje nové reťazce; toto pohyblivé miesto rozštiepenia pôvodnej molekuly DNA a syntézy nových reťazcov sa označuje ako replikačná vidlica.

Rozhodujúci enzým z celého komplexu je DNA-polymeráza, ktorá na hydroxylovú skupinu na 3' uhlíku posledného nukleotidu novovytváraného reťazca pripája nové deoxynukleotidy a tým ho postupne predlžuje. Presnejšie, ako materiál pre budovanie nového reťazca slúžia molekuly deoxynukleozid-trifosfátov (dNTP), ktoré na 5' uhlíku majú pripojené tri fosfátové zvyšky, navzájom prepojené energeticky bohatými (makroergickými) väzbami; energia týchto väzieb slúži na pripojenie deoxynukleotidu na 3' koniec reťazca. DNA-polymeráza však potrebuje začiatok nového reťazca (teda existujúcu časť nového reťazca s 3' koncom), nedokáže pripojiť nový nukleotid na voľné miesto jednoreťazcovej molekuly DNA. Preto po prerušení nízkoenergetických vodíkových väzieb medzi oboma reťazcami pôvodnej molekuly helikázou sa na každý z pôvodných reťazcov DNA pripojí krátky komplementárny reťazec RNA označovaný ako primer (tento proces sprostredkováva enzým DNA primáza), ktorý slúži ako iniciálne miesto predlžovania. Po ukončení replikácie sa RNA primer odbúra a nahradí deoxynukleotidmi. Jednoreťazcové úseky DNA chráni pred spontánnym spárovaním špecifické proteíny (*single-strand binding protein*; SSB proteín). Keďže DNA je

špirálovitá molekula, pri jej odvíjaní vzniká mechanické napätie (podobné ako keď sa rozpletá lano spletené z viacerých špirálovite stočených prameňov); toto mechanické napätie odstraňuje enzým topoizomeráza. Reťazce materskej molekuly DNA sú orientované protismerne. Preto ak sa replikačná vidlica po molekule posúva jedným smerom, znamená to, že syntéza jedného z nových reťazcov môže prebiehať kontinuálne, pretože je orientovaný „správne“ ($5' \rightarrow 3'$; *leading strand*), ale druhý reťazec je orientovaný opačne proti smeru pohybu replikačnej vidlice ($3' \rightarrow 5'$; *lagging strand*). Tento reťazec je syntetizovaný v krátkych fragmentoch (Okazakiho fragmenty), z viacerých iníciačných miest postupne vytváraných v smere pohybu replikačnej vidlice, fragmenty sú následne spájané enzýmom ligázou (obr. 10 treba chápať ako schematický, RNA primery aj Okazakiho fragmenty sú v skutočnosti podstatne dlhšie).

Oba pôvodné reťazce sú navzájom komplementárne ($A=T$, $G \equiv C$) a do novovytváraných reťazcov sa takisto môže zaradiť len nukleotidy, ktoré sú komplementárne k pôvodným. Preto týmto spôsobom vznikajú dve úplne identické molekuly, v ktorých vždy jeden reťazec je pôvodný a druhý je novo doplnený, obe sú teda spoločne originálom a spoločne kópiou.



Obr. 10 Schéma semikonzervatívnej replikácie DNA. Orientácia helikázy naznačuje smer pohybu replikačnej vidlice, prázdne šípky určujú smer syntézy nového reťazca. (<http://www.clker.com>, upravené)

Takýmto relatívne komplikovaným spôsobom za účasti veľkej enzymatickej výbavy prebieha replikácia obrovských DNA molekúl v živých bunkách (*in vivo*). Na replikáciu krátkych DNA fragmentov v umelých podmienkach stačí samotná DNA polymeráza; na tejto skutočnosti stavia metóda široko využívaná v biologickom výskume a diagnostickej praxi, polymerázová reťazová reakcia (PCR); podrobnejší popis PCR je uvedený v ďalšom texte.

Expresia génu

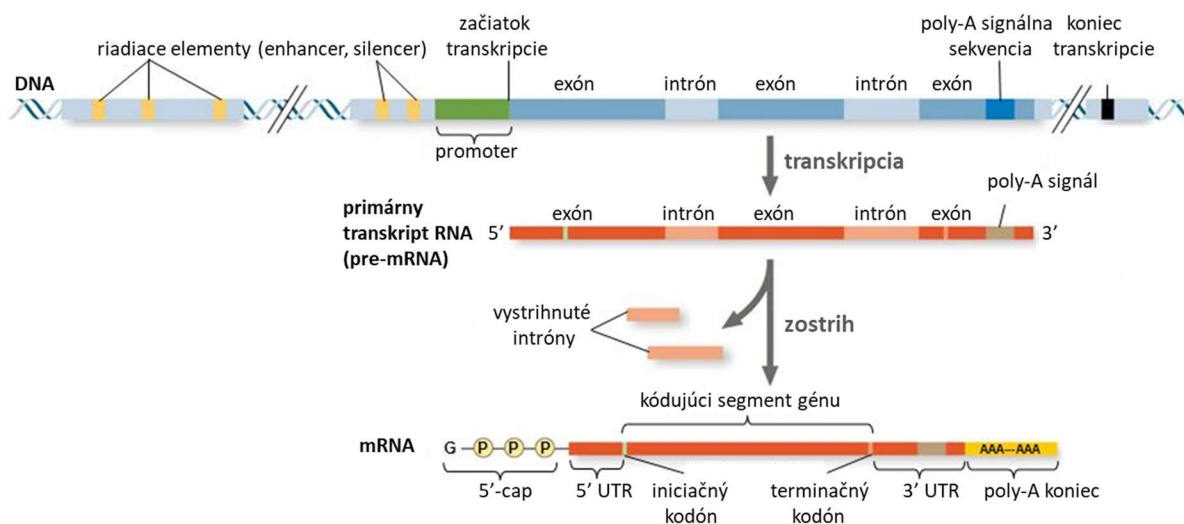
Pod génom rozumieme jednotku genetickej informácie, ktorá zodpovedá za nejaký fenotypový znak v najširšom slova zmysle; fenotypovým znakom môže byť morfológická vlastnosť, fyziologický proces, ale aj prítomnosť konkrétnej látky v organizme. Existuje celý rad definícií génu, pre náš účel postačí chápať gén ako úsek na molekule DNA, ktorý nesie túto informáciu. Každá molekula DNA obsahuje množstvo génov: v prípade človeka je v jadre

bunky prítomných 23 párov molekúl DNA (chromozómov), pričom celkový počet génov je cca 30000, na jednu molekulu teda rádovo pripadajú stovky až tisíce génov. Miesto na molekule DNA (chromozóme alebo organelárnej molekule DNA), kde je gén alebo iný konkrétny úsek lokalizovaný, sa označuje termínom lokus.

Expresia génu je jeho vyjadrenie v nejakom fenotypovom znaku, ktorý daný gén kontroluje. Na molekulárnej úrovni je výsledkom expresie molekula RNA alebo (častejšie) molekula proteínu, ktorá tým že plní svoju enzymatickú, stavebnú, prenosovú či inú funkciu v bunke prispieva k vytvoreniu daného fenotypového znaku.

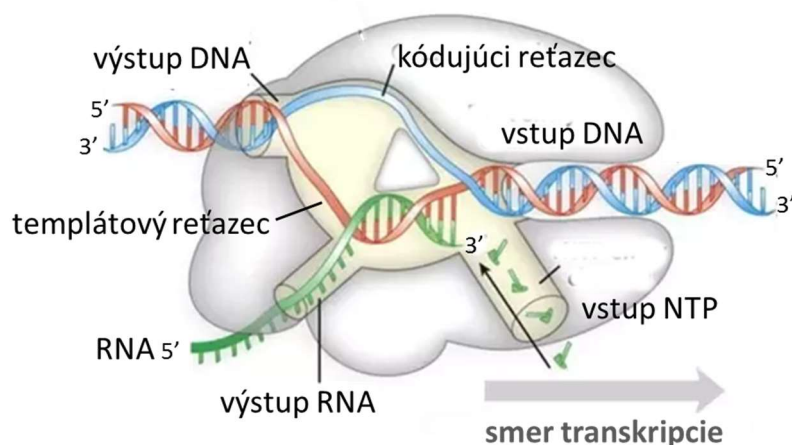
Syntéza proteínu prebieha v dvoch krokoch. Prvým krokom je prepis (transkripcia), pri ktorom sa na základe informácie uloženej v DNA syntetizuje molekula RNA. Môže ísť o molekulu, ktorá sama plní nejakú funkciu v bunke (ribozómová, transferová, malá jadrová RNA, mikroRNA, ribozým...) alebo nesie informáciu pre syntézu polypeptidu (mediátorová RNA; mRNA; messenger RNA). Druhým krokom je preklad (translácia), pri ktorom sa na základe informácie v mRNA syntetizuje molekula polypeptidu.

Gény nie sú exprimované trvale, ich aktivita je riadená. Pred samotnou sekvenciou génu je priradený (t.j. nachádza sa na 5' konci génu) úsek označený ako promotor, ktorý má funkciu „vypínača“ génu. Okrem toho sa ako pred 5' koncom tak aj za 3' koncom génu môžu nachádzať úseky, ktorú zosilňujú či zoslabujú mieru expresie (*enhancers, silencers*; obr. 11). Tieto sekvencie sú súčasťou nekódujúcich úsekov, ktoré oddeľujú jednotlivé gény, označovaných ako oddeľovače (*spacers*).



Obr. 11 Štruktúra eukaryotického génu a schéma transkripcie a zostrihu RNA

Transkripciu riadi enzým RNA polymeráza. Iniciáciu transkripcie umožňuje naviazanie sa malého proteínu (transkripčného faktora) na sekvenciu v promotore génu s priemernou dĺžkou okolo 10 bp (5–30 bp), transkripčný faktor následne umožňuje naviazanie RNA polymerázy. Jej fungovanie je obdobné ako DNA polymerázy pri replikácii. Enzým najprv rozdelí vodíkové mostíky medzi reťazcami DNA, a následne syntetizuje molekulu RNA tak, že ku každému nukleotidu v jednom z reťazcov DNA, označovanom ako templátový resp. antikódujúci reťazec, priraduje nukleotid RNA, ktorý je k nemu komplementárny. Rovnako ako pri replikácii platí, že ako materiál slúžia molekuly nukleozidtrifosfátov (NTP), energia pre pripojenie nového nukleotidu k reťazcu RNA kovalentnou väzbou je čerpaná z makroergických väzieb medzi fosfátovými zvyškami. Keďže neprepisovaný reťazec DNA je tiež komplementárny k templátovému reťazcu, má rovnakú sekvenciu nukleotidov ako novovytvorená RNA, preto sa označuje ako kódujúci reťazec.



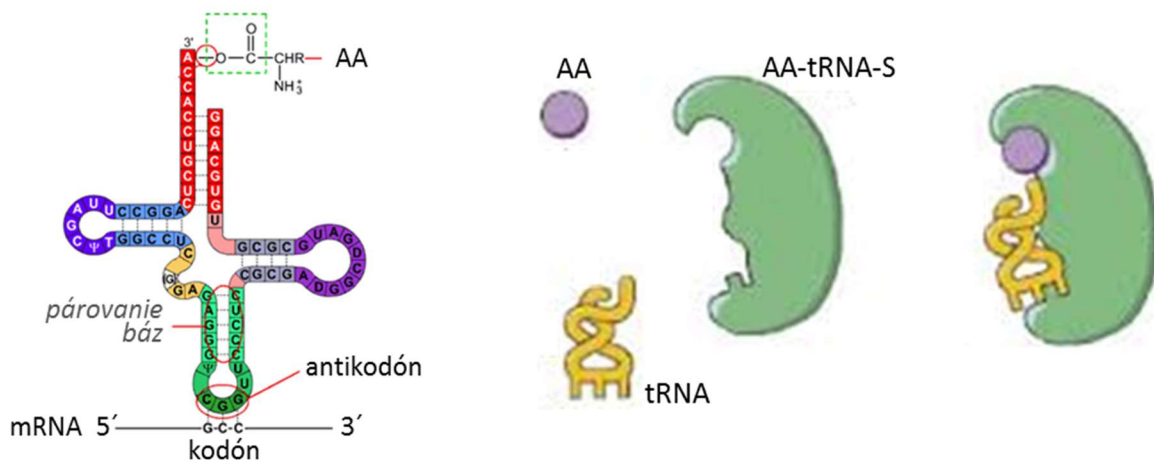
Obr. 12 Schéma fungovania RNA polymerázy pri transkripcii.

Funkčné molekuly RNA spravidla nepodliehajú rozsiahlejším ďalším úpravám, na rozdiel od molekúl RNA nesúcich informáciu pre syntézu proteínov. U eukaryotov gény spravidla nie sú súvislé, obsahujú úseky ktoré sú podkladom pre transláciu (exóny) a úseky, ktoré nie sú prekladané (intróny). Prepisovaný je vždy celý úsek génu. Produktom prepisu ale nie je hotová molekula mRNA, ale tzv. primárny transkript (pre-mRNA), ktorý podlieha ďalším úpravám. V procese zostrihu (*splicing*) sa intróny z primárneho transkriptu vystrihnú a do mRNA sa spoja len exóny (obr. 11). Pri niektorých molekulách RNA prebieha zostrih autokatalyticky (teda bez účasti ďalších molekúl), ale väčšinou ho zabezpečuje útvar označovaný ako spliceozóm, zložený z komplexov malej jadrovej RNA (*small nuclear RNA*; snRNA) a proteínov. Okrem toho sa na 5' koniec pripojí 7-metyl-guanozínový nukleotid (5'-cap), a na 3' koniec sa pripojí reťazec adenínových nukleotidov (polyadenylový reťazec), ktorý mRNA chráni pred rýchlym odbúraním RNázami. Až takto opravená (zrelá) mRNA môže slúžiť ako zdroj informácie pre preklad. U prokaryotov sa v génoch kódujúcich proteíny intróny spravidla nevyskytujú, celý gén je tvorený jediným exónom. V sekvenciách kódujúcich funkčné molekuly RNA (najmä tRNA) sa intróny vyskytnúť môžu, v tom prípade zostrih prebieha autokatalyticky.

U eukaryotov môže dochádzať k alternatívnemu zostrihu: z rovnakého primárneho transkriptu môžu byť vytvorené rôzne alternatívne molekuly mRNA z dôsledku zaradenia resp. vynechania konkrétnych exónov, čo vedie k vytvoreniu viacerých rozdielnych produktov (polypeptidov), odrážajúcich sa na rozdielom fenotypu buniek.

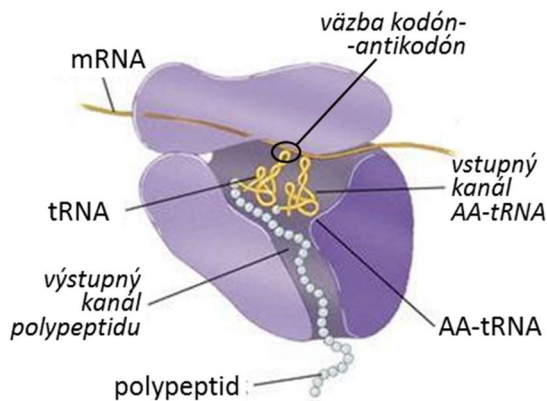
Pri translácii (preklade) je na základe informácie v mRNA syntetizovaná molekula polypeptidu, ktorá tvorí základ proteínu. Prebieha v cytoplazme, čo v prípade eukaryotov znamená, že molekula mRNA musí byť transportovaná cez jadrovú membránu do cytoplazmy. Transláciu katalyzuje útvar označovaný ako ribozóm, ktorý je tvorený z cca 2/3 nukleovou kyselinou (ribozómovou RNA; rRNA) a z 1/3 proteínmi, s dvomi podjednotkami. Veľkosť týchto podjednotiek je kvantifikovaná na základe sedimentačnej rýchlosti pri cetrifugácii v Svedbergových jednotkách ($1S=10^{-13}s^{-1}$); sedimentačná rýchlosť závisí od veľkosti a tvaru, teda Svedbergove jednotky nie sú aditívne, súčet zložiek nedáva veľkosť celého útvaru. Veľkosť ribozómov u prokaryotov je cca 70S (malá podjednotka 16S rRNA + 21 proteínov = 30S, veľká podjednotka 23S + 5S rRNA + 31 proteínov = 50S), u eukaryotov je cca 80S (malá podjednotka 18S rRNA + 33 proteínov = 40S, veľká podjednotka 28S + 5,8S + 5S rRNA + 49 proteínov = 60S). U prokaryotov sú ribozómy lokalizované voľne v cytoplazme, u eukaryotov je väčšina naviazaná na membrány endoplazmatického retikula. Mitochondrie a chloroplasty majú vlastné ribozómy, ktorých veľkosť je rovnaká ako u prokaryotov (cca 60S); preklad génov týchto organel prebieha priamo v nich.

Informáciu pre zaradenie aminokyseliny do vytváraného polypeptidového reťazca predstavuje trojica nukleotidov mRNA, označovaná triplet resp. kodón. Na translácii sa zúčastňuje ďalší typ RNA molekúl, transferová RNA (tRNA), ktorá má tvar deformovaného trojlístka (schematické zobrazenie, skutočný trojrozmerný tvar je odlišný) s tromi slučkami a štyrmi stopkami, ktoré sú stabilizované párovaním komplementárnych báz. Na 3' koniec tRNA (koniec tzv. akceptorovej stopky) sa naväzuje aminokyselina, a na strednej slučke (antikodónová slučka) je trojica báz, predstavujúca tzv. antikodón. Spárovanie aminokyseliny so správnou tRNA (tRNA so správnym antikodónom) zabezpečujú enzýmy aminoacyl-tRNA-syntetázy, ktoré majú dve aktívne miesta. Tvar jedného z nich umožňuje pripojenie výlučne tRNA so správnym antikodónom, do druhého zapadne výlučne aminokyselina so správnym zvyšok (bočným reťazcom). Následne enzým prepojí obe molekuly esterovou väzbou (obr. 13).



Obr. 13 Tvar molekuly tRNA (vľavo) a schéma fungovania aminoacyl-tRNA-syntetázy. AA – aminokyselina, AA-tRNA-S – aminoacyl-tRNA-syntetáza

Pri preklade sa malá podjednotka ribozómu pripojí na molekulu mRNA v mieste, kde sa nachádza trojica nukleotidov AUG, ktorá predstavuje tzv. iniciačný kodón. Následne sa na celý komplex pripojí tRNA, ktorá má komplementárny antikodón CAU (pripomínam, že sekvencia sa zapisuje vždy v smere 5'→3') a na 3' konci má naviazanú aminokyselinu metionín. Následne sa pripojí veľká podjednotka ribozómu a celý komplex je pripravený na ďalšie predlžovanie. Na každom ribozóme sú tri miesta pre väzbu tRNA: aminoacylové miesto (A), do ktorého vstupuje tRNA a naviazanou aminokyselinou (komplex AA-tRNA), peptidylové miesto (P), v ktorom sa nachádza tRNA s narastajúcim peptidovým reťazcom, a výstupné miesto (E), v ktorom tRNA po odpojení aminokyseliny a jej pripojení k peptidovému reťazcu opúšťa chromozóm. Genetická informácia v mRNA je čítaná po kodónoch, teda trojicích za sebou nasledujúcich báz. mRNA sa posúva cez ribozóm analogicky ako dierna páska cez čítaciu hlavicu v starých počítačoch (presnejšie, ribozóm sa posúva po molekule mRNA). Keď sa ribozóm posunie o jeden kodón, do miesta A môže vstúpiť a na mRNA sa naviazať len taká tRNA, ktorá má presne komplementárny antikodón; žiadna iná tRNA by s kodónom nedokázala vytvoriť vodíkové väzby a ani priestorovo by nezapadla do kodónu mRNA v aminoacylovom mieste. Po pripojení AA-tRNA na mRNA aminokyselina na nej prepojí s peptidovým reťazcom, naviazaný na tRNA v peptidylovom mieste, čím sa tRNA v P mieste uvoľní. Následne sa ribozóm posunie po mRNA o ďalší kodón. Uvoľnená tRNA sa tým dostáva do výstupného miesta (E) a ribozóm opúšťa, a uvoľní sa A miesto, do ktorého môže vstúpiť ďalšia AA-tRNA (obr. 14).



Obr. 14 Schéma priebehu translácie. AA-tRNA – komplex tRNA s naviazanou aminokyselinou

Poradie kodónov v mRNA teda určuje poradie aminokyselín v polypeptide. Bázy mRNA (predstavujúce písmená genetického kódu) sú štyri (A, C, G, U). Kodón predstavuje vždy trojica báz, genetický kód teda poskytuje $4^3 = 64$ možných kombinácií. Esenciálnych aminokyselín je len 20. Genetický kód je teda nadbytočný (redundantný), viaceré triplety môžu kódovať rovnakú aminokyselinu, pričom ich počet sa môže pohybovať od 1 (napr. UGG – tryptofán) po 6 (UUA, UUG, CUN – leucín). U mnohých tripletov rozhoduje o naviazanej aminokyseline len prvá dvojica báz (od 3'-konca), tretia báza už nemá informačný význam (tab. 2). Tri kodóny (UAA, UAG, UGA) sú terminačné, signalizujú ukončenie translácie a odpojenie produkovanej molekuly polypeptidu od ribozómu. Kodón AUG kódujúci metionín je iniciačný, signalizuje začiatok translácie. Preklad u eukaryotov vždy začína od tripletu AUG, ktorý sa nachádza najbližšie k 5' koncu mRNA. Preto polypeptidy začínajú metionínom, niekedy je táto aminokyselina pri posttranslačných úpravách z reťazca odštiepená.

Tab. 2 Genetický kód (kodóny v mRNA, im zodpovedajúce aminokyseliny, ich trojpísmenové a jednopísmenové skratky)

Báza													
1.	2.											3.	
	U			C			A			G			
U	UUU	Fenylalanín	Phe (F)	UCU	Serín	Ser (S)	UAU	Tyrozín	Tyr (Y)	UGU	Cysteín	Cys (C)	U
	UUC	Fenylalanín		UCC	Serín		UAC	Tyrozín		UGC	Cysteín		C
	UUA	Leucín	Leu (L)	UCA	Serín		UAA	term		UGA	term		A
	UUG	Leucín		UCG	Serín		UAG	term		UGG	Tryptofán		Trp (W)
C	CUU	Leucín	Leu (L)	CCU	Prolín	Pro (P)	CAU	Histidín	His (H)	CGU	Arginín	Arg (R)	U
	CUC	Leucín		CCC	Prolín		CAC	Histidín		CGC	Arginín		C
	CUA	Leucín		CCA	Prolín		CAA	Glutamín		CGA	Arginín		A
	CUG	Leucín		CCG	Prolín		CAG	Glutamín		CGG	Arginín		G
A	AUU	Izoaucín	Ile (I)	ACU	Treonín	Thr (T)	AAU	Asparagín	Asn (N)	AGU	Serín	Ser (S)	U
	AUC	Izoaucín		ACC	Treonín		AAC	Asparagín		AGC	Serín		C
	AUA	Izoaucín		ACA	Treonín		AAA	Lyzín		AGA	Arginín		A
	AUG	Metionín (ini)		Met (M)	ACG		Treonín	AAG		Lyzín	AGG		Arginín
G	GUU	Valín	Val (V)	GCU	Alanín	Ala (A)	GAU	Kys. asparágová	Asp (D)	GGU	Glycín	Gly (G)	U
	GUC	Valín		GCC	Alanín		GAC	Kys. asparágová		GGC	Glycín		C
	GUA	Valín		GCA	Alanín		GAA	Kys. glutámová		GGA	Glycín		A
	GUG	Valín		GCG	Alanín		GAG	Kys. glutámová		GGG	Glycín		G

ini – iniciačný kodón začína transláciu, term – terminačný (STOP) kodón ukončuje transláciu

kyslý zvyšok
 bázičný zvyšok
 polárny zvyšok
 nepolárny zvyšok
 term terminačný kodón

Čítanie reťazca mRNA je nepretržité, ak teda dôjde k vsunutiu alebo vypadnutiu (inzercia/delécia) jedného alebo niekoľkých nukleotidov, zmenia sa od miesta mutácie všetky triplety a teda aj všetky aminokyseliny zaradené do polypeptidu. Ak dôjde k bodovej mutácii (zámena

jedného nukletidu za iný), zmení sa len jedna zaradená aminokyselina, aj to len v prípade, že mutácia nie je synonymná (t.j. že pôvodný a mutovaný triplet kódujú rôzne aminokyseliny). Na jednu molekulu mRNA je často naviazaných viac ribozómov súčasne, takže paralelne prebieha syntéza viacerých identických polypeptidových molekúl. Pri prokaryotoch často prebiehajú všetky procesy súčasne: na ešte nedokončený reťazec mRNA sa pripájajú ribozómy, syntetizujú molekuly polypeptidov, a od 5' konca súčasne je mRNA postupne degradovaná RNázou.

Genetický kód je univerzálny, teda funguje rovnako pri všetkých organizmoch, čo je tiež jednou zo známk spoločného evolučného pôvodu všetkých živých organizmov. U viacerých taxónov existujú síce výnimky, ale platia vždy pre celý taxón a uplatňujú sa zákonite (napr. v mitochondriách stavovcov triplet UGA kóduje tryptofán, čo ale platí pre všetky stavovce a výlučne pre mitochondrie, pri jadrových génoch je kodón UGA vždy terminačný).

Translácia mRNA teda vždy prebieha rovnakým spôsobom, z rovnakej predlohy nemôžu vznikáť rozdielne produkty. Mnohé proteíny sú však následne modifikované v rámci posttranslačných úprav: je z nich odbúraná určitá časť, môžu sa na ne naväzovať nepeptidové zložky a pod. Mnohé proteíny sú oligomérne, alebo sú funkčné len po naviazaní ďalších organických molekúl (sacharidov, lipidov, hému atď.), anorganickej zložky (spravidla kationy kovov) a pod. Posttranslačné úpravy, konformačné zmeny (alosterické proteíny), alternatívny zostrih a ďalšie procesy vysvetľujú, prečo je počet rozličných proteínov v bunkách spravidla rádovo vyšší, než počet kódujúcich génov. Napr. u človeka sa počet génov odhaduje na cca 30000, počet RNA transkriptov na cca 100000, ale počet proteínov na cca 1000000. Výskyt a množstvo konkrétnych proteínov môže byť modifikovaný často v závislosti na prostredí, v ktorom sa jedinec vyskytuje. Preto sa len určitá časť proteínov s konštantnou úrovňou expzie a stabilnou štruktúrou hodí ako identifikačné markéry.

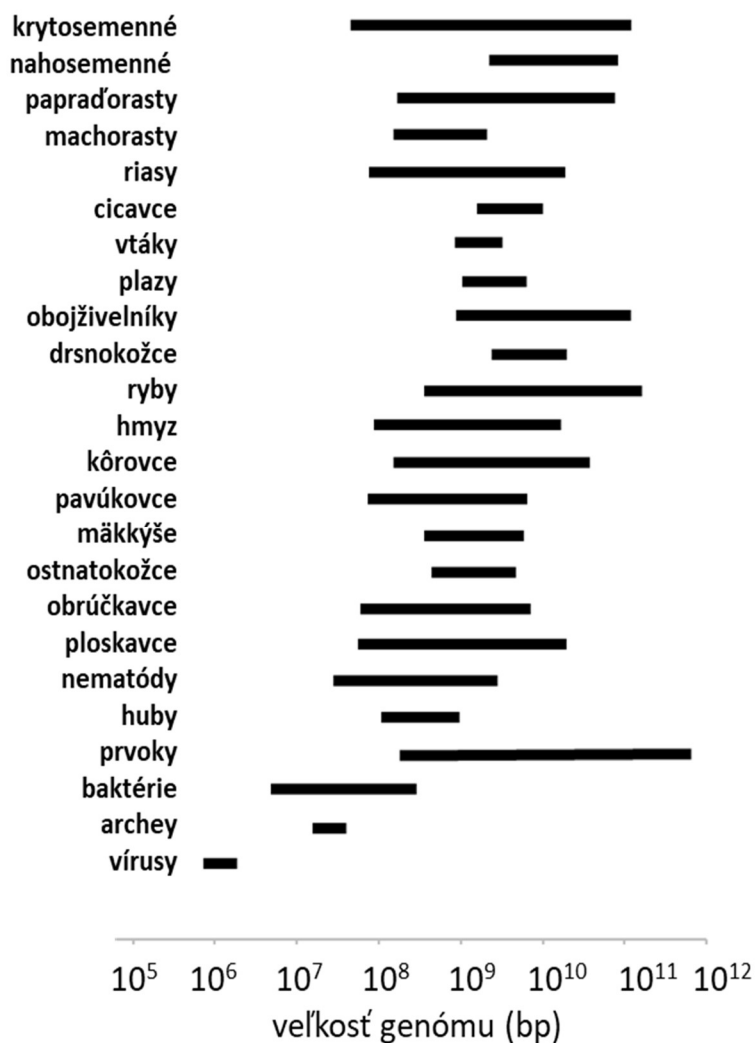
Genóm

Celková genetická informácia bunky (kódujúce aj nekódujúce úseky) sa označuje termínom genóm. Veľkosť genómu a aj počet génov v jeho rámci závisí čiastočne od komplexnosti organizmu, ale čiastočne je výsledkom evolúcie genómu, organizmy s rovnakou zložitosťou stavby tela môžu mať zásadne rozdielne veľkosti genómu (obr. 15, tab. 3).

Nie všetky nukleotidy v reťazci DNA majú informačný význam. Sekvencie, ktoré sú exprimované do fenotypových znakov predstavujú u mnohobunkových eukaryotov len zlomok z celkovej dĺžky DNA (tab. 3). Súčasťou genómu sú predovšetkým rôzne typy repetitívnych (opakovaných) sekvencií. Ich podstatnú časť tvoria transponibilné elementy (transpozóny), ktoré predstavujú „molekulárne parazity“, pravdepodobne pôvodne vírusy (RNA vírusy pri

Tab. 3 Veľkosť genómu vybraných organizmov

Organizmus	Veľkosť genómu (bp)	Odhadovaný počet génov
Vírus λ	48500	50
<i>Escherichia coli</i>	$4,6 \cdot 10^6$	4300
<i>Saccharomyces cerevisiae</i>	$1,3 \cdot 10^7$	6200
<i>Aspergillus fumigatus</i>	$1,58 \cdot 10^7$	14000
<i>Caenorhabditis elegans</i>	$9,7 \cdot 10^7$	19000
<i>Drosophila melanogaster</i>	$1,8 \cdot 10^8$	13600
<i>Mus musculus</i>	$2,7 \cdot 10^9$	22–30000
<i>Homo sapiens</i>	$2,9 \cdot 10^9$	28–35000
<i>Arabidopsis thaliana</i>	$1,25 \cdot 10^9$	25500
<i>Populus trichocarpa</i>	$4,85 \cdot 10^8$	45000
<i>Zea mays</i>	$2,2 \cdot 10^9$	42–56000
<i>Triticum aestivum</i>	$1,6 \cdot 10^{10}$	107000
<i>Pinus sylvestris</i>	$2,5 \cdot 10^{10}$	30000



Obr. 15 Rozsah veľkostí genómu jednotlivých vyšších taxónov

Tab. 4 Charakteristika genómu smreka obyčajného (*Picea abies* Karst.) (Nystedt et al. 2013)

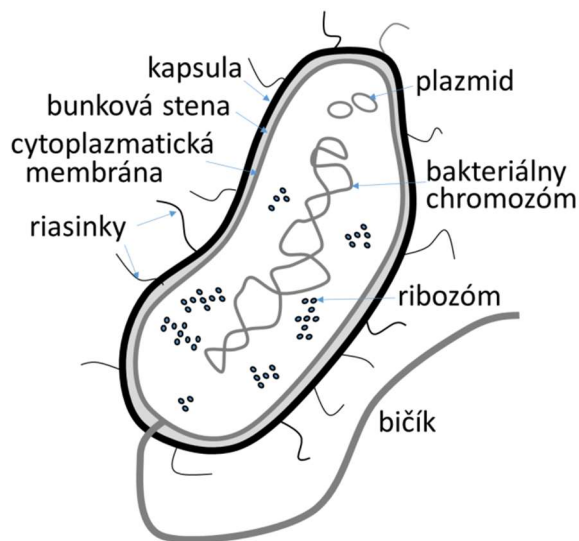
Charakteristiky genómu	
Veľkosť haploidného genómu	19,6 Gbp
Karyotyp	2n = 24
Podiel párov GC	37,9%
Repetitívne sekvencie	
LTR retrotranspozóny, z toho	
<i>Gypsy</i>	35%
<i>Copia</i>	16%
ostatné	7%
LINE retrotranspozóny	1%
DNA transpozóny	1%
Ostatné	10%
Sekvencie génov	2,4%
Anotácia genómu	
Počet génov	28354
Priemerná dĺžka exónov	312 bp
Priemerná dĺžka intrónov	1017 bp
Priemerná hustota génov	1,418 Mbp ⁻¹
Počet transpozónových sekvencií	284587

retrotranspozónoch, DNA vírusy pri DNA transpozónoch), ktoré sa v priebehu evolúcie zabudovali do genómu hostiteľa. Sú schopné presúvať sa na nové miesta v genóme a zmnožovať sa, a obsahujú gény resp. sekvencie, ktoré im presun a zmnožovanie umožňujú (počet transpozónových sekvencií často násobne prevyšuje počet vlastných génov organizmu, vid' tab. 4). Napríklad element *mariner* sa vyskytuje u väčšiny živočíchov (v genóme človeka sa počet kópií odhaduje na 14000 s celkovou dĺžkou 2,6 Mbp) a dokonca aj u niektorých prvokov. Ďalšiu skupinu repetitívnych sekvencií predstavujú tandemové opakovania krátkych sekvenčných motívov (minisatelity a mikrosatelity).

Štruktúra bunky a bunkový cyklus

Prokaryotická bunka

Bunka prokaryotov (archeí a baktérií) je relatívne jednoduchý útvar. Od prostredia je oddelená bunkovou membránou zloženou z dvojvrstvy fosfolipidov, a bunkovou stenou, ktorá je pri baktériách tvorená mureínom (peptidoglykán; polymér sacharidov a aminokyselín), u archeí pseudomureínom (analogický komplex s mierne odlišnou chemickou štruktúrou). Cytoplazma nie je rozdelená na kompartmenty, teda bunka nemá diferencované jadro ani organely. DNA je obsiahnutá v tzv. bakteriálnom chromozóme (nukleoide), a je tvorená jednou kruhovou molekulou DNA, ktorá je voľne uložená v cytoplazme. Okrem nej sa v bunkách prokaryotov často nachádzajú plazmidy, malé kruhové molekuly DNA, ktoré sú tiež obsahujú gény schopné expresie (obr. 15).



Obr. 15 Schéma štruktúry typickej bakteriálnej bunky

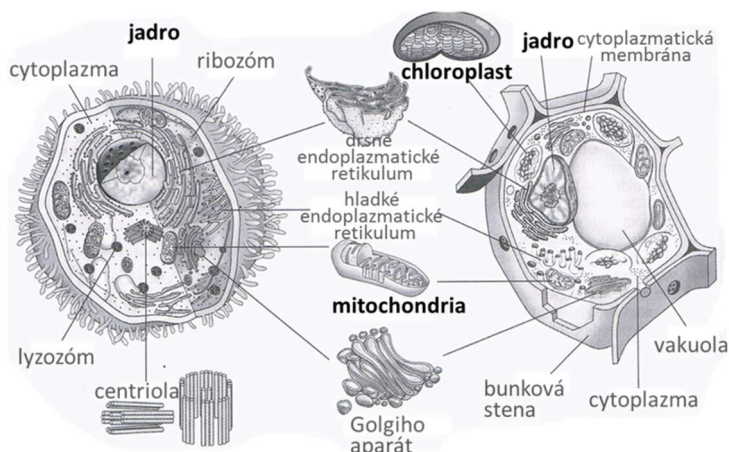
Prokaryoty sú spravidla jednobunkové, niektoré sú schopné vytvárať mnohobunkové kolónie (cyanobaktérie, myxobaktérie). Aj keď v rámci kolónií môže dochádzať k čiastočnej diferenciácii buniek, nevytvárajú špecializované zoskupenia (pletivá, orgány, organizmy) porovnateľné s mnohobunkovými eukaryotmi. Ich bunkový cyklus je relatívne jednoduchý. Bunky rastú až do dosiahnutia kritickej veľkosti. Následne replikujú svoju DNA: replikácia začína na špecifickom mieste na chromozóme (lokus *ori*), v ktorom je molekula DNA prichytená na bunkovú membránu. Syntéza vonkajšej membrány prebieha paralelne s replikáciou, fosfolipidové molekuly dopĺňané medzi oba lokusy *ori* s postupujúcou replikáciou odťahujú dcérske molekuly DNA k opačným pólom bunky. Bunka sa potom fyzicky rozdelí vytvorením deliacej priehradky (septum; bunková membrána a bunková stena). Rozdelenie buniek (cytokinéza) začína vchlípením plazmatickej membrády po obvode bunky, pričom novosyntetizovaný materiál septa je postupne pridávaný v rovine buncného delenia. Tento proces sa označuje ako binárne delenie. Druhým mechanizmom je pučanie: na

jednom konci materskej bunky sa vytvorí púčik, ktorý postupne narastá, a keď dorastie do veľkosti materskej bunky, oddelí sa.

Eukaryotická bunka

Štruktúra bunky

V porovnaní s prokaryotickou bunkou je bunka eukaryotov komplikovaný a vysoko organizovaný útvar. Má samostatné jadro, oddelené od cytoplazmy fosfolipidovou jadrovou membránou, ktoré obsahuje najväčšiu časť DNA. Eukaryotická bunka je vnútorne rozdelená na kompartmenty a obsahuje množstvo organel, ktoré predstavujú relatívne samostatné súčasti so špecializovanými funkciami. Špecifický typ organel predstavujú (semi)autonómne organely, ktoré sú vybavené vlastnou DNA. Patria k nim mitochondrie, zabezpečujúce aerobnú respiráciu a produkujúce ATP ako hlavný zdroj energie pre biochemické procesy. Ďalšiu skupinu autonómnych organel, ktorá sa vyskytuje len v bunkách rastlín, rias a u niektorých skupín jednobunkových organizmov tvoria plastidy, predovšetkým chloroplasty, vykonávajúce fotosyntézu. Predstavujú zbytky endosymbiotických baktérií (v prípade mitochondrií α -proteobaktérií, v prípade chloroplastov cyanobaktérií), ktoré eukaryotická bunka v dávnej evolúcii pohltila a ktoré sa stali jej súčasťou. O bakteriálnom pôvode svedčí štruktúra ich molekuly DNA, ktorá je rovnako ako u baktérií kruhová. V priebehu evolúcie sa časť génov súvisiacich s aktivitou organel presunula do jadra, v súčasnosti je funkčnosť organel závislá na interakcii s jadrovými génmi. Autonómne organely sú schopné autoreprodukcie; nevznikajú *de novo*, ale delením existujúcich organel (obr. 16). Mechanizmus delenia u nich je rovnaký ako u prokaryotickej bunky.



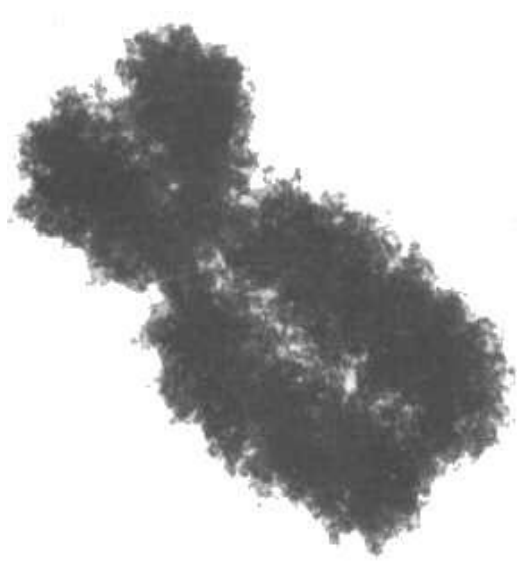
Obr. 16 Schematické zobrazenie živočíšnej (vľavo) a rastlinnej (vpravo) bunky (www.biocyclopedia.com, upravené)

Chromozóm

Rozhodujúcim znakom eukaryotickej bunky je prítomnosť jadra, ktoré predstavuje samostatný útvar odelený od cytoplazmy jadrovou membránou a obsahuje genetický materiál. Molekula DNA v jadre eukaryotických buniek je na rozdiel od prokaryotov lineárna, teda nie je uzavretá. Je organizovaná do vyšších štruktúr, aby s ňou bunkový aparát mohol pohybovať pri delení bunky. Je navinutá na nukleozómy, malé komplexy histónových proteínov s priemerom 11 nm (8 molekúl histónov vytvára nukleozóm). Nukleozómy sú následne spakované (hyperšpiralizované) do 30 nm hrubého chromatinového vlákna, ktoré vytvára slučky s dĺžkou cca 300 nm. V jednej slučke je cca 1000 nukleozómov. V tejto štruktúre je molekula DNA uložená v jadre v interkinéze, teda 'kludovom' období, keď sa bunka nedelí. Počas delenia dochádza k ďalšej kondenzácii vlákna DNA vytvára útvar s hrúbkou cca 0,7 μm a dĺžkou rádovo niekoľko mikrometrov označovaný ako chromozóm (obr. 17). Časti DNA kódujúce aktívne prepisované gény sú spakované voľnejšie a tvoria tzv.

euchromatín, naopak neaktívne časti sú tesnejšie asociované s podpornými proteínmi a tvoria heterochromatín.

Na chromozóme sú pozorovateľné niektoré typické útvary. V rámci chromozómu sa nachádza úsek označovaný ako centroméra, ktorý sa javí ako zúžené miesto. Pri bunkovom delení sa molekula DNA musí replikovať, chromozóm sa v tomto štádiu skladá z dvoch sesterských chromatíd, ktoré zostávajú spojené v centromére až do delenia, chromozóm má teda tvar X (obr. 17). Centroméra delí chromozóm na dve ramená, ktoré môže byť približne rovnaké (metacentrický chromozóm) alebo je centroméra umiestnená v blízkosti teloméry (akrocentrický chromozóm), prípadne na konci chromozómu (telocentrický chromozóm). Ramená sú ukončené telomérami, ktoré chránia chromozóm pred postupným odbúraním pri každej replikácii. Teloméry majú špecifickú štruktúru a podobne ako centroméra sú tvorené tandemovo opakovanými neexprimovanými sekvenciami.



Obr. 17 Snímka z elektrónového mikroskopu zobrazujúca chromozóm v štádiu maximálnej kondenzácie, počas metafázy mitotického delenia (<http://bioweb.wku.edu/courses/Biol115/wyatt/wku/mitosisa.htm>)

Počet chromozómov v jadre bunky určuje jej ploidiu. U eukaryotov dochádza ku striedaniu haploidnej generácie (bunky ktorej obsahujú jednu sadu chromozómov, a diploidnej generácie, kde obsahujú dve sady chromozómov. U živočíšnych druhov je haploidná generácia zredukovaná na jednu bunku (pohlavná bunka; gaméta). Pri väčšine rastlín je haploidná generácia mikroskopický viacbunkový útvar (gametofyt), pričom jedna z buniek plní funkciu gaméty. Splynutím samčej a samičej gaméty pri oplodnení vzniká diploidná zygota, z ktorej pri živočíchoch a väčšine rastlín vznikajú postupným bunkovým delením somatické bunky tvoriace samotný makroskopický organizmus. Výnimku tvoria napr. machorasty, kde makroskopické štádium predstavuje haploidný gametofyt.

U diploidných organizmov somatické bunky obsahujú po dvoch exemplároch každého chromozómu, teda dve kópie každej molekuly DNA a teda aj po dve kópie každého génu resp. nekódujúcej sekvencie. Chromozómy tvoriace takúto dvojicu sa označujú ako homologické, spravidla obsahujú rovnakú sadu génov (ale nie nutne rovnaké varianty každého génu) resp. iných sekvencií, v rovnakom usporiadaní. Organizmy ale môžu vykazovať aj vyššie stupne ploidiu, t.j. môžu obsahovať viac ako dve chromozómové sady; tento jav sa označuje všeobecne ako polyploidia.

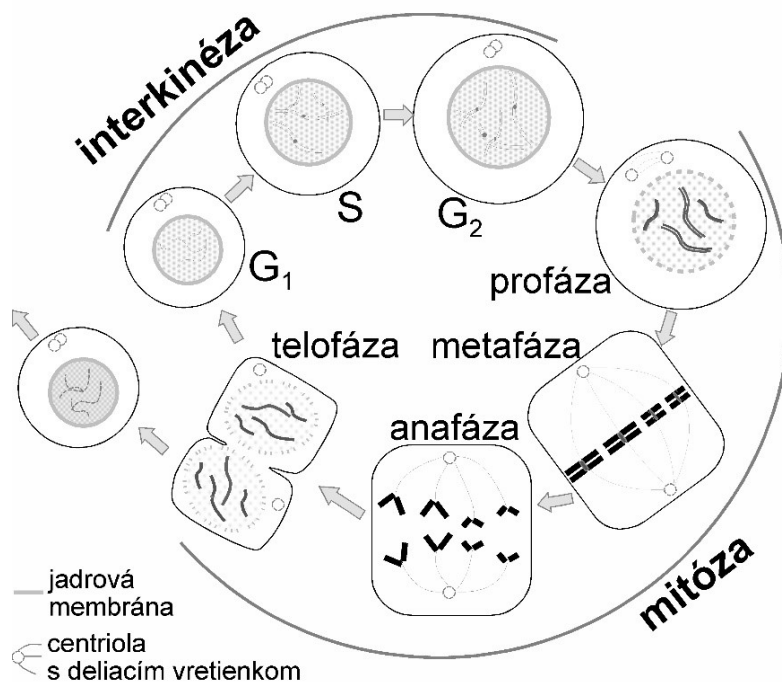
Na rozdiel od jadrovej DNA organelárna DNA sa dedí iba od jedného z rodičov, každá bunka teda obsahuje len po jednej kópii každého mitochondriálneho či chloroplastového génu. Autonómne organely sú samozrejme obsiahnuté v oboch gamétach (samičej aj samčej), ale po oplodnení sú organely od „nesprávneho“ rodiča odbúrané a bunka si ponecháva len organely jedného z rodičov. Vo veľkej väčšine prípadov je tým rodičom matka, ale existujú

aj výnimky; napr. v celej čeľadi Pinaceae (všetky európske hospodársky významné ihličnany) sa chloroplastová DNA dedí po otcovi.

Bunkový cyklus a mitóza

Jednobunkové eukaryoty sa často rozmnožujú aj klonálne, nepohlavnou cestou. U mnohobunkového organizmu je rast nutne spojený so zväčšovaním počtu somatických (telových) buniek, čo si vyžaduje ich delenie. V oboch prípadoch každá novovytvorená dcérska bunka musí získať celú a nezmenenú dedičnú informáciu, ktorá bola obsiahnutá v materskej bunke. Takéto rozdelenie dedičného materiálu zabezpečuje mechanizmus delenia, označovaný ako mitóza.

Bunka eukaryotov v priebehu svojho života prechádza viacerými fázami (obr. 18). Štádium medzi dvomi mitotickými deleniami sa označuje ako interfáza alebo interkinéza, počas ktorej sú molekuly DNA relaxované, rozvinuté, a svetelným ani elektrónovým mikroskopom v jadre nie sú viditeľné. Nové bunky vzniknuté delením obsahujú po jednej kópii každej molekuly jadrovej DNA (t.j. každý chromozóm je tvorený jednou dvojzávitnicou DNA navinutou na nukleozómy). Toto štádium sa označuje ak G_1 (z angl. *gap*, t.j. medzera), v jeho priebehu bunky rastú, zväčšujú svoj objem. Počas G_1 dochádza k expresii génov a prebieha enzymatická aktivita nutná pre syntézu stavebných a ďalších látok nutných pre život bunky. Následne bunka vstupuje do fázy S (*synthesis*), v ktorej dochádza k replikácii (syntéze) nových molekúl DNA. Na konci S fázy sa každý chromozóm skladá z dvoch sesterských chromatíd, teda dvoch dcérskych, replikovaných molekúl DNA, ktoré zostávajú spojené v mieste centroméry; spojenie sprostredkujú špecifické proteíny, viažuce sa na tandemovo opakované sekvencie v centromére. Expresia génov je utlmená s výnimkou génov pre históny, nutné pre stabilizáciu vytvorených dcérskych DNA molekúl. Ďalšou fázou interkinézy je G_2 fáza, v rámci ktorej d'alej narastá objem bunky, obnoví sa expresia génov a tým aj enzymatická aktivita nutná pre ďalšiu syntézu látok potrebných pre fungovanie a stavbu bunky. G_2 je ukončená mitózou.



Obr. 18 Schéma priebehu eukaryotického bunkového cyklu. Zobrazenie interfázových chromozómov treba brať len ako ilustračné, v skutočnosti nie sú v jadre viditeľné.

Mitotické delenie prebieha v štyroch fázach. Keďže v S-fáze prebehla replikácia molekúl DNA, bunka dočasne obsahuje po dvoch kópiách každej molekuly DNA (každý chromozóm tvoria dve chromatidy), ktoré zostávajú spojené v oblasti centroméry. V prvej fáze mitózy, profáze, sa rozpadá jadrová membrána a chromozómy začínajú kondenzovať (nukleozómy, na ktorých je DNA navinutá, sa začínajú zhlukovať a organizovať do vyšších štruktúr) a vytvárajú sa z nich viditeľné vlákna. Zároveň sa z centriol umiestnených na póloch bunky začínajú predlžovať vlákna deliaceho vretienka, tvorené pružnou bielkovinou, aktínom. V metafáze sa chromozómy usporiadajú do ekvatoriálnej („rovníkovej“) roviny bunky a vlákna deliaceho vretienka sa ukotvia na proteíny spájajúce chromatidy v centromére. V anafáze sa aktínové vlákna deliaceho vretienka začínajú skracať, čím oddelia sesterské chromatidy (repliky molekúl DNA) každého chromozómu od seba a každú z nich priťahujú k opačnému pólu bunky. Vďaka tomu sú v telofáze kompletne sady molekúl nahromadené každá pri opačnom póle bunky. Chromozómy sa opäť dešpiralizujú a okolo každej sady sa obnovuje jadrová membrána, čím je ukončené delenie jadier. Následne počas cytokinézy dochádza k deleniu cytoplazmy sformovaním deliacej priehradky medzi jadrami (lipidová membrána) a následne sformovaním bunkovej steny. Počet molekúl jadrovej DNA v bunke počas bunkového cyklu sa teda mení $2n \rightarrow 4n \rightarrow 2n$. Autonómne bunkové organely mnohobunkových eukaryotov sa počas cytokinézy (rozdelenia buniek po ukončení mitotického delenia jadier) rozdelia do oboch dcérskych buniek náhodne. Vzhľadom na veľký počet mitochondrií (v priemere 200) aj chloroplastov (20–100) v bunke sa prakticky s istotou dostanú do oboch dcérskych buniek. V nich sa ďalej množia rovnakým spôsobom ako ich prokaryotickí predchodcovia: binárnym delením alebo pučaním. Jednobunkové eukaryoty často obsahujú mitochondrie resp. chloroplasty len v jedinom exemplári, ich rozmnožovanie je synchronizované s bunkovým cyklom. V niektorých prípadoch nedôjde po delení jadier k cytokinéze, takže vznikajú bunky s viacerými jadrami. Multinukleárne bunky sú bežné u húb, v niektorých tkanivách živočíchov (napr. pečeň stavovcov), ale nachádzajú sa aj v apikálnych meristémoch rastlín.

V mnohobunkovom organizme sa trvale delia len bunky rastových meristémov resp. kmeňové bunky (a v prípade patogenézy aj rakovinové bunky). S postupnou diferenciáciou (teda nadobúdaním špecifickej funkcie, ktorú bunka v organizme plní) sa schopnosť delenia stráca a bunka (spravidla) po ukončení G_1 vstupuje do kludového stavu (G_0), v ktorom bunka plní svoju funkciu v organizme až do ukončenia svojej životnosti apoptózou (programovaná smrť) spojená s organizovanou deštrukciou bunky) alebo nekrózou („násilná“ smrť bunky) v prípade poškodenia).

Z hľadiska forenzných analýz je relevantným aspektom mechanizmov replikácie a mitotického delenia skutočnosť, že zaisťujú, aby každá bunka v tele jedinca obsahovala v princípe presne rovnakú genetickú informáciu. Materiál odobraný z ktorejkoľvek časti tela (vrátane biologickej stopy, ktorú jedinec zanechá) je teda geneticky identický. Pri replikácii dochádza k presnému kopírovaniu poradia nukleotidov v materskej molekule DNA do oboch dcérskych. Pochopiteľne, mechanizmus replikácie nie je dokonalý, dochádza pri ňom k chybám (viď kap. Mutácie). Aj keď bunka má reparačné mechanizmy na vyhľadávanie a odstraňovanie týchto chýb, časť z nich sa v molekulách DNA uchová a ďalej prenáša pri bunkovom delení. Tieto chyby sú však zriedkavé a náhodne rozdelené v genóme. Ak teda dôjde napr. k substitúcii nukleotidu na konkrétnej pozícii v jednej molekule DNA v konkrétnej bunke, táto sa prenáša len do dcérskych molekúl, ktoré vznikli jej replikáciou, teda dedí sa len v bunkovej línii, ktorá vznikla delením postihnutej bunky. Iné bunky túto konkrétnu chybu obsahovať nebudú. V DNA izolovanej z tkaniva či biologickej stopy (rádovo z tisícov až miliónov buniek, v závislosti na množstve materiálu) budú teda molekuly s touto konkrétnou bodovou mutáciou predstavovať zanedbateľný podiel, ktorý neovplyvní výsledok analýzy. Skutočnosť, že po replikácii DNA zostávajú obe dcérske molekuly fyzicky

spojené v mieste centroméry, a že mechanizmus mitózy zabezpečí, že každá z nich je odtiahnutá k opačným pólom bunky, zase zaisťuje, že každá bunka nezávisle na počte cyklov delenia, ktorými prešla, bude obsahovať kompletnú sadu molekúl DNA.

Meióza, väzba génov

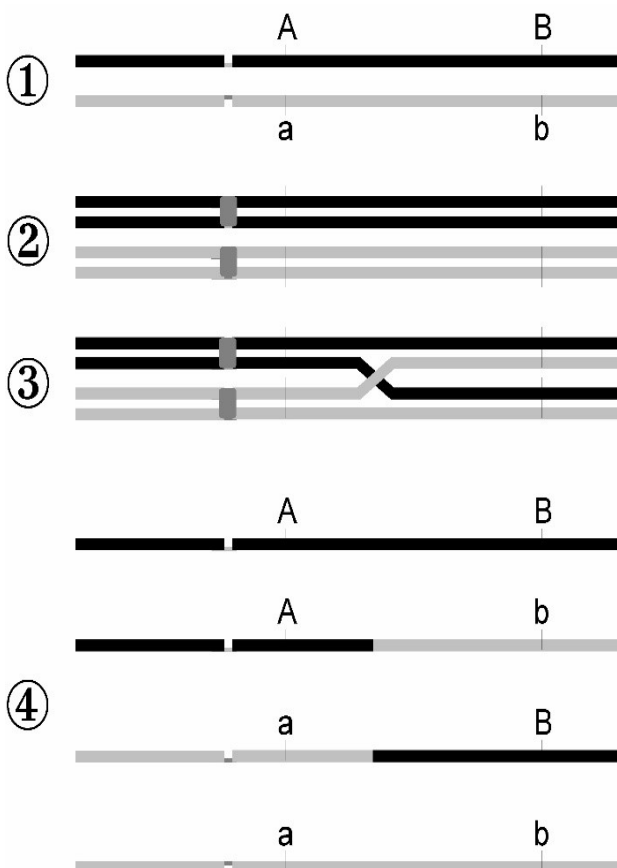
Gaméty (pohlavné bunky) musia na rozdiel od somatických buniek obsahovať len po jednom exemplári každej molekuly DNA. Ak by gaméty boli diploidné, počet chromozómov v jadre by sa každým cyklom reprodukcie (pri každom vzniku zygoty) exponenciálne zvyšoval ($2n \rightarrow 4n \rightarrow 8n \rightarrow 16n \rightarrow \dots$). Rozdelenie dvojice chromozómových sád (diploidného počtu) v materskej somatickej na haploidné sady v novovznikajúcich gamétach zabezpečuje mechanizmus delenia buniek nazývaný meióza. Meióza predstavuje sled dvoch delení jadra s analogickým priebehom fáz ako pri mitóze, k redukcii počtu chromozómov dochádza pri prvom delení. Rovnako ako pri mitotickom delení musí ešte v interkinéze dôjsť k replikácii DNA, opäť sa vytvoria dve chromatidy navzájom spojené v centromére. Profáza prvého delenia (profáza I) je časovo predĺžená a pozostáva z piatich štádií. V leptoténe chromozómy kondenzujú. V ďalšom štádiu, zygoténe, sa homologické chromozómy navzájom párujú, spájajú sa prostredníctvom proteínového spojovacieho komplexu; tento proces sa označuje ako synapsa, počas ktorej sa chromozómy kondenzujú do stále kratších a hrubších útvarov (pre pripomenutie: homologické sú chromozómy, ktoré obsahujú rovnaké gény resp. iné sekvencie, usporiadané v rovnakom poradí; jeden z dvojice homologických chromozómov jedinec získava od matky, druhý od otca. Homologické chromozómy teda obsahujú rovnaké gény, ale nie nutne rovnaké varianty týchto génov). V ďalšom štádiu, pachyténe, spárované chromozómy vytvárajú tvar označovaný ako bivalent resp. tetráda, tvorený 4 molekulami DNA, z ktorých vždy dve a dve sú identické repliky (sesterské chromatidy). Počas pachyténu sa môžu nesesterské chromatidy (teda chromatidy dvoch rozdielnych homologických chromozómov) prekrižiť, spojiť, a navzájom si vymeniť časti. Tento jav sa označuje anglickým termínom crossing-over (slov. prekríženie) a vedie k rekombinácii alel, teda vzniku takých kombinácií alel, ktoré sa na pôvodných chromozómoch materskej bunky nevyskytovali. Ku crossing-overu môže dôjsť aj medzi sesterskými chromatidami, ale tam nemá žiadne genetické následky, keďže sesterské chromatidy sú identické. Prekríženie je dobre viditeľné počas ďalšieho štádia profázy I, diploténu, počas ktorého sa spárované chromozómy trochu od seba oddialia, ale zostávajú spojené v miestach, v ktorých došlo ku crossing-overu a ktoré sa označujú termínom chiazma (keďže pripomínajú grécke písmeno χ – χ ; obr. 19). V poslednom štádiu profázy I, v diakinéze, chromozómy naďalej kondenzujú a pohybujú sa k ekvatoriálnej rovine bunky. Jadrová membrána sa rozpadá a začína sa vytvárať deliace vretienko. V metafáze I sa páry homologických chromozómov usporiadajú do ekvatoriálnej roviny tak, že ich centroméry smerujú každá k opačnému pólu bunky, a ukotvia sa na ne aktínové vlákna deliaceho vretienka. Ich ťahom vlákien sa chiazmy postupne od centroméry smerom k teloméram oddeľujú, k definitívnemu rozdeleniu dôjde počas anafázy I, keď sa homologické chromozómy (stále tvorené dvoma chromatidami, ktoré ale môže mať rekombinované segmenty) od seba odtrhnú a sú odtiahnuté k opačným pólom bunky. Počas anafázy I teda dochádza k redukcii počtu chromozómov; po jednej sade chromozómov sa nahromadí pri každom z dvoch pólom bunky. Priebeh telofázy I závisí od druhu: pri niektorých organizmoch deliace vretienko zanikne, obnoví sa jadrová membrána okolo oboch dcérskych jadier a chromozómy opäť dekondenzujú, pri iných sa nové jadrá nediferencujú, chromozómy sa dešpiralizujú len čiastočne, a dcérske bunky priamo vstupujú do ďalšieho delenia. V každom prípade výsledkom I. meiotického delenia je vznik dvoch dcérskych buniek, ktoré obsahujú po dvoch sádach molekúl DNA. Meióza II má priebeh analogický ako mitóza. Počas profázy II chromozómy opäť kondenzujú a pripájajú sa k vláknam nového deliaceho vretienka, v metafáze II sa usporiadajú do novej ekvatoriálnej

roviny (kolmej na pôvodnú), v anafáze II sú sesterské chromatidy priťahované každá k opačnému pólu bunky, a v telofáze II sa zhlukujú pri opačných póloch, dekondenzujú, a tvorbou jadrových a následne cytoplazmatických membrán sa tvoria samostatné dcérske bunky. Počet molekúl DNA v bunke počas gametogenézy sa teda mení $2n \rightarrow 4n \rightarrow 2n \rightarrow 1n$. Na rozdiel od mitózy, produktom ktorej sú dve nové geneticky identické dcérske bunky (majú rovnaký genotyp ako mala materská bunka), meiózou sa vytvárajú štyri haploidné bunky, ktoré nie sú geneticky identické (majú rozdielny haplotyp). Dvojice homologických chromozómov, ktoré nesú tie isté gény, ale môžu niesť rozdielne varianty (alely) týchto génov, sa pri meióze rozchádzajú do rôznych gamét, tento proces sa označuje ako segregácia chromozómov resp. génov pri gametogenéze (obr. 20). Gény, ktoré sa nachádzajú na rozdielnych chromozómových pároch, sa kombinujú náhodne, nezávisle na tom, či konkrétny variant zdedil jedinec od otca alebo od matky.



Obr. 19 Bivalent tvorený štyrmi chromatidami dvoch homologických chromozómov v štádiu diploténu

(http://www.scilproj.org/IBHbio2_knowledge.html)



Obr. 20 Schéma segregácie a rekombinácie génov pri meióze. ① dvojica homologických chromozómov v G_1 -fáze interkinézy diploidnej zárodočnej bunky. Jedinec je heterozygotný v génoch A a B , ktoré sú vo vzájomnej väzbe: obe dominantné alely (AB) sa nachádzajú na jednom chromozóme, obe recesívne alely (ab) na druhom. Pre názornosť sú homologické chromozómy a ich časti zobrazené rôznymi farbami. ② stav po S-fáze interkinézy: oba homologické chromozómy sú replikované, každý je tvorený 2 chromatidami, spojenými v mieste centroméry, znázornenej tmavosivým obdĺžnikom. ③ bivalent tvorený replikovanými homologickými chromozómami v pachyténe profázy I: dve nesesterské chromatidy sa prekrížili medzi génmi A a B a vymenili si medzi sebou úseky od miesta prekríženia až po teloméry. ④ stav po telofáze II: vznikli 4 haploidné bunky, z ktorých dve majú pôvodnú *cis*-konšteláciu alel (AB a ab) a dve rekombinovanú *trans*-konšteláciu alel (Ab a aB).

Chromozóm pri meióze zostáva jedným celkom. Gény, ktoré sa na ňom nachádzajú, sa preto nemôže voľne kombinovať, ale správajú sa ako jeden súbor. Tomuto javu hovoríme väzba génov (angl. *linkage*), a súbor génov, nachádzajúci sa na rovnakom chromozóme sa

označuje ako väzbová skupina. Jedinou možnosťou, ako sa môžu génu na jednom chromozóme rekombinovať, je crossing-over. Dvojica nesesterských chromatíd sa medzi dvomi génmi môže prekrížiť aj dvakrát, v tomto prípade sa obnoví pôvodná (nerekombinovaná) konštelácia alel, takisto môže dôjsť k viacnásobným prekríženiam medzi rôznymi dvojicami nesesterských chromatíd v rámci tetrády. V každom prípade pravdepodobnosť, že medzi dvomi génmi dôjde ku crossing-overu je tým väčšia, čím ďalej od seba sa na chromozóme nachádzajú. Táto skutočnosť sa využíva pri mapovaní polohy génov na chromozómoch. Miesto na chromozóme, kde sa konkrétny gén (alebo iná nukleotidová sekvencia) nachádza, sa označuje termínom lokus.

Chromozómová sada sa naopak pri delení buniek ako celok nespráva. Heterologické chromozómy sa môžu pri tvorbe gamét dostať do dcérskych buniek v ľubovoľnej kombinácii, preto gény lokalizované na rozdielnych chromozómoch sa rekombinujú voľne. Pri diploidných organizmoch (väčšina rastlín a živočíchov) je teda počet možných kombinácií chromozómov v gamétach 2^n (kde n je veľkosť haploidnej chromozómovej sady). V prípade buka lesného, ktorý má haploidný počet chromozómov $n = 12$, je teda aj pri zanedbaní možnosti rekombinácie génov crossing-overom možnosť vytvorenia $2^{12} = 4096$ haplotypových kombinácií v gamétach, tvorených jediným jedincom. V prípade ženy s 23 dvojicami homologických chromozómov (u muža pohlavné chromozómy nie sú homologické; viď nižšie) je tento počet $2^{23} = 8388608$.

Väzba génov je opäť relevantným faktorom z hľadiska forenzných analýz. Určenie pravdepodobnosti opakovania genotypu v populácii pri fingerprintingu alebo pravdepodobnosti príbuzenstva vychádza z predpokladu voľne sa kombinujúcich markérov. Pokiaľ sú markérové lokusy viazané, odhad výsledku nie je nestranný.

Gonozómy

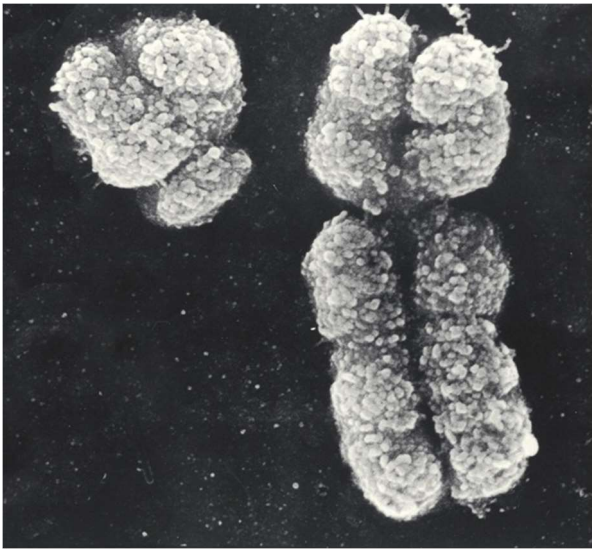
Väčšina chromozómov sa v diploidnej bunke vyskytuje v pároch, v rámci ktorých sú oba chromozómy plne homologické, označujú sa ako autozómy. Pri mnohých organizmoch je ovšem pohlavie určené prítomnosťou špecifických pohlavných chromozómov (gonozómov), ktoré sú na rozdiel od autozómov homologické len z malej časti alebo vôbec. V prípade cicavcov a ďalších organizmov (typ dedičnosti *Drosophila*) sa dvojica pohlavných chromozómov označuje ako chromozómy X a Y. Jedinec s konšteláciou gonozómov XY je samec (heterogametické pohlavie, vytvárajúce dva rôzne typy gamét, v tomto prípade spermii, určujúcich teda pohlavie potomstva), jedinec s konšteláciou XX je samica (homogametické pohlavie; samice cicavcov tvoria len jeden typ vajíčok).

Na rozdiel od dvojíc homologických autozómov, chromozómy X a Y majú odlišnú morfológiu. Chromozóm X je dlhší a metacentrický (centroméra je približne v strede), takže v mikroskope pripomína písmeno X; chromozóm Y je kratší a akrocentrický (centroméra je posunutá ku koncu chromozómu), tvar pripomína písmeno Y (obr. 21). Medzi dvojicou chromozómov X v bunke samice môže dochádzať ku crossing-overu, gény na nich sa môžu rekombinovať. V bunke sa ale nemajú ako nachádzať dva chromozómy Y, a v bunke samca (konštelácia XY) sa chromozóm Y nemôže kombinovať s chromozómom X. Chromozóm Y sa teda správa ako celok bez rekombinácie.

U vtákov, niektorých plazov, motýľov a ďalších organizmov (typ dedičnosti *Abraxas*) sa gonozómy označujú Z a W; homogametické pohlavie je samec (ZZ), heterogametické samica (ZW). U viacerých druhov hmyzu (šváby, rovnokrídlovce, niektoré blanokrídlovce) jeden chromozóm chýba úplne: jedinec s konšteláciou XX je samica, jedinec s konšteláciou X0 samec (autozómy u samca sú v pároch, ale gonozóm má len jeden: 0 znamená chýbajúci gonozóm), niektoré mory majú konšteláciu opačnú, teda ZZ je samec, Z0 samica. U sociálnych blanokrídlovcov (včely, osy, mravce), ale aj niektorých vošiek, sa vyskytuje haplodiploidné určenie pohlavia: z neoplodnených vajíčok sa liahnu haploidné samce

(na rozdiel od konstelácie XO majú aj autozómy len v haploidnej sade), z oplodnených samice.

Obr. 21 Snímka metafázového chromozómu Y (vľavo) a chromozómu X (vpravo) u človeka (www.berkeley.edu)



Dedičnosť fenotypových znakov

Autozomálna dedičnosť kvalitatívnych znakov

V kapitole *Expresia génu* bol gén definovaný ako úsek na molekule DNA, ktorý zodpovedá za nejaký fenotypový znak. Pojem 'fenotyp' sa vzťahuje na akúkoľvek vlastnosť, ktorú je možné na jedincovi pozorovať, hodnotiť či už kvalitatívne alebo kvantitatívne. Pre účely forenzných analýz je potrebné chápať pojem 'fenotypový znak' v najširšom možnom zmysle; nemusí sa nutne jednať o pozorovateľný (napr. morfológický) znak, ale rovnako môže ísť o krvnú skupinu či prítomnosť konkrétneho fragmentu DNA či frakcie proteínu na géle po elektroforetickej separácii. Pri takejto definícii bude pojem 'gén' zahŕňať nielen sekvencie DNA prepisované do RNA a prekladané do proteínov, ale aj nekódujúce sekvencie typu mikrosatelitov, minisatelitov, intrónov, spacerov atď.

Ako bolo spomenuté, somatické bunky väčšiny organizmov, ktoré sú potenciálnym predmetom forenzných genetických analýz, sú diploidné, teda obsahujú dve sady homologických molekúl jadrovej DNA (chromozómov), z ktorých jedna je zdedená od matky a druhá od otca. Z toho vyplýva, že veľká väčšina jadrových génov sa v genóme bunky vyskytuje v dvoch exemplároch, ktoré môžu ale nemusia byť identické, môžu ale nemusia predstavovať rovnaké varianty génu (*alely*). Jediniec, ktorý od oboch rodičov získal rovnaký variant génu (*alelu*), sa označuje ako homozygot v danom géne, nositeľ dvoch rôznych alel rovnakého génu je heterozygot. Kombináciu alel v konkrétnom géne označujeme termínom 'genotyp'. Súbor alel, ktoré sa nachádzajú na rovnakom úseku DNA a dedia sa z rodičov na potomkov ako jeden celok (t.j. bez rekombinácie) sa označuje termínom 'haplotyp'.

Jednotlivé gény a ich alelické varianty nemusia byť z hľadiska fenotypu rovnocenné, fenotypový prejav závisí nielen od prítomnosti alel v genotype, ale aj od ich vzájomného vzťahu. V prípade úplnej dominancie fenotypový účinok jednej alely prekrýva fenotypový účinok druhej. Alela, ktorá sa označuje ako dominantná, sa fenotypovo prejaví vždy, ak je v genotype jedinca prítomná v homozygotnom či v heterozygotnom stave. Fenotypový prejav recesívnej alely je dominantnou alelou u heterozygota potlačený, môže sa teda prejaviť len v homozygotnom stave. Pri úplnej dominancii je teda fenotyp dominantného homozygota a heterozygota totožný, odlišuje sa len fenotyp recesívneho homozygota. Príkladom môže byť Rh-faktor. Krvné skupiny vo všeobecnosti predstavujú antigény na povrchu červených krvi-

niek, väčšina z týchto antigénov sú proteíny. V prípade Rh-faktora ide o skupinu antigénov, najdôležitejší je antigén D – ak sa na krvinke vyskytuje, ide o skupinu Rh+, ak nie, ide o Rh-. Syntéza tohoto proteínu je riadená génom s 2 alelami, z ktorých jedna je funkčná (syntetizuje antigén D; rh^+), druhá nefunkčná (nesyntetizuje nič; rh^-). Nezáleží teda na tom, či má jedinec genotyp (kombináciu alel) rh^+/rh^+ (dominantný homozygot) alebo rh^+/rh^- (heterozygot), v oboch prípadoch má aspoň jednu alelu, ktorá je exprimovaná do antigénu D a má teda fenotyp Rh+. Ak má genotyp rh^-/rh^- (recesívny homozygot), nemá ani jednu alelu, ktorá by bola schopná syntetizovať antigén D, a teda má fenotyp Rh-.

V prípade kodominancie je možné u heterozygota rozoznať prítomnosť oboch alel v genotype. Príkladom môžu byť opäť krvné skupiny systému AB0: opäť ide o systém kontrolovaný jedným génom s tromi alelami I^A , I^B a O , pričom I^A a I^B sú kodominantné, kódujúce antigény A a B, a O je nefunkčná a preto recesívna. Ak je jedinec heterozygot $I^A I^B$, obe alely sú exprimované do dvoch rozdielnych antigénov A a B, jedinec má teda krvnú skupinu AB. V prípade krvnej skupiny A nie je možné rozoznať, či ide o homozygota $I^A I^A$ alebo heterozygota $I^A O$, v oboch prípadoch má aspoň jednu alelu produkujúcu antigén A a preto krvnú skupinu A. To isté platí pre krvnú skupinu B ($I^B I^B$, $I^B O$). Iba jedinec s genotypom OO má krvnú skupinu 0, teda žiadny antigén na povrchu krviniek.

V niektorých prípadoch je heterozygot fenotypovo intermediárny, teda predstavuje prechodný typ medzi oboma heterozygotmi. Napríklad u papuľky väčšej (*Antirrhinum majus*) existujú červenokveté a bielokveté odrody, ich hybrid (heterozygot) je ružový.

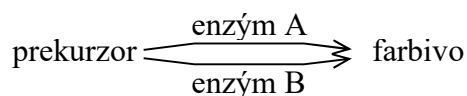
Pri kvantitatívnych znakoch je možné fenotypové hodnoty presne zmerať, a teda určiť aj mieru fenotypovej podobnosti či nepodobnosti jednotlivých genotypov. Pri neúplnej dominancii leží fenotypová hodnota heterozygota medzi fenotypmi oboch homozygotov. Ak dĺžka šišky jedného homozygota je 15 cm, druhého homozygota 10 cm a heterozygota 14 cm, tak alela pre väčšiu dĺžku šišky prejavuje istú mieru dominancie, ale nie úplnú: heterozygot je bližší homozygotovi s dlhšou šiškou, ale nie je s ním úplne totožný. Špecifickým prípadom neúplnej dominancie je aditivita – ak hodnota heterozygota je presne v strede medzi fenotypovými hodnotami oboch homozygotov (v tomto ilustračnom prípade by heterozygot mal dĺžku šišiek 12,5 cm), nie je možné rozlišovať dominantnú a recesívnu alelu, pretože žiadna z nich vo fenotype neprevažuje (nedominuje), ale jedna z alel zvyšuje fenotypovú hodnotu, druhá nie. Účinky alel sa sčítavajú. Pri superdominancii heterozygot prevyšuje svojím fenotypovým prejavom fenotypové hodnoty oboch homozygotov (ak by heterozygot v ilustračnom prípade mal dĺžku šišky napr. 20 cm, išlo by o superdominanciu).

Názvy génov sa spravidla označujú šikmým písmom (kurzívou). Existuje viacero konvencií pre označovanie alel. Jedna z nich je označovanie dominantných alel veľkým písmenom (A) a recesívnych alel malým písmenom (a), prípadne u multialelických lokusov rozlišovanie alel spodným indexom ($A_1, A_2, A_3...$). Druhá rozoznáva bežný typ v populácii (najčastejšia alela, ktorá je spravidla dominantná) označovaný ako 'wildtype' (wt); pričom zriedkavé mutácie (spravidla recesívne) sa buď označujú symbolom +, alebo sú rozlišované akronymom podľa fenotypového účinku (napríklad u psov mutácia spôsobujúca skrátenie spodnej čeľuste ako sh z angl. *short*). Pre označenie ľubovoľnej alely sa používa spodná pomlčka $_$: ak je genotyp označený ako A_- , znamená to, že jedna z alel je dominantná a druhá môže byť akákoľvek (dominantná či recesívna).

Niektoré znaky môžu byť kontrolované aj viac ako jedným génom. V tom prípade môžu existovať interakcie nielen medzi alelami v rámci génov, ale aj medzi génmi navzájom; tieto interakcie sa označujú termínom epistáza. Ak je napríklad nejaká látka v organizme (napr. červené farbivo listu) syntetizovaná z prekursora v dvoch reakciách, katalyzovaných enzýmami A a B, ktorých expresia je riadená génmi A a B :



potom pre vytvorenie farbiva je potrebné, aby prebehli obe reakcie. Ak dôjde k mutácii, ktorá znefunkční gén *A* (teda nesyntetizuje sa enzým A), nevytvorí sa medziprodukt, teda niet z čoho syntetizovať farbivo; ak dôjde k mutácii v géne *B* (t.j. chýba enzým B), z medziproduktu sa nebude vytvárať farbivo. Na syntézu farbiva je teda potrebné, aby fungovali oba gény. Defektná alela sa bude správať ako recesívna (u heterozygota je jedna z dvojice alel funkčná, teda priebeh reakcie zabezpečí). Vo všetkých genotypoch, kde je prítomná aspoň jedna dominantná alela v oboch génoch (*A_B_*) môžu prebiehať oba kroky syntézy farbiva, farbivo teda v liste bude prítomné, list bude červený. Ak je jedinec recesívne homozygotný v ktoromkoľvek géne (alebo v oboch), je zablokovaný prvý alebo druhý krok syntézy (alebo oba), syntéza farbiva teda neprebíha a list zostane zelený (tento prípad sa označuje ako dvojnásobná recesívna epistáza). Opačným príkladom môže byť, ak je to isté farbivo vytvárané z jedného prekursoru len v jednom kroku, ktorý ale môže prebiehať dvomi nezávislými metabolickými dráhami, riadenými dvomi génmi:



V tomto prípade bude farbivo syntetizované u každého genotypu, ktorý obsahuje aspoň jednu funkčnú (dominantnú) alelu ktoréhokoľvek génu (*A_B_*, *A_bb*, *aaB_*). Rozličné typy dvojgénovej epistázy a im zodpovedajúce štiepne pomery sú uvedené v tab. 5.

Tab. 5 Zhodnosť fenotypov pri rôznych typoch epistázy u génov s úplnou dominanciou (spoločná bunka značí, že fenotypy sú zhodné)

Účinok génov	Genotypy			
	<i>A_B_</i>	<i>A_bb</i>	<i>aaB_</i>	<i>aabb</i>
Komplementarita				
Dominantná epistáza				
Recesívna epistáza				
Kumulatívny účinok génov				
Dvojnásobná dominantná epistáza				
Dvojnásobná recesívna epistáza				

Dedičnosť konkrétneho fenotypového znaku u konkrétneho druhu sa zisťuje genetickou analýzou. Jej cieľom je identifikovať, koľko génov kontroluje daný znak, aké alelické varianty sa u nich vyskytujú, a akým spôsobom ovplyvňujú fenotyp. Pri genetickej analýze je možné použiť viacero postupov: kontrolované kríženie, analýzu rodokmeňa, pri rastlinách analýzu voľnoopeleného potomstva známeho materského jedinca (spol'ahlivá genetická analýza je možná len pri neúplnej dominancii alebo kodominancii), paralelnú analýzu haploidného endospermu a diploidného embrya nahosemenných rastlín (týka sa molekulárnych alebo biochemických znakov). V prípade markérov používaných pre forenzné analýzy je overenie dedičnosti predpokladom ich použiteľnosti v praxi.

Gonozomálna dedičnosť kvalitatívnych znakov

V prípade cicavcov gény na chromozóme X nemajú svoj náprotivok na oveľa kratšom chromozóme Y. Jedinice sú teda v takomto géne hemizygotné. Táto skutočnosť ovplyvňuje aj mechanizmus dedičnosti.

Aj v prípade pohlavných chromozómov platí, že defektné mutácie na chromozóme X (označujú sa spravidla X^+) majú tendenciu správať sa ako recesívne. Ak je nositeľom takejto defektnej mutácie samec, poškodenie sa u neho prejaví, keďže na chromozóme Y nemôže

mať dominantnú alelu, ktorá by vplyv defektnej mutácie prekryla (genotyp X^+Y). U samice sa defektná vloha prejaví, len ak je v nej homozygotná (X^+X^+), inak je len jej prenášačkou (X^+X).

Dedičnosť kvantitatívnych znakov

Názor na vzťah medzi génom a znakom sa v priebehu vývoja genetiky menil. Pôvodný náhľad 1 gén = 1 znak sa s rozvojom biochémie zmenil na 1 gén = 1 enzým, neskôr 1 gén = 1 polypeptid, a dnes sa zdá, že už ani táto hypotéza neplatí (viď informácie o alternatívnom zstrihu mRNA).

Gén môže v mnohých prípadoch súčasne ovplyvňovať viacero znakov, čo sa označuje ako pleiotropný účinok génu. Naopak, iné znaky môžu byť kontrolované polygénné, teda na kontrole sa podieľa viacero (niekedy aj niekoľko desiatok) génov, z ktorých každý prispieva svojou časťou (ktorá nemusí byť pri všetkých alelách rovnako veľká). Typická je táto situácia pri kvantitatívnych znakoch. Čím vyšší je počet kontrolujúcich génov, tým vyšší je počet genotypových kombinácií a teda aj fenotypových tried; naopak medzialelické a medzigénové interakcie počet fenotypových tried znižujú. Pri malom počte kontrolujúcich génov sa fenotypové triedy spravidla dajú rozoznať, ale s narastajúcim počtom génov sú ich hranice stále menej zreteľné a rozdelenie fenotypových tried sa postup mení z diskrétného na spojité. Využitie takýchto znakov pre akékoľvek forenzné aplikácie je značne obmedzené.

Mutácie ako zdroj dedičnej premenlivosti

Génové mutácie

Génové mutácie zasahujú jeden alebo niekoľko nukleotidov v rámci jedného génu, menia teda kvalitu génu. Zmena sekvencie môže spočívať v substitúcii jedného alebo viacerých nukleotidov za iný (iné), alebo v prírastku či úbytku počtu nukleotidov, teda zmene dĺžky sekvencie génu.

Zámena jedného nukleotidu za iný je označovaná ako bodová mutácia alebo substitúcia. Bodové mutácie sa rozlišujú na tranzície (zámena purínovej bázy za purínovú, teda $A \leftrightarrow G$, alebo pyrimidínovej za pyrimidínovú, teda $C \leftrightarrow T$; tento typ substitúcií je častejší) a transverzie (zámena purínovej bázy za pyrimidínovú alebo naopak, teda $C/T \leftrightarrow A/G$; zriedkavejší typ substitúcie). Takáto zámena môže nastať spontánne alebo byť indukovaná chemickým mutagénom, najčastejšie ide o mutagénne analógy báz (5-bromouracil, 2-aminopurín), alkylačné činidlá (yperit, alkylsulfáty) spôsobujúcimi alkyláciu báz a následné chybné párovanie, dusitany a reaktívne formy kyslíka (peroxydy, superoxid, ozón) spôsobujúce oxidačnú deamináciu adenínu, guanínu a cytozínu. Bodové mutácie spravidla zasahujú jediný nukleotid, a v závislosti na tom, na ktorej pozícii v rámci kodónu k nim dôjde, môžu viesť k zámene aminokyseliny zaradovanej do polypeptidového reťazca za inú. Ich dopad na fenotyp môže byť preto rozmanitý. Veľkú časť predstavujú synonymné mutácie (*synonymous mutation*), ktoré nemenia štruktúru produkovaného polypeptidu. Genetický kód je nadbytočný, tá istá aminokyselina môže byť kódovaná 1–6 rôznymi kodónmi. Zámena bázy na 3. pozícii kodónu často nemení jeho zmysel a vedie k zaradeniu rovnakej aminokyseliny do produkovaného polypeptidu. V iných prípadoch mutácia mení zmysel kodónu (*missense mutation*). Nie vždy takáto mutácia mení funkciu polypeptidu (viď príklad). Zmena funkcie závisí od chemických vlastností pôvodnej a zmenenej aminokyseliny, dopadu na trojrozmernú (terciárnu) štruktúru proteínu, dopadu na jeho afinitu voči iným štruktúram bunky (cytoplazma, fosfolipidové membrány a pod.) a teda lokalizácie v bunke atď. Proteín s pozmenenou štruktúrou môže dokonca nadobudnúť úplne novú funkciu. Bodové mutácie touto cestou prispievajú k tvorbe novej užitočnej genetickej premenlivosti. Vzhľadom na skutočnosť, že mutácia je náhodná zmena, je pochopiteľne jej efekt spravidla opačný: mutovaný polypeptid obvykle stratí schopnosť plniť akúkoľvek

funkciu. Poslednou možnosťou je zámena kódujúceho tripletu za niektorý z troch terminačných kodónov UAA, UAG, UGA (*nonsense mutation*), čo vedie k skráteniu polypeptidu a spravidla jeho nefunkčnosti.

Vsunutie alebo vypadnutie nukleotidov (inzercia alebo delécia) spôsobujú posun čítacieho rámca; takéto mutácie sa označujú ako posunové (*frameshift mutation*). Pokiaľ nevieme určiť, ktorá z alel je pôvodná a ktorá mutovaná, ako skratka pre posunovú mutáciu sa používa termín indel (*insertion/deletion*). V dôsledku posunu čítacieho rámca sa zmenia všetky aminokyseliny od miesta mutácie po koniec reťazca. Zároveň sa zmení pozícia terminačného kodónu (pôvodný zanikne a nový sa utvorí na inom mieste), v dôsledku čoho sa reťazec skráti alebo predĺži. Takáto masívna zmena štruktúry takmer vždy vedie k nefunkčnosti polypeptidu produkovaného mutovaným génom.

Chybovosť DNA polymeráz pri replikácii je rádovo 1 na 10^4 až 10^5 nukleotidov, ďalšie chyby môžu byť indukované exogénnymi faktormi (mutagénmi). Bunky si preto vytvárajú reparačné mechanizmy na vyhľadávanie a odstraňovanie týchto chýb. Časť z nich je odstránená priamo DNA polymerázou, ďalšie mechanizmy fungujú po ukončení replikácie. Reparačné mechanizmy znižujú výslednú frekvenciu mutácií na úroveň 10^{-10} – 10^{-11} na nukleotid a generáciu.

Mutácie spôsobujúce nefunkčnosť produkovaného proteínu majú tendenciu byť recesívne: každý jedinec má 2 kópie jadrového génu, takže ak jedna z nich nefunguje, druhá je spravidla schopná ju zastúpiť. Vo všeobecnosti sa nové mutácie v genofonde populácie spravidla dlho neudržia, vypadávajú v dôsledku náhodných procesov (genetického driftu) a pokiaľ sú škodlivé (čo je väčšina mutácií), ich frekvencia je prírodným výberom udržiavaná na nízkych hodnotách. Mutácie v nekódujúcich úsekoch (intróny, spacers, pseudogény, repetitívne sekvencie) neovplyvňujú životaschopnosť svojho nositeľa a nie sú preto predmetom prírodného výberu; majú preto vyššiu šancu zachovať sa v genofonde po dlhšiu dobu. Rôzne úseky DNA teda vykazujú rozdielnu mieru variability.

Z hľadiska forenzných analýz je otázka dopadu génovej mutácie na fenotyp irelevantná; pre identifikáciu jedinca či iné účely rovnako dobre slúžia synonymné či škodlivé mutácie ako užitočné.

Chromozómové mutácie

Chromozómové mutácie postihujú väčší úsek chromozómu, spravidla obsahujúci niekoľko génov. V tomto prípade sa teda nejedná o zmenu kvality, alebo o zmenu kvantity alebo usporiadania génov. Dlhší úsek chromozómu môže vypadnúť (delécia) alebo naopak byť zdvojený (duplikácia). Najjednoduchšie sa duplikujú tandemové opakovania génov, t.j. ak sa úsek už vyskytuje vo viacerých kópiách, ľahko k nim pribudne alebo ubudne ďalšia. Opakovanými duplikáciami vznikajú génové rodiny (*gene families*). Génová dávka, teda počet kópií génuv genóme, môže spoluurčovať biochemickú funkčnosť génu. Bodová mutácia môže poškodiť promotor niektorého z génov génovej rodiny, čo zabráni jeho expresii; takéto úseky sa označujú ako pseudogény.

Zmena pozície génu alebo úseku na chromozóme sa označuje ako *translokácia*, ktorá je spravidla dôsledkom aktivity transponibilných elementov (*transpozónov*), „molekulárnych parazitov“, ktoré sa v genóme často množia v obrovskom rozsahu (viď tab. 4). Ďalším typom chromozómovej mutácie je inverzia, pri ktorej je segment chromozómu vystrihnutý a vložený na rovnaké miesto s opačnou orientáciou.

Aj keď sú chromozómové mutácie významné z evolučného hľadiska, pre forenzné analýzy majú minoritný význam. Komplikáciu môže predstavovať existencia duplikovaných sekvencií (paralógov).

Genómové mutácie

Genómové mutácie postihujú celé chromozómy resp. celé chromozómové sady, menia ich počet a tým zásadne menia veľkosť genómu. Vznikajú v dôsledku nepravidelnosti bunkového delenia, ak sa chromozómy nerozídu správne do dcérskych jadier. Pri aneuploidii jeden exemplár chromozómu buď chýba (monozómia) alebo je navyše (trizómia). U živočíchov je aneuploidia spravidla letálna alebo vedie k závažným vývojovým poruchám; rastliny tolerujú aneuploidiu podstatne lepšie, nie vždy má zmena počtu chromozómov aj dopad na fenotyp.

Pri euploidii (polyploidii) je znásobená celá chromozómová sada. Polyploidia je u živočíchov výnimočná, u rastlín naopak pomerne bežná, mnohé rastlinné rody obsahujú sériu druhov s rozdielnymi stupňami ploidie. Takisto vysoký počet chromozómov niektorých rodov (*Tilia*, $n=81$) naznačuje ich polyploidný pôvod. Chromozómové sady polyploidov nemusia byť identické pôvodom. Pri autopolyploidii bunky obsahujú nadbytočné sady plne homologických chromozómov, spravidla v dôsledku toho, že nedôjde k rozdeleniu jadra alebo k cytokinéze. Allopolyploidia (alloplóidia) je spojená s hybridizáciou, jedinci obsahujú zmmozžené chromozómové sady, pochádzajúce od rozdielných rodičovských druhov. K vzniku alloplóidov môže dôjsť prostredníctvom mitózy (absencia rozdelenia diploidnej hybridnej zygoty) alebo meiotickou cestou (vnik neredukovaných diploidných gamét). Keďže chromozómy pochádzajúce od rôznych rodičovských druhov nie sú plne homologické a majú rozdielnu štruktúru, z funkčného hľadiska sa takéto jedince správajú ako diploidy (sú označované ako amfidiploidy) a sú spravidla plne fertílné. Tento proces sa môže opakovať aj viackrát a môže sa na ňom zúčastňovať viac ako dva rodičovské druhy.

Polyploidia (a do istej miery aj aneuploidia) predstavuje pre forenzné genetické analýzy problém: pokiaľ sa konkrétna markérová sekvencia vyskytuje v genóme viac ako dvakrát, môže vzniknúť technický problém pri určení presného genotypu. Navyše, väčšina matematických modelov, ktoré sú základom jednotlivých typov forenzných analýz, je formulovaná pre diploidné organizmy.

Genetika populácií

Populácia a jej štruktúra

Populácia je súbor jedincov rovnakého druhu, obývajúcich konkrétny biotop v konkrétnom čase, ktorí tvoria reprodukčnú jednotku, teda sú schopní sa navzájom medzi sebou krížiť za vzniku životaschopného a plodného potomstva. Napriek takejto jednoduchej definícii je praktické vymedzenie populácie v realite spravidla zložitú. Aj u druhov s fragmentovaným areálom si jednotlivé ostrovčeky (subpopulácie, demy) do určitej miery môžu vymieňať gény pri reprodukcii. V praxi sa teda pojem populácia spravidla používa ako operačný koncept pre skupinu jedincov rovnakého druhu, žijúcu na konkrétnej lokalite alebo v konkrétnej oblasti – v tomto zmysle možno hovoriť o lokálnej populácii, regionálnej populácii, ale aj populácii druhu ako takého.

Vlastnosti populácie v značnej miere ovplyvňujú vývoj jej genetickej štruktúry. Základným populačným parametrom je veľkosť, teda počet jedincov, z ktorých sa populácia skladá. Z hľadiska početnosti je ale potrebné si uvedomiť, že nie všetky jedince prítomné v populácii sa aj musia podieľať na reprodukcii a teda v rovnakej miere odovzdávať svoju dedičnú informáciu generácii potomstva. Preto populačno-genetické modely vychádzajú a efektívnej veľkosti, ktorá predstavuje ekvivalent reálnej početnosti vyjadrený hypotetickým počtom jedincov rovnomerne sa podieľajúcich na reprodukcii.

Ďalšou vlastnosťou, určujúcou vývoj zastúpenia génov, genotypov a fenotypov v populácii, je systém reprodukcie, t.j. spôsob odovzdávania genetickej informácie z jednej generácie na nasledujúcu. Z hľadiska systému reprodukcie možno rozoznávať dva typy párovania. Panmixia je náhodné párovanie, pri ktorom pravdepodobnosť spojenia ktorýchkoľvek dvoch gamét je nezávislá od ich pôvodu a génov, ktoré sú v nich zastúpené. Pre výber

partnera pre párovanie neexistujú žiadne kritériá, je úplne náhodný. Pri výberovom párovaní (angl. *assortative mating*) naopak takéto kritériá existujú, jedince, ktoré ich splňajú, majú väčšiu šancu spojiť sa pri reprodukcii, než jedince, ktoré ich neplnia (napr. stromy, kvitnúce v rovnakej dobe, majú vyššiu šancu vzájomného opelenia než stromy, ktoré sú fenologicky asynchrónne; sobášnosť medzi príslušníkmi rovnakej cirkvi či náboženskej skupiny je vyššia než medzi rôznymi vierovyznaniami a pod.). Kritériom výberu partnera môže byť aj príbuzenstvo, príbuzenské kríženie sa označuje aj termínom inbreeding.

Genetickú štruktúru populácie definujú podiely jednotlivých genotypov (genotypová štruktúra) a podiely jednotlivých génov resp. ich variantov – alel (alelická štruktúra), čiže genotypové a alelické frekvencie. Frekvencie sa pri známej malej veľkosti populácie sa dajú vyjadriť aj absolútnou početnosťou, vo veľkých populáciách je zmysluplnejšie odhadovať ich relatívny podiel (ak v populácii s počtom členov 1000 je počet jedincov genotypu AA je 20 jedincov, ich podiel predstavuje 2%, t.j. frekvencia je $P_{AA} = 0,02$). Frekvencia konkrétneho genotypu teda vyjadruje pravdepodobnosť (viď Box I), s akou náhodne vybraný jedinec z populácie bude jeho nositeľom.

Box I Pojem pravdepodobnosti

Každá minca má dve strany (označme ich tradične hlava a orol, ako to bolo za Rakúsko-Uhorska, keď na líci bol portrét cisára a na rube štátny znak s habsburskou dvojhlavou orlicou). Ak ňou hodíme len raz, padnúť môže len jedna z nich. Ak ňou hodíme za sebou päťkrát, nie je vylúčené, že päťkrát za sebou padne hlava, ale každý intuitívne cíti, že skôr môžeme očakávať, že aspoň pri niektorom z hodov padne aj orol. Inými slovami, oba prípady ($5 \times$ hlava vs. aspoň raz orol) nie sú rovnako pravdepodobné. Ak budeme mincou hádzať povedzme miliónkrát, podiel prípadov, keď padne hlava, sa rovnako ako podiel prípadov, keď padne orol, bude približovať 50%. Toto je tradičné chápanie pojmu pravdepodobnosti: pravdepodobnosť udalosti A predstavuje podiel prípadov, keď by udalosť A nastala pri veľkom (teoreticky nekonečnom) počte nezávislých pokusov. Podstata je, že pokusy musia byť **náhodné**. Pokiaľ budete mincou hádzať zámerné tak, aby padala jedna strana, alebo bude minca deformovaná či nevyvážená, výsledok sa bude líšiť a pravdepodobnosť sa nedá kvantifikovať.

Existujú aj iné prístupy k definícii tohoto pojmu, z praktického hľadiska sú ale relevantné skôr jej vlastnosti, ktoré nezávisia na filozofickom prístupe:

- Ak pravdepodobnosť udalosti A označíme $P(A)$, potom pravdepodobnosť alternatívy \bar{A} , teda množiny všetkých ostatných možností, ktoré sa s udalosťou A navzájom vylučujú, je rovná $P(\bar{A}) = 1 - P(A)$ (aj v ďalšom texte bude symbol \bar{A} používaný pre negáciu).
- Súčet pravdepodobností všetkých N navzájom sa vylučujúcich možností je 100%: $\sum_i^N P(A_i) = 1$. Pravdepodobnosť, že pri hode padne hlava + pravdepodobnosť, že padne orol je 100%; možnosti sú len dve ($N = 2$), žiadny tretí prípad nastať nemôže.
- Pravdepodobnosť, že súčasne nastanú dve navzájom nezávislé udalosti A a B je daná súčinom pravdepodobností každej z nich: $P(A \cap B) = P(A)P(B)$. Pravdepodobnosť, že človek náhodne vybraný z populácie na Slovensku bude mať krvnú skupinu $A+$ (t.j. A a **súčasne** $Rh+$) je daná súčinom frekvencie krvnej skupiny A a frekvencie $Rh+$ v slovenskej populácii, pretože gény pre ABO a Rh -faktor sú lokalizované na rozdielnych chromozómoch a teda rekombinujú voľne. Toto pravidlo je možné rozšíriť na ľubovoľný počet nezávislých udalostí.
- Pravdepodobnosť, že nastane aspoň jedna z dvoch nezávislých udalostí, ktoré sa navzájom nevylučujú, je daná súčtom pravdepodobnosti prvej, pravdepodobnosti druhej, zmenšená o pravdepodobnosť, že nastanú obe súčasne: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Pravdepodobnosť, že človek náhodne vybraný z populácie na Slovensku bude mať krvnú skupinu A **alebo** $Rh+$ je daná súčtom frekvencie krvnej skupiny A (nezávisle

Box I pokračovanie

na Rh-faktore) a frekvencie Rh⁺ (vo všetkých krvných skupinách AB0 spolu), zmenšeným o podiel nositeľov kombinácie A⁺. Pravdepodobnosť, že pri hode mincou padne hlava alebo orol je súčet pravdepodobností jednej + druhej možnosti, obe súčasne nastať nemôžu.

- Udalosti sa môžu navzájom podmieňovať, jedna môže byť závislá na druhej. Človek nemôže ochoriť na maláriu, ak ho nepoštípe infikovaný komár, iný spôsob prenosu neexistuje. Podmienená pravdepodobnosť $P(A|B)$ určuje, s akou pravdepodobnosťou nastane závislá udalosť A za predpokladu, že nastala nezávislá udalosť B . Určí sa ako $P(A|B) = P(A \cap B)/P(B)$. Pravdepodobnosť, že dostanete maláriu (závislá udalosť), je podiel frekvencie ľudí s maláriou (čo automaticky znamená, že boli poštípaní komárom, inak by ju nedostali) z celkového počtu ľudí poštípaných komárom (nezávislá udalosť).

Okrem pojmu 'pravdepodobnosť' štatistika používa aj termín 'vierohodnosť' (*likelihood*). Aj keď sa oba termíny často používajú ako synonymá, istý významný rozdiel medzi nimi tu je. Pravdepodobnosť sa v princípe týka budúcich udalostí, predikuje aká udalosť môže nastať. Vierohodnosť sa vzťahuje k minulosti; určuje, s akou pravdepodobnosťou udalosť, ktorá už nastala, priniesla konkrétny výsledok.

Iným spôsobom vyjadrenia pravdepodobnosti je šanca (*odds*), ktorá vyjadruje pomer počtu prípadov, keď udalosť nastane k počtu prípadov, keď nenastala; ak je pravdepodobnosť nejakej udalosti 10%, znamená to, že je šanca 1:9, že udalosť nastane.

Pri konečnom a nízkom počte jedincov v populácii a možnosti ogenotypovať všetkých z nich (čo je v praxi zriedkavá situácia vzhľadom na technické možnosti a spravidla obrovské veľkosti populácií) je možné určiť frekvencie genotypov presne, ale väčšinou sa používajú výberové postupy a genotypové či alelické frekvencie sa odhadujú na základe výberovej vzorky (viď Box II). Z praktického hľadiska sa spravidla používa bodový odhad, frekvencia genotypu A_iA_j sa dá odhadnúť ako $P(A_iA_j) = n(A_iA_j) / \sum_k \sum_l n(A_kA_l) = n(A_iA_j) / N$, kde $n(A_iA_j)$ je počet jedincov genotypu A_iA_j vo výberovej vzorke ($k, l = 1 \dots$ celkový počet alel) a N je celkový rozsah výberu (celková početnosť všetkých genotypov spolu vo výberovej vzorke). Bodový odhad frekvencie i . alely sa určí z podielov genotypov, v ktorých je daná alela zastúpená: $p(A_i) = P(A_iA_i) + \sum_j 0,5P(A_iA_j); j \neq i$ (v genotype homozygota je zastúpená len sledovaná alela, v genotype heterozygotov je to vždy jedna z dvoch, teda ich frekvencie sa berú s polovičnou váhou). Relatívne početnosti (frekvencie) majú binomické rozdelenie (viď Box III). Skupina tesne viazaných génov (väzbová skupina) sa môže z generácie na generáciu odovzdávať spolu, tvorí tzv. haplotyp. To isté platí pre mitochondriálnu a chloroplastovú DNA, kde sa kruhová molekula DNA sa pri reprodukcii odovzdáva potomstvu ako celok len od jedného z rodičov, a teda nezávisle na fyzickej vzdialenosti dvoch génov nemá ako dôjsť k ich rekombinácii. Pre odhady haplotypových frekvencií a ich stredné chyby platia rovnaké vzťahy ako pre genotypové frekvencie. Pri dvojiciach alebo skupinách voľne sa rekombinujúcich génov (napr. lokalizovaných na rôznych chromozómoch) platia pre odhad frekvencie kombinácie génov zákony pravdepodobnosti: frekvencia kombinácie génov v gamétach kombinácie alebo genotypov je súčinom frekvencií génov/genotypov: $P(A_iB_k) = P(A_i) \cdot P(B_k)$; $P(A_iA_jB_kB_l) = P(A_iA_j) \cdot P(B_kB_l)$ (viď Box I). Napríklad gén pre systém krvných skupín AB0 je lokalizovaný na chromozóme 9, gén pre Rh-faktor na chromozóme 1, kombinujú sa teda nezávisle na sebe. Ak je v konkrétnej populácii frekvencia krvnej skupiny AB (teda heterozygotov $I^A I^B$) 8%

a frekvencia nositeľov Rh⁻ (teda homozygotov rh^-rh^-) je 15%, frekvencia kombinácie AB bude $0,08 \times 0,15 = 0,012$, teda 1,2% (čo zhruba zodpovedá pomerom strednej Európy).

Box II Výber a odhady parametrov základného súboru

V biológii rovnako ako v ekonómii, sociológii atď. nás zaujímajú vlastnosti základného súboru (ZS; *statistical population*), teda súboru všetkých mysliteľných pozorovaní. Základný súbor môže mať konečnú početnosť (súbor študentov TU vo Zvolene k 1.1.2020), môže sa limitne blížiť nekonečnu (súbor atómov vodíka vo vesmíre), dokonca ani fyzicky nemusí existovať (ak testujeme efekt liečiva, zaujíma nás efekt na akéhokoľvek potenciálneho pacienta, ktorý liek mohol či môže zobrať v minulosti, prítomnosti či budúcnosti). Ak chceme určiť vlastnosti (parametre) základného súboru, väčšinou sme odkázaní na to, že z neho vyberieme **náhodnú** vzorku, označovanú ako výberový súbor (VS; *sample*), a vlastnosti zistené na základe vzorky následne zovšeobecňujeme na celý ZS. Tento proces sa označuje ako odhad parametrov ZS.

Kľúčovým slovom pri výbere je **náhoda**. Akonáhle je výber deformovaný preferenciou toho, kto vyberá, neumožňuje spoľahlivý odhad parametrov ZS. Ak chce antropológ určiť priemernú náchylnosť slovenskej populácie na diabetes, ale je rasista či nacionalista, a preto zo svojej vzorky vylúči Rómov, Vietnamcov, Maďarov a Čechov, nezíska správny odhad.

Hodnoty akéhokoľvek znaku v základnom aj výberovom súbore sú variabilné (inak by nemalo zmysel ich merať), teda možno ich zoradiť od najmenej po najväčšiu, alebo rozdeliť celú škálu hodnôt na jednotlivé na seba naväzujúce intervaly a kvantifikovať, koľko hodnôt (alebo aký podiel hodnôt) sa vyskytuje v jednotlivých intervaloch; inými slovami, je možné zostrojiť rozdelenie hodnôt resp. rozdelenie frekvencií hodnôt znaku. Predmetom záujmu sú vlastnosti rozdelenia v ZS (ako symboly sa pre ne používajú písmená gréckej abecedy), ale väčšinou ich odhadujeme na základe VS (označované písmenami latinskej abecedy).

Z vlastností rozdelenia frekvencií či početností hodnôt nás spravidla zaujímajú dva aspekty: jedným je poloha rozdelenia, teda kde má rozdelenie stred či nejakú charakteristickú hodnotu; druhým je variabilita, teda aká je šírka rozdelenia, v akom intervale sa hodnoty pohybujú. Najbežnejšou a intuitívne najľahšie uchopiteľnou charakteristikou polohy je aritmetický priemer: $\bar{x} = \sum_i^N x_i / n$ (priemer ZS sa označuje μ ; angličtina rozlišuje medzi priemerom VS: *average*, a priemerom ZS: *mean*). V prípade, že je rozdelenie asymetrické (teda prevládajú v ňom malé alebo naopak veľké hodnoty), zmysluplnejšími charakteristikami polohy sú medián (stredná hodnota, od ktorej 50% hodnôt je nižších a 50% je vyšších) a modus, teda najčastejšia hodnota v súbore – hodnota s najvyššou početnosťou. Základnou charakteristikou premenlivosti je variancia (rozptyl; *variance*): $s^2 = \sum_i^N (x_i - \bar{x})^2 / n$ (v ZS označovaná σ^2). Odmocnina z rozptylu sa označuje smerodajná odchýlka ($s = \sqrt{s^2}$; *standard deviation*).

Každý odhad parametra je spojený s konkrétnym počtom stupňov voľnosti (*degrees of freedom*; d.f.), čo je počet hodnôt vo výberovej vzorke, ktoré sa môžu voľne meniť pri odhade variancie spojenej s daným parametrom (spravidla počet pozorovaní mínus počet parametrov pre výpočet variancie).

Dobry odhad parametra ZS by mal byť nestranný (nevychýlený), teda nemal by hodnotu parametra systematicky pod- ani nadhodnocovať. Inými slovami, ak by sme zo ZS vybrali veľký (teoreticky nekonečný) počet vzoriek s rozsahom n jedincov, a v každej z nich spočítali odhad, mal by sa priemer odhadov približovať parametru ZS. Tento hypotetický priemer z veľkého počtu odhadov sa označuje ako očakávaná hodnota (*expectation*; E). Odhad rozptylu ZS na základe horeuvedeného vzorca je podhodnotený, pre odhad σ^2 na základe výberovej vzorky je menovateľ potrebné použiť $(n - 1)$ namiesto n . Mieru, v akej sa očakávanie približuje skutočnej hodnote parametra ZS, je možné posúdiť na základe variancie týchto odhadov – čím je väčšia, tým je miera neurčitosti vyššia. Odmocnina z tejto variancie sa označuje ako stredná chyba odhadu (*standard error*).

Box II pokračovanie

Rozlišujú sa bodové odhady, ktoré popisujú parameter ZS jednou hodnotou, a intervalové odhady, ktoré stanovujú rozsah hodnôt, v ktorom sa parameter nachádza s nejakou definovanou pravdepodobnosťou (interval spoľahlivosti; *confidence interval*). Odhady priemeru a variancie uvedené v predchádzajúcom odstavci sú bodové.

Box III Pravdepodobnostné rozdelenia

V rozdelení frekvencií hodnôt frekvencia predstavuje podiel, ktorý predstavujú nositelia danej hodnoty znaku zo všetkých jedincov súboru. Zároveň teda predstavuje pravdepodobnosť, s akou náhodne vybraný jedinec zo súboru bude nositeľom práve tejto hodnoty. Rozdelenie početností či frekvencií teda možno interpretovať aj ako rozdelenie pravdepodobnosti.

- Kontinuálne rozdelené kvantitatívne znaky (výška, krvný tlak, obsah dusíka v listoch, fluorescencia chlorofylu) zvyknú mať normálne rozdelenie, ktoré je popísané Gaussovou funkciou: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(x-\mu)^2/\sigma^2}$. Funkcia $f(x)$ je tzv. hustota pravdepodobnosti, teda pravdepodobnosť, s akou sa v súbore vyskytuje hodnota x , a je definovaná dvoma parametrami: priemerom (μ) a varianciou (σ^2). Pre odhad týchto parametrov platí $E(\bar{x}) = \mu$; $E(s^2) = \sigma^2 n / (n - 1)$. Gaussova funkcia je symetrická, medián a modus sú teda totožné s priemerom.
- Rozdelenie počtov (*counts*), ktoré sú z definície vždy celé čísla, je spravidla asymetrické. Ak je povedzme priemerná hustota populácie medveďa na Slovensku $0,025 \text{ ks.km}^{-2}$, potom ak na Slovensku založíme napr. 1000 štvorcových pokusných plôch, na ktorých budeme spočítavať medvede, na väčšine nenájdeme žiadneho, na niekoľkých jedného, a plochy, na ktorých sa nájde povedzme 7 medveďov budú výnimkou. Teda počet 0 sa bude vyskytovať často, počet 5 zriedka, a počet 100 prakticky vôbec. Takýto typ popisuje Poissonovo rozdelenie: $P(x=r) = e^{-\lambda} \cdot \lambda^r / r!$, kde $P(x=r)$ je pravdepodobnosť výskytu počtu r (inými slovami frekvencia, s akou sa medzi nájdenými hodnotami počtov bude vyskytovať počet r). Poissonovo rozdelenie má len jeden parameter λ , ktorý je súčasne priemerom aj varianciou rozdelenia ($\lambda = \mu = \sigma^2$).
- V prírode sa často vyskytujú procesy, ktoré môžu mať len dva rôzne výsledky: jedinec môže byť živý alebo mŕtvy, samec alebo samica, patriť alebo nepatriť k danému druhu a pod. Pravdepodobnosť výsledku takéhoto procesu popisuje binomické rozdelenie: $P(x=r) = [n! / r!(n-r)!] \cdot \pi^r (1-\pi)^{n-r}$, kde $P(x=r)$ je pravdepodobnosť, že náhodný pokus prinesie výsledok r , ktorý je jednou z dvoch možností (napr. $r =$ náhodne vybraný jedinec je živý), a π (jediný parameter rozdelenia) je frekvencia výsledku r v celom súbore (napr. $\pi =$ frekvencia živých v populácii).

S ďalšími rozdeleniami pravdepodobnosti sa stretneme v kapitolách venovaných samotným forenzným analýzám.

Príbuzenstvo a príbuzenské križenie

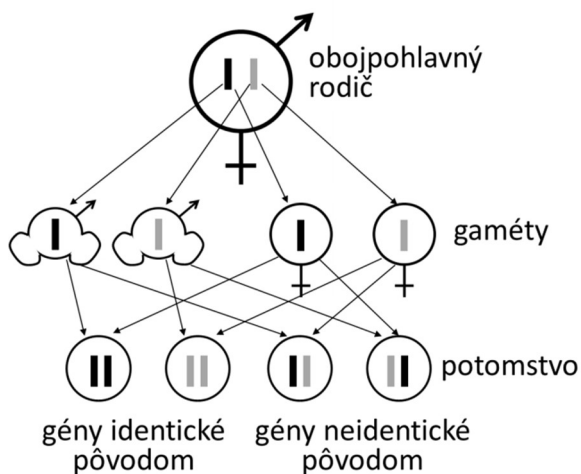
Príbuzenstvom (*coancestry*, *consanguinity*, *relatedness*) sa rozumie skutočnosť, že dvaja jedinci zdieľajú spoločného predka alebo predkov. Terminológia, ktorú používame vo vzťahu k príbuzenským väzbám v ľudskej populácii, síce popisuje tieto väzby zrozumiteľnou formou (vlastný súrodenec, nevlastný súrodenec, rodič, potomok, starý rodič, strýko atď.), ale hodnotí sociálny, nie biologický aspekt a neumožňuje kvantifikáciu miery príbuznosti.

Exaktná definícia príbuznosti dvoch jedincov takisto vychádza zo zdieľania spoločného predka, od ktorého obaja jedinci zdedili časť svojej genetickej výbavy. Pochopiteľne, čím

viac takýchto spoločných predkov dvaja jedinci zdieľajú a čím menej generácií ich od nich delí, tým je tento podiel spoločnej dedičnej informácie vyšší. Príbuzenstvo možno kvantifikovať mierou identity génov pôvodom (*identity by descent*), teda pravdepodobnosťou, s akou je dvojica homologických génov alebo nekódujúcich sekvencií náhodne vybraných z genómu jedinca predstavuje dve repliky tej istej sekvencie, ktoré vznikli replikáciou konkrétneho úseku na rovnakej molekule DNA v genóme spoločného predka. Výpočet príbuzenstva zohľadňuje len autozomálne sekvencie, teda jadrový genóm dedený po oboch rodičoch. Gonozómy rovnako ako uniparentálne dedené mitochondriálne a chloroplastové sekvencie nie sú brané do úvahy, jednak kvôli absencii rekombinácie, jednak kvôli skutočnosti, že predstavujú len malú časť celkovej dedičnej výbavy, ktorú je možné zanedbať.

Každá somatická bunka (vrátane zárodočných buniek, z ktorých sa tvoria gaméty) obsahuje dvojicu autozómov. Pravdepodobnosť, že dve náhodne vybrané gaméty (samčie alebo samičie) budú obsahovať kópie rovnakého chromozómu a budú teda obsahovať identické alely každého génu, je preto 50% (obr. 22). Koeficient príbuzenstva dvoch jedincov zo samoopelenia je teda len 0,5, nie 1, napriek tomu, že pri samoopelení otec a matka je jeden a ten istý jedinec. Analogicky koeficient príbuzenstva medzi plnosesterskými príbuznými (“vlastní súrodenci”) alebo medzi rodičom a potomkom je 0,25, pri polosesterských príbuzných (“nevlastní súrodenci”, ktorí majú spoločného jedného z rodičov), druhostupňových súrodencoch (bratrance, sesternice) a vzťahu starý rodič – vnuk je 0,125, a rovnako je možné kvantifikovať mieru akéhokoľvek príbuzenského vzťahu.

Aj keď obr. 22 ilustruje koncept identity pôvodom na chromozómoch, v skutočnosti sa týka génov: aj gény na rovnakom chromozóme sa môžu rekombinovať vďaka crossing-overu. Chromozóm sa nededí ako celok, ktorýkoľvek jeho segment môže byť medzi homologickým párom vymenený.



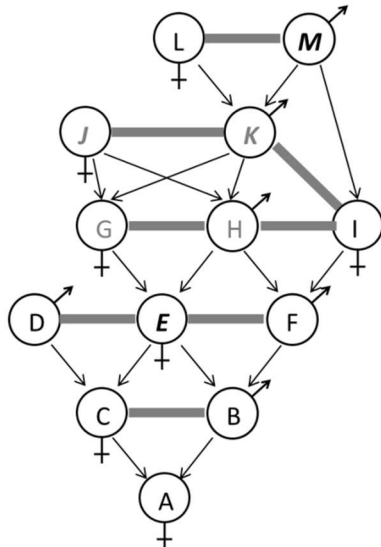
Obr. 22 Ilustrácia konceptu identity génov pôvodom: schematické znázornenie segregácie homologických chromozómov (repliky rovnakého chromozómu rodičovského jedinca sú zobrazené rovnakou farbou) do gamét a ich spojenie do zygot potomstva.

Príbuzenské kríženie sa označuje termínom inbreeding. Miera inbreedingu jedinca sa meria rovnakým spôsobom ako príbuzenstvo jeho rodičov: koeficient inbreedingu jedinca je rovný koeficientu príbuznosti rodičov. Keďže v rodokmeni jedinca sa môže vyskytnúť viacero spoločných predkov jeho rodičov a každý z nich prispieva k miere inbreedingu, je možné vypočítať koeficient inbreedingu ako sumu ich príspevkov:

$$F = \sum_i 0.5^{n_i + m_i + 1} \cdot (1 + F_i)$$

kde n_i a m_i je počet generácií, ktoré delia inbredného jedinca od i . spoločného predka po materskej a otcovskej línii, a F_i je koeficient inbreedingu i . spoločného predka; ak už genómy spoločných predkov samy o sebe obsahujú gény identické pôvodom, môžu byť odovzdané obom rodičom jedinca, a teda zvyšuje sa pravdepodobný podiel IBD génov v jeho genóme.

Príklad: Na obrázku je rodokmeň jedinca A. Aký je jeho koeficient inbreeding, teda aký je podiel génov, u ktorých obe kópie v jeho genotype sú replikami rovnakého originálu v genotype niektorého z jeho predkov, ktorý je spoločným predkom jeho rodičov?



Spoloční predkovia jedinca A sú vyznačení *tučnou kurzívou*. Rodičia jedinca A sú nevlastní súrodenci, majú spoločnú matku *E*. Tá je sama inbredná, jej rodičia sú vlastní súrodenci, teda má spoločných oboch starých rodičov *J* a *K*. Ďalším spoločným predkom rodičov jedinca A je jedinec *M*; jeho predkov nepoznáme, nie je dôkaz, že je inbredný. Medzi jedincami *K* a *I* je vzťah strýko-neter.

Koeficient inbreedingu jedinca *E* je $0,5^{(1+1+1)} + 0,5^{(1+1+1)} = 0,25$. Príspevok jedinca *E* k inbreedingu jedinca *A* je $0,5^{(1+1+1)}(1+0,25) = 0,15625$. Koeficient inbreedingu jedinca *M* je 0. Počet generácií medzi jedincom *A* a *M* po materskej línii (*A*-*CEHK*-*M*) je 4, po otcovskej línii (*A*-*BFI*-*M*) je 3, príspevok jedinca *M* je teda $0,5^{(4+3+1)} (1+0) = 0,00391$. Suma za oboch spoločných predkov je **0,16016**.

Malá početnosť populácie logicky vedie k zvýšenej miere príbuzenského kríženia. Vo veľkej populácii, pokiaľ nie je priestorovo alebo inak štrukturovaná, vedie náhodný výber spravidla k voľbe nepríbuzného partnera (viď ľudská populácia v súčasnosti). Pokiaľ je však populácia početne obmedzená, vytvorí sa v nej už v priebehu generácií sieť príbuzenských vzťahov, takže jedinec v princípe nemá možnosť vybrať si partnera, s ktorým by nebol v príbuzenskom vzťahu (viď populácie malých dedín ešte v nedávnej minulosti). Medzi veľkosťou populácie a priemerným koeficientom inbreedingu pri náhodnom párovaní existuje funkčný vzťah. *N* jedincov obsahuje $2N$ exemplárov každého autozómu. Pravdepodobnosť, že sa v zygote potomstva náhodne stretnú dve repliky toho istého chromozómu je teda $1/(2N)$. Ostatných $1 - 1/(2N)$ zygote potomstva teda obdrží kópie rozdielnych chromozómov, ovšem už tieto môžu obsahovať gény identické pôvodom vďaka inbreedingu v predchádzajúcich generáciách s pravdepodobnosťou, ktorú kvantifikuje koeficient inbreedingu rodičovskej generácie. Koeficient inbreedingu v *t*. generácii F_t je teda súčet pravdepodobností oboch prípadov:

$$F_t = 1/(2N) + [1 - 1/(2N)] \cdot F_{t-1}$$

(vzťah vychádza z veľkosti populácie *N*, ktorá sa v čase nemení). Prírastok inbreedingu za jednu generáciu ($\Delta t = 1$) sa teda dá vypočítať ako $\Delta F = 1/(2N)$. Čím menšia populácia (*N*), tým je prírastok inbreedingu väčší, teda tým rýchlejšie sa v nej vytvára príbuzenstvo a hromadí inbreeding.

Panmiktická rovnováha – Hardy-Weinbergov zákon

Panmixia je definovaná ako úplne náhodné párovanie, teda absencia akýchkoľvek kritérií pre výber partnera pre párovanie. Aj v panmiktických populáciách dochádza k príbuzenskému kríženiu, ale len v miere, ktorá zodpovedá miere príbuzenstva v populácii, teda podielu príbuzenských dvojíc na všetkých možných pároch jedincov opačného pohlavia v populácii.

Základnou zákonitosťou, ktorá popisuje prenos genetickej informácie medzi generáciami v panmiktickej populácii, je Hardy-Weinbergov zákon: zastúpenie jednotlivých alelických variantov (alelické frekvencie) autozomálnych lokusov zostáva z generácie na generáciu nezmenené, ak v populácii nepôsobí žiadny z faktorov, ktoré tento stav môžu zmeniť: prírodný výber, mutácie, tok génov a náhodné zmeny (genetický drift). Zároveň stačí jediná generácia náhodného párovania, aby sa ustanovili aj konštantné podiely jednotlivých genotypov, ktoré zodpovedajú binomickému (pri 2 alelách) resp. multinomickému (pri viacerých alelách) pravdepodobnostnému rozdeleniu, ktoré sa už v ďalších generáciách takisto nemenia. Panmiktická populácia je teda z genetického hľadiska v rovnováhe, nedochádza v nej k žiadnym zmenám genetickej štruktúry na úrovni jednotlivého génu.

Hardy-Weinbergov zákon popisuje idealizovaný stav populácie, platí len v efektívne nekonečne veľkých populáciách za predpokladu oddelených generácií. Drobné odchýlky od týchto podmienok (prekryv generácií, konečné ale veľké populácie) spravidla nevyvolávajú merateľné odchýlky od rovnovážneho stavu: odchýlka spôsobená nedodržaním podmienok je spravidla menšia ako výberová chyba odhadu alelických alebo genotypových frekvencií.

Priebeh náhodného párovania sa u sedentérnych organizmov (rastliny, morské hubky etc.) odlišuje od pohyblivých organizmov (väčšina živočíchov). V prípade usadlých organizmov sa náhodne párujú gaméty (resp. gametofyty), teda haploidná generácia. Sporofyty papraďorastov a cievnatých rastlín, hubiek, koralov a pod. sú nepohyblivé, a s nimi aj samičie gaméty, viazané na materský organizmus. Nech je však vektor prenosu samčích pohlavných buniek akýkoľvek (vietor, hmyz, voda), pohybujú sa v priestore, a teda ak ich výber materského organizmu a teda aj samičej gaméty je náhodný, populácia sa správa ako panmiktická. V najjednoduchšom prípade bialelického génu je možné Hardy-Weinbergov princíp popísať nasledovne:

V generácii t je v populácii sú genotypy v autozomálnom géne A s úplnou dominanciou zastúpené nasledovne: dominantní homozygoti AA podielom P , heterozygoti Aa podielom Q , recesívni homozygoti aa podielom Q . Pri rozmnožovaní dominantní homozygoti AA tvoria len gaméty s dominantnou alelou A , recesívni homozygoti aa len gaméty s recesívnou vlohou a . U heterozygotov Aa alely segregujú do gamét (samčích aj samičích) v pomere 1:1, teda polovica gamét, ktoré vytvárajú, bude obsahovať dominantnú alelu A a druhá polovica recesívnu alelu a . Podiel dominantnej vlohy v gamétach je preto $p = P + \frac{1}{2}H$, podiel recesívnej alely $q = Q + \frac{1}{2}H$ ($p_{\delta} = p_{\varnothing}$, $q_{\delta} = q_{\varnothing}$). V prípade že gaméty sa do genotypov potomstva spájajú **náhodne** (čo je z definície podmienka panmixie) možno očakávané podiely genotypov v generácii $t + 1$ určiť na základe kombinatoriky. Frekvencia dominantne homozygotného genotypu AA , t.j. podiel prípadov, keď sa do zygoty spoja dve gaméty obsahujúce dominantnú alelu A , je $p_{\delta} \times p_{\varnothing} = p^2$. Rovnako frekvencia recesívne homozygotného potomstva aa je $q_{\delta} \times q_{\varnothing} = q^2$. Pravdepodobnosť spojenia dvoch gamét s rozdielnymi alelami A a a do heterozygotnej zygoty Aa je $p_{\delta} \times q_{\varnothing} + p_{\varnothing} \times q_{\delta} = 2pq$. Rozdelenie frekvencií genotypov potomstva (v generácii $t + 1$) je teda binomické: $p^2 AA : 2pq Aa : q^2 aa$.

Pri reprodukcii generácie $t + 1$ (potomstva generácie t) dominantní homozygoti AA , zastúpení v potomstve podielom p^2 , tvoria len gaméty s dominantnou alelou A , heterozygoti Aa , ktorí sú v populácii zastúpení podielom $2pq$, gaméty s dominantnou a recesívnou vlohou v pomere 1:1, recesívni homozygoti len gaméty s recesívnou alelou a . Analogicky ako v generácii t , podiel gamét s dominantnou alelou teda bude $p' = p^2 + \frac{1}{2} \cdot 2pq = p^2 + pq = p(p + q) = p$ a podiel gamét s recesívnou alelou $q' = q^2 + \frac{1}{2} \cdot 2pq = q^2 + pq = q(p + q) = q$ (ak sú v populácii zastúpené len dve alely A a a , súčet ich podielov je 100%, teda $p + q = 1$, čoho dostávame opäť frekvencie $p \times 1 = p$; $q \times 1 = q$).

Pre prehľadnosť je možné celý mechanizmus popísať tabelárnou formou (tab. 6)

Tab. 6 Vývoj alelických a genotypových frekvencií v panmiktickej populácii sedentérneho diploidného druhu.

Generácia	Genotypové frekvencie				Frekvencie alel v gamétach		
	<i>AA</i>	<i>Aa</i>	<i>aa</i>	Spolu	<i>A</i>	<i>a</i>	Spolu
0	<i>P</i>	<i>H</i>	<i>Q</i>	1			
					$p = P_{(AA)} + \frac{1}{2}H_{(Aa)}$	$q = \frac{1}{2}H_{(Aa)} + Q_{(aa)}$	1
1	p^2	$2pq$	q^2	1			
					$p^2_{(AA)} + \frac{1}{2}2pq_{(Aa)} = p^2 + pq = p(p+q) = p$	$q^2_{(aa)} + \frac{1}{2}2pq_{(Aa)} = q^2 + pq = q(p+q) = q$	1
2	p^2	$2pq$	q^2	1	...		

Tab. 7 Vývoj alelických a genotypových frekvencií v panmiktickej populácii pohyblivého diploidného druhu.

		Genotypové frekvencie rodičov			Alelické frekvencie		
		<i>AA</i>	<i>Aa</i>	<i>aa</i>	<i>A</i>	<i>a</i>	
Generácia 0		<i>P</i>	<i>H</i>	<i>Q</i>	$p=P+H/2$	$q=Q+H/2$	
Kríženie	Frekv	Genotypové frekvencie potomstva					
<i>AA</i> × <i>AA</i>	P^2	P^2	0	0			
<i>AA</i> × <i>Aa</i>	$2PH$	$\frac{1}{2}2PH$	$\frac{1}{2}2PH$	0			
<i>AA</i> × <i>aa</i>	$2PQ$	0	$2PQ$	0			
<i>Aa</i> × <i>Aa</i>	H^2	$\frac{1}{4}H^2$	$\frac{1}{2}H^2$	$\frac{1}{4}H^2$			
<i>Aa</i> × <i>aa</i>	$2QH$	0	$\frac{1}{2}2QH$	$\frac{1}{2}2QH$			
<i>aa</i> × <i>aa</i>	Q^2	0	0	Q^2			
Σ		$P^2+PH+\frac{1}{4}H^2 = (P+\frac{1}{2}H)^2 = p^2$	$PH+2PQ+QH+\frac{1}{2}H^2 = 2(P+\frac{1}{2}H)(Q+\frac{1}{2}H) = 2pq$	$Q^2+QH+\frac{1}{4}H^2 = (Q+\frac{1}{2}H)^2 = q^2$	$p^2+\frac{1}{2}\cdot 2pq = p^2+pq = p$	$q^2+\frac{1}{2}\cdot 2pq = q^2+pq = q$	
Generácia 1							
Kríženie	Frekv	Genotypové frekvencie potomstva					
<i>AA</i> × <i>AA</i>	$p^2 \cdot p^2$	p^4	0	0			
<i>AA</i> × <i>Aa</i>	$2(p^2 \cdot 2pq)$	$2p^3q$	$2p^3q$	0			
<i>AA</i> × <i>aa</i>	$2(p^2 \cdot q^2)$	0	$2p^2q^2$	0			
<i>Aa</i> × <i>Aa</i>	$2pq \cdot 2pq$	P^2q^2	$2p^2q^2$	p^2q^2			
<i>Aa</i> × <i>aa</i>	$2(2pq \cdot q^2)$	0	$2pq^3$	$2pq^3$			
<i>aa</i> × <i>aa</i>	$q^2 \cdot q^2$	0	0	q^4			
Σ		$p^4+2p^3q+p^2q^2 = p^2(p^2+2pq+q^2) = p^2 \cdot 1 = p^2$	$2(p^3q+2p^2q^2+pq^3) = 2pq(p^2+2pq+q^2) = 2pq \cdot 1 = 2pq$	$p^2q^2+2pq^3+q^4 = q^2(p^2+2pq+q^2) = q^2 \cdot 1 = q^2$	$p^2+\frac{1}{2}\cdot 2pq = p^2+pq = p$	$q^2+\frac{1}{2}\cdot 2pq = q^2+pq = q$	
Generácia 2							
...							

V prípade väčšiny živočíchov pohyblivé štádium predstavujú spravidla diploidné dospelé organizmy, a aj keď rozmnožovanie nie je nutne spojené s kopuláciou, k spojeniu samčích a samicích gamét dochádza v mieste, ktoré je dané pozíciou samčieho a samicieho jedinca.

To, čo sa náhodne vyhladáva a páruje, sú teda dospelé jedince s konkrétnymi genotypmi, nie haploidné gaméty. Aj tu však platí, že pravdepodobnosť párovania dvoch genotypov je úmerná ich zastúpeniu v populácii. Očakávaná frekvencia pohlavného spojenia dvoch genotypov je rovná súčinu ich podielov v populácii: pravdepodobnosť párenia dvoch rovnakých genotypov (napr. AA) je rovná štvorcu frekvencie genotypu ($P(AA \times AA) = P_{\delta} \times P_{\text{♀}} = P^2$), pravdepodobnosť párenia dvoch rozdielnych genotypov (napr. AA a Aa) je rovná dvojnásobku súčinu ich frekvencií ($P(AA \times Aa) = P_{\delta} \times H_{\text{♀}} + P_{\text{♀}} \times H_{\delta} = 2PH$). Pri popise panmixie v populácii pohyblivého druhu a odhade očakávaných podielov alel a genotypov je potrebné zohľadniť zastúpenie genotypov v rodičovskej generácii a genotypové štiepne pomery pri jednotlivých typoch kríženia (tab. 7).

U mnohých jednobunkových diploidných eukaryotov je pohyblivé ako diploidné (nepohlavné) tak aj haploidné (pohlavné) štádium; vzťahuje sa na ne teda jeden aj druhý popis panmixie.

Príklad: U silenky *Silene littorea* sa vyskytujú tri farebné variety: homozygoty sú fialové (AA) alebo biele (aa), heterozygoty sú svetlomodré (Aa). Ak sa v populácii vyskytujú jednotlivé genotypy v podieloch $P(AA) = 40\%$, $P(Aa) = 50\%$, $P(aa) = 10\%$, ako sa bude vyvíjať alelická a genotypová štruktúra tejto populácie, ak je panmiktická?



Generácia	Genotypové frekvencie				Frekvencie alel v gamétach		
	AA	Aa	aa	Spolu	A	a	Spolu
0	0,40	0,50	0,10	1			
					$p = 0,40 + \frac{1}{2} \cdot 0,50 = \mathbf{0,65}$	$q = 0,10 + \frac{1}{2} \cdot 0,50 = \mathbf{0,35}$	1
1	$0,65^2 = \mathbf{0,4225}$	$2 \cdot 0,65 \cdot 0,35 = \mathbf{0,4550}$	$0,35^2 = \mathbf{0,1225}$	1			
					$p = 0,4225 + \frac{1}{2} \cdot 0,455 = \mathbf{0,65}$	$q = 0,1225 + \frac{1}{2} \cdot 0,455 = \mathbf{0,35}$	1
2	$0,65^2 = \mathbf{0,4225}$	$2 \cdot 0,65 \cdot 0,35 = \mathbf{0,4550}$	$0,35^2 = \mathbf{0,1225}$	1	...		

V prípade, že sú všetky genotypy rovnako plodné, prispieva každý z nich k tvorbe peľových zŕn a vajíčok úmerne ich zastúpeniu v populácii, teda fialové homozygoty AA tvoria 40% samčích a samičích gamét, ktoré všetky obsahujú alelu A , modré heterozygoty Aa 50% gamét, z ktorých polovica (25%) obsahuje A a zvyšných 25% a , a biele homozygoty aa 10% gamét s alelou a , z čoho vyplýva podiel alely A $p = 0,40 + 0,25 = 0,65$, podiel alely a $q = 0,10 + 0,25 = 0,35$. Ak sa gaméty kombinujú náhodne, pravdepodobnosť vzniku zygoty s genotypom AA je úmerná súčinu

pravdepodobností, že (náhodne vybrané) peľové zrno a vajíčko obe obsahujú alelu A (viď Box I), teda $0,65 \times 0,65 = 0,4225$; analogickou dostaneme podiel heterozygotov Aa $0,4550$ a homozygotov aa $0,1225$. Podiel genotypov sa teda zmenil oproti východiskovému stavu. Pokiaľ sú ale všetky genotypy rovnako životaschopné, ich podiely sa až do dospelosti nezmenia (to znamená, že sa všetky jedince dospelosti dožijú, ale z každého genotypu vypadne rovnaké percento). Keď teda ďalšia generácia silieniek začne kvitnúť, bude podiel alely A opäť zodpovedať celému podielu gamét tvorených homozygotmi AA + polovici podielu gamét tvorených heterozygotmi Aa , čo je $0,4225 + \frac{1}{2} \cdot 0,455 = 0,65$, a podiel alely a celému podielu gamét tvorených homozygotmi aa + polovici podielu gamét tvorených heterozygotmi Aa , čo je $0,1225 + \frac{1}{2} \cdot 0,455 = 0,35$, teda zostane nezmenený, z čoho vyplynie nezmenený podiel genotypov v ďalšej generácii atď.

Ako pri usadlých (tab. 6), tak aj pri pohyblivých organizmoch (tab. 7) je zrejmé, že stačí jedna generácia náhodného kríženia, a v populácii sa nezávisle na východiskových frekvenciách genotypov (P , H , Q) ustanoví stála genotypová štruktúra závislá výlučne na frekvenciách alel: podiel p^2 homozygotov AA , $2pq$ heterozygotov Aa a q^2 homozygotov aa . Túto štruktúru však možno rozšíriť na lokus s ľubovoľným počtom alel: ak frekvenciu alely A_i označíme p_i , potom frekvencia ľubovoľného homozygota A_iA_i je p_i^2 , frekvencia ľubovoľného heterozygota A_iA_j je $2p_i p_j$. Takisto v populácii zostávajú frekvencie všetkých alel nemenné. Táto konštantnosť alelickej a genotypovej štruktúry je dôvodom, prečo sa panmixia spravidla berie ako štandard, voči ktorému sa porovnáva genetická štruktúra reálnej populácie.

Komplikovanejšia je situácia v prípade gonozomálnych génov. Ak sa jedná o gén, lokalizovaný v heterologickej časti pohlavného chromozómu Y , dedí sa výlučne otcovi (pri dedičnosti pohlavia typu *Drosophila*), a jeho frekvencia zostáva konštantná. Pri gène viazanom na chromozóm X sa môže dediť od oboch rodičov:

Gény viazané na chromozóm X sa vyskytujú u homogametického pohlavia (v prípade cicavcov samice) v 2 exemplároch, zatiaľ čo u heterogametického pohlavia len v jednom, teda samice prispievajú k frekvencii génu v populácii dvomi tretinami, samce jednou tretinou. Ak pri bialelickom lokuse budú genotypové frekvencie u samíc P homozygotov AA , H heterozygotov Aa a Q homozygotov aa , zatiaľ čo frekvenciu hemizygotov A označíme R a hemizygotov a označíme S , potom frekvencia alely A u samíc je $p_f = P + H/2$, u samcov $p_m = R$; analogicky $q_f = Q + H/2$, $q_m = S$. Vzhľadom na rozdielny príspevok samíc a samcov k alelickým frekvenciám bude zastúpenie alely A v celej populácii

$$p = \frac{2}{3} p_f + \frac{1}{3} p_m = \frac{1}{3}(2p_f + p_m) = \frac{1}{3}(2P + H + R)$$

(analogicky $q = \frac{1}{3}(2Q + H + S)$).

Samce s chromozómovou konšteláciou XY dedia chromozóm X od matky a Y od otca. Gény viazané na gonozóm X teda môže zdediť len od matky, teda frekvencia alely v samčom potomstve je rovná frekvencii u samíc v rodičovskej generácii ($p'_m = p_f$, $q'_m = q_f$). Samice (XX) získavajú jeden chromozóm X od matky, druhý od otca, teda nezávisle na pomere pohlaví frekvencia alely v samičom potomstve je priemerom frekvencií v oboch pohlaviach rodičovskej generácie ($p'_f = \frac{1}{2}(p_f + p_m)$; $q'_f = \frac{1}{2}(q_f + q_m)$).

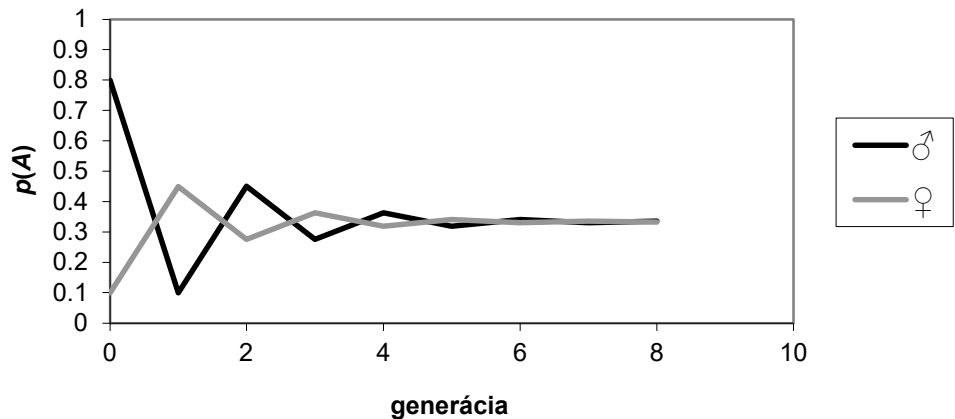
Priemerná frekvencia sa teda nemení, ale frekvencia v rámci pohlaví osciluje okolo rovnovážnej frekvencie a pomerne rýchlo sa k nej približuje (obr. 23). Rovnaké vzťahy samozrejme platia pri nielen pri určení pohlavia typu *Drosophila* (XY), ale pri všetkých typoch dedičnosti pohlavia, u ktorých o pohlaví rozhodujú nehomologické pohlavné chromozómy alebo absencia pohlavného chromozómu (ZW , $X0$, $Z0$, haplodiploidné určenie).

Príklad: U niektorých plemien hovädzieho dobytku je farba srsti riadená autozomálnym génom s kodominanciou: homozygoty sú hnedé (AA) alebo biele (aa), heterozygoty sú strakaté (Aa). Ak sa vo voľne žijúcom stáde vyskytujú jednotlivé genotypy v podieloch $P(AA) = 50\%$, $P(Aa) = 20\%$, $P(aa) = 30\%$, ako sa bude vyvíjať alelická a genotypová štruktúra tejto populácie, ak je panmiktická a v oboch pohlaviach sú genotypy zastúpené rovnako?



		Genotypové frekvencie rodičov			Alelické frekvencie	
		AA	Aa	aa	A	a
Generácia 0		$P = 0,50$	$H = 0,20$	$Q = 0,30$	$p = P + H/2$ $= 0,5 + 0,2/2$ $= 0,60$	$q = Q + H/2$ $= 0,3 + 0,2/2$ $= 0,40$
Kríženie	Frekv	Genotypové frekvencie potomstva			$p^2 + \frac{1}{2} \cdot 2pq$ $= 0,36 + \frac{1}{2} \cdot 0,48$ $= 0,60 = p$	$q^2 + \frac{1}{2} \cdot 2pq$ $= 0,16 + \frac{1}{2} \cdot 0,48$ $= 0,40 = q$
$AA \times AA$	0,25	0,25				
$AA \times Aa$	0,20	0,10	0,10			
$AA \times aa$	0,30		0,30			
$Aa \times Aa$	0,04	0,01	0,02	0,01		
$Aa \times aa$	0,12		0,06	0,06		
$aa \times aa$	0,09			0,09		
Σ		$0,36 =$	$0,48 =$	$0,16 =$		
Generácia 1		$= p^2$	$= 2pq$	$= q^2$		
Kríženie	Frekv	Genotypové frekvencie potomstva			$p^2 + \frac{1}{2} \cdot 2pq$ $= 0,36 + \frac{1}{2} \cdot 0,48$ $= 0,60 = p$	$q^2 + \frac{1}{2} \cdot 2pq$ $= 0,16 + \frac{1}{2} \cdot 0,48$ $= 0,40 = q$
$AA \times AA$	0,1296	0,1296				
$AA \times Aa$	0,3456	0,1728	0,1728			
$AA \times aa$	0,1152		0,1152			
$Aa \times Aa$	0,2304	0,0576	0,1152	0,0576		
$Aa \times aa$	0,0768		0,0384	0,0384		
$aa \times aa$	0,0256			0,0256		
Σ		$0,3600 =$	$0,4800 =$	$0,1600 =$		
Generácia 2		$= p^2$	$= 2pq$	$= q^2$		
...						

Ak sa jedinci v stáde pária náhodne (bez ohľadu na genotyp), pravdepodobnosť, že sa stretnú samec aj samica s dominantne homozygotným genotypom (AA) bude $P \times P = P^2 = 0,50 \times 0,50 = 0,25$, ich potomstvo musí byť tiež dominantne homozygotné. Pravdepodobnosť kríženia $AA \times Aa$ bude $2PH = 2 \times 0,50 \times 0,20 = 0,20$ ($\text{♀}AA \times \text{♂}Aa + \text{♀}Aa \times \text{♂}AA$), polovica potomstva z tohoto kríženia budú dominantné homozygoty AA , druhá polovica heterozygoty Aa . Rovnakou logikou sa dajú odvodiť pravdepodobnosti kríženia ktorejkoľvek dvojice rodičovských genotypov a frekvencie genotypov ich potomstva. Po ich sčítaní dostaneme sumárne frekvencie genotypov v potomstve: 36% AA , 48% Aa a 16% aa , čo zodpovedá podielom odvodeným z alelických frekvencií. Tieto podiely sa ustanovia po 1 generácii náhodného párovania a už sa ďalej nemenia, rovnako ako sa nemenia alelické frekvencie: 60% A a 40% a .



Obr. 23 Vývoj frekvencie alely viazanej na X chromozóm u samcov a samíc v panmiktickej populácii.

Väzbová rovnováha

Ak sú dva lokusy v populácii vo väzbovej rovnováhe, znamená to, že sa ich alely rekombinujú voľne. Akákoľvek nenáhodná asociácia medzi alelami znamená existenciu väzbovej nerovnováhy (*linkage disequilibrium*), ktorá napriek svojmu názvu môže byť podmienená mnohými faktormi, nielen väzbou medzi lokusmi (viď podkap. *Meióza, väzba génov*), ale takisto prírodným výberom, genetickým driftom, populačnou štruktúrou a pod.

V prípade nezávislých lokusov sa ich alely kombinujú náhodne, výskyt jednej nie je korelovaný s výskytom druhej. V prípade dvoch alel *A* a *B*, ak je ich výskyt nezávislý, pre pravdepodobnosť ich kombinácie platí zákonitosť uvedená v Boxe I: $p_{AB} = p_A \cdot p_B$. Koefficient väzbovej nerovnováhy je teda definovaný ako rozdiel týchto hodnôt: $D_{AB} = p_{AB} - p_A \cdot p_B$. Vzhľadom na to, že takto definovaná miera závisí od frekvencií alel v populácii (a teda nie je porovnateľná medzi lokusmi), je spravidla vyjadrovaná formou korelačného koeficienta medzi alelami: $r_{AB} = D_{AB} / \sqrt{p_A(1 - p_A)p_B(1 - p_B)}$.

V čerstvo etablovanej populácii (napr. pri kolonizácii nového stanovišťa malým počtom jedincov) je väzbová nerovnováha spravidla vysoká, keďže u zakladajúcich členov populácie sa vyskytujú konkrétne kombinácie lokusov na ich chromozómoch. Striedaním generácií sa nerovnováha postupne zmiernuje, keďže iniciálne kombinácie alel postupne zanikajú opakovaným crossing-overom. Veľké dlho existujúce panmiktické populácie (ľudské, populácie boreálnych drevín a pod.) spravidla vykazujú tendenciu k rovnováhe alel. Tento stav môže narúšať napr. prírodný výber: ak sú niektoré konkrétne kombinácie alel pre organizmus výhodné (napr. ich produkty dokážu dobre spolupracovať v metabolizme), výber má tendenciu ich v genofonde udržiavať, aj keď môžu byť lokalizované na rôznych chromozómoch.

ŠPECIÁLNA ČASŤ

Páchateľom trestného činu môže byť z definície tohoto pojmu len osoba s právnou subjektivitou, teda človek (aby sme boli úplne presní, len trestne zodpovedný človek; osoby, ktoré ešte nenadobudli trestnú zodpovednosť, napr. deti, alebo osoby, ktoré boli súdnym rozhodnutím zbavené spôsobilosti pre právne úkony, napr. kvôli psychickej chorobe, sem nespádajú). Obetou trestného činu môže byť človek (pre zjednodušenie sem zahrňme fyzické aj právnické osoby) alebo životné prostredie (kam zase pre zjednodušenie zahrňme aj trestné činy voči iným druhom než *Homo sapiens*). Kvôli jednoduchosti bude tento text pre objasňovanie trestných činov prvej kategórie používať termín 'humánna kriminalistika', a pri druhej kategórii termín 'environmentálna kriminalistika'.

Je vcelku pochopiteľné, že nástroje forenzných genetických analýz (analytické metódy, databázy a pod.) boli vyvíjané primárne pre potreby humánnej kriminalistiky. Na druhej strane, aj pri prečinoch voči životnému prostrediu či iným organizmom môže byť samotná humánna kriminalistika nápomocná. Aj keď sme v úvode spomínali len príklady, keď biologický materiál určený pre kriminalistickú expertízu pochádzal z divožijúceho druhu, môže byť predmetom expertízy v environmentálnej kriminalistike aj materiál pochádzajúci z človeka. Pri vnútornostiach zastreleného a vyvrhnutého rýsa sa môže nájsť ohorok z cigarety s epitelovými bunkami ústnej sliznice na povrchu, ktorý je potrebné stotožniť s osobou podozrivou z pytlactva. Samozrejme, škála nástrojov použiteľných v tejto oblasti je užšia než pri trestných činoch voči človeku: mnohé štáty napríklad budujú databázy DNA profilov páchatel'ov trestných činov, tie sa však obmedzujú na tie najťažšie trestné činy: prečiny proti štátu či násilné trestné činy. Odsúdený terorista či vrah sa do nich dostane, odsúdený pytlík či pokutovaný šofér nákladného auta, ktorý vysypal odpad do potoka, nie. Pre úplnosť treba dodať, že v niektorých krajinách (USA, Veľká Británia) sa postupne budujú databázy DNA profilov obyvateľstva, ale vzhľadom na ochranu osobných údajov a občianske slobody ich využitie nie je vždy jednoduché. Pri hľadaní možného páchatel'a v prípade vraždy či znásilnenia je teda možné tieto zdroje využiť, v prípade pašovania chránených živočíchov či neoprávneného nakladania s odpadmi o túto výhodu prichádzame.

Forenzná prax v prípade prečinov voči iným organizmom než človek si musí vystačiť s obmedzeným rozsahom nástrojov a metód. Jeho šírka závisí od konkrétneho druhu: v niektorých prípadoch bola genetická premenlivosť monitorovaná pre vedecké a výskumné účely a k dispozícii sú ako metodické postupy, tak aj pomerne bohaté údaje. V prípade druhov, ktoré zostali bokom tohoto záujmu, si často musíme vystačiť s primitívnejšími postupmi a porovnávacie údaje chýbajú. Nie všetky postupy uvedené v ďalších kapitolách sa teda dajú využiť u každého druhu.

Biologická stopa

Biologickým materiálom je čokoľvek, čo bolo alebo je súčasťou alebo produktom živého organizmu. Ako biologickú stopu označujeme biologický materiál zaistený kriminalistom, ktorý je následne predmetom znaleckého hodnotenia.

V humánnej kriminalistike môže byť biologickou stopou skutočne takmer čokoľvek. Nemusí sa jednať o kúsok tkaniva, stopou môže byť krvná škvrna, biologický materiál zachytený na nedopalku cigarety (bunky v slinách, epitel slizníc), zaschnutý ejakulát, vlasové koreničky a pod. V princípe na vytvorenie DNA profilu stačí cca 15 nedegradovaných buniek. Problém spočíva najmä v slove 'nedegradovaný'; napriek tomu, že DNA je chemicky pomerne stabilná molekula, podlieha s postupom času a v závislosti na podmienkach chemickým zmenám, ktoré môže znemožniť jej využitie. Zároveň treba pripomenúť, že analytické metódy boli vyvíjané a optimalizované primárne pre analýzu ľudského materiálu. Pre kriminalistiku, ktorá sa zaoberá činnosťami proti človeku, sú k dispozícii špecializované kity, ktoré umožňujú získať relevantné výsledky aj v podmienkach, ktoré majú ďaleko k optimálnym.

Do istej (menšej) miery to platí aj pre druhy, na ktorých má človek bezprostredný záujem (pes, hospodárske zvieratá, mikrobiálne patogény človeka a pod.). V prípade voľne žijúcich druhov sú k dispozícii len postupy využívané v botanickom, zoologickom, mykologickom etc. výskume, ktoré často vyžadujú dostatočné množstvo kvalitnej DNA, na izoláciu ktorej je potrebné dostatočné množstvo biologického materiálu v relatívne dobrom stave. Ani to však neplatí absolútne. S ohľadom na obtiažny odchyt živých jedincov, ich zákonnú ochranu a ďalšie faktory boli aj v prípade druhov z prírody vypracované postupy izolácie DNA zo srsti, krvných stôp, a dokonca aj čerstvý trus môže na povrchu obsahovať živé bunky črevného epitelu. Problémom môže byť konzervácia, preparácia, tepelná či chemická úprava biologického materiálu – nie je možné izolovať použiteľnú DNA z drevotrieskovej dosky.

Pri odbere materiálu je potrebné minimalizovať potenciálnu kontamináciu. Pochopiteľne, pri práci s ľudskou DNA je problém kontaminácie akútnejší, keďže tu hrozí kontaminácia stopy biologickým materiálom kriminalistu, ktorý vykonáva odber. Preto sú v kriminalistickej praxi vypracované veľmi prísne postupy odberu stôp. V prípade voľne žijúcich organizmov je v situáciách prichádzajúcich v praxi do úvahy spravidla k dispozícii kúsok tkaniva či pletiva. Aj tu sa však stretávame so situáciami, keď k dispozícii je len krvná škvrna či iný materiál, obsahujúci minimálne množstvá DNA. V týchto prípadoch môže byť kontaminácia problémom ani nie tak pre hrozbu falošne pozitívneho výsledku, ako skôr možnosť zlyhania analýzy. PCR amplifikuje fragmenty zo všetkej DNA, ktorá je v reakčnej zmesi k dispozícii. Ak v nej prevláda kontaminantná DNA nad DNA zo samotnej stopy (čo je ľahko predstaviteľné napr. pri analýze trusu, obsahujúceho malé množstvo buniek črevnej výstelky), môže sa stať, že prioritne sa budú amplifikovať fragmenty kontaminantných molekúl. Problémom nie je možnosť, že by si ich niekto pomýlil s fragmentmi z cieľového druhu, ale môže prekryť signál. Nie je nutné ani účelné pri odbere materiálu pre tento typ analýz používať tak prísne kritériá ako v kriminalistike (ochranný oblek, rúško, sterilné rukavice etc.). Odber často uskutočňuje laik, takže ten ani postupy odberu nemusí ovládať, ale určité minimálne pravidlá je potrebné dodržiavať už z hygienických dôvodov: materiál odoberať v rukaviciach, eliminovať prímes iného materiálu v odobranej vzorke (rastlinné či živočíšne zvyšky, pôda a pod.). Extrémne dôležité je správne a jednoznačné označenie vzoriek. Materiál následne musí byť uskladnený v suchom a studenom prostredí chrániacom pred mikrobiálnou degradáciou, spravidla v mrazničke pri $-20\text{ }^{\circ}\text{C}$, alebo v hlbokomraziacom boxe pri $-80\text{ }^{\circ}\text{C}$.

Postupy izolácie DNA z biologického materiálu sa líšia podľa jeho charakteru. Vo všeobecnosti je materiál najprv homogenizovaný, aby sa narušili bunkové membrány a DNA sa uvoľnila z buniek. Následne sú odbúrané lipidy, ktoré tvoria podstatnú časť sušiny bunky (všetky membránové systémy), pomocou detergentov a povrchovo aktívnych látok. Niektoré procedúry si vyžadujú aj odbúranie proteínov a RNA (proteázami a RNázami). Tieto odpadové látky sú následne vyzrážané koncentrovaným soľným roztokom a odstránené centrifugáciou. DNA je následne purifikovaná; najbežnejšie postupy sú

- vyzrážanie etanolom alebo izopropanolom s následnou centrifugáciou,
- fenol-chloroformová extrakcia, pri ktorej sú proteíny denaturované fenolom, pri následnej centrifugácii zostávajú v organickej fáze, pričom zvyšky fenolu sú z vodnej fázy obsahujúcej DNA odstránené chloroformom
- adsorpcia DNA na kremelinu v prítomnosti chaotropných solí, ktoré sprostredkovávajú väzby medzi DNA a povrchom kremeliny; extrakcia prebieha v minikolóne, v ktorej je kremelina so vzorkou DNA premývaná etanolom obsahujúcim chaotropné látky, a následne extrahovaná z kolóny a vyzrážaná.

Vyextrahovaná DNA je následne spravidla rozpustená v mierne alkalickom tlmiacom roztoku a skladovaná.

Nástroje forenznej genetickej analýzy

Genetické a génové markéry

Pre popis a kvantifikáciu genetickej premenlivosti v rámci populácie a charakteru genetickej variability medzi populáciami potrebujeme znaky, u ktorých z ich fenotypového prejavu sme schopní odvodiť ich genotyp. Takéto znaky nazývame genetickými markérami. Ide o fenotypové znaky, ktoré nie sú ovplyvnené prostredím (resp. vplyv prostredia je len minimálny), t.j. sú prakticky úplne kontrolované geneticky. Z hľadiska hodnotenia genetickej premenlivosti sú najdôležitejšou podskupinou genetických markérov tzv. génové markéry: znaky, ktoré majú jednoduchú genetickú kontrolu (malý počet génov, v ideálnom prípade len jeden), ktorú je možné metódami genetickej analýzy určiť, t.j. každému fenotypu sme schopní jednoznačne priradiť genotyp. Podrobnejšie informácie o princípoch analýzy a použitia jednotlivých typov markérov sú uvedené v návodoch na cvičenia, tu uvádzame len základný všeobecný popis.

Spôsob dedičnosti má pre použiteľnosť génových markérov rozhodujúci význam, v prípade úplnej dominancie (dominantný homozygot a heterozygot nie sú fenotypovo rozlíšiteľní) sú možnosti použitia markéra obmedzené, naopak výhodná je neúplná dominancia alebo kodominancia alel.

Podľa charakteru možno genetické markéry klasifikovať na:

- morfológické (u rastlín najčastejšie morfológické odchýlky v tvare a farbe listov či kvetov/strobilov, habite, u živočíchov albinizmus, vývojové anomálie a pod.)
- biochemické a fyziologické (sekundárne metabolity – monoterpény, sesquiterpény, fenolické látky, fyziologické parametre a pod.)
- molekulárne
 - proteínové
 - izoenzýmy
 - krvné skupiny
 - DNA markéry
 - založené na reštrikčných fragmentoch
 - založené na amplifikácii PCR
 - založené na sekvenovaní
 - založené na iných princípoch

Z uvedených skupín predovšetkým molekulárne markéry spĺňajú požiadavky použiteľnosti na väčšinu praktických účelov, aktuálny štandard forezných analýz spĺňa len tento typ markérov.

Proteínové markéry

Najstarší používaný typ proteínových markérov už bol spomenutý v kap. Dedičnosť fenotypových znakov: krvné skupiny sú antigény na povrchu červených krviniek, väčšinou ide o monogénne kontrolované proteíny alebo glykoproteíny. Počet krvných skupín je cca 50. Ich nevýhodou je častá úplná dominancia, ktorá znemožňuje presnú identifikáciu všetkých genotypov na základe fenotypu. Krvné skupiny sú určované aglutinačným testom: do krvi sa pridá protilátka proti konkrétnemu proteínu a sleduje sa, či dochádza k zhlukovaniu červených krviniek a vyzrážaniu krvi.

Nie všetky živé organizmy majú krv. Univerzálnejším typom proteínových markérov sú izoenzýmy, čo sú rôzne molekulárne formy toho istého enzýmu. Enzýmy ako bielkovinné molekuly predstavujú priamy produkt transkripcie a translácie genetickej informácie, vzťah medzi primárnou štruktúrou enzýmu a poradím nukleotidov v kódujúcej časti DNA je beprostredný. Nesynonymné mutácie v DNA vyvolávajú zmeny proteínu: jeho veľkosti, tvaru, prítomnosti elektricky nabitých funkčných skupín vo zvyškoch aminokyselín ($-\text{COO}^-$, $-\text{NH}_3^+$, $=\text{NH}_2^+$, $-\text{O}^-$, $-\text{S}^-$) a z nej vyplývajúceho elektrického náboja vo vodnom prostredí (v závislosti na pH). Na základe týchto vlastností je možné pôvodné a mutované enzýmové

molekuly oddeliť elektroforeticky, t.j. separáciou v elektrickom poli. Extrakt z živočíšneho či rastlinného materiálu sa vloží do polotuhého média (škrobový alebo polyakrylamidový gél) s konkrétnym pH a pripojí sa na jednosmerný elektrický prúd, teda vloží do elektrického poľa, v ktorom sa elektricky nabité molekuly proteínov pohybujú rýchlosťou úmernou ich náboju. Po ukončení migrácie sa gél sa vloží do farbivaceho roztoku, obsahujúceho substrát, na ktorý zisťovaný enzým pôsobí, kofaktory nutné pre priebeh biochemickej reakcie ktorú enzým katalyzuje (NAD, NADP, ATP, Mg²⁺ a pod.) a farbivá reagujúce s produktmi reakcie. Keďže biochemické reakcie neprebiehajú spontánne, reakcia prebehne len v miestach, kde sa v géli nachádzajú molekuly enzýmu – tu dôjde k vyzrážaniu farbiva, teda tieto miesta sa na géle prejavajú ako farbené pružky, kde každý pružok zodpovedá jednej frakcii enzýmu. Genetická interpretácia zymogramu sa hľadá metódami genetickej analýzy. V najjednoduchšom prípade (monomérny enzým produkovaný len 1 génom) každý pružok zodpovedá jednej alele génu, ktorý kontroluje syntézu enzýmu. Počet alelických variant sa pri izoenzýmoch pohybuje spravidla v rozmedzí 1–4 (výnimočne nad 10) alel na populáciu. Počet izoenzýmových génov, u ktorých je možné enzýmy extrahovať, separovať a farbiť bez straty biochemickej aktivity, a pre ktoré sú vypracované laboratórne protokoly, je spravidla 10–20, v závislosti na organizme. Kľúčovým bodom je extrakcia enzýmov; ďalšie kroky (elektroforéza, farbenie) sú často univerzálne použiteľné pre všetky druhy. Nevýhodou izoenzýmov z hľadiska forenzných analýz je, že proteíny rýchlo podliehajú degradácii, spravidla sa dajú úspešne extrahovať len z čerstvého materiálu. V súčasnosti sú proteínové markéry vo forenznej praxi využívané len výnimočne.

DNA markéry

Princípy analýzy DNA

DNA markéry vychádzajú priamo z analýzy DNA, ktorá je nositeľkou dedičnej informácie. V súčasnosti už je technicky možná analýza celého genómu, ale pre väčšinu praktických aplikácií (vrátane forenzných aplikácií) by to bolo príliš technicky a najmä finančne náročné podujatie. Predmetom analýzy sú preto spravidla vybrané úseky (fragmenty) DNA.

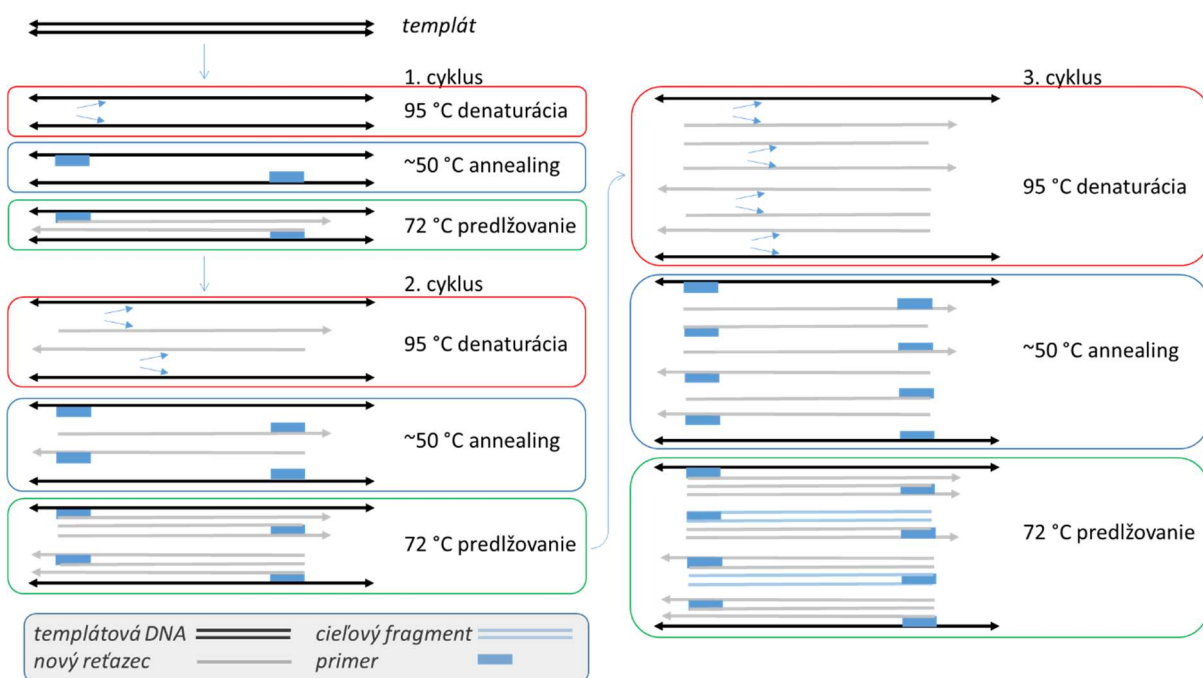
Vytvorenie týchto fragmentov môže byť založené na dvoch postupoch:

- štiepenie molekuly DNA na definované úseky pomocou špecifických enzýmov, reštrikčných endonukleáz (restriktáz). Ide o enzýmy, ktoré prokaryoty používajú ako jeden z mechanizmov obrany proti vírusovej infekcii. Na molekule DNA vyhľadávajú konkrétny krátky sekvenčný motív (sekvenciu nukleotidov s dĺžkou cca 4–8 bp), a pokiaľ ho nájdu, molekulu DNA rozdelia (napr. endonukleáza *EcoRI*, izolovaná z črevnej baktérie *Escherichia coli*, vyhľadáva sekvenciu G↓AATTC/CTTAA↓G a delí molekulu DNA na oboch komplementárnych reťazcoch medzi guanínom a adenínom). Názov konkrétnej endonukleázy je odvodený od skratky rodového a druhového názvu baktérie (*Eco* = *Escherichia coli*), prípadne konkrétneho bakteriálneho kmeňa, z ktorého bola restriktáza izolovaná (kmeň **R**), a číslované sú rímskymi číslami v poradí, ako boli popísané (poradové číslo **I**). Existujú 3 typy reštrikčných endonukleáz, pre účely molekulárnej biológie aj génového inžinierstva sa najčastejšie využívajú endonukleázy typu II, ktoré vyhľadávajú konkrétny sekvenčný motív a reťazec DNA prerušia v jeho rámci. Väčšina endonukleáz vyhľadáva tzv. palindromické sekvencie, t.j. motívy, ktoré sú v smere čítania (5'→3') identické (napr. spomínaná endonukleáza *EcoRI* vyhľadáva sekvenciu 5'GAATTC3', t.j. komplementárna sekvencia v druhom reťazci 3'CTTAAG5' je v smere čítania 5'→3' (t.j. odzadu) identická). Niektoré z endonukleáz prerušujú obe vlákna DNA na rovnakom mieste, teda vytvárajú tzv. hladký koniec (*blunt end*), iné prerušujú vlákna DNA na rôzdielnych miestach, na konci teda vytvárajú jednoreťazcový previs (*overhang*). Enzýmy, schopné spájať prerušenia na reťazci DNA (ligázy) vyžadujú práve previs (táto vlastnosť sa využíva pri niektorých typoch markérov, napr. AFLP). Stríhaním reštrikčnými

endonukleázami sa celý reťazec DNA rozdelí na fragmenty, ktoré je možné separovať elektroforeticky a následne farbiť rôznymi technikami. Dĺžka fragmentov sa meria počtom bázových párov (bp).

Pri jadrovej DNA je množstvo fragmentov získaných štiepením ktoroukoľvek z možných endonukleáz veľmi veľké, pričom závisí od dĺžky sekvenčného motívu. Čisto štatisticky, pri dĺžke n nukleotidových párov v sekvenčnom motíve endonukleáza strihá v priemere každých 4^n nukleotidov (t.j. endonukleáza *EcoRI*, u ktorej je vyhľadávaný motív 6-bázový, vytvára fragmenty s priemernou dĺžkou $4^6 = 4096$ fragmentov).

Pre restriktívnu analýzu sa používajú aj akronym RFLP (*restriction fragment length polymorphism* – polymorfizmus dĺžky restriktívnych fragmentov); toto označenie sa ovšem používa aj pre konkrétny typ analýzy založenej na štiepení, ktorá je technicky náročná a spojená s používaním rádioaktívne značených oligonukleotidov. V minulosti bola vo forennej praxi široko využívaná pre fingerprinting.



Obr. 24 Schéma amplifikácie fragmentov prostredníctvom PCR: počínajúc 3. cyklom v zmesi exponenciálne narastá počet cieľových fragmentov. Šípka na konci reťazca znamená, že reťazec v danom smere pokračuje.

- umelá syntéza úsekov DNA (amplifikácia) metódou PCR (*polymerase chain reaction* – polymerázová reťazová reakcia). PCR vlastne kopíruje mechanizmus replikácie DNA. Táto metóda predstavuje cyklické opakovanie troch reakcií: denaturácie DNA vysokou teplotou (*denaturation*; rozdelenie dvojitého reťazca na dve jednoreťazcové molekuly pri cca 95 °C), pripojenie dvojice oligonukleotidov, tzv. primery, s dĺžkou cca 10–20 bp (*annealing*; pri teplote ~45–60 °C, optimálna teplota závisí od dĺžky a zloženia primerov), a syntéza komplementárneho reťazca (*extension*; 72 °C, čo je optimálna “pracovná” teplota termostabilnej DNA polymerázy). Táto trojica reakcií sa cyklicky opakuje cca. 40–60-krát. Reakčná zmes teda okrem analyzovanej DNA (tzv. templát) a primerov musí obsahovať voľné nukleotidy (presnejšie deoxynukleozidtrifosfáty – dATP, dTTP, dGTP, dCTP) a DNA polymerázu, ktorá ovšem musí byť schopná uchovať si aktivitu aj po zahriatí na teplotu blízku bodu varu počas denaturačného kroku. Najčastejšie sa pre tento účel používa termostabilná *Taq*-polymeráza, izolovaná z archebaktérie *Thermus*

aquaticus, žijúcej v horúcich prameňoch. Primery sa na denaturovanom reťaci DNA pripoja na miesta s komplementárnym poradím nukleotidov a od týchto miest sa jedným smerom (5' → 3', t.j. na -OH skupinu na 3'-konci posledného nukleotidu primera) začne dopĺňať druhý reťazec. Od druhého cyklu týmto spôsobom exponenciálne narastá počet fragmentov s dĺžkou zodpovedajúcou vzdialenosti medzi dvomi komplementárnymi miestami (obr. 24). Fragменты je opäť možné separovať elektroforézou a identifikovať farbivami, ktoré sa viažu na DNA.

Pre separáciu fragmentov DNA sa rovnako ako pri proteínoch používa pohyb v elektrickom poli, čiže elektroforéza. Zdrojom elektrického náboja nukleových kyselín a oligonukleotidov vo vodnom prostredí sú voľné elektrónové páry na atómoch dusíka a kyslíka v heterocyklických bázach nukleotidov (A, T, G, C). Sumárny náboj dvojice A=T a G=C je prakticky rovnaký, takže náboj pripadajúci na jednotku dĺžky DNA fragmentu je v podstate konštantný nezávisle na sekvencii. Fragменты sú teda pri pohybe triedené podľa dĺžky, dlhšie fragменты sa cez póry gélu predierajú menšou rýchlosťou ako kratšie.

Ako gélový nosič sa pri DNA používa najčastejšie agaróza alebo polyakrylamid. Agaróza je vhodným médiom pre rýchle a hrubšie analýzy, keďže separácia v nej je menej dokonalá. Je možné ju použiť aj pre rutinné analýzy, pokiaľ rozdiely vo veľkosti medzi fragmentami, ktoré sú predmetom záujmu, sú veľké (viac ako 10 bp). Pre jemnejšie analýzy sa ako gélový nosič používa elektroforéza v polyakrylamidovom géli (PAGE). DNA je možné aj farbiť fluorescenčnými farbivami alebo striebrom. Veľmi efektívna je separácia fragmentov DNA v automatických DNA-sekvenátoroch. Elektroforéza tu neprebíha na plochom géli, ale v dlhých kapilárach naplnených polyakrylamidovým gélom. Pri PCR sa používajú značené primery (oligonukleotidy s naviazaným farbivom) a farbivo je vybudzované laserovým lúčom.

Oproti iným typom markérov majú DNA markéry tú výhodu, že pre konkrétny účel možo vybrať konkrétny markér, ktorého vlastnosti sa preň najlepšie hodia. Rôzne molekuly DNA a ich rôzne segmenty sa odlišujú spôsobom dedičnosti (biparentálna, maternálna, paternálna), mierou polymorfizmu, ktorá sa odvíja od rýchlosti mutácií, adaptívnym významom a pod. Kombináciou horeuvedených základných postupov analýzy DNA možno vybrať markér alebo skupinu markérov, optimálne slúžiacich k definovanému účelu.

DNA markéry založené na fragmentačnej analýze

V prípade elektroforetickej separácie fragmentov DNA nepoznáme obsah fragmentu, len jeho dĺžku. Potenciálna variabilita sekvencie nukleotidov v rámci fragmentu sa na základe samotnej jeho dĺžky identifikovať nedá. Variabilita dĺžky vyplýva buď z rozdielov v sekvencii na okraji fragmentov (prítomnosti/absencii restričného miesta, rozdielov v sekvenciách pre pripojenie primerov) alebo z insercií/delécií v jeho rámci (spravidla rozdielných počtov tandemových opakovaní sekvenčných motívov). Pre väčšinu účelov vo forenznej analýze je to však plne postačujúce.

Spomenutá RFLP technika *sensu stricto* bola zrejme prvou používanou technikou pre vytvorenie DNA profilu, teda pre fingerprinting. DNA je pri nej štiepená restričnou endonukleázou. Vzhľadom na veľkosti genómov ktoré sa u eukaryotov pohybujú v rozsahu rádovo $10^8 - 10^{11}$ bp, vzniká pri štiepení obrovské množstvo fragmentov rôznych veľkostí. Po separácii na géle a farbení ich nie je možné odlíšiť. Preto sa po elektroforéze prenesú z gélu na nitrocelulóзовú alebo nylonovú membránu, denaturujú (rozpad vodíkových mostíkov a teda rozpad dvojreťazcových fragmentov na dvojice jednoreťazcových) a následne hybridizujú s rádioaktívne značenou próbou, teda jednoreťazcovým oligonukleotidom so sekvenciou vybranou z niektorého konkrétneho génu (Southernov hybridizácia; *Southern blotting*). Próba sa dokáže pripojiť len na tie fragmenty, na ktorých sa vyskytuje sekvencia presne komplementárna k sekvencii próby. Poloha týchto fragmentov sa následne určí

autorádiograficky: na gél sa položí svetlocitlivý film a nechá sa ožiarit rádioaktívnymi fragmentmi DNA, takže sa získa 'fotografia' gélu s pozíciami fragmentov z konkrétneho cieľového génu. Vzhľadom na komplikovanosť celej procedúry a nutnosť používania rádioaktívne značeného materiálu (ktorá si vyžaduje príslušne vybavené a úradne certifikované laboratórium) bola táto technika z praxe vytlačená metódami založenými na PCR.

Ako bolo spomenuté, len malú časť DNA eukaryotov tvoria sekvencie, zodpovedajúce funkčným génom, teda kódujúce nejaký funkčný produkt (polypeptid, tRNA, rRNA, snRNA...). Značnú časť genómu predstavujú nekódujúce sekvencie, ktoré spravidla vykazujú väčšiu premenlivosť. Ich funkcia nie je objasnená, môžu sa podieľať na regulácii génovej aktivity či stabilizácii štruktúry chromozómu. Je však možné, že viaceré takéto sekvencie žiadnu funkciu nemajú a predstavujú len balast evolúcie. Mutácie, ktoré sa v nich hromadia, neovplyvňujú životaschopnosť jedincov a nie sú z populácií vylučované prirodzeným výberom. K nekódujúcim sekvenciám patria predovšetkým intróny, spacers a pseudogény. Veľkú časť DNA tvoria repetitívne, (tandemovo sa opakujúce) sekvencie. K nim patria tandemové opakovania dlhších úsekov cca 10–40 báz (VNTR – *variable number of tandem repeats*; minisatelity), ktoré sa vyskytujú predovšetkým v oblasti centroméry a telomér. Vykazujú vysokú variabilitu, ktorá spočíva spravidla v rozdielom počte opakovaní (rôzne alely sú rozdielne dlhé) a dajú sa veľmi výhodne využiť pre identifikáciu jedincov. Druhú skupinu tvoria tzv. mikrosatelity (SSR – *simple sequence repeats* resp. STR – *short tandem repeats*), t.j. tandemové opakovania krátkeho motívu cca. 1–6 báz, ktoré sa vyskytujú nielen v jadrovej, ale aj v mitochondriálnej a chloroplastovej DNA. Jednotlivé varianty sa u týchto markérov takisto odlišujú počtom opakovaní motívu, a tiež vykazujú veľkú variabilitu (nezriedka 15–30 alel v populácii). Nákladná a náročná je u nich identifikácia okrajových sekvencií, ale následne sa analyzujú samotnou PCR, sú často použiteľné u viacerých blízkych druhov a veľmi dobre reprodukovateľné.

Typ makrórov náhodne reprezentujúcich celý jadrový genóm je metóda AFLP (*amplified fragment length polymorphism* – polymorfizmus dĺžky amplifikovaných fragmentov), ktorá je založená na kombinácii restriktívnej analýzy a PCR amplifikácie: templátová DNA sa štiepi na fragmenty kombináciou dvoch endonukleáz vytvárajúcich previs, jednej štiepiacej DNA na krátke fragmenty, s dĺžkou rádovo niekoľko sto bp, ktorá je optimálna pre elektroforetickú separáciu, druhej vytvárajúcej dlhé a teda zriedkavejšie fragmenty. Tým vzniká menší počet fragmentov, ktoré majú rozdielne konce. Následne sa pomocou DNA ligázy na jeden aj druhý koniec týchto fragmentov pripoja adaptéry, teda krátke oligonukleotidy so známou sekvenciou a komplementárnym previsom, a fragmenty sa namnožia PCR (sekvencie primerov = sekvencie adaptérov + previs). Polymorfizmus AFLP (teda rozdiely medzi jedincami) vyplýva z bodovej mutácie v restriktívnom mieste, ktorá spôsobí, že endonukleáza dané miesto neštiepi, prípadne (zriedkavo) z rozsiahlejšej posunovej mutácie (insercie/delécie) medzi dvoma miestami štiepenia restriktívnu endonukleázou. AFLP sú dominantné markéry, na základe prítomnosti konkrétneho fragmentu sa nedá určiť, či sa nachádza na oboch homologických chromozómoch (homozygot), alebo len na jednom (heterozygot). Jedná sa o anonymné markéry – nevieme, z ktorej časti genómu ktorý z nich pochádza, a jedna analýza poskytuje naraz rádovo desiatky fragmentov. Metóda je technicky náročná ale reprodukovateľná, a ako rezervný postup v prípade organizmov, u ktorých nie sú k dispozícii lepšie markéry, použiteľná aj pre forenzné účely.

Pre identifikáciu taxónov alebo určenie pôvodu zo širších geografických oblastí sa používajú maternálne dedené fragmenty mimojadrovej (mitochondriálnej a chloroplastovej) DNA. Molekuly organelárnej DNA sa nerekombinujú, haplotyp sa dedí ako celok. Keďže organelárny genóm je neporovnateľne menší v porovnaní s jadrovým, pri analýze organelárnej DNA možno kombinovať PCR a RFLP: použitím špecifických primerov sa amplifikuje príslušný úsek cpDNA alebo mtDNA, a pokiaľ nevykazuje premenlivosť (t.j.

u rôznych organizmov má rovnakú dĺžku), možno polymorfizmus v jeho rámci hľadať jeho štiepením restriktívnou endonukleázou. Sekvencie v mtDNA a cpDNA sú spravidla evolučne konzervatívne, teda menej variabilné v porovnaní s jadrovými sekvenciami; vďaka tejto skutočnosti sú primery vyvinuté pre konkrétny druh často použiteľné pre celý rad príbuzných taxónov.

Sekvenovanie DNA

Pre niektoré účely je potrebné poznať priamo poradie báz v konkrétnom úseku DNA. Najčastejšie používanou metódou sekvenovania (teda postupu určenie poradia báz) je Sangerova metóda založená na PCR, používajúca ukončovanie reťazca dideoxynukleotidmi (*chain-termination method*). Sekvenovanie využíva skutočnosť, že pri predlžovaní reťazca DNA, či už *in vivo* pri replikácii alebo *in vitro* pri PCR, DNA-polymeráza buduje komplementárny reťazec smerom 5'→3', teda voľné nukleotidy vždy pripája na –OH skupinu na 3' uhlíku deoxyribózy posledného nukleotidu. Pri sekvenovaní reakčná zmes pre PCR obsahuje okrem normálnych fosforylovaných 2'-deoxynukleotidov (dNTP; vid' popis metódy PCR) aj 2'3'-dideoxynukleotidy (ddNTP; nukleozidtrifosfáty, ktorým chýba –OH skupina nielen na 2' ale aj na 3' uhlíku). Pokiaľ je do reťazca zaradený ddNTP, ďalšie predlžovanie sa zastaví, pretože na 3' konci chýba –OH skupina, na ktorú by sa ďalší nukleotid mohol pripojiť. Voľba medzi dNTP a ddNTP je náhodná. Pri vhodnej koncentrácii dideoxynukleotidov sa teda budú vytvárať reťazce rozličných dĺžok postupne narastajúcich po jednom báзовom páre, ktoré sú vždy ukončené dideoxynukleotidom. Klasický Sangerov postup využíval rádioaktívne (izotopom fosforu ³²P) značené ddNTP alebo značené primery, vyžaduje teda 4 samostatné reakcie PCR, každú s pridaním iného dideoxynukleotidu, po ktorých boli fragmenty separované elektroforézou v štyroch rozdielnych dráhach a identifikované autorádiograficky. Poloha fragmentov v jednotlivých dráhach postupne narastá po jednom báзовom páre. V súčasnosti sa využívajú automatické prístroje (DNA-sekvenátory), ktoré pracujú na mierne odlišnom princípe. Dideoxynukleotidy sú značené každý iným farbivom, ktoré je vybudené laserovým lúčom. PCR sa teda robí naraz (reakčná zmes obsahuje všetky 4 typy rôzne farebne značených dideoxynukleotidov), fragmenty sú následne separované elektroforézou (staršie sekvenátory používali ploché gély, v moderných prístrojoch je elektroforéza robená v kapilárach) a identita dideoxynukleotidu, ktorý ukončuje každý fragment, je identifikovaná na základe farebného signálu.

Pre potreby sekvenovania celých genómov alebo ich rozsiahlych častí je sekvenovanie Sangerovou metódou príliš pomalé, drahé a pracovne náročné. Preto sa objavili metódy druhej generácie (*next-generation sequencing*; NGS), ktoré umožňujú v jednom behu analýzy paralelne sekvenovať rádovo milióny až stá milióny krátkych úsekov, ktoré sa následne zostavili do súvislej sekvencie (*alignment*). Chybovosť týchto postupov je zákonite vyššia v porovnaní so Sangerovou metódou, ale vzhľadom na to, že každý úsek je týmto postupom osekvenovaný mnohokrát, dá sa väčšina chýb identifikovať a vylúčiť. Pri všetkých týchto postupoch je jednoreťazcová molekula DNA imobilizovaná naviazaním na magnetické guľôčky potiahnuté streptavidínom a pri sekvenovaní sa postupne vymieňajú reagenty. Jedným z postupov NGS je sekvenovanie syntézou, napríklad pyrosekvenovanie (*pyrosequencing*), ktoré je založené na detekcii aktivity DNA-polymerázy chemoluminiscenciou. Jednoreťazcová molekula DNA je hybridizovaná s primerom, ktorý predstavuje začiatok komplementárneho reťazca. Následne je inkubovaná s reakčnou zmesou obsahujúcou enzýmy DNA polymerázu, ATP sulfurylázu, luciferázu a apyrázu a substráty adenosín 5'-fosfosulfát (APS) a luciferín. Do reakčnej zmesi sa postupne pridávajú jednotlivé dNTP. Pokiaľ je pridaný taký dNTP, ktorá je komplementárny k prvému voľnému nukleotidu na templátovom reťazci, zaradí sa do novobudovaného reťazca za uvoľnenia pyrofosfátu (~PP, dva fosfátové zvyšky prepojené vysokoenergetickou väzbou). ATP sulfuryláza pripojí

pyrofosfát na APS za uvoľnenia sulfátového zvyšku a vytvorenia molekuly ATP, ktorý je využitý ako zdroj energie pre oxidáciu luciferínu luciferázou (ide o rovnakú reakciu aká prebieha v telách svetielkujúcich živočíchov, napr. svetlušiek). Táto reakcia je sprevádzaná svetelným zábleskom (luminiscenciou), ktorý je detekovaný CCD kamerou. Apyráza následne odbúra nevyužitú dNTP a ATP a reakcia sa môže zopakovať ďalšou sériou. Ďalším, v súčasnosti asi najviac využívaným postupom sekvenovania syntézou je metóda založená na využití reverzibilných terminátorov reťazca (napr. technológia Illumina). Reakčná zmes obsahuje dNTP s naviazaným fluorescenčne značeným terminátorom blokujúcim ďalšie predlžovanie reťazca. Po zabudovaní takejto nukleotidu sa zaznamená jeho fluorescenčný signál a následne je terminátor odbúraný, čo umožní predĺženie reťazca zabudovaním ďalšieho nukleotidu. Všetky spomínané metódy vyžadujú zastavenie reakcie po zabudovaní nukleotidu, výmenu reaktantov a prípadnú deaktiváciu nadbytočných fluorescenčných značiek. Ide o metódy, ktoré sú v porovnaní so Sangerovou metódou podstatne rýchlejšie a v prepočte na 1 bázu aj podstatne lacnejšie. Umožňujú sekvenovať veľké množstvo DNA, dokonca aj celé genómy vyšších organizmov.

V súčasnosti sú vyvíjané metódy tretej generácie, ktoré na určovanie poradia báz využívajú jednotlivé molekuly DNA bez nutnosti zastavenia procesu medzi jednotlivými krokmi detekcie. Patria sem metódy sekvenovania pomocou nanopórov, využívajúce prechod fragmentov DNA cez nanopóry membrán a poradie báz detekujú na základe ich vplyvu na elektrické pole, priame zobrazovanie pomocou pokročilých mikroskopických techník (elektrónová mikroskopia, skenovacia tunelová mikroskopia) a ďalšie techniky, ktoré by v budúcnosti mohli podstatne rozšíriť možnosti analýzy genómov.

Z hľadiska samotných forenzných analýz je genotypovanie sekvenovaním (*genotyping by sequencing*) spravidla technicky a finančne náročné a pre väčšinu účelov je informácia o kompletnej sekvencii zbytočne podrobná. Sekvenovanie ale môže byť nevyhnutným krokom pre vývoj postupov používaných vo forenzných analýzach, napr. vývoj primerov pre analýzu mikrosatelitov, organelárnych markérov a pod.

SNP markéry

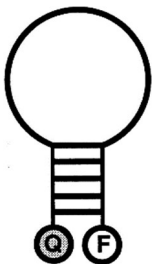
Bodové polymorfizmy (*single-nucleotide polymorphism*; SNP) sú výsledkom bodových génových mutácií, teda zámenny jedného nukleotidu za iný. Sú základom genetickej variability, ktorá sa odráža vo fenotypovej premenlivosti, teda za rozdielmi vo fenotypových znakoch sa skrývajú predovšetkým nesynonymné bodové mutácie v kódujúcich úsekoch génov, ktoré dané znaky kontrolujú, alebo v ich riadiacich sekvenciách (promotery, zosilňovače a pod.). Vyskytujú sa samozrejme aj v nekódujúcich sekvenciách. Spravidla majú dve alternatívne alely; pravdepodobnosť viacnásobnej bodovej mutácie v tom istom mieste DNA je nízka.

Na identifikáciu SNP je samozrejme možné využívať sekvenovanie – celej DNA, náhodne vybraných fragmentov alebo konkrétnych fragmentov. Sekvenovanie kompletného genómu je síce technicky dostupnou možnosťou, ale pre forenzné účely ďaleko presahujúcou ako finančné možnosti, tak aj potrebu. Druhou možnosťou je sekvenovanie (spravidla veľkého počtu) náhodne vybraných fragmentov; najčastejšie používaným postupom je RAD-Seq (*restriction-site associated DNA sequencing*), dnes najčastejšie vo variante ddRAD-Seq (*double-digest RAD-Seq*), pri ktorom je DNA štiepená dvojicou restriktčných endonukleáz vytvárajúcich previs, na ktoré sú následne ligáciou pripojené adaptéry, slúžiace pre fixáciu fragmentov na pevnú fázu (streptavidínom potiahnuté magnetické guľôčky alebo dostička v prietokovej kyvete) a fragmenty sú následne sekvenované spravidla pyrosekvenovaním alebo technológiou Illumina. Týmto postupom je možné identifikovať rádovo stotisíce až milióny SNP, pre ich identifikáciu je však potrebné mať referenčný genóm (teda kompletne zosekvenovanú sekvenciu celeého jadrového genómu) daného alebo blízko príbuzného

druhu. Poslednou možnosťou je sekvenovanie fragmentu konkrétneho génu, ktorý sa spravidla najprv amplifikuje PCR a následne sekvenuje. Tento postup sa využíva pri genetickom barcodingu (sekvenovanie *cox1* u živočíchov, húb, prvokov, ITS2 pri rastlinách a pod.).

Ďalšie možnosti identifikácie konkrétnych SNP, ktoré sú predmetom záujmu (napr. vykazujú rozdiely medzi druhmi alebo výskyt v konkrétnej geografickej oblasti), sú založené na iných princípoch než sekvenovanie. Niektoré z nich vychádzajú z chemických vlastností DNA. Konformačný polymorfizmus jednoreťazcovej DNA (*single strand conformation polymorphism*; SSCP) je založený na skutočnosti, že jednoreťazcové molekuly nukleových kyselín majú tendenciu zbalit' sa do trojrozmerného tvaru, ktorý závisí od párovania báz. Aj zmena jedného nukleotidu môže tento tvar zmeniť. DNA je teda denaturovaná a jednoreťazcové fragmenty sú elektroforeticky separované, pričom výskyt SNP v ich rámci mení ich mobilitu. Elektroforéza v teplotnom gradiente (*temperature gradient gel electrophoresis*; TGGE) alebo denaturačnom gradiente (*denaturation gradient gel electrophoresis*; DGGE) sú založené na rozdielnej migračnej rýchlosti normálneho a čiastočne denaturovaného fragmentu: fragment z analyzovanej DNA je zmiešaný s referenčným fragmentom, ktorý má rovnakú sekvenciu okrem potenciálneho bodového polymorfizmu, denaturujú sa na jednoreťazcové molekuly nechajú sa znovu renaturovať. Pri renaturácii vznikajú heteroduplexy, teda hybridné molekuly tvorené jedným reťazcom referenčného a druhým cieľového fragmentu, ktoré sú následne elektroforeticky separované v gradiente teploty alebo chemicky definovaných denaturačných podmienok. Keďže majú rozdielnu teplotu či rozdielne podmienky denaturácie, budú sa pohybovať rozdielnymi rýchlosťami. Heteroduplexy od normálnych fragmentov je možné separovať aj vysokovýkonnou kvapalinovou chromatografiou (HPLC).

Iná skupina metód je založená na hybridizácii jednoreťazcových molekúl DNA so špecifickými oligonukleotidmi. Hybridizačné próby (*molecular beacons*) sú krátke jednoreťazcové oligonukleotidy, ktoré v strede obsahujú sekvenciu komplementárnu k miestu s predpokladaným výskytom SNP a po okrajoch sekvencie, ktoré sú navzájom komplementárne, a preto sa navzájom párujú za vzniku štruktúry, ktorá sa označuje ako špendlík (*hairpin*; obr. 25). Na 5' konci próby je naviazaný fluorofor schopný vydávať svetelný signál, na 3' konci tzv. zhášač (*quencher*), ktorý tlmí jeho aktivitu. Próba sa dokáže hybridizovať len s miestom na cieľovej DNA, ktoré je presne komplementárne k stredovej sekvencii (už jeden odlišný nukleotid tomu bráni) – ak k hybridizácii dôjde, špendlíková štruktúra sa tým rozpadne, a tým, že fluorofor sa vzdiali od zhášača, začne vydávať signál. Na rovnakom princípe sú založené DNA čipy (*DNA chips*, *SNP microarrays*), na ktorých môžu byť týmto spôsobom fixované stovky až tisíce oligonukleotidov, ktoré pri hybridizácii s cieľovou DNA umožňujú identifikovať jednu analýzou veľké množstvá konkrétnych SNP v genotype jedinca. Aj keď sa primárne využívajú pre medicínske účely (identifikácia mutácií potenciálne spôsobujúcich dedičné choroby alebo náchylnosť na iné patogenézy), v princípe sú využiteľné aj pre forenzné účely.



Obr. 25 Štruktúra oligonukleotidovej próby (molecular beacon)

Ďalšou často používanou metódou je SNaPshot metóda, ktorá je analógiou PCR. DNA je denaturovaná a hybridizovaná s primerom, komplementárnym k sekvencii bezprostredne

predchádzajúcej polymorfnému miestu. Na predlžovanie reťazca sa následne použije zmes fluorescenčne značených dideoxynukleotidov (ddNTP), teda reťazec sa predĺži len o jeden nukleotid (podobne ako pri Sangerovom sekvenovaní). Podľa farebného signálu sa následne určí, ktorý z dvojice nukleotidov sa do reťazca zaradil.

Identifikácia druhovej príslušnosti

Princíp a používané markéry

Identifikácia druhovej príslušnosti (genetický barcoding) vychádza z analogického princípu ako čiarový kód (*barcode*) v obchode. Pod DNA barcodingom rozumieme identifikáciu druhu alebo taxónu na inej taxonomickej úrovni na základe jedného alebo viacerých krátkych úsekov DNA, ktoré pochádzajú zo špecifického lokusu (génu, medzigénového oddeľovača resp. inej sekvencie). Barcoding vychádza z predpokladu, že daná sekvencia je druhovo špecifická, a teda na základe porovnania s inými sekvenciami z ortologického lokusu u iných organizmov je možné ho jednoznačne priradiť ku konkrétnemu taxónu.

Barcoding má samozrejme množstvo aplikácií aj mimo forenznej praxe. Základným cieľom je identifikácia druhu. Molekulárne metódy je možné s úspechom využiť tam, kde tradičné metódy determinácie môžu zlyhávať, teda pri spoľahlivom určovaní taxonomickej identity jedincov, u ktorých nie sú vyvinuté zreteľné determinačné znaky. Môže ísť o rastliny, ktoré ešte nezačali kvitnúť, larvy či vajíčka hmyzu a pod. Zároveň prichádzajú do úvahy ako jediné riešenie, ak ide o fragment organizmu, krvnú či inú biologickú stopu. Z hľadiska potenciálnych prečinov proti životnému prostrediu môže ísť aj o určenie druhového pôvodu potravín či iných spracovaných produktov zo živých organizmov, pokiaľ pri technologickom spracovaní nedošlo k degradácii DNA. Do úvahy ďalej prichádza identifikácia inváznych druhov, kde môže byť problémom podobnosť k domácim druhom a nedostatok skúseností lokálnych expertov.

Konkrétne markéry využívané pre tento účel závisia od konkrétneho organizmu. Vhodný „čiarový kód DNA“ by mal vykazovať minimálnu (ideálne žiadnu) vnútrodruhovou premenlivosť a veľkú medzidruhovou variabilitu. Vzhľadom na skutočnosť, že v súčasnosti sa pre amplifikáciu cieľového fragmentu používa PCR, mali by byť okrajové sekvencie cieľového fragmentu (teda sekvencie, ktoré s ním susedia na 5' a 3' konci) vysoko konzervované, teda nevykazujúce premenlivosť, aby sa na ich základe dali definovať primery pre čím širšie skupiny organizmov. Dĺžka cieľového fragmentu by mala umožniť bezproblémové sekvenovanie. Gén či iný lokus, ktorý by spĺňal všetky tieto kritériá a vyskytoval sa ako ortológ u všetkých organizmov (zrejme) neexistuje, takže pre rôzne taxonomické skupiny sa využívajú rôzne barcodes. V prípade mnohobunkových eukaryotov (rastliny, živočíchy, huby) sa spravidla využívajú sekvencie organelárnej DNA. Vzhľadom na to, že spravidla je predmetom záujmu jedna konkrétna vzorka a vyžaduje sa čím vyššia presnosť sekvenovania, využíva sa skôr klasická Sangerova metóda než NGS.

Najbežnejšie využívanými génmi u živočíchov sú mitochondriálne gény: ich výhodou je maternálna dedičnosť prakticky bez rekombinácie, a dostatok mtDNA v bunke. Živočíšna bunka spravidla obsahuje tisíce mitochondrií, čo umožňuje predstavuje veľký počet DNA molekúl na bunku a teda umožňuje analýzu aj malého množstva materiálu. Najčastejšie sa analyzuje už spomínaný *cox1*, teda gén pre 1. podjednotku mitochondriálnej cytochróm c oxidázy, ale využívajú sa aj gény pre ribozómovú RNA (12S, 16S) alebo gén pre cytochróm b (*cytb*).

U rastlín sú bežnejšími nástrojmi skôr chloroplastové sekvencie, takisto s uniparentálnou dedičnosťou. Najčastejšie využívaný je medzigénový oddeľovač medzi génmi pre malú (5.8S) a veľkú (25S) podjednotku rRNA (ITS2; *internal transcribed spacer*), v menšej miere ITS1 (medzi 18S a 5.8S rRNA). Ďalšími génmi používanými pre barcoding sú gén pre veľkú podjednotku ribulózo-1,5-bisfosfát karboxylázy/oxygenázy (RuBisCO), čo je napriek

zložitému názvu jeden zo základných enzýmov fotosyntézy riadiaci fixáciu CO₂ (*rbcL*), gén pre maturáciu K (*matK*), gén pre transferovú RNA histidínu (*trnH*) alebo gén pre jeden z proteínov fotosystému II (*psbA*). Najčastejšie sa pre spoľahlivú identifikáciu využíva kombinácia markérov.

V prípade húb je situácia komplikovanejšia, spravidla sa využíva kombinácia markérov. U niektorých skupín dobre funguje *cox1*, takisto sa využívajú ITS1 a ITS2 alebo sekvenovanie génu pre 28S či 18S rRNA. Takisto sa používa gén pre translačný elongačný faktor 1 α (TEF1 α) a gény pre podjednotky RNA polymerázy II (RPB1, RPB2).

U jednobunkových eukaryotov (Protista) sa používajú niektoré úseky génov pre 28S či 18S rRNA, ITS a *cox1*, u fotosynteticky aktívnych jednobunkovcov aj chloroplastové gény (*rbcL*, 23S rRNA).

Pre prokaryoty (baktérie, archej) sa spravidla používa gén pre malú podjednotku ribozómu (16S rRNA). Pre niektoré skupiny možno využiť *cox1*, gény pre β podjednotku RNA polymerázy (*rpoB*) a gén pre bakteriálny chaperonin, teda proteín riadiaci zbaľovanie bielkovín (*cpn60*), v špecifických prípadoch gén pre glukonát-6-fosfát dehydrogenázu (*gnd*) a gén pre elongačný faktor Tu (*tuf*).

Bioinformatické spracovanie

Použitelnosť barcodingu je pochopiteľne závislá na možnosti porovnať sekvenciu vo vzorke s referenčnými údajmi. Existujú databázy referenčných sekvencií, v ktorých sú uložené dáta rôznych druhov. Samozrejme nie sú úplné a systematicky sa dopĺňajú. Kvalita údajov v databázach je úmerná prístnosti procesu, ktorý sprevádza ich uloženie: spravidla sa vyžadujú presné informácie o mieste a čase zberu, autorovi, a najmä pri ťažko určiteľných taxónoch (mikroorganizmy, mnohé huby) aj potvrdenie autoritatívnej osoby o taxonomickej identite. Samozrejmosťou je požiadavka na uloženie referenčnej vzorky v herbároch a podobných zbierkach. Niektoré z databáz sú budované na národnej úrovni, iné sú medzinárodné a spravované veľkými konzorciami inštitúcií, napr. International Barcode of Life Project (iBOL). Rovnako sa líši ich zameranie – niektoré sú otvorené pre všetky organizmy, iné sú zamerané na špecifické skupiny: v databáze Unite (unite.ut.ee) sú uložené ITS sekvencie húb. Najširším zdrojom je databáza BOLD (Barcode of Life Data System), otvorená pre všetky druhy a markéry, aj keď v súčasnosti obsahuje najmä *cox1* sekvencie živočíchov. Ďalšiu rozsiahlu databázu spravuje americké Národné centrum biotechnologických informácií (NCBI; www.ncbi.nlm.nih.gov/genbank/barcode).

Na priradenie vzorky ku konkrétnemu taxónu sa používa program BLAST (Basic Local Alignment Search Tool), vyvinutý ako bioinformatický nástroj práve pre spracovávanie a porovnávanie sekvencií. BLAST vyhodnotí mieru homológie sekvencie vzorky so sekvenciami uloženými v databáze. Pokiaľ databáza obsahuje záznam príslušného druhu, je možné identifikovať taxonomickú príslušnosť na druhovej úrovni. V mnohých prípadoch, najmä pri menej skúmaných skupinách alebo skupinách, kde si taxonomická identifikácia vyžaduje expertnú skúsenosť (riasy, jednobunkovce, mnoho húb, morských organizmov a pod.) je možné identifikovať príslušnosť len k vyšším taxónom (rod, čeľaď, rad, niekedy dokonca len trieda). V konkrétnych prípadoch environmentálnej kriminality však aj takého určenie môže byť v závislosti na situácii postačujúce.

Individuálna identifikácia

Princíp a používané markéry

Pod identifikáciou konkrétneho jedinca (genetickým fingerprintingom, individualizáciou) rozumieme stanovenie genetického profilu jedinca, ktorý umožní v ideálnom prípade odlíšiť ho od všetkých ostatných jedincov daného druhu. V súčasnosti sa pre tento účel využívajú takmer výlučne nástroje založené na analýze DNA, teda skôr by bolo možné hovoriť o

stanovení DNA profile (*DNA profiling*), teda súboru charakteristík DNA jedinca. Z forenzného hľadiska porovnanie DNA profilu biologickej stopy s podozrivým objektom umožňuje stanoviť pravdepodobnosť ich identity. Presnejšie: v prípade, že sa stopa s objektom nezhoduje, je ich identita vylúčená jednoznačne. V prípade, že je zhoda úplná, ich identita jednoznačne potvrdená nie je, a dá sa stanoviť len s určitou pravdepodobnosťou, ktorá závisí od sady konkrétnych markérov použitých pre profilovanie a ich variability v rámci populácie. Z uvedeného zároveň vyplýva, že požiadavky na markéry sú pri fingerprintingu opačné než pri barcodingu: vhodné markéry sú tie, ktoré vykazujú v rámci druhu čo najväčšiu mieru polymorfizmu, teda majú čo najväčší počet variantov (alel) s čo najrovnomernejším zastúpením.

Najstarší typ markérov použitých v kriminalistike pre stotožnenie jedinca a biologickej stopy boli krvné skupiny. Aglutinačná analýza bola prvý raz použitá v súdnom prípade už na začiatku 20. storočia a následne využívaná až do vývoja DNA markérov.

Pre forenzné účely je využívaná široká paleta DNA analýz:

- minisatelity (VNTR) boli historicky prvým typom DNA markérov pre forenzné analýzy. Na ich identifikáciu sa používala RFLP technika spojená so Southernovou hybridizáciou. Tento typ markérov sa vyznačuje vysokou mierou polymorfizmu v ľudskej populácii a už niekoľko lokusov poskytuje pomerne spoľahlivú identifikáciu. Pre forenzné účely jedinca spravidla neboli genotypovaní lokus po lokuse, ale blotting sa vykonával pre všetky lokusy spoločne. Výhodou VNTR je kodominantná dedičnosť, teda heterozygota vieme odlíšiť od oboch homozygotov. Postupne bola RFLP analýza vytlačená postupmi, ktoré sú technicky jednoduchšie, nevyžadujú prácu s rádioaktívnym materiálom, a nepotrebujú také veľké množstvo kvalitnej nedegradovanej DNA.
- mikrosatelity (SSR, STR), teda tandemové opakovania krátkych sekvenčných motívov, u ktorých sa jednotlivé varianty (alely) odlišujú počtom opakovaní, teda celkovou dĺžkou, sú v súčasnosti štandardným nástrojom na individuálnu identifikáciu. Identifikácia ich okrajových sekvencií nutná pre stanovenie primerov je síce obtiažna, ale následne je samotná analýza technicky nenáročná, založená len na PCR s následnou elektroforetickou separáciou. Opäť ide o kodominantné markéry, a vzhľadom na to, že ide o nekódujúce sekvencie, vyznačujú sa značným polymorfizmom, a už pri pomerne malom počte lokusov spoľahlivú identifikáciu. Štandardné postupy využívané pre budovanie databáz DNA profilov ľudskej DNA sú spravidla založené na 15–20 lokusoch.

Pri iných organizmoch než človek je situácia komplikovanejšia. U mnohých druhov boli SSR primery vyvinuté pre výskumné účely. Sekvencie primerov sú spravidla prenositeľné medzi druhmi v rámci rodu alebo aspoň nižších taxónov (podrody, sekcie), takže z technického hľadiska sa stanovenie samotného DNA profilu vykonať dá. U málokterých druhov bol ale vykonávaný systematický genetický monitoring, databázy DNA profilov analogické databázam ľudskej DNA teda spravidla nie sú k dispozícii. Pri druhoch, ktoré boli predmetom intenzívneho výskumu (veľké cicavce, hospodársky významné dreviny a pod.) bolo mapovanie genetickej variability vykonávané často pomocou rozsiahlej sady SSR markérov v značnej časti areálu, teda poznáme ich populačný kontext v dostatočnom rozsahu. U menej emblematických druhov však táto informácia chýba, teda nevieme odhadnúť frekvencie jednotlivých mikrosatelitných alel, ktoré (ako ukážeme neskôr) sú nevyhnutné pre stanovenie pravdepodobnosti zhody dvoch genotypov.

V prípade forenzných analýz môže byť problémom degradácia DNA. Mikrosatelity síce nie sú tak náročné na množstvo a kvalitu DNA ako analýza VNTR/RFLP, ale prílišná fragmentácia DNA môže viesť k tomu, že predovšetkým dlhšie alely nie sú amplifikované (*allelic dropout*), takže heterozygoty sú chybné identifikované ako homozygoty, alebo je DNA profil nekompletný. Jedným z riešení tohoto problému sú miniSTR markéry, založené na hľadaní primerov, lokalizovaných bližšie k tandemovému opakovaniu a teda

amplifikujúce kratšie fragmenty, čo znižuje pravdepodobnosť zlomu v ich rámci a zlepšuje šancu na amplifikáciu.

- AFLP ako vysoko polymorfné ale anonymné markéry sú náhradným riešením, ak nie sú k dispozícii lepšie nástroje. Konkrétna kombinácia restriktčných endonukleáz umožňuje skórovanie cca 50–100 fragmentov v rámci jednej analýzy; použitie viacerých kombinácií umožňuje tento počet znásobiť, čo dáva dobrý predpoklad pre vytvorenie jedinečného DNA profilu.

Nevýhodou AFLP je, že ide o dominantné markéry; na základe prítomnosti fragmentu v géle nevieme určiť, či je produktom jednej alely daného lokusu (len na jednom chromozóme) alebo oboch. Naopak, ich veľkou výhodou je, že nevyžadujú znalosť sekvencií cieľového organizmu. Aj keď sú fragmenty amplifikované PCR, primery nie sú definované podľa sekvencie v cieľovom organizme, ale podľa známej sekvencie pripájaných adaptérov. To umožňuje analyzovať v princípe akúkoľvek DNA, pochádzajúcu z ktoréhokoľvek organizmu. Jej degradácia je problémom aj v tomto prípade, ale menším v porovnaní s mikrosatelitmi či inými DNA markérmami. Pri organizmoch, ktoré neboli v minulosti predmetom výskumu, takže u nich chýba „infraštruktúra“ v podobe primerov alebo aspoň údajov o sekvenciách segmentov DNA, môžu byť AFLP použiteľnou alternatívou.

Problémom AFLP je porovnateľnosť, a teda možnosť budovať databázy. Pri ktorejkoľvek fragmentačnej analýze je odhadovaná veľkosť fragmentov do istej miery závislá od prístrojového vybavenia: rôzne sekvenátory, PCR cyklery a rôzna chémia môžu viesť k drobným rozdielom vo výsledkoch. V prípade markérov typu nSSR sa tieto technické rozdiely dajú eliminovať a výsledky z rôznych laboratórií navzájom štandardizovať a zlúčiť, keďže tu sa jedná o sekvencie zo známych lokusov, takže ak v rôznych laboratóriách bol analyzovaný identický materiál alebo aspoň populácie daného druhu z rovnakej oblasti, spravidla sa na základe porovnania frekvencií alel dá opraviť ich identita. Napriek slušnej reprodukovateľnosti je u AFLP štandardizácia týmto spôsobom problematická kvôli ich neznámej identite a pôvodu z neznámych miest genómu. V literatúre je zároveň zriedkavo možno nájsť zdrojové dáta, takže odhad alelických frekvencií nutný pre stanovenie pravdepodobnosti zhody dvoch vzoriek niet na čom postaviť. Jednou z možností je analyzovať nielen cieľovú vzorku, ale zozbierať materiál z celej lokálnej populácie v dostatočnom rozsahu (>100 jedincov), tento postup je ale ťažko použiť v prípade zriedkavejších druhov či organizmov podliehajúcich druhovej ochrane.

- SNP markéry boli donedávna považované sa príliš náročnú metódu pre individuálnu identifikáciu, ale s rozvojom metód NGS a technológií skórovania SNP (SnaPshot, TaqMan atď.) sa začínajú považovať za perspektívnu náhradu mikrosatelitov ako štandardu pre fingerprinting, prinajmenšom v humánnej kriminalistike. Vzhľadom na menšiu mieru polymorfizmu (takmer vždy 2 alely) a nižšiu diverzitu je pre spoľahlivú identifikáciu potrebný vyšší počet SNP v porovnaní s SSR lokusmi, ale to technicky nie je problém. Pre budovanie databáz využiteľných pre forenzné účely je ale nutné vybrať sadu štandardných SNP, ktoré sú skórované v rámci druhu. V prípade ľudskej DNA to opäť nie je problém a viaceré takéto sady už pre národné databázy boli definované. Pre iné organizmy je to mysliteľné u druhov, ktoré sú predmetom intenzívneho (najmä hospodárskeho) záujmu – to ale nie je prípad väčšiny chránených druhov. Aktuálne je využitie SNP v environmentálnej forenznej analýze obmedzené, ale potenciál pre ich využitie nepochybne existuje a s rozvojom technických nástrojov a automatizáciou analýz bude nepochybne narastať. Pokiaľ cena analýz poklesne natoľko, že bude porovnateľná napr. s AFLP, bude možné vybudovať databázy aspoň pre najdôležitejšie chránené druhy, alebo využiť analogický prístup ako pri AFLP, teda analýzu populačnej vzorky pre odhad

frekvencií polymorfizmov tam, kde je to možné s ohľadom na výskyt druhu a spôsob odberu materiálu (pokiaľ je pre získanie dostatočného množstva DNA nutné organizmus usmrtiť, nie je tento prístup zmysluplný).

Štatistické hodnotenie

Rovnako, ako je pri druhovej identifikácii dôležitý kontext ostatných druhov, je pre stotožnenie biologickej stopy s podozrivým jedincom dôležitý kontext celej populácie, z ktorej stopa hypoteticky môže pochádzať. V prípade nezahody je dôkazná sila jednoznačná. V prípade zhody naopak dôkazná sila závisí od pravdepodobnosti, s akou môže byť zhoda dielom náhody.

Toto tvrdenie samozrejme platí za predpokladu, že je vylúčená možnosť technickej chyby, napr. nesprávneho genotypovania, zámény vzoriek a pod. Tento predpoklad u viacerých typov markérov nemusí byť nutne úplne realistický. V prípade metód založených na PCR, najmä tandemových opakovaní (VNTR, SSR), sa v okrajových sekvenciách, pre ktoré sú definované primery, môže vyskytnúť bodová mutácia, ktorá zabráni annealingu primerov a teda amplifikácii príslušnej alely. Takého alely sa označujú ako nulové (*null alleles*) a spôsobujú, že heterozygotné jedince môžu byť chybné identifikované ako homozygoty. V prípade profilu podozrivého to nemusí byť nutne problém (v prípade zhody sa mutácia vyskytuje v oboch profiloch), ale skresľuje to odhad alelických frekvencií v populácii. Vážnejším problémom môže byť vypadok alely (dropout) spôsobený nízkou kvalitou alebo malým množstvom DNA vo vzorke; v tomto prípade je profil vzorky určený chybné. Tieto technické problémy môžu viesť k dvom typom logických chýb. Výsledok môže byť s pravdepodobnosťou *v* falošne negatívny (*false negative*): stopa v skutočnosti pochádza z podozrivého, ale ich profily odlišujú, alebo s pravdepodobnosťou *u* falošne pozitívny (*false positive*): stopa v skutočnosti nepochádza z podozrivého, ale ale ich profily zhodujú. Hodnota $(1 - v)$ sa označuje ako citlivosť (senzitivita) DNA testu, hodnota $(1 - u)$ ako špecificita testu. Oba typy chýb sú na sebe nezávislé a ich pravdepodobnosti sú rozdielne. V prípade DNA testov je špecificita spravidla podstatne vyššia než senzitivita: možnosť chybného genotypovania povedzme v jednom z 12 nSSR lokusov (alebo 50 AFLP lokusov) je podstatne pravdepodobnejšia, než že vo viacerých z týchto lokusov dôjde náhodne práve k takým chybám, ktoré vyprodukujú genotyp biologickej stopy či podozrivého. Podiel falošne pozitívnych výsledkov sa spravidla dá očakávať na úrovni frekvencie profilu v populácii. Tab. 8 uvádza možné scenáre zhody/rozdielnosti skutočných DNA profilov stopy a podozrivého s výsledkami DNA testu.

Tab. 8 Právne relevantné výsledky¹⁾ zhody DNA profilu X biologickej stopy a podozrivého s výsledkami testu a im zodpovedajúce pravdepodobnosti

Zhoda ²⁾	Skutočný profil		Výsledok testu		Pravdepodobnosť
	stopa	podozrivý	stopa	podozrivý	
áno	X	X	X	X	$P_G(1 - v)^2$
áno	nie X	nie X	X	X	$(1 - P_G)u^2$
nie	X	X	X	X	$P_G^2(1 - v)^2$
nie	X	nie X	X	X	$P_G(1 - P_G)(1 - v)u$
nie	nie X	X	X	X	$(1 - P_G)P_Gu(1 - v)$
nie	nie X	nie X	X	X	$(1 - P_G)^2u^2$

¹⁾Pokiaľ je ktorýkoľvek výsledok testu „nie X“, zhoda je vylúčená, teda výsledok nie je relevantný z hľadiska práva, keďže cieľom testu je preukázať zhodu

²⁾Skutočná fyzická zhoda: biologickej stopa naozaj pochádza od podozrivého nie X – iný DNA profil než profil X; P_G – frekvencia DNA profilu v populácii (viď ďalej)

Z praktického hľadiska sú relevantné len prípady, uvedené v tab. 8, teda ak sa výsledky testu stopy a podozrivého zhodujú. Pokiaľ medzi výsledkami testu nie je zhoda (teda aspoň jeden z výsledkov testu je „nie X“), nemá zmysel používať ich ako dôkaz, resp. ide o dôkaz v prospech obvineného (nezávisle na oprávnenosti obvinenia). Do úvahy prichádza 6 scenárov, z ktorých dva svedčia v neprospech obvineného. Len prvý scenár je z hľadiska správnosti obvinenia ideálny: biologická stopa pochádza od podozrivého, teda ich DNA profily sú skutočne zhodné, a sú zhodné aj s oboma výsledkami testu. Pravdepodobnosť tohoto výsledku zodpovedá očakávanej frekvencii DNA profilu v populácii P_G (viď ďalej), zmenšenej o falošne negatívne výsledky. Ak podiel falošne negatívnych výsledkov je v , podiel správnych výsledkov (senzitivita) je zvyšok, teda $(1 - v)$. DNA testy stopy a podozrivého sú dve na sebe nezávislé procedúry, z ktorých každá poskytne správny výsledok s frekvenciou $(1 - v)$, preto pravdepodobnosť, že oba budú správne je ich súčin, teda $(1 - v)^2$ (viď Box I). Z toho je odvodená výsledná pravdepodobnosť tohoto scenára $P_G(1 - v)^2$. Analogickou logikou je možné odvodiť výsledné pravdepodobnosti aj pre ostatné scenáre.

Dostupnosť údajov o podieloch falošne pozitívnych a falošne negatívnych výsledkov opäť závisia od organizmu a markéra. V prípade človeka či iných intenzívne študovaných druhov sú tieto údaje spravidla k dispozícii, v prípade málo skúmaných druhov sa buď dá vychádzať zo všeobecnej chybovosti daného typu markéra, alebo sa pri fingerprintingu možnosti technických chýb neberú do úvahy.

Ďalšiu komplikáciu pre fingerprinting predstavujú klonálne sa množiace organizmy. Jednovaječné dvojčatá cicavcov (vrátane človeka) tiež predstavujú v istom zmysle klon. U rastlín je relatívne bežné vegetatívne rozmnožovanie – mnohé druhy sa rozmnožujú poplazmi, koreňovými výmladkami, adventívnymi výhonmi, hlúzami, špecializovanými orgánmi (pacibuľky *Dentaria bulbifera*) a pod. Pri poľnohospodárskych plodinách a dfevinách sa vegetatívne rozmnožovanie používa často a rozmanitými spôsobmi – odrezky, štepenie, hlúzy a cibulky, exantátové kultúry na živných médiách atď. Špecifickým spôsobom rozmnožovania je apomixia resp. agamospermia, teda tvorba semien bez oplodnenia – embryo sa nevyvíja zo zygoty ale zo somatickej bunky. Niektoré živočíchy, huby a lišajníky sú schopné rozmnožovať sa fragmentáciou. U húb a rias prichádza do úvahy vznik asexuálnych spór. Jednobunkové eukaryoty, kvasinky a primitívne živočíchy sa môžu množiť pučaním. Vo všetkých týchto prípadoch sú jedinci pochádzajúci z klonálneho rozmnožovania geneticky identickí, nástrojmi analýzy DNA nie je možné ich rozlíšiť.

Samostatnú problematiku najmä v prípade rastlín predstavujú polyploidné druhy. Aj keď mnohé z nich sú apomiktické (hlavne v prípade nepárneho počtu chromozómových sád), niektoré druhy sú schopné množiť sa aj pohlavnou cestou. Chromozómy alloplodných hybridov sú často málo homologické, takže alely pochádzajúce od rôznych rodičovských druhov sa technicky dajú odlíšiť. Problémom sú predovšetkým autoploidy, kde odhad alelických frekvencií potrebný pre stanovenie dôkaznej sily DNA profilu je podstatne ložitejší než pri diploidoch.

Bez ohľadu na tieto problémy vo forenzej praxi existuje viac prístupov ku kvantifikácii sily DNA dôkazu.

Pravdepodobnosť náhodnej zhody

Určenie pravdepodobnosti náhodnej zhody (*random match probability*) vychádza z kvantifikácie pravdepodobnosti, že v populácii sa vyskytujú dvaja alebo viac jedincov so zhodným DNA profilom.

V prípade použitia kodominantných jadrových markérov (SSR, SNP) vieme genotyp jedinca určiť presne. Stanovenie očakávanej frekvencie daného genotypu v populácii (teda

zároveň pravdepodobnosti, že náhodne vybraný jedinec bude nositeľom tohoto genotypu vychádza z Hardy-Weinbergovho zákona, teda predpokladu, že populácia je panmiktická; očakávaná frekvencia homozygota A_iA_i je štvorec frekvencie alely A_i : $P(A_iA_i) = p(A_i)^2$, očakávaná frekvencia heterozygota A_iA_j je dvojnásobok súčinu alelických frekvencií: $P(A_iA_j) = 2p(A_i)p(A_j)$. Ak v populácii dochádza aj k príbuzenskému kríženiu, je tento odhad v prípade homozygotov podhodnotený a v prípade heterozygotov nadhodnotený; keďže pri inbreedingu sa kombinujú v genotype potomstva alely identické pôvodom (a teda aj identické stavom), vedie príbuzenské kríženie k nárastu podielu homozygotov. Preto je potrebné pri odhade očakávaných genotypových podielov zaradiť aj koeficient inbreedingu:

$$P(A_iA_i) = p(A_i)^2 + F \cdot p(A_i) \cdot p(A_{j \neq i}); \quad p(A_{j \neq i}) \text{ je sumárny podiel všetkých alel odlišných od } A_i, \\ F \text{ je koeficient inbreedingu}$$

$$P(A_iA_j) = 2p(A_i)p(A_j) \cdot (1 - F)$$

Koeficient inbreedingu je možné v populácii odhadnúť na základe porovnania skutočného podielu heterozygotov h_O a podielu očakávaného pri panmiktickej rovnováhe:

$$F = h_O / [1 - \sum_i p(A_i)^2]$$

V prípade intenzívne študovaných druhov (vrátane človeka) sú spravidla k dispozícii odhady alelických frekvencií a koeficientu inbreedingu či už pre druh ako celok, alebo pre jeho regionálne či lokálne populácie. Pri druhoch, u ktorých takéto údaje nie sú k dispozícii, je možné pre účely forenznej analýzy vykonať štúdiu v lokálnej mierke na rozumne veľkej vzorke (cca 100 jedincov), pokiaľ je to technicky možné (nejde napr. o chránený alebo extrémne zriedkavý a teda ťažko vzorkovateľný druh). Pre stanovenie miery inbreedingu je možné vychádzať z biologických charakteristík: napríklad pri rastlinách majú populácie vetroopelivých druhov spravidla bližšie k panmixii, u hmyzoopelivých druhov väčšinou dochádza aj k príbuzenskému kríženiu, ak nemajú vyvinutý niektorý zo systémov autoinkompatibility. Dôležitá je aj veľkosť populácie: malé populácie majú tendenciu akumulovať inbreeding. Pokiaľ informácie chýbajú a nie sú k dispozícii ani iné údaje pre racionálny odhad inbreedingu, nezostáva iná možnosť než vychádzať pri odhade očakávaných genotypových frekvencií z panmiktickej rovnováhy.

Pri použití dominantných markérov (AFLP) nie je možné odlišiť dominantných homozygotov a heterozygotov. Keďže ide o bialelické lokusy (absencia fragmentu je interpretovaná ako neprítomnosť restriktívneho miesta, ktorá sa prejavuje ako recesívna alela), je možné frekvenciu recesívnej alely odhadnúť za predpokladu panmiktickej rovnováhy na základe frekvencie recesívnych homozygotov: $p(A_0) = \sqrt{P_0}$; kde P_0 je podiel prípadov, keď je konkrétny fragment v populácii neprítomný.

Potenciálnym problémom pre určenie koeficientu inbreedingu je subštruktúra populácie, t.j. ak sa populácia skladá z viacerých subpopulácií, medzi ktorými je výmena génov obmedzená. V prípade ľudskej populácie môže ísť o segmenty navzájom izolované kvôli rasovým, etnickým, náboženským či iným rozdielom. Aj keď sa jedná primárne o kultúrne príčiny, môžu mať biologický dôsledok v tom, že v rámci populácie vytvárajú biologicky navzájom izolované skupiny s rozdielnou genetickou štruktúrou. V tomto prípade sú očakávané podiely heterozygotov na základe priemerných alelických frekvencií za celú populáciu (teda všetky skupiny spolu) nadhodnotený, a v populácii sa to prejavuje ako nadbytok homozygotov. Táto zdanlivá odchýlka od panmiktickej rovnováhy sa označuje ako Wahlundov efekt, a môže skresliť odhad očakávaných frekvencií genotypov.

Ziadny markér nie je natoľko polymorfný, aby dokázal viesť k jednoznačnej identifikácii jedinca, vždy sa používa kombinácia viacerých lokusov. Pre odhad očakávanej frekvencie takejto kombinácie sa vychádza z predpokladu, že lokusy sú vo väzbovej rovnováhe, teda ktorákoľvek kombinácia alel sa v gaméte môže vyskytnúť s rovnakou pravdepodobnosťou. Tento predpoklad nie je úplne realistický ani v prípade lokusov, ktoré nie sú vo vzájomnej väzbe: selekcia, tok génov či náhodné zmaný (genetický drift) môžu viesť k zvýšenému či

zníženému výskytu konkrétnych kombinácií. Presná kvantifikácia zohľadňujúca väzbovú nerovnováhu pri veľkom počte lokusov je ale ťažko riešiteľným matematickým problémom. Z praktického hľadiska sa odporúča používať lokusy lokalizované na rozdielnych chromozómoch, čo je jednoduché u človeka či iných intenzívne skúmaných druhov, ale u menej známych druhov spravidla táto informácia o markéroch chýba. Jednou z možností je otestovať väzbovú nerovnováhu na populačnej vzorke (opäť, ak je to technicky vykonateľné) a prípadne niektoré lokusy radšej vylúčiť.

Pokiaľ je genotyp určený kombináciou N nezávislých lokusov, pravdepodobnosť výskytu kombinácie P_G je daná súčinom ich frekvencií (viď Box I): $P_G = \prod_k P(A_i A_j)_k$; $k=1\dots N$, kde $P(A_i A_j)_k$ je frekvencia genotypu $A_i A_j$ k . lokusu.

Pravdepodobnosť náhodnej zhody je daná prevrátenou hodnotou očakávanej frekvencie genotypu: $P = 1 : 1/P_G$. Dáva odpoveď na otázku: „aká je pravdepodobnosť, že jedinec náhodne vybraný z populácie bude mať daný DNA profil?“.

Príklad: V sade 5 nezávislých mikrosatelitných markérov sa genotyp nájdenej biologickej stopy zhoduje s genotypom podozrivého organizmu. Genotyp je $A_1 A_2 B_5 B_9 C_3 C_3 D_2 D_4 E_1 E_1$. Koeficient inbreedingu F v populácii je 0,1. Alelické frekvencie sú uvedené v tabuľke:

	Lokus									
	A		B		C		D		E	
alela	A_1	A_2	B_5	B_9	C_3	\bar{C}_3	D_2	D_4	E_1	\bar{E}_1
frekvencia	0,10	0,40	0,05	0,01	0,20	0,80	0,15	0,05	0,01	0,99

\bar{C}_3, \bar{E}_1 – sumárna frekvencia ostatných alel než C_3 resp. E_1 v rámci lokusu.

Aká je pravdepodobnosť zhody jedinca náhodne vybraného z populácie s biologickou stopou, teda pravdepodobnosť, že stopa a podozrivý sú dva rozdielne organizmy?

Očakávané frekvencie genotypu po jednotlivých lokusoch:

$$A_1 A_2: P(A_1 A_2) = 2p(A_1)p(A_2) \cdot (1 - F) = 2 \times 0,10 \times 0,40 \times (1 - 0,1) = 0,072$$

$$B_5 B_9: P(B_5 B_9) = 2p(B_5)p(B_9) \cdot (1 - F) = 2 \times 0,05 \times 0,01 \times (1 - 0,1) = 0,0009$$

$$C_3 C_3: P(C_3 C_3) = p(C_3)^2 + F \cdot p(C_3) \cdot p(\bar{C}_3) = 0,20^2 + 0,1 \times 0,20 \times 0,80 = 0,056$$

$$D_2 D_4: P(D_2 D_4) = 2p(D_2)p(D_4) \cdot (1 - F) = 2 \times 0,15 \times 0,05 \times (1 - 0,1) = 0,0135$$

$$E_1 E_1: P(E_1 E_1) = p(E_1)^2 + F \cdot p(E_1) \cdot p(\bar{E}_1) = 0,01^2 + 0,1 \times 0,01 \times 0,99 = 0,00109$$

Ak sú tieto lokusy vo väzbovej rovnováhe (t.j. rekombinujú sa nezávisle na sebe a gametické frekvencie nie sú posunuté oproti očakávaniam náhodnej kombinácie selekciou alebo iným mechanizmom) je výsledná očakávaná frekvencia ich kombinácie súčinom frekvencií po lokusoch:

$$P_G = P(A_1 A_2) \times P(B_5 B_9) \times P(C_3 C_3) \times P(D_2 D_4) \times P(E_1 E_1) = 0,072 \times 0,0009 \times 0,056 \times 0,0135 \times 0,00109 = \mathbf{5,3398.10^{-11}}$$

Pravdepodobnosť zhody náhodne vybraného genotypu s biologickou stopou je prevrátenou hodnotou frekvencie:

$$P = 1 : 1/P_G = 1 : 1/5,3398.10^{-11} = 1 : \mathbf{18\ 727\ 366\ 105}$$

Ak by sa jednalo o človeka, je toto číslo 2,4-násobkom celkovej populácie sveta.

V prípade humánnej kriminalistiky je všeobecne akceptovaná pravdepodobnosť $1 : 10^9$. Štandardne využívané nástroje v súčasnosti umožňujú dosahovať aj čísla rádovo vyššie. V prípade environmentálnej kriminalistiky je to samozrejme závislé na konkrétnej situácii: u druhov, ktoré boli v minulosti predmetom intenzívneho výskumu, sú pre ne k dispozícii ako technické nástroje, tak aj odhady alelických frekvencií v podstatnej časti areálu, je možné sa

dopracovať k porovnateľnej dôkaznej sile; u zriedkavých a málo študovaných druhov situácia môže byť horšia.

Vierohodnosť navzájom sa vylučujúcich hypotéz

Logickejšim a ľahšie interpretovateľným spôsobom prezentácie dôkazu založeného na DNA profile je porovnanie dvoch navzájom sa vylučujúcich hypotéz o zhode profilu biologickej stopy sa profilom podozrivého objektu: hypotézy, že oba profily pochádzajú od rovnakého organizmu (nulová hypotéza H_A) vs. hypotézy, že profily pochádzajú od rôznych organizmov a ich zhoda je náhodná (alternatívna hypotéza $H_{\bar{A}}$). Pomer vierohodnosti (*likelihood ratio*) dáva síce rovnaký číselný výsledok ako predchádzajúci prístup, ale jeho interpretácia je priamočiarejšia, keďže berie do úvahy aj potenciálny alternatívny scenár. Princípy tohoto prístupu sa dajú vyjadriť nasledovne:

- pri hodnotení dôkazu zloženého na DNA profile (E ; *evidence*) treba zvažovať obe horeuvedené možnosti
- je potrebné kvantifikovať podmienenú pravdepodobnosť, že DNA profil stopy a podozrivého bude/nebude zhodný, za predpokladu platnosti nulovej hypotézy [$P(E|H_0)$] aj alternatívnej hypotézy [$P(E|H_A)$]
- rozhodnutie má byť založené na pomere oboch pravdepodobností.

Pomer vierohodnosti je daný podielom oboch pravdepodobností: $L = P(E|H_0)/P(E|H_A)$. V princípe odpovedá na otázku „koľkokrát pravdepodobnejšie je pozorovať zhodu profilu stopy a podozrivého za predpokladu, že platí nulová hypotéza, než za predpokladu, že platí alternatívna hypotéza?“. Je zrejmé, že pravdepodobnosť nulovej hypotézy môže nadobúdať len dve hodnoty: ak sa profily nezhodujú, $P(E|H_0) = 0$, ak sa zhodujú, $P(E|H_0) = 1$ (nie je možné, aby sa profily nezhodovali, ak H_0 platí). Prvý prípad je z praktického hľadiska nezaujímavý (nikto nepríde na súd s dôkazom vo svoj neprospech), takže v realite pravdepodobnosť v čitateli je vždy rovná 1. V prípade menovateľa (alternatívnej hypotézy) je pravdepodobnosť opäť daná frekvenciou profilu v populácii: čím väčší je podiel nositeľov daného DNA profilu v populácii, tým vyššia je pravdepodobnosť, že náhodne vybraný jedinec bude jeho nositeľom (aby sme boli presní, po správnosti by bolo potrebné pri výpočte frekvencie vylúčiť z populácie podozrivého, ale to s výnimkou extrémne malých populácií na výsledku nič nemení). Z toho vyplýva, že pomer vierohodnosti je opäť prevrátenou hodnotou frekvencie profilu: $L = 1/P_G$.

Príklad: Aký je pomer vierohodnosti pri rovnakom zadaní ako v predchádzajúcom príklade?

DNA profil biologickej stopy a podozrivého sa zhodujú, preto pravdepodobnosť nulovej hypotézy $P(E|H_0) = 1$.

Hypotetická frekvencia, s akou sa profil vyskytuje v populácii (teda pravdepodobnosť alternatívnej hypotézy), je $P(E|H_A) = P_G = 5,3398 \cdot 10^{-11}$.

Pomer vierohodnosti je

$$L = \frac{P(E|H_0)}{P(E|H_A)} = \frac{1}{5,3398 \cdot 10^{-11}} = \mathbf{18\ 727\ 366\ 105}$$

Inými slovami, pravdepodobnosť, že biologická stopa pochádza z podozrivého, je $18 \cdot 10^9$ -krát vyššia než pravdepodobnosť, že pochádza z iného jedinca.

Dôkaz založený na Bayesovskej logike

V úvode tejto podkapitoly bolo uvedené že hodnotenie dôkaznej sily či už v prípade zhody alebo nezahody profilu biologickej stopy s profilom podozrivého môže byť zaťažené technickou alebo biologickou chybou.

Okrem informácie o možných technických chybách sa v priebehu vyšetrovania môžu objaviť aj iné informácie relevantné pre stanovenie pravdepodobnosti, s akou biologická stopa pochádza z podozrivého objektu (alebo subjektu, ak sa jedná o páchatel'a). Prístup, ktorý umožňuje takéto dodatočné či doplňujúce informácie zohľadniť, vychádza z Bayesovej teóremy, ktorá predstavuje samostatnú vetvu myslenia v matematickej štatistike (Box IV).

Box IV Bayesova teórema a Bayesovská štatistika

Dodatočná informácia B môže pomôcť spresniť odhad pravdepodobnosti udalosti A . Spresnenie vychádza z Bayesovej teóremy:

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

kde $P(A)$ je pravdepodobnosť udalosti A pred zohľadnením dodatočnej informácie (pravdepodobnosť *a priori*), zatiaľ čo $P(A|B)$ je spresnený odhad podmienený informáciou B (pravdepodobnosť *a posteriori*), pričom kvocient $P(B|A)/P(B)$ kvantifikuje mieru podpory informácie B pre realizáciu udalosti A (pre pripomenutie: \bar{A} znamená negáciu A).

Autorom teóremy je Thomas Bayes; jeho interpretácia je mierne odlišná od klasického chápania pravdepodobnosti, pretože interpretuje pravdepodobnosť subjektívne, ako odôvodnené očakávanie nejakej udalosti resp. stupeň osobnej domnienky či dôvery v realizáciu udalosti (*degree of belief*). Teórema v zmysle Bayesovskej interpretácie dáva väzbu medzi kvantifikáciou dôvery pred a po zohľadnení dodatočných informácií. Ak napr. na TU študuje cca 33% externých študentov (2018), pravdepodobnosť, že za predsedu študentskej časti AS TU bude zvolený externý študent, je *a priori* 33%, keďže externí študenti sú tiež členmi akademickej obce a ich voliteľnosť do orgánov akademickej samosprávy nie je nijako obmedzená. Ovšem ak zohľadníme skutočnosť, že externí študenti sa nevyznačujú mimoriadnou angažovanosťou v akademickom živote (napr. spočítaním miery ich zastúpenia v AS v minulých rokoch), pravdepodobnosť po jej započítaní (*a posteriori*) dramaticky poklesne. Je evidentné, že v tomto prípade chápanie pojmu „pravdepodobnosť“ v klasickom zmysle frekvencie, v akej sa udalosť udeje v dlhodobom priemere (Box I) nie je použiteľné: počet doterajších volieb v AS je príliš nízky na to, aby sa z neho dala odvodzovať akákoľvek zákonitosť (a navyše, externista nebol zatiaľ zvolený nikdy); napriek tomu má zmysel hovoriť o pravdepodobnosti voľby externistu v zmysle oprávnenej domnienky.

Bayesovský prístup je tiež alternatívou pre odhad parametrov, najmä v prípade, ak ich pravdepodobnostné rozdelenie nie je známe. V genetike (včetně forenznej genetiky) sa využívajú v zložitých modeloch, kde sa parametre nedajú odhadnúť priamo, na ich numerické aproximácie pomocou Monte Carlo algoritmov založených na Markovovských reťazcoch (*Monte Carlo Markov Chain*; MCMC).

Bayesovský prístup stavia na pomere vierohodnosti nulovej vs. alternatívnej hypotézy, ale umožňuje zohľadnenie ďalších faktov vo forme pomeru šancí týchto hypotéz *a priori* (*prior odds*) pre určenie definitívneho pomeru šancí *a posteriori* (*posterior odds*):

$$L = \frac{P(H_0)}{P(H_A)} \times \frac{P(E|H_0)}{P(E|H_A)} = \frac{P(H_0|E)}{P(H_A|E)}$$

kde $P(H_0)$ resp. $P(H_A)$ predstavuje pravdepodobnosť nulovej resp. alternatívnej hypotézy *a priori*, teda bez zohľadnenia akýchkoľvek dodatkových informácií, $P(E|H_0)$ resp. $P(E|H_A)$ pravdepodobnosť pozorovaného dôkazu E za predpokladu platnosti H_0 resp. H_A , a $P(H_0|E)$ resp. $P(H_A|E)$ pravdepodobnosť *a posteriori*, teda po zohľadnení dôkazu E .

Príklad: AFLP profil biologickej stopy $X=110010111100\dots$ sa zhoduje s profilom podozrivého objektu. V populácii sa vyskytuje s frekvenciou $P_G = 10^{-4}$. Na základe predchádzajúcich skúseností s AFLP analýzou u rovnakého druhu bola senzitivita AFLP testu odhadnutá na 95% (teda frekvencia falošne negatívnych výsledkov je $v = 0,05$), špecificita na 99,99% (teda frekvencia falošne pozitívnych výsledkov je $u = 0,0001$). Aká je pomer pravdepodobností, že biologickú stopu zanechal podozrivý vs. že stopa pochádza z iného objektu?

Nulová hypotéza H_0 je, že biologická stopa pochádza od podozrivého, alternatívna hypotéza H_A je, že zhoda je náhodná a stopa pochádza od iného jedinca. Apriórna pravdepodobnosť alternatívnej možnosti je 10^{-4} (= pravdepodobnosť, že náhodne vybraný jedinec bude mať profil X), a keďže nulová a alternatívna hypotéza sa navzájom vylučujú, apriórna pravdepodobnosť H_0 je $(1 - 10^{-4}) = 0,9999$.

V zmysle tab. 8 môže zhoda profilu stopy a podozrivého za predpokladu platnosti nulovej hypotézy (teda stopa = podozrivý), pravdepodobnosť ktorej je $P(E|H_0)$, nastať v dvoch prípadoch:

i) skutočný profil stopy aj podozrivého je X, pričom tento profil sa v populácii vyskytuje s frekvenciou P_G , a AFLP test ho správne určí v oboch vzorkách, čo nastane s pravdepodobnosťou $(1 - v) \times (1 - v)$; teda tento prípad nastane s pravdepodobnosťou $P_G (1 - v)^2 = 0,0001 \times (1 - 0,05)^2 = 9,025 \cdot 10^{-5}$.

ii) skutočný profil stopy aj podozrivého nie je X, pričom tento prípad v populácii nastáva s frekvenciou $(1 - P_G)$, a AFLP test ho chybné určí ako X v oboch vzorkách (falošne pozitívny výsledok), čo nastane s pravdepodobnosťou $u \times u$; teda tento prípad nastane s pravdepodobnosťou $(1 - P_G) u^2 = 0,9999 \times 0,0001^2 = 9,999 \cdot 10^{-9}$.

Suma pravdepodobností týchto dvoch prípadov je $P(E|H_0) = 9,026 \cdot 10^{-5}$.

Alternatívna hypotéza H_A je, že biologická stopa pochádza od iného jedinca, než je podozrivý. Zhoda profilov za predpokladu platnosti alternatívnej hypotézy, pravdepodobnosť ktorej je $P(E|H_A)$, môže nastať v štyroch prípadoch:

iii) skutočný profil stopy aj podozrivého je X, ale táto zhoda je náhodná. Ak sa tento profil sa v populácii vyskytuje s frekvenciou P_G , pravdepodobnosť, že dvaja nezávislí jedinca budú jeho nositeľmi, je P_G^2 . AFLP test ho správne určí v oboch vzorkách, čo nastane s pravdepodobnosťou $(1 - v) \times (1 - v)$; teda tento prípad nastane s pravdepodobnosťou $P_G^2 (1 - v)^2 = 0,0001^2 \times (1 - 0,05)^2 = 9,025 \cdot 10^{-9}$.

iv) skutočný profil stopy je X a test ho určí správne [pravdepodobnosť $P_G (1 - v)$], profil podozrivého nie je X ale test ho určí falošne pozitívne ako X [pravdepodobnosť $(1 - P_G) u$]; teda tento prípad nastane s pravdepodobnosťou $P_G (1 - v) (1 - P_G) u = 0,0001 \times (1 - 0,05) \times 0,9999 \times (1 - 0,05) = 9,499 \cdot 10^{-9}$.

v) opačný prípad ako iv) – správny výsledok v prípade podozrivého, falošne pozitívny v prípade stopy; prípad nastáva s rovnakou pravdepodobnosťou $(1 - P_G) u P_G (1 - v) = 0,9999 \times (1 - 0,05) \times 0,0001 \times (1 - 0,05) = 9,499 \cdot 10^{-9}$.

vi) falošne pozitívne výsledky sa vyskytnú u stopy [pravdepodobnosť $(1 - P_G) u$] aj podozrivého [pravdepodobnosť rovnako $(1 - P_G) u$]; tento prípad nastane s pravdepodobnosťou $(1 - P_G)^2 u^2 = 0,9999^2 \times 0,0001^2 = 9,998 \cdot 10^{-9}$.

Suma pravdepodobností týchto štyroch prípadov je $P(E|H_A) = 3,802 \cdot 10^{-8}$.

Podľa Bayesovej teóremy platí

$$L = \frac{P(H_0|E)}{P(H_A|E)} = \frac{P(E|H_0) \times P(H_0)}{P(E|H_A) \times P(H_A)} = \frac{9,025 \cdot 10^{-5} \times 0,9999}{3,802 \cdot 10^{-8} \times 0,0001} = 23\,737\,075$$

Inými slovami, pri zohľadnení možných falošne negatívnych či falošne pozitívnych výsledkov testov je možnosť, že biologická stopa skutočne pochádza od podozrivého cca 24 miliónkrát pravdepodobnejšia, než možnosť, že pochádza od iného jedinca.

Hodnotenie príbuznosti

Princíp a používané markéry

Podľa evolučnej teórie ktorákoľvek dvojica jedincov rovnakého druhu (dokonca akákoľvek dvojica živých organizmov) je si navzájom príbuzná v tom zmysle, že zdieľa spoločných predkov, od ktorých zdedila časť svojej genetickej výbavy. Otázkou je len to, ako hlboko v ich rodokmeňoch títo spoloční predkovia ležia a koľkým mutačným zmenám došlo v tej časti dedičnej výbavy, ktorú od nich zdedili.

Kvantifikácia miery príbuzenstva vychádza z konceptu identity alel pôvodom (*identity by descent*; IBD), teda pravdepodobnosti, že dve náhodne vybrané alely rovnakého génu sú kópiami (replikami) tej istej sekvencie, ktorá sa vyskytovala v genóme ich spoločného predka. Z toho zároveň vyplýva, že IBD alely sú zároveň identické stavom (*identity in state*; IIS), teda že sú zhodné, za predpokladu, že pominieme veľmi zriedkavý prípad gametickej mutácie. Teoreticky by alely identické stavom mali byť identické aj pôvodom. V realite to úplne neplatí; pri genetických markéroch záleží na tom, čo označujeme termínom „alela“. Za zhodné sú v praxi brané markéry s rovnakým fenotypom, teda pri molekulárnych markéroch spravidla s rovnakou elektroforetickou mobilitou. Dva izoenzymové gény však nemusia mať zhodnú sekvenciu, aby boli exprimované do proteínu s rovnakou migračnou rýchlosťou. Nielen synonymné mutácie, ale ani mutácie nemeniace náboj proteínu resp. nemeniace zásadne tvar a veľkosť proteínu pri zachovaní jeho enzymatickej aktivity, sa na elektroforetickej mobilite nedajú rozoznať. To isté platí o mikrosatelitoch: ich alely sú rozlišované na základe celkovej dĺžky, ktorá ale neznamená zhodnú sekvenciu ich okrajových segmentov. Naviac rýchlosť migrácie v géli neodráža možnosť spätných mutácií (t.j. jeden mutačný krok náboj či dĺžku o jednu jednotku zväčší a následná mutácia ich vráti do pôvodného stavu). Fenotyp jedinca je ovšem to jediné, čo poznáme. Pôvod sa dá určiť spoľahlivo len z rodokmeňa, a ak by sme poznali rodokmeň, nepotrebovali by sme odhadovať príbuzenstvo, poznali by sme ho. Metódy odhadu príbuzenstva na základe porovnávania genotypov sú teda založené na identite stavu, nie na identite (neznámeho) pôvodu.

Zrejme prvou oblasťou vo forenzej praxi, v ktorej išlo o určovanie príbuzenstva, bolo testovanie otcovstva. *Mater certa, pater semper incertus est*; touto zásadou sa riadilo už rímske právo. V štandardnom prípade je matka známa, ale spor môže byť o určenie otca. To sa nemusí týkať len človeka, ale aj mnohých iných (najmä živorodých) živočíchov, najmä starajúcich sa o juvenilné potomstvo (aj keď pri niektorých druhoch samica starajúca sa o potomstvo nemusí byť nutne biologickou matkou). V environmentálnej forenzej praxi však predmetom záujmu môže byť aj miera príbuznosti dvoch jedincov, vyjadrená koeficientom príbuznosti.

Nástroje, využívané na hodnotenie príbuznosti, sú analogické ako nástroje pre genetický fingerprinting. Opäť prax začínala u človeka pri testovaní paternity krvnými skupinami, keďže lepšie nástroje neboli k dispozícii, a opäť súčasná prax využíva prakticky výlučne DNA markéry. Vysoká miera polymorfizmu je aj v tomto prípade výhodou. Na rozdiel od fingerprintingu, kde dominancia nie je zásadnou prekážkou, pri určovaní príbuznosti predstavuje výraznú komplikáciu. Výhody a nevýhody jednotlivých typov markérov sú rovnaké ako pri fingerprintingu, nemá zmysel ich opakovať. V súčasnosti štandardný nástroj predstavujú mikrosatelity (SSR, STR), pokiaľ sú pre príslušný druh k dispozícii sekvencie primerov pre ich identifikáciu. Ako náhradný nástroj možno využiť AFLP, ktoré sú ovšem dominantnými markérmami, teda odporúčateľné sú len v núdzovom prípade.

Štatistické hodnotenie

Testovanie paternity

Určenie otcovstva je tiež jedným z prípadov, keď identitu otca možno na základe porovnania genotypu matky, potomka a označeného otca (v konkrétnom lokuse alebo v sade lokusov)

možno vylúčiť jednoznačne, ale potvrdiť len s definovanou pravdepodobnosťou. Ak je identita matky známa a genotyp označeného jedinca nevylučuje, že môže byť otom, rozhodnutie o uznaní otcovstva vychádza rovnako ako pri genetickom fingerprintingu z porovnania vierohodností nulovej hypotézy (H_0 : označený samec je otec) a alternatívnej hypotézy (H_A : otcom je iný jedinec z populácie). Pri testovaní otcovstva sa tento pomer označuje ako index paternity:

$$PI = \frac{P(G_C|G_M, G_P, H_0)}{P(G_C|G_M, G_P, H_A)}$$

kde $P(G_C|G_M, G_P, H_0)$ resp. $P(G_C|G_M, G_P, H_A)$ je pravdepodobnosť, že genotyp potomka bude G_C za predpokladu rodičovských genotypov G_M a G_P pri platnosti nulovej resp. alternatívnej hypotézy (podmienené pravdepodobnosti). Pre ich určenie je najjednoduchšie použiť Punnettov štvorec, teda tabelárny diagram určujúci výsledky konkrétneho kríženia. Ak je napr. genotyp potomka A_1A_2 , genotyp matky A_1A_3 a genotyp jedinca označeného za otca A_2A_2 , potom pravdepodobnosti výskytu genotypových kombinácií v potomstve vychádzajú z možných gamét, ktoré rodičia tvoria: matka dva typy vajíčok (s alelami A_1 resp. A_3), otec len jeden typ spermii resp. peľu (s alelou A_2). Punnettov štvorec obsahujúci možné genotypy vznikajúceho potomstva teda vyzerá nasledovne:

		Otcovský príspevok	
		A_2	A_2
Materský príspevok	A_1	A_1A_2	A_1A_2
	A_3	A_2A_3	A_2A_3

Tab. 9 Pravdepodobnosti výskytu konkrétneho genotypu potomka G_C za predpokladu kombinácie materského (G_M) a otcovského (G_P) genotypu v prípade, že označený jedinec je otcom (H_0) a v prípade, že nie je otcom (H_A), a index paternity, teda pomer vierohodností oboch hypotéz (PI) (Evet a Weir 1998, upravené)

G_C	G_M	G_P	$P(G_C G_M, G_P, H_0)$	$P(G_C G_M, G_P, H_A)$	PI
A_iA_i	A_iA_i	A_iA_i	1	p_i	$1/p_i$
		$A_iA_j; j \neq i$	$1/2$	p_i	$1/2p_i$
		$A_jA_k; k \neq i, j$	0	p_i	0
	$A_iA_j; j \neq i$	A_iA_i	$1/2$	$p_i/2$	$1/p_i$
		$A_iA_j; j \neq i$	$1/4$	$p_i/2$	$1/2p_i$
		$A_iA_k; k \neq i, j$	$1/4$	$p_i/2$	$1/2p_i$
		$A_jA_k; k \neq i, j$	0	$p_i/2$	0
$A_iA_j; j \neq i$	A_iA_i	A_jA_j	1	p_j	$1/p_j$
		$A_iA_j; j \neq i$	$1/2$	p_j	$1/2p_j$
		$A_jA_k; k \neq i, j$	$1/2$	p_j	$1/2p_j$
		$A_kA_i; k, l \neq j$	0	p_j	0
$A_iA_j; j \neq i$	A_iA_i	A_iA_i	$1/2$	$(p_i + p_j)/2$	$1/(p_i + p_j)$
		$A_iA_j; j \neq i$	$1/2$	$(p_i + p_j)/2$	$1/(p_i + p_j)$
		$A_iA_k; k \neq i, j$	$1/4$	$(p_i + p_j)/2$	$1/2(p_i + p_j)$
		$A_kA_i; k, l \neq i, j$	0	$(p_i + p_j)/2$	0
$A_iA_k; k \neq i$	A_iA_i	A_iA_i	0	$p_j/2$	0
		A_jA_j	$1/2$	$p_j/2$	$1/p_j$
		$A_iA_j; j \neq i$	$1/4$	$p_j/2$	$1/2p_j$
		$A_jA_k; k \neq j$	$1/4$	$p_j/2$	$1/2p_j$
		$A_jA_i; l \neq j$	$1/4$	$p_j/2$	$1/2p_j$
		$A_kA_i; k, l \neq j$	0	$p_j/2$	0

je zrejme, že alelu A_1 musel potomok zdediť od matky. Ak je označený jedinec otcom, alelu A_2 musel potomok zdediť od neho, čo predstavuje 2 zo 4 prípadov, teda podmienená pravdepodobnosť výskytu konkrétneho genotypu potomstva pri platnosti H_0 je 0,5. Aj v prípade, že platí alternatívna hypotéza, mohol od matky potomok zdediť len alelu A_1 (alelu A_2 matka v genotype nemá) s pravdepodobnosťou 0,5, a alelu A_2 zdedil od iného jedinca v populácii, kde sa táto alela vyskytuje s frekvenciou $p(A_2)$ (pre skrátenie označme ako p_2). Pravdepodobnosť výskytu genotypu A_1A_2 u potomstva pri platnosti H_0 je teda $0,5p_2$. Index paternity je pomer oboch pravdepodobností, teda $PI = 0,5/0,5p_2 = 1/p_2$. Touto logikou je možné odvodiť všetky kombinácie genotypov a im prislúchajúce indexy paternity uvedené v tab. 9 (Goodwin et al. 2007).

Keďže index paternity je pomerom dvoch pravdepodobností, vzťahujú sa naň zákonitosti uvedené v Boxe I. Ak teda pre hodnotenie príbuzenstva použijeme viacero štatisticky nezávislých lokusov (v princípe lokusy vo väzbovej rovnováhe), je výsledný index paternity súčinom indexov za jednotlivé lokusy.

Príklad: Genotypy potomka, matky a potenciálneho otca, ako aj frekvencie alel v 5 SSR lokusoch v populácii vlka dravého v Nízkych Tatrách sú uvedené v tabuľke (alely sú označené veľkosťou fragmentov v bp). Frekvencie alel, relevantných pre výpočet konkrétneho PI , sú podčiarknuté. Označenie alel (i, j, k, l) zodpovedá tab. 9.

Lokus	G_C	G_M	G_P	frekvencie alel			
				i	j	k	l
CPH12	191/191	191/197	191/197	191: <u>0,0328</u>	197: 0,4303		
CPH2	93/93	91/93	93/93	093: <u>0,2131</u>	091: 0,0041		
CPH4	139/147	147/147	139/139	147: 0,7131	139: <u>0,0164</u>		
CPH5	111/113	111/113	113/115	111: <u>0,0574</u>	113: <u>0,7705</u>	115: 0,0615	
FH2088	125/127	125/131	109/127	125: 0,0043	127: <u>0,2521</u>	131: 0,1154	109: 0,0043

Aký je index paternity potenciálneho otcovského jedinca?

Indexy paternity podľa jednotlivých lokusov sú vypočítané v nasledujúcej tabuľke:

Lokus	G_C	G_M	G_P	$P(G_C G_M, G_P, H_0)$	$P(G_C G_M, G_P, H_A)$	PI
CPH12	ii	ij	ij	0,25	$p_i/2=0,0164$	$1/2p_i=15,24$
CPH2	ii	ij	ii	0,5	$p_i/2=0,1066$	$1/p_i=4,69$
CPH4	ij	ii	jj	1	$p_j/2=0,0082$	$1/p_j=121,95$
CPH5	ij	ij	ik	0,25	$(p_i+p_j)/2=0,4140$	$1/2(p_i+p_j)=0,60$
FH2088	ij	ik	jl	0,25	$p_l/2=0,1261$	$1/2p_l=1,98$
celkovo				0,00781250	0,00000075	10449,32

Logika výpočtu sa dá ilustrovať na príklade lokusu FH2088. Rodičia majú genotypy 125/131 a 109/127. V ich potomstve sa ktorákoľvek kombinácia materskej a otcovskej alely môže objaviť s rovnakou pravdepodobnosťou: 109/125, 125/127, 109/131, 127/131. Za predpokladu platosti H_0 (teda označený jedinec je skutočne otcom) je kombinácia, ktorá sa objavila u potomka (125/127) jedna zo štyroch možných, teda $P(G_C|G_M, G_P, H_0) = 1/4 = 0,25$. Zároveň je zrejme, že alelu 125 potomok musel zdediť od matky, s pravdepodobnosťou 0,5 (je to jedna z dvoch alel v genotype matky), potenciálny otec ju v genotype nemá, teda od otca musel zdediť alelu 127. Ak platí H_A , teda otcom je iný jedinec z populácie, je pravdepodobnosť, že náhodne vybraný jedinec z populácie odovzdá potomkovi alelu 127 úmerná jej frekvencii v populácii, ktorá je 0,2521. Keďže získanie konkrétnej alely od matky a od otca sú dva nezávislé javy, pravdepodobnosť ich

súčasného výskytu je daná súčinom pravdepodobnosti jedného a pravdepodobnosti druhého, teda $0,5 \text{ (matka)} \times 0,2521 \text{ (otec)} = 0,1261 = P(G_C|G_M, G_P, H_A)$. Index paternity n ziklade samotného lokusu FH2088 je teda $0,25/0,1261 = 1,98$. Ak by sme teda posudzovali paternitu len na základe FH2088, pravdepodobnosť, že potenciálny jedinec nie je otcom, je cca 2-krát vyššia než pravdepodobnosť, že otcom je.

Ani jeden z lokusov teda otcovstvo označeného jedinca nevyvylúčil. Pravdepodobnosť, že otcom skutočne je, je 10449-krát vyššia, než že otcom nie je. Z výsledku je zrejmé, že nie všetky lokusy rovnako prispievajú k preukázaniu otcovstva. V prípade lokusu CPH5 je pravdepodobnosť nulovej hypotézy (je otec) dokonca menšia, než pravdepodobnosť alternatívnej hypotézy (nie je otec), čo je spôsobené skutočnosťou, že v genotype potomstva sa vyskytuje veľmi častá alela (131 bp, s frekvenciou 0,7705). Pre určovanie paternity je teda výhodné používať čím polymorfnejšie lokusy – čím viac alel, tým menšie sú v priemere ich frekvencie.

Odhad koeficientu príbuznosti

Často môže byť potrebné odhadnúť mieru príbuznosti dvojice jedincov bez akejkoľvek vedomosti o ich rodokmeňoch; v prípade jedincov z prirodzených populácií je to bežná situácia.

Mieru príbuznosti medzi dvoma jedincami kvantifikuje koeficient príbuznosti (*coancestry coefficient, kinship coefficient*), ktorý stanovuje pravdepodobnosť, že dvojica alel rovnakého génu vybraných z genotypov jedného a druhého jedinca je identická pôvodom (IBD). V prípade dvojice diploidných organizmov s genotypmi A_iA_j a A_kA_l je koeficient príbuznosti aritmetickým priemerom pravdepodobností IBD pre všetky možné dvojice alel:

$$r = \frac{1}{4}P(A_i \equiv A_k) + \frac{1}{4}P(A_i \equiv A_l) + \frac{1}{4}P(A_j \equiv A_k) + \frac{1}{4}P(A_j \equiv A_l)$$

kde operátor \equiv označuje identitu alel pôvodom.

Ak nepoznáme rodokmene jedincov, pravdepodobnosť IBD nemôžeme priamo vypočítať, len odhadnúť na základe zhody ich genotypov (identita alel stavom, IIS). Alely, ktoré sú zhodné stavom, sú identické aj pôvodom, ale naopak to nutne neplatí.

Odhad stupňa príbuznosti na základe zdieľania IIS alel vychádza z toho, že dva diploidné jedince (nezávisle na tom, či sú príbuzné alebo nie) môže zdieľať 0, 1 alebo 2 alely v genotype ktoréhokoľvek lokusu. Koeficienty príbuznosti pre jednotlivé typy vzťahov uvedené v podkap. *Príbuzenstvo a inbreeding* sú priemerné hodnoty za celý genóm. V konkrétnom lokuse nemusia ani dvaja vlastní súrodenci (plnosesterské potomstvo) zdieľať žiadnu alelu: pokiaľ sú obaja rodičia v tomto lokuse heterozygoti, môže jeden potomok získať inú kombináciu ich alel než druhý. Hodnotenie príbuznosti teda musí vychádzať z väčšej sady markérových lokusov. Koeficient príbuznosti r_{ij} medzi jedincami i a j je možné odhadnúť nasledovne (Weir a Goudet 2017):

$$r_{ij} = \frac{M_{ij} - M_S}{1 - M_S}$$

kde M_{ij} je počet IIS alel v genotypoch L markérových lokusov jedincov i a j : $M_{ij} = \sum_l^L [1 + (k_{li} - 1)(k_{lj} - 1)]/2$ (k_{li} – počet IIS alel v l . lokuse i . jedinca) a M_S je priemerná zhoda v IIS alelách v súbore n jedincov:

$$M_S = \frac{1}{n(n-1)} \sum_i^n \sum_{j \neq i}^n M_{ij}$$

Tento spôsob odhadu však neberie do úvahy alelické frekvencie. Odhad adjustovaný na kontext populácie je možné vypočítať nasledovne (VanRaden 2008):

$$r_{ij} = \frac{\sum_l^L (k_{li} - 2p_l)(k_{lj} - 2p_l)}{2 \sum_l^L p_l(1 - p_l)}$$

kde p_l je frekvencia referenčnej alely v l . lokuse.

Určovanie geografického pôvodu

Princíp a používané markéry

Vo forennej praxi sa môžeme stretnúť s otázkou odkiaľ biologická stopa pochádza. Existuje mnoho oblastí, v ktorých je geografický pôvod relevantný, od sledovania ilegálneho obchodu s drevom, cez ilegálny obchod so zvieratami alebo produktmi z nich (slonovina, suveníry) až po pašovanie chránených rastlín. V prípadoch, že chránené sú len konkrétne populácie druhu, ide o rozlíšenie ilegálne získaného materiálu od legálneho (napr. mnohé tropické dreveniny ohrozené v mieste ich prirodzeného výskytu sú pestované a hospodársky využívané v iných oblastiach). U druhov, ktoré sú chránené ako celok, môže byť cieľom identifikácia „ohnísk pytláckenia“, teda oblastí, kde je odchyt, lov či zber obzvlášť intenzívny.

Pri otázke na pôvod je však rozhodujúca otázka mierky. Pýtame sa na to, či povedzme konkrétna vzorka bukového dreva pochádza z Európy? Zo strednej Európy? Zo Slovenska? Z okresu Zvolen? Z porastu 5117a na LC Ihráč? Nie na všetky tieto otázky umožňuje forenzná genetika dať rozumnú odpoveď, akceptovateľnú v prípadnom právnom spore.

Už intuitívne je zrejmé, že na rozlíšenie, či vzorka pochádza z tej či onej populácie (ak na toto rozlíšenie majú byť použité genetické nástroje), je nutné, aby sa populácie svojou genetickou štruktúrou odlišovali. Mechanizmov, ktoré utvárajú genetickú štruktúru, je viacero: niektoré z nich prispievajú ku genetickej divergencii, teda zväčšovaniu rozdielov (prírodný výber, náhodné zmeny, vznik nových mutácií), iné naopak rozdiely zotierajú (migrácia, tok génov prostredníctvom prenosu peľu semien a pod.). Pojem „populácia“ sa v genetickom výskume často zvykol používať len v operatívnom zmysle, teda na označenie väčšej či menšej lokálnej skupiny jedincov na konkrétnej lokalite. Pre forenzné účely je však nutné populáciu biologicky aj geograficky vyčleniť. Populácia predstavuje reprodukčnú jednotku, súbor jedincov, medzi ktorými dochádza ku genetickej výmene v procese reprodukcie, a ktorá je prinajmenšom čiastočne izolovaná od iných takýchto súborov jedincov. Akékoľvek postupy identifikácie príslušnosti k populácii vychádzajú z týchto predpokladov. Zároveň je dôležitý historický kontext. Každá populácia nejako vznikla, vyvíjala sa a diferencovala sa od iných populácií rovnakého druhu. Vývoj rôznych súčastí genómu bol pritom rozdielne rýchly, v závislosti od rýchlosti mutácií v danej časti genómu, miery rekombinácie, vektorov prenosu génov medzi populáciami a ďalších faktorov. Čím väčšia je mierka, v ktorej identifikáciu pôvodu uskutočňujeme, tým výhodnejšie je použiť konzervatívne, málo mutujúce sekvencie, a zároveň tým spoľahlivejšie je možné pôvod určiť. Pôvod vzorky z konkrétnej časti kontinentu sa často dá určiť jednoznačne alebo s vysokou pravdepodobnosťou. Naopak, určenie pôvodu z konkrétnej lokality je spravidla problematické (s výnimkou endemických druhov alebo druhov s extrémne fragmentovaným areálom), vyžaduje naopak vysoko polymorfné markéry, a spravidla aj vylúčenie pôvodu je možné len s relatívne nízkou pravdepodobnosťou.

V kontexte Európy či Severnej Ameriky je dôležitým faktorom určujúcim genetickú štruktúru populácií postglaciálna migrácia. V pleistocéne (staršie štvrtohory) bola značná časť oboch kontinentov pokrytá kontinentálnym ľadovcom; v prípade Slovenska končil tesne za našou severnou hranicou, takže aj klíma nášho územia pripomínala skôr sever Sibíri než dnešný stav. Väčšina rastlinných druhov sa stiahla na stanovištia, na ktorých dokázala prežiť (glaciálne refúgiá), spravidla v južnej časti kontinentu (Balkán, Apeninský a Iberský poloostrov). Živočíšne druhy, ktoré na ich boli závislé, ich nasledovali. Refugiálne populácie boli spravidla malé a výrazne fragmentované, s minimálnou vzájomnou výmenou génov. V takýchto podmienkach dochádzalo k rýchlej diferenciácii medzi refúgiami resp. refugiálnymi oblasťami predovšetkým v dôsledku náhodných zmien (genetického driftu), keď mnohé alely z genofondu populácií vypadli. Aj keď sa následne v holocéne rastlinné a živočíšne druhy rozšírili do ich súčasných areálov, stopy tento diferenciácie sa na ich genofonde dajú rozoznať dodnes. V prípade rastlín sa to týka predovšetkým génov resp. iných sekvencií

DNA, ktoré sa dedia len po matke. Takéto sekvencie sa môžu šíriť len prostredníctvom semien alebo vegetatívnou cestou; oba spôsoby šírenia sú podstatne menej efektívne a fungujú v podstatne menšej priestorovej mierke než šírenie peľom. Jedná sa o mitochondriálne a chloroplastové sekvencie väčšiny rastlinných druhov (aj keď existujú významné výnimky – napr. v čeľadi Pinaceae, ku ktorej patria všetky naše významné ihličnany, sa chloroplastová DNA dedí po otcovi a teda šíri peľom). U živočíchov, kde tok génov primárne prebieha prostredníctvom migrácie dospelých jedincov a ich následného zapojenia do reprodukcie, nie je pohlavie migrujúceho jedinca dôležité (pokiaľ u daného druhu nie je významný rozdiel v migračnej schopnosti medzi pohlaviami). Na hraniciach migračných prúdov vznikajú hybridné zóny, v ktorých sa výskyty špecifických alel prekrývajú. Ich šírka býva rozdielna a závisí od vzdialenosti šírenia semien (rastliny) či migračných vzdialeností (živočíchy).

Opäť tu platí, že rozdiely medzi populáciami pochádzajúcimi z rozdielnych refúgií sú najvýraznejšie v uniparentálne (t.j. len po jednom z rodičov) dedených sekvenciách, teda mitochondriálnych (maternálna dedičnosť) resp. sekvenciách chromozómu Y (paternálna dedičnosť). Výhodou uniparentálne dedených sekvencií je chýbajúca rekombinácia (keďže sa dedia len po jednom z rodičov, nemajú sa s čím rekombinovať). Je teda možné identifikovať príslušnosť k populácii na základe stálej kombinácie alel (haplotypu).

Biparentálne dedené jadrové gény sú medzi migračnými prúdmi premiešané v podstatne väčšom rozsahu. Navyiac, vďaka crossing-overu sa alely jadrových génov rekombinujú. Pokiaľ sa jedná o kódujúce sekvencie (gény *sensu stricto*), ich zastúpenie v populáciách môže byť potenciálne modifikované aj prírodným výberom spojeným s adaptáciou na lokálne podmienky prostredia. Variabilita takýchto sekvencií sa spravidla prejavuje aj v podstatne menšej mierke, sú teda využiteľné pre identifikáciu pôvodu z lokálnych populácií.

Pre identifikáciu geografického pôvodu vo veľkej mierke je v princípe možné využiť spektrum markérov, využívaných v identifikácii druhovej príslušnosti. U živočíchov a húb ide primárne o sevenovanie mitochondriálnych génov (*cox1*, *cytb*, 12S, 16S, 18S, 28S rRNA). Tieto gény však často vykazujú príliš malú vnútrodruhovou variabilitu, v poslednej dobe je preferovaný skôr hypervariabilný región D-slučky mtDNA (*D-loop*), čo je oblasť DNA, v ktorej sú komplementárne reťazce separované na určitom úseku tretím reťazcom, ktorý je čiastočne komplementárny s jedným z pôvodných – celá štruktúra tvarom pripomína písmeno D. U rastlín sa využívajú skôr chloroplastové sekvencie (ITS1, ITS2, *rbcL*, *matK*, *psbA*).

Sekvenovanie je ovšem pre prevádzkové účely relatívne náročný a drahý postup. V mnohých prípadoch môžu rovnakú službu poskytnúť jednoduchšie a lacnejšie markéry. V minulosti bola najmä u rastlín široko využívaná PCR-RFLP analýza organelárnej DNA (u rastlín predovšetkým chloroplastovej): pomocou PCR je amplifikovaný úsek (alebo viacerých úsekov) cpDNA a následne sú získané fragmenty štiepené restriktívnou endonukleázou. Kombinácia restriktívnych miest cez viaceré markéry poskytuje haplotypy charakteristické pre rôzne genetické línie. V niektorých prípadoch si vystačí len so samotnou PCR: u jedle bielej sa genetická línia pochádzajúca z refúgií na Apeninskom poloostrove odlišuje od balkánskej dlhým intronom (80 bp) v mitochondriálnom gène pre NADH-dehydrogenázu (*nad5-4*). Príslušný fragment stačí amplifikovať PCR a následne sa obe varianty dajú odlíšiť aj po separácii na agarózovom géle; cena analýzy tu predstavuje niekoľko centov na vzorku. Hranica medzi genetickými líniami prebieha Východnými Karpatmi v blízkosti ukrajinskej dediny Jasiňa, a je veľmi ostrá. Identifikácia je ale možná len v hrubom merítku: výsadbu materiálu jedle z Rumunska (ktorá je v rozpore so slovenskou legislatívou) by bolo možné touto cestou jednoducho zistiť, výsadbu českých či francúzskych jedlí (rovnako ilegálnu) už nie.

Historická fragmentácia zanecháva stopy aj v jadrovom (biparentálne dedenom) genóme. V prípade rastlín sa však jadrové gény šíria aj prostredníctvom samčích gamét, teda peľu,

u ktorého je nezávisle na vektore šírenia (vietor, hmyz, voda) priestorový dosah podstatne väčší v porovnaní so semenami. U jadrových markérov preto spravidla nie sú viditeľné žiadne ostré hranice medzi oblasťami s rozdielou štruktúrou genofondu, prechody medzi nimi (hybridné zóny) sú široké a kontinuálne. Pre identifikáciu genetickej štruktúry je v tomto prípade nutné využívať pokročilejšie matematicko-štatistické postupy.

Pri určovaní pôvodu v lokálnom merítke, teda z konkrétnej populácie či malej geografickej oblasti, je rozhodujúcim faktorom charakter areálu. V prípade druhov s výrazne fragmentovaným areálom, kde každý fragment predstavuje samostatnú populáciu izolovanú od zvyšku druhu (stenoendemity, lokálne rasy), je spravidla možné použiť rovnaké nástroje ako v celoareálovej mierke, teda spravidla maternálne dedené markéry. U druhov s kontinuálnym areálom je tento prístup aj v prípade malej populačnej hustoty problematický. Definovať hranice populácií je v tomto prípade ťažké, medzi jednotlivými výskytmi môže prebiehať intenzívny tok génov. Určenie pôvodu je v týchto prípadoch spravidla nejednoznačné a nie vždy poskytujú genetické nástroje dôkaznú silu vyžadovanú z právneho hľadiska.

Na rozdiel od fingerprintingu či určovania príbuznosti možno pri určovaní geografického pôvodu použiť len markéry, o ktorých niečo vieme, ktoré boli v minulosti použité pri skúmaní variability v rámci celého areálu druhu alebo jeho podstatnej časti (spravidla ide o vedecký výskum), a ktoré sú v dostatočnej miere reprodukovateľné. Z toho vyplýva, že použiteľnosť anonymných markérov typu AFLP pre tieto účely je výrazne obmedzená; na to, aby sme genotyp alebo haplotyp jedinca vedeli zaradiť do kontextu doterajších poznatkov a geografickom rozdelení genetickej variability, musíme si byť istí, že ho porovnávame s identickým lokusom resp. lokusmi. Z dostupných technických nástrojov sú preto v súčasnosti štandardom okrem organelárnej DNA aj jadrové mikrosatelity.

Štatistické a bioinformatické hodnotenie

Pre výber štatistického postupu pre geografické priradenie vzorky sú dôležité dve otázky:

- je cieľom potvrdenie resp. vylúčenie pôvodu z konkrétneho známeho zdroja alebo odhad, z ktorej oblasti (širšie či užšie definovanej) vzorka pochádza?
- jedná sa o konkrétneho jedinca alebo populačnú vzorku (súbor jedincov náhodne vybraných z populácie)?

V prípade porovnania vzorky s deklarovaným zdrojom je možné len vylúčenie, ale nie potvrdenie pôvodu. V istom zmysle najjednoduchšia je situácia pri použití vysoko diferencovaných organelárných markérov, u ktorých sú hranice medzi rozšírením jednotlivých genetických línií spravidla ostré. Vo vzorke (aj v prípade, ak ide o populačnú vzorku, teda viaceru jedincov) sa tu spravidla vyskytuje jediný haplotyp, ktorý možno porovnať s haplotypom potenciálneho zdroja. Opäť platí: ak sa nezhoduje, je pôvod zo zdroja vylúčený, ak sa zhoduje, neznamená to ešte, že pôvod zo zdroja je potvrdený. Komplikovanejšia situácia nastáva len vtedy, ak spadá poloha deklarovaného zdroja zhodou okolností do hybridnej zóny. V tomto prípade je potenciálna zdrojová populácia zmeskou dvoch (prípadne viacerých) haplotypov. Pravdepodobnosť vylúčenia pôvodu konkrétneho jedinca z takéhoto zdroja je rovná doplnku k frekvencii jeho haplotypu ($1 - p_i$) v zdrojovej populácii, ktorá sa spravidla pohybuje v jednotkách alebo desiatkach percent, čo spravidla nie je použiteľné ako dôkaz v právnom spore (malá sila dôkazu). V prípade populačnej vzorky je možnosť zozbierať reprezentatívny výber materiálu z deklarovaného zdroja a následne porovnať alelické frekvencie vzorky a zdroja štatistickým testom (Box V). Táto možnosť nezávisí na druhu použitého markéra, prichádza do úvahy ako pri organelárných, tak aj pri kodominantných jadrových markéroch. Pre testovanie rozdielov frekvencií sa používa χ^2 -test; testovania charakteristika s $(k - 1) \times (2 - 1)$ stupňami voľnosti (k – počet riadkov kontingenčnej tabuľky = počet alel, počet stĺpcov je 2 = počet porovnávaných vzoriek) sa počíta

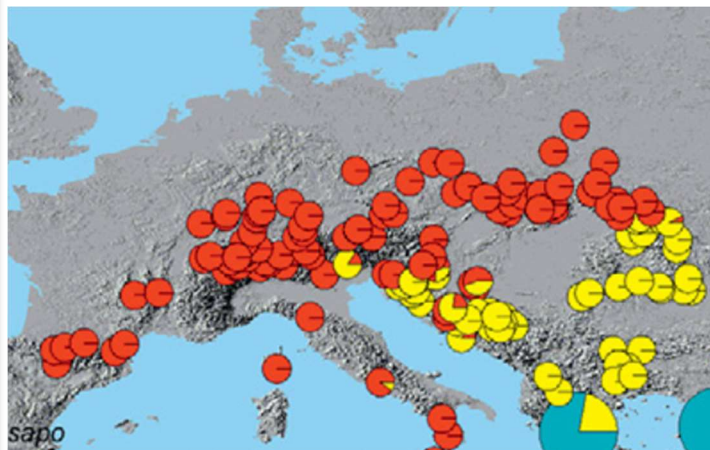
$$\text{ako } \chi^2 = \sum_i^k \sum_1^2 \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

kde n_{ij} je skutočný a n'_{ij} očakávaný počet exemplárov i . alely v j . vzorke. Očakávaný počet sa vypočíta ako $n'_{ij} = n_{i.} \times n_{.j} / n_{..}$, kde $n_{i.}$ je suma i . riadku ($n_{i.} = \sum_j n_{ij}$), $n_{.j}$ je suma j . stĺpca kontingenčnej tabuľky a $n_{..}$ je celková suma.

Príklad: Pri zalesňovaní na lokalite Čertovica v NP Nízke Tatry bol použitý reprodukčný materiál s deklarovaným pôvodom z uznaného porastu all 16LM-015 9severoslovenská semenárska oblasť, 6. l.v.s. Pre overenie pôvodu bolo testovaných 50 vysadených jedincov a 60 jedincov z uznaného porastu (UP) použitím mitochondriálneho markéra *nad5-4* a nSSR lokusu *SF333*. Zistené frekvencie alel sú uvedené v tabuľke. Je dôvod domnievať sa, že výsadba pochádza z iného zdroja?

lokus	alela (bp)	výsadba	UP
<i>nad5-4</i>	150	0	0
	230	1,00	1,00
<i>SF333</i>	168	0,24	0,15
	170	0,01	0,05
	172	0,59	0,50
	174	0,00	0,05
	176	0,01	0,05
	178	0,15	0,20

V mitochondriálnom lokuse *nad5-4* sa v stredoeurópskych populáciách vyskytuje výlučne alela 230 bp, čo potvrdzuje aj výsledok z UP. Vo výsadbe sa tiež u všetkých jedincov vyskytla alela 230 bp. Na základe tejto skutočnosti možno vylúčiť, že materiál pochádza z Balkánu, ale ešte to nedokazuje, že pochádza z konkrétneho UP; aj keby pochádzal z Pyrenejí či Sicílie, vykazoval by tento haplotyp.



Frekvencie alel *nad5-4* v areáli jedle bielej: žltá 150 bp, červená 230 bp. Liepelt et al. (2009) upravené

Alelické frekvencie v mikrosatelitnom lokuse *SF333* sú na prvý pohľad odlišné, ale otázkou je, či je táto odlišnosť výsledkom náhodného výberu vzoriek jedincov alebo je dôsledkom rozdielu v genetickej štruktúre medzi výsadbou a UP. Pravdepodobnosť existencie rozdielu možno otestovať χ^2 -testom. Vo vzorkách je zastúpených 50 resp. 60 diploidných jedincov, čo predstavuje $2 \times 50 = 100$ resp. $2 \times 60 = 120$ kópií nSSR fragmentu. Počet alel je daný súčinom frekvencií

a celkového počtu, napr. počet exemplárov alely 168 vo výsadbe je $0,24 \times 100 = 24$ ks. Očakávaný počet exemplárov sa vypočíta ako $n'_{ij} = n_i \times n_j / n_{..}$, napr. u alely 168 vo výsadbe (v prípade rovnosti výsadby a UP; skutočný počet je podčiarknutý) je $42 \times 100 / 220 = 19,09$; rovnako možno vypočítať ostatné očakávané počty.

alela (bp)	skutočné počty		Σ	očakávané počty	
	výsadba	UP		výsadba	UP
168	<u>24</u>	18	42	19,09	22,91
170	1	6	7	3,18	3,82
172	59	60	119	54,09	64,91
174	0	6	6	2,73	3,27
176	1	6	7	3,18	3,82
178	15	24	39	17,73	21,27
Σ	100	120	220		

Testovacia charakteristika je $\chi^2 = \sum_i \sum_j \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} = \frac{(24 - 19,09)^2}{19,09} + \frac{(1 - 3,18)^2}{3,18} + \dots + \frac{(24 - 21,27)^2}{21,27} = 14,386$

čo pre $(6-1) \times (2-1) = 5$ stupňov voľnosti dáva $\alpha = 0,0133$. S pravdepodobnosťou 98,67% je teda možné tvrdiť, že genetická štruktúra výsadby sa odlišuje od udávaného zdroja, teda deklarovaný údaj o pôvode nie je pravdivý. Vo výskume je takáto pravdepodobnosť akceptovaná; či by stačila aj súdu, je otázne.

Iným typom problému je, ak je zdrojová populácia neznáma, a cieľom je odhadnúť, z ktorej časti areálu vzorka (jedinec alebo populačná vzorka) pochádza. V tomto prípade je priradenie závislé na dostupnosti referenčného datasetu, teda údajov o tom, ako je genetická variabilita v zvolených markérových lokusoch organizovaná v rámci areálu. Akékoľvek DNA dáta sú užitočné len vtedy, ak ich je s čím porovnať, a pri určovaní geografického pôvodu to platí dvojnásobne. Na rozdiel od humánnej kriminalistiky, kde v mnohých štátoch existujú DNA databázy budované pre forenzné účely, pri divožijúcich organizmoch je zaistenie referenčného materiálu najdôležitejším praktickým problémom. Forenzné analýzy sa v praxi primárne týkajú chránených druhov, teda organizmov, ktoré sú väčšinou zriedkavé a zber ich vzoriek je obtiažny. Forenzná prax je odkázaná na údaje, získavané pre iné (spravidla výskumné) účely, ktorých kvalita a reprezentatívnosť nie je vždy zaručená. Samozrejme platí, že čím vyššia je hustota zberu dát a čím väčšie sú populačné vzorky v referenčnom datasete, tým spoľahlivejšie je geografické priradenie. Pre relatívne spoľahlivý odhad alelických frekvencií v populácii sa vyžaduje zber cca 30–50 jedincov na populáciu (aj v závislosti na type markéra), čo pri výskumných prácach nie je vždy dodržané.

Napriek týmto skutočnostiam zostávajú publikované dáta vo väčšine prípadov jediným použiteľným referenčným zdrojom pre geografické priradenie. Organizovať reprezentatívny zber materiálu pre účely jednej forenznej analýzy by bolo finančne a logisticky príliš náročným podujatím. Istú alternatívu poskytujú zbierky materiálu (muzeálne zbierky, génové banky, herbáre a pod.), ale aj v tomto prípade by bolo potrebné zorganizovať a najmä zaplatiť analýzy materiálu, čo môže byť považované za neefektívne. Publikované údaje sú pravidla verejne prístupné; väčšina vedeckých časopisov dnes podmieňuje zverejnenie vedeckej práce sprístupnením surových údajov (genotypy skúmaných jedincov, alelické frekvencie a pod, vrátane geografických údajov) vo verejných databázach (repozitóriách) a povolením ich využitia nielen pre vedecké, ale aj iné účely, vrátane forezných. Najznámejšie repozitóriá sú

DRYAD (datadryad.org), GBIF (Global Biodiversity Information Facility; <https://www.gbif.org>), Zenodo (<https://zenodo.org/>) a ďalšie.

V prípade určenia zdrojovej populácie pre jedinca prichádza do úvahy viacero postupov (Cornuet et al. 1999). Jedným je frekvenčná metóda, ktorá priraduje jedinca do populácie, v ktorej je výskyt jeho genotypu najpravdepodobnejší. Ak frekvenciu k . alely j . lokusu v i . referenčnej populácii označíme ako p_{ijk} , pravdepodobnosť výskytu konkrétneho genotypu vychádza zo zákonitostí populačnej genetiky: pravdepodobnosť výskytu homozygota A_kA_k je p_{ijk}^2 , pravdepodobnosť výskytu heterozygota A_kA_l je $2p_{ijk}p_{ijl}$. Pokiaľ použijeme viacero štatisticky nezávislých lokusov (t.j. v princípe lokusov vo väzbovej rovnováhe), pravdepodobnosť výskytu multilokusového genotypu je v zmysle Boxu I súčinom pravdepodobností z jednotlivé lokusy. Jediniec je priradený referenčnej populácii, pre ktorú má výskyt jeho genotypu najvyššiu pravdepodobnosť. Problémom tohoto prístupu je odhad alelických frekvencií v referenčných populáciách. Zriedkavé alely nemusia byť referenčnou vzorkou zachytené, ale môžu sa vyskytnúť v genotype klasifikovaného jedinca. V tomto prípade je ich frekvencia v referenčnej populácii nulová, čo automaticky vedie k diskvalifikácii tejto populácie, aj keď je v skutočnosti zdrojovou (ak $p_{ijk}=0$ pre ktorýkoľvek lokus, tak aj súčin pravdepodobností cez všetky lokusy je automaticky rovný nule). Riešením je buď pridať klasifikovaného jedinca postupne ku každej referenčnej populácii (čím sa eliminuje nulová frekvencia ktorejkoľvek alely v genotype jedinca), alebo nahradením nulových frekvencií malou konštantnou hodnotou, prípadne prevrátenou hodnotou rozsahu výberu v referenčnej populácii.

Druhou možnosťou je Bayesovský prístup (Rannala a Mountain 1997), ktorý ovšem vyžaduje surové dáta (genotypy jedincov) referenčných populácií. Pravdepodobnosť výskytu homozygota A_kA_k v j . lokuse v i . populácii je

$$P(A_kA_k|i) = \frac{(n_{ijk} + \frac{1}{K_j} + 1)(n_{ijk} + \frac{1}{K_j})}{(\sum_k n_{ijk} + 2)(\sum_k n_{ijk} + 1)}$$

pravdepodobnosť výskytu heterozygota A_kA_l v j . lokuse v i . populácii je

$$P(A_kA_l|i) = \frac{2(n_{ijk} + \frac{1}{K_j})(n_{ijl} + \frac{1}{K_j})}{(\sum_k n_{ijk} + 2)(\sum_k n_{ijk} + 1)}$$

kde n_{ijk} je počet kópií k . alely a K_j je celkový počet alel v j . lokuse vo všetkých populáciách spolu. Jediniec je opäť priradený populácii, v ktorej sa jeho genotyp vyskytne s najvyššou pravdepodobnosťou.

V prípade populačných vzoriek je možné použiť prístupy založené na genetických vzdialenostiach alebo na simulačných modeloch. V prvom prípade postup spočíva vo využití niektorého z veľkého počtu koeficientov genetickej podobnosti resp. nepodobnosti populácií (genetických vzdialeností), kvantifikujúcich mieru genetickej diferenciacie medzi populáciami. Zrejme najbežnejšou je genetická vzdialenosť podľa Neia (1972):

$$D = -\ln \frac{J_{ij}}{\sqrt{J_i J_j}}, \text{ kde } J_i = \sum_k p_{ik}^2, J_j = \sum_k p_{jk}^2 \text{ a } J_{ij} = \sum_k p_{ik} p_{jk}, \text{ pričom } p_{ik} \text{ je frekvencia } k.$$

alely v i . populácii. Opäť platí, že sa vypočítajú genetické vzdialenosti medzi klasifikovanou vzorkou a každou z referenčných populácií a vzorka je priradená k tej referenčnej populácii, s ktorou vykazuje najmenšiu genetickú vzdialenosť. V princípe tento prístup možno použiť aj pre identifikáciu pôvodu jedinca, kde alelická frekvencia p_{ik} môže nadobudnúť len hodnoty 0 (absentuje), 0,5 (heterozygot) alebo 1 (homozygot v danej alele).

Box V Testovanie štatistických hypotéz

Ako bolo konštatované v Boxe II, zaujímajú nás vždy vlastnosti základného súboru, ktorý ale spravidla nemáme k dispozícii, a preto sme odkázaní na výberovú vzorku z neho. Dva výbery hoci aj z toho istého súboru sa budú navzájom líšiť: ak povedzme z lesného porastu vyberieme náhodne 10 stromov, zmeriame ich výšky a vypčítame z nich priemer, je extrémne málo pravdepodobné, že presne rovnakú priemernú výšku dosiahne iných 10 stromov náhodne vybraných z rovnakého porastu. Pokiaľ teda porovnávame dva navzájom sa líšiace výberové súbory, v skutočnosti nás zaujíma, či sa odlišujú aj parametre základných súborov, z ktorých boli vybrané. Analogickým problémom je aj porovnanie parametra základného súboru oproti nejakej očakávanej konštante či porovnávanie parametrov viac než dvoch základných súborov. Na základe toho, o aký parameter ide, aké veľké sú výberové vzorky, a nakoľko sa líšia, sa dá stanoviť pravdepodobnosť, že sa líšia aj základné súbory – stanovenie tejto pravdepodobnosti je princípom testovania štatistických hypotéz.

Postup testovania je nasledovný:

- Stanovenie dvoch navzájom sa vylučujúcich hypotéz o testovaných parametroch. Ako nulová hypotéza (H_0) sa vždy berie tá, ktorú netreba dokazovať a platí implicitne, t.j. hypotéza, že dva parametre ZS sa navzájom nelíšia resp. že parameter ZS sa nelíši od očakávanej konštantnej hodnoty. To, čo treba dokázať, je opak, alternatívna hypotéza (H_A), teda hypotéza, že rozdiel existuje.
- Stanovenie pravdepodobnostného rozdelenia testovaného parametra (viď Box III). Napr. priemery majú normálne rozdelenie, variance majú χ^2 -rozdelenie atď.
- Výber a výpočet testovacej charakteristiky, ktorá má známe pravdepodobnostné rozdelenie (ktoré nemusí byť zhodné s testovaným parametrom: pre testovanie rozdielu dvoch variancií s χ^2 -rozdelením sa používa F -test, kde testovacia charakteristika má Fisher-Snedecorovo rozdelenie). Testovacia charakteristika je číslo, ktoré sa vypočíta na základe charakteristiky výberovej vzorky (vzoriek) a má vzťah k testovaným parametrom základného súboru.
- Porovnanie vypočítanej testovacej charakteristiky s jej pravdepodobnostným rozdelením. Na základe tohoto porovnania sa stanoví pravdepodobnosť, že výberový súbor poskytne túto hodnotu testovacej charakteristiky za predpokladu platnosti H_0 . Pravdepodobnosť platnosti H_A je doplnok: $P(H_0) = 1 - P(H_A)$.
- Rozhodnutie o výsledku: ak je asociovaná pravdepodobnosť pre platnosť H_A dostatočne veľká, prijatie H_A , inak zostáva v platnosti implicitne prijímaná H_0 . Vo výskume sa ako konvenčná hranica pre prijatie H_A berie pravdepodobnosť 95%.

Je zrejmé, že pri tomto postupe sa môžeme dopustiť dvoch chýb. Chyba I. typu (*type I error*) znamená, že zamietneme nulovú hypotézu, hoci je správna. Jej pravdepodobnosť sa označuje α , a nazýva sa významnosť testu. Napríklad porovnávame priemerný tlak krvi Slovákov a Maďarov, pričom v celej populácii každého štátu sa priemery nelíšia, ale my robíme porovnanie na základe dvoch výberových vzoriek s rozsahom 1000 jedincov, pričom náhodnou zhodou okolností na Slovensku vyberieme jedincov s normálnym tlakom, zatiaľ čo v Maďarsku samých hypertonikov. V tomto prípade, ak na základe testovacej charakteristiky dôjdeme k záveru, že Slováci majú v priemere nižší tlak ako Maďari, dopúšťame sa chyby I. typu. Chyba II. typu (*type II error*) znamená, že nulovú hypotézu prijmeme, hoci je nesprávna, jej pravdepodobnosť sa označuje β , a nazýva sa sila testu. Aby sme použili rovnaký príklad, ak majú Slováci skutočne nižší krvný tlak ako Maďari, ale my zhodou okolností vyberieme dve rovnaké výberové vzorky, potom prijmeme na základe nášho výberu nesprávny záver, že obe populácie sú rovnaké, a dopúšťame sa chyby II. typu.

Vzhľadom na to, že chyba I. typu je z praktického hľadiska závažnejšia, štatistické testy stanovujú práve pravdepodobnosť chyby I. typu (α).

Príklad: Pri zalesňovaní na lokalite Čertovica v NP Nízke Tatry bol použitý reprodukčný materiál jedle bielej neznámeho pôvodu. Pre overenie pôvodu bolo genotypovaných 50 vysadených jedincov použitím nSSR lokusu *SF333* a porovnaných s 10 uznanými porastmi (0 – klasifikovaná populácia/výsadba, 1–10 – referenčné populácie/potenciálne zdrojové uznané porasty). Zistené frekvencie alel sú uvedené v tabuľke. Z ktorého zdroja výsadba pochádza?

Lokus/alela	Populacia											
	0	1	2	3	4	5	6	7	8	9	10	
<i>SF333</i> 168	0,24	0,35	0,45	0,21	0,30	0,24	0,21	0,16	0,15	0,41	0,25	
170	0,01	0,00	0,10	0,02	0,05	0,01	0,05	0,09	0,11	0,00	0,06	
172	0,59	0,40	0,20	0,54	0,45	0,68	0,50	0,50	0,45	0,36	0,42	
174	0,00	0,00	0,20	0,01	0,09	0,00	0,10	0,02	0,06	0,11	0,06	
176	0,01	0,25	0,05	0,00	0,00	0,00	0,03	0,06	0,11	0,11	0,06	
178	0,15	0,00	0,00	0,22	0,11	0,07	0,11	0,17	0,12	0,01	0,15	
<i>J_i</i>	0,428	0,345	0,295	0,385	0,315	0,525	0,320	0,317	0,267	0,322	0,272	
<i>J_{ij}</i>		0,323	0,228	0,402	0,355	0,469	0,363	0,360	0,322	0,313	0,332	
<i>D</i>		0,1757	0,4464	0,0092	0,0359	0,0103	0,0200	0,0216	0,0504	0,1698	0,0297	

Výpočet genetickej vzdialenosti medzi výsadbou a potenciálnymi zdrojmi je ilustrovaný na príklade 1. uznaného porastu:

$$J_0 = \sum_k p_{0k}^2 = 0,24^2 + 0,01^2 + \dots + 0,15^2 = \mathbf{0,428}$$

$$J_1 = \sum_k p_{1k}^2 = 0,35^2 + 0,00^2 + \dots + 0,00^2 = \mathbf{0,345}$$

$$J_{01} = \sum_k p_{0k}p_{1k} = 0,24 \times 0,35 + 0,01 \times 0,00 + \dots + 0,15 \times 0,00 = \mathbf{0,323}$$

$$D = -\ln \frac{J_{ij}}{\sqrt{J_i J_j}} = -\ln \frac{0,323}{\sqrt{0,428 \times 0,345}} = -\ln(0,8389) = \mathbf{0,1757}$$

Analogickým spôsobom možno spočítať geneticke vzdialenosti k ostatným referenčným UP. Najnižšia genetická vzdialenosť je v prípade populácie 3 (0,0092), preto je z celej sady referenčných populácií najpravdepodobnejším zdrojom.

Bayesovský modelový prístup (procedúra Structure; Pritchard et al. 2000) je založený na priradení multilokusových genotypov do vopred definovaného počtu K geneticky homogénnych skupín tak, aby to optimálne zodpovedalo odchýlkam od panmixie a väzbovej rovnováhy v rámci skupín. Postup využíva náhodnú simuláciu založenú na markovovských reťazcoch (Monte Carlo Markov chain; MCMC) a je založený na Bayesovskom prístupe pre stanovenie posteriórnej pravdepodobnosti priradenia jedinca do každej zo skupín. Simulácia prebehne pre postupne rastúci počet skupín (spravidla 1–10, málokedy má zmysel testovať vyšší počet skupín), a následne je možné určiť optimálny počet skupín na základe priebehu priemernej posteriórnej pravdepodobnosti dát za podmienky príslušného počtu skupín $P(X|K)$ a priemerného prírastku $P(X|K)$ váženého jej varianciou ΔK . Tento prístup opäť možno použiť ako pre jedinca, tak aj pre populačnú vzorku.

Aj v prípade sady referenčných populácií je možné použiť aj prístup založený na testovaní hypotézy o zhode rozdelenia alelických frekvencií (χ^2 -test). Samotná testovacia charakteristika χ^2 okrem toho, že ju možno využiť pre stanovenie štatistickej významnosti

prípadeného rozdielu, je aj mierou zhody (*goodness of fit*); inými slovami, čím je testovacia charakteristika menšia, tým lepšie sa zhoduje rozdelenie alelických frekvencií medzi populačnou vzorkou a referenčnou populáciou.

Pravdepodobnosti priradenia sa však spravidla buď nedajú stanoviť (genetické vzdialenosti), alebo sa pohybujú rádovo pod 95%, takže ich použiteľnosť ako súdneho dôkazu je otázna. Sú užitočné skôr ako nástroj na stanovenie iniciálnej hypotézy, ktorú je následne nutné podoprieť dôkazmi, založenými na iných než genetických prístupoch.

Literatúra

- Cornuet, J.-M., Piry, S., Luikart, G., Estoup, A., Solignac, M., 1999: New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153: 1989–2000.
- Evett, I.W., Weir, B.S., 1998: *Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists*. Sinauer Associates, Sunderland, Massachusetts. ISBN 0-87893-155.
- Goodwin, W., Linacre, A., Hadi, S., 2007: *An Introduction to Forensic Genetics*. John Wiley & Sons, Chichester, UK, ISBN 978-0-470-01025-9.
- Nei, M., 1972: Genetic distance between populations. *American Naturalist* 106: 283–292.
- Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.C., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Källner, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, Å., Rilakovic, N., Ritland, C., Rosselló, J.A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, B., Garcia Gil, R., Hvidsten, T.R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Lee Thompson, S., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P.K., Lundeberg, J., Jansson, S., 2013: The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000: Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Rannala, B., Mountain, J.L. 1997: Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* 94: 9197–9201.
- VanRaden, P.M., 2008: Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91: 4414–4423.
- Weir, B.S., Goudet, J. 2017: A unified characterization of population structure and relatedness. *Genetics* 206: 2085–2103.

Literatúra odporúčaná pre štúdium

- Evett, I.W., Weir, B.S., 1998: *Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists*. Sinauer Associates, Sunderland, Massachusetts. ISBN 0-87893-155.
- Goodwin, W., Linacre, A., Hadi, S., 2007: *An Introduction to Forensic Genetics*. John Wiley & Sons, Chichester, UK. ISBN 978-0-470-01025-9.