

Genetická kartografie

Vazba

U klasického dihybridismu podle Johanna Gregora Mendela segregují různé alely dvou genů nezávisle. Rekapitulace dihybridismu je na obrázku 8.1. Toto schéma platí jen pro lokusy na různých chromozomech (záměrně se vyhýbám slovu gen, neboť významná část polymorfismů, které jsou používány pro vazbu a vazebné mapování se nachází mimo geny. Průměrný gen je také významně větší než interval mezi dvěma polymorfními lokusy. Vzhledem k tomu, že locus (lat.)=místo, dá se tento termín bez potíží použít i pro polymorfismy v jednom nukleotidu, SNPs). Pokud jsou však dva lokusy velmi blízko u sebe, u savců řekněme v měřítku kilobází (kb), alely těchto dvou lokusů vůbec nese segregují a dědí se tedy jako fixní kombinace neboli haplotyp (haplotyp je kombinace určitých alel dvou nebo více lokusů na jednom chromozomu. Genotyp pro danou chromozomální oblast se pak u většiny lidí skládá ze dvou haplotypů). Tato situace (fig. 8.2) se nazývá úplná vazba. Nicméně, díky crossing-overu při meióze existuje kontinuum mezi těmito dvěma okrajovými situacemi. Tato neúplná vazba je popsána na obr. 8.3. Pozorujeme 4 genotypové a fenotypové skupiny jako u klasického dihybridismu, ale štěpný poměr již není 1:1:1:1, neboť množství jedinců s genotypem vzniklým na základě crossing-overů (tedy rekombinantů) je menší než množství jedinců s původními parentálními chromosomy (non-rekombinantů). Obecně můžeme předpokládat, že čím blíže leží oba lokusy na chromozomu, tím spíše se budou dědit spolu a tím menší bude množství rekombinantů.

Můžeme tedy stanovit genetickou (vazebnou) vzdálenost, pokud změříme, jak často se alely dvou lokusů dědí spolu a jak často jsou odděleny crossing-overem. Nejlépe se měří vazebná vzdálenost u experimentálního zpětného křížení. Příklad založený na datech Herrona a spol. je na obrázku 8.4. Myší autosomálně dominantní mutace "opakovaná epilace", která porušuje strukturu epidermis, byla lokalizována na myší chromozom 4, díky vazbě s mikrosatelitním markerem D4Mit204, respektive s jeho alelou a2. Z 632 zvířat bylo 10+19=29 rekombinantů a 318+285=603 non-rekombinantů. Vazebná vzdálenost se měří jako rekombinační zlomek (θ), což je podíl rekombinantů v celkovém množství potomků:

(8.1)

$$\theta = \frac{\text{počet rekombinantů}}{\text{počet všech jedinců}}$$

v našem případě

$$\theta = \frac{29}{603 + 29} = 0.046 \text{ or } 4.6\%$$

Minimum rekombinačního zlomku je 0 - žádný rekombinant při vazbě úplné. Maximum je 50%, v případě, že jsou sledované geny na různých chromozomech nebo i na stejném chromozomu, ale velmi daleko od sebe. 50% maximum se dá vysvětlit skutečností, že při profázi 1. zracího dělení dochází ke crossing-overu duplikované dvojice chromozomů, proto i pokud mezi dvěma lokusy dochází vždy ke vzniku crossing-overu, 2 řetězce jsou stále bez rekombinace (podrobnější vysvětlení, které bere v úvahu i dvojitě rekombinace najdete v angličtině na <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.figgrp.1383>).

Mapová funkce

Při konstrukci genetických map (tedy map chromozomů založených na pořadí a vzdálenosti polymorfních markerů a znaků, viz dále, třibodový pokus) dochází k určitým problémům při použití rekombinačního zlomku jako měřítka vzdálenosti. Uvažujme například rekombinační zlomek mezi A a B je 0,3 (30%) a mezi B a C 0,4 (40%). Jaký bude rekombinační zlomek mezi A a C? Jistě je 0,7, neboť maximum je 0,5 neboli 50%. Rekombinační zlomky tedy nejsou aditivní, což je pro mapování nepraktické. Rekombinační zlomek také nereflektuje interferenci chiasmat (viz dále). Proto bylo navrženo mnoho matematických transformací rekombinačního zlomku, aby byla umožněna aditivita a korigována nenáhodná distribuce chiasmat. Nejznámější je Haldanova a Kosambiho mapová funkce:

Haldane (8.2):

$$W = -\frac{1}{2} \ln(1 - 2\theta)$$

Kosambi (8.3):

$$W = \frac{1}{4} \ln \left(\frac{1 - 2\theta}{1 - 2\theta'} \right)$$

Jednotka mapové vzdálenosti je Morgan (M) nebo častěji odvozený centimorgan (1 cM = 1/100 M). Pro malé vazebné vzdálenosti odpovídá 1 cM 1% rekombinací.

Vazba ve fázi cis a trans

U dihybridismu není možné rozlišit tyto dvě možné kombinace parentálních genotypů v genotypu první a druhé filialní generace:

AA⁺BB x aabb
AaBb

nebo

AA⁺bb x aaBB
AaBb

V obou případech mají F1 hybridní stejný genotyp.

Představte si ale, že lokusy A a B leží na stejném chromozomu. Pak haplotypy prvního F1 hybridního budou AB/ab, zatímco druhý bude mít genotyp Ab/aB. A nyní provedme křížení těchto F1 hybridů s parentálním kmenem aabb. Je jasné, že v prvním případě budou genotypy potomstva s původními haplotypy AB/ab (AaBb) a ab/ab (aabb), rekombinanti budou mít genotypy Ab/ab (Aabb) nebo aB/ab (aaBb), tak jak je uvedeno na obr. 8.3. V druhém případě jsou genotypy rekombinantů a non-rekombinantů přesně naopak (obr. 8.5). První možnost (fig. 8.3), kdy jsou dominantní respektive recesivní alely obou lokusů na tomtéž chromozomu se nazývá vazebná fáze cis neboli coupling. Druhý případ (fig. 8.5), tedy dominantní alela lokusu A s recesivní alelou lokusu B na jednom chromozomu a obrácená kombinace alel na druhém chromozomu F1 hybridního, se nazývá vazebná fáze trans čili repulsion.

Vazba u interkrosu

Vazba může být hodnocena také v F2 generaci (interkrosu). Na rozdíl od zpětného křížení nastává crossing-over s rekombinací u obou rodičů. Výhodou je zvýšená efektivita vazebné analýzy pro kodominantní alely, neboť každý F2 hybrid je výsledkem dvou informativních meióz. Nicméně i pro kodominantní markery existuje určitá nejistota, pokud jsou u daného jedince F2 generace oba lokusy heterozygotní (tabulka 8.1). Naštěstí je tato nejistota malá pro lokusy, které jsou blízko sebe. Rekombinační zlomek se počítá takto:

(8.4)

$$\theta = \frac{\text{počet rekombinantů}}{\text{počet meióz}} \approx \frac{2 \times (aaBB + AA bb) + AABb + AaBB + aaBb + Aabb}{2 \times \text{počet jedinců}}$$

Jestliže jsou některé z alel dominantní, mnoho F2 hybridů je neinformativních - nemůžeme rozhodnout, jestli jsou rekombinanti nebo nikoli. Vazebná analýza je tedy možná jen u informativní části F2 populace. Pro autozomálně recesivní znaky můžeme například použít pouze postižené jedince (recesivní homozygoty), kteří by měli být homozygotní i pro další lokusy nacházející se v těsné vazbě. Tento přístup se jmenuje homozygotní nebo autozygotní mapování.

U člověka nelze předem zajistit správný typ "křížení". Je však možné analyzovat rodokmeny, které odpovídají backcrossu nebo interkrosu. Rodokmeny s autosomálně dominantními znaky odpovídají prakticky vždy zpětnému křížení, u autosomálně recesivních znaků se zase jedná většinou o interkros. Rodiny jsou většinou malé a mnoho lokusů, které se používají jako genetické markery, není vždy polymorfních (viz sekci polymorfismy). Proto se obvykle slučují data z více rodin, aby vzrostla statistická síla, a testuje se více polymorfních markerů, aby byla větší pravděpodobnost nalezení informativních genotypů.

LOD skóre

Pro stanovení vazby je velice důležité rozlišit mezi náhodnými fluktuacemi počtu potomků v každé skupině potomků a skutečnou vazbou. Vzhledem k tomu, že patřičné statistické testování se provádí poněkud odlišně od jiných často používaných statistických metod, budu se tímto problémem zabývat více do hloubky. LOD skóre, nejčastěji používaná statistika, je založeno na přímém porovnání pravděpodobnosti nulové hypotézy, která praví, že vazba neexistuje (rekombinační zlomek 1/2), s alternativní hypotézou, která tvrdí, že existuje vazba s určitým rekombinačním zlomkem $\theta < 1/2$. Jak se to dělá? Spočítáme přesnou pravděpodobnost získání našich dat za předpokladu platnosti alternativní ($\theta < 1/2$) a nulové ($\theta = 1/2$) hypotézy, $P(\theta)$ respektive $P(1/2)$. Poměr $P(\theta)$ a $P(1/2)$ představuje pak "kurs" pro vazbu. Pro jednoduchost se většinou používá logaritmus se základem 10 (logarithm of the odds ratio = **LOD** score). Pro zpětné křížení jsou vzorce pro výpočet $P(\theta)$ a $P(1/2)$ relativně jednoduché. Mějme zpětné křížení AaBb x aabb ve fázi cis. Máme r rekombinantů z celkového množství N jedinců. Z toho plyne rekombinační zlomek $d = r/N$. Pro jednoduchost předpokládejme, že d je nejlepší aproximací reálné frekvenci rekombinace θ (tento předpoklad jsme již uplatnili v rovnici 8.1, ale je jasné, že zvláště v malých výběrech je na tento předpoklad třeba pohlížet s opatrností). Pravděpodobnost, že náhodně vybraný jedinec bude mít rekombinantní genotyp Aabb je d/2, podobně pro rekombinanta aaBb. Nerekombinantní genotyp AaBb bude mít pravděpodobnost (1-d)/2, stejně pro aabb. Pravděpodobnost našich dat bude rovna pravděpodobnosti současného výskytu r rekombinantů, každý s pravděpodobností d/2, dohromady $(d/2)^r$, a N-r non-rekombinantů, každý s pravděpodobností (1-d)/2, dohromady $((1-d)/2)^{N-r}$. Tedy:

$$P(\theta) = P(d) = (d/2)^r((1-d)/2)^{N-r}$$

Jestliže platí nulová hypotéza, a lokusy nejsou ve vazbě, každý genotyp má stejnou pravděpodobnost, což dává 1/4 pro 4 genotypy, které pozorujeme u zpětného křížení. $P(1/2)$ je součinem N pravděpodobností pro každého jedince:

$$P(1/2) = (1/4)^N$$

LOD skóre je tedy (po úpravě):

(8.5)

$$LOD = \log_{10} \left(\frac{P_d}{P_{1/2}} \right) = \log_{10} \left(\frac{d^r (1-d)^{N-r}}{1/2^N} \right)$$

Jsou situace, kdy je správné připustit, že d není správný odhad hodnoty θ - zejména u malých lidských rodokmenů. V tomto případě je možné najít lepší odhad θ pomocí rovnice pro LOD skóre, jako lokální maximum LOD skóre (pro zpětné křížení popsané výše je však LOD skóre maximální právě pro $d=r/N$).

Pro náš příklad s kožní abnormalitou je LOD skóre:

$$LOD = \log_{10} \left(\frac{0.046^{29} (1-0.046)^{632-29}}{1/2^{632}} \right) = 29 \log_{10}(0.046) + 603 \log_{10}(1-0.046) + 632 \log_{10}(2)$$

$$LOD = 242,4$$

Když jsme získali hodnotu LOD skóre, jako další krok je nezbytné učinit rozhodnutí, jestli připustit nebo zamítnout vazbu studovaných lokusů. Pravděpodobnost nulové hypotézy nebude nikdy nula, čili je zde vždy možnost, že se ve skutečnosti nezávislé lokusy budou jevit ve vazbě. Otázka tedy zní, jak rozumně malá musí být pravděpodobnost nulové hypotézy, abychom ji mohli zamítnout ve prospěch alternativní. Tuto hranici si musíme stanovit sami. Musíme však mít na paměti, že je zde malá pravděpodobnost omylu.

Ve vazebné analýze se LOD skóre větší než 3 považuje za signifikantní, to znamená pokud je pravděpodobnost vazby 1000 krát větší než pravděpodobnost nulové hypotézy. Často je také třeba vyloučit vazbu, například aby bylo možno snížit množství kandidátních genů pro další studie. Pro vyloučení vazby je požadováno LOD skóre menší než -2 (vazba je 100x méně pravděpodobná než žádná vazba). Náš příklad s myší genetiky vede díky vysokému počtu jedinců k velmi vysokému LOD skóre 242,4, takže o vazbě zde nemůže být pochyb. V lidských rodokmenech je ale situace odlišná. I relativně velká rodina, jako ta na obr. 8.6, s lokusy v těsné vazbě (jen jeden rekombinant) nestačí k potvrzení vazby. Tento problém je ještě výraznější, jestliže má rodina jen 2 informativní generace, jako na obr. 8.6B, neboť neznáme vazebnou fázi, a tedy musíme započítat jak cis tak trans fázi do výpočtu LOD skóre.

Pořadí lokusů a tříbodový pokus

Vazba může být využita k nalezení genu mutovaného u genetického onemocnění, jak v experimentu, tak u člověka v rodokmenu. V tomto případě představuje jeden polymorfni lokus postižený (nebo zdravý) fenotyp, který je testován na vazbu s polymorfním DNA markerem (Fig. 8.6) nebo v praxi spíše se sadou polymorfních markerů rozmístěných na všech chromozomech. Nalezení takové vazby nemoc-DNA marker může být užitečné -

protože je sekvence lidského genomu, až na nepatrné výjimky známa, je známa i poloha DNA markeru. Jestliže je těsná vazba mezi tímto markerem a nemocí, je gen pro onemocnění nedaleko. Lze tedy hledat v sekvenci v okolí tohoto markeru kandidátní geny a testovat je, jestli neobsahují mutace, a to i v případě, že o patogenezi onemocnění a možné funkci mutovaného genu není nic známo, víme jen chromozomální polohu. Tato metoda se proto nazývá "poziční klonování".

Jednoduchá vazba nemoc-marker má jednu zřetelnou nevýhodu - víme jen, že gen pro nemoc musí být v blízkosti markeru, ale daleko lepší by bylo znát přesný úsek DNA, kde se musí daný gen nacházet. Pak by byl úspěch v hledání mutace prakticky zaručen, například při sekvenaci celého takového úseku u postiženého jedince bychom museli najít zodpovědnou mutaci. Tento úkol (nalezení kritického úseku DNA pro hledaný gen) je nejnáze splnitelný pomocí stanovení pořadí lokusů - pokud bychom byli schopni provést vazebnou analýzu tak, abychom obdrželi pořadí lokusů ve vazbě, pak by měl lokus pro nemoc "padnout" mezi dva markery - a máme náš segment.

Minimální počet lokusů pro stanovení pořadí je samozřejmě 3. Provedeme tedy "tříbodový pokus". Vezměme tři lokusy 1, 2 a 3 v uvedeném pořadí. Crossing-over může nastat mezi lokusem 1 a 2 nebo mezi lokusem 2 a 3, nebo mohou vzniknout 2 crossing-overy najednou, oddělující tak alelu lokusu 2 od alel 1 a 3. Pravděpodobnost takové dvojité rekombinace bude relativně malá - teoreticky součin pravděpodobností jednoduchých crossing-overů. Jestliže je vzdálenost 1-2 je například 5 cM a vzdálenost 2-3 10 cM, součin je $0,05 \times 0,1 = 0,005$. Na obrázku 8.7 vidíte zpětné křížení u potkana, kde gen H s mutantní alelou h způsobující samčí neplodnost segreguje se dvěma mikrosatelitními markery A a B. Očekáváme 8 genotypových skupin, které můžeme seskupit do 4 kategorie. Můžeme vidět, že mimo nerekombinantní genotypy jsou zde 3 skupiny rekombinantů. Ve skupině ABh nebo abH (gen H je rekombinován od A a B) je jen jeden jedinec. To musí být dvojitý rekombinant a pořadí je tedy A-H-B nebo ekvivalentní B-H-A. Pokud známe fyzickou polohu markerů A a B na chromozomu, jeden z genů, které se nacházejí v genomické DNA mezi A a B musí být mutovaný gen H.

Správné pořadí bylo určeno splněním podmínky minimálního počtu dvojitých rekombinantů. Tento přístup může být rozšířen na libovolné množství lokusů. Stanovení pořadí se nyní může zdát komplikovanější. Nicméně, i v tříbodovém pokusu, správné pořadí znamená minimum dvojitých rekombinací a následně minimální délku genetické mapy. Správná genetická mapa je tedy ta nejkratší - a hledání takové mapy se dá pro snadno zautomatizovat pro mnoho lokusů. Příklad takových výsledků ukazuje obrázek 8.7C pro mapování myší mutace luxoid, způsobující poruchu vývoje končetin a infertilitu. Všimněte si, že zobrazený úsek chromozomu je tak malý, že zde není ani jeden dvojitý rekombinant. Přesto je pořadí lokusů stanovené autory (Buaas a spol., 2004) nejlepší možné (nepočítáme-li možnost otočit celou mapu vzhůru nohama).

Nyní, podíváme-li se zpět na příklad na obr. 8.7A a B, pravděpodobnost rekombinace mezi A a H a zároveň mezi H a B, tedy dvojitá rekombinace, která by měla být současným výskytem dvou nezávislých rekombinací, by měla být rovna součinu pravděpodobností jednoduchých rekombinací. V našem případě $0,127 \times 0,088 = 0,0112$. Skutečná frekvence je však nižší, 0,0029. To je běžné pozorování, které bylo potvrzeno i na úrovni celého genomu. Z těchto dat vyplývá, že výskyt crossing-overu ovlivňuje negativně vznik druhého crossing-overu v jeho blízkosti. Tento jev se nazývá interference. Míru interference se možno spočítat jako koeficient interference, který nám říká, jak velký podíl potenciálních dvojitých crossing-overů byl inhibován přítomností prvního crossing-overu:

(8.6)

$$i = 1 - \frac{\text{pozorovaný podíl dvojitých rekombinantů}}{\text{očekávaná pravděpodobnost dvojité rekombinace}}$$

Poměr aktuální a očekávané frekvence dvojitých rekombinantů se nazývá koeficient koincidence (coc). Pro náš příklad je $i=0,74$ a $coc=0,26$. Máme tedy pouze 26% očekávaných dvojitých crossing-overů, jinými slovy 74% možných dvojitých rekombinací bylo inhibováno po vytvoření prvního chiasmatu. Interference vede k rovnoměrnějšímu rozložení crossing-overů podél chromozomu, což je pravděpodobně funkčně významné, například vzhledem k tomu, že chiasmata vytvořená procesem crossing-overu jsou místem koheze homologních chromozomů v meióze a substituují tak kohezní funkci, kterou mají v mitóze centromery. Rovnoměrné rozložení chiasmat může tak vést k menšímu počtu poruch segregace a zajistit výskyt alespoň jednoho chiasmatu na chromozom, jako nutnou podmínku pro správnou segregaci.

Polymorfismy

Polymorfní lokusy neboli polymorfismy se v populaci vyskytují nejméně ve dvou alelách. Některé mají fenotypové projevy, ale všechny jsou vposledku variace na úrovni DNA, v genotypu. Minimální polymorfismus je pak polymorfismus v jednom nukleotidu, SNP.

Pro vysvětlení, všichni lidé mají například gen pro angiotenzin konvertující enzym. Ale někteří mohou mít tento gen mírně odlišný od druhých, mohou mít odlišný nukleotid, který se může překládat do odlišné aminokyseliny, případně tyto varianty enzymu mohou konvertovat angiotenzin I na angiotenzin II různou rychlostí, což může vést k odlišnému krevnímu tlaku u těchto 2 skupin lidí. Ale pozor, tento řetěz příčin a následků je zřídka úplný až na úroveň fenotypu.

Technicky považujeme lokus za polymorfní, pokud má vzácnější alela ve studované populaci frekvenci alespoň 1%. Jestliže je frekvence alely menší, hovoříme o vzácné alele. Rozlišení je to poněkud svévolné, ale pomáhá zjednodušit populační i jiné genetické studie.

Jednotlivé druhy polymorfismů

- monogenně podmíněné fenotypové znaky
- krevní skupiny
- sérové proteiny
- HLA antigeny
- tandemové repetice - minisatelity, mikrosatelity
- SNPs - RFLP a ostatní SNPs

Počet viditelných nebo snadno detekovatelných polymorfních monogenních fenotypů (například 2 nebo více mendelovských znaků jako je barva květu u hrachu) je limitován na malé množství. Pro uspokojení potřeby více polymorfních lokusů je nezbytné prozkoumat organismus na hlubší úrovni.

Historicky prvními takovými polymorfními lokusy byly krevní skupiny, vysoce polymorfní proteiny (MN, Ss, Rh) nebo sacharidy (ABO) na buněčné membráně červených krvinek, snadno detekovatelné aglutinací erytrocytů specifickými protilátkami.

Varianty hojných sérových proteinů se dají vyšetřovat pomocí rozdílné pohyblivosti při elektroforéze.

MHC (HLA) antigeny, nejvíce polymorfní skupina proteinů, jsou velice vhodné pro vazebné studie, které jsou ale limitovány na krátký úsek chromozomu 6, kde se tyto geny nachází. Daleko více polymorfismů existuje a může se stanovovat na úrovni DNA.

VNTRs = variable number of tandem repeats, variabilní počet tandemových repetice jsou polymorfismy minisatelitních sekvencí, které se většinou analyzují pomocí Southernova blotu. Pokud na obou stranách fragmentu DNA obsahujícího tandemovou repetici a také unikátní sekvenci leží restriční místa, můžeme vyrobit z unikátní sekvence značenou sondu, která se může hybridizovat se Southernovými membránami, kde se variabilní množství repetice projeví jako různá elektroforetická mobilita fragmentů. Pokud je sonda vyrobena z vlastní repetice, budou zviditelněny všechny podobné minisatelity z celého genomu. Dostaneme tak dosti komplikovaný "otisk", díky vysoké míře polymorfismu unikátní pro každého jedince. Tato metoda byla užívána pro určování identity. Nyní jsou však minisatelity prakticky vytěsněny použitím mikrosatelitů jako genetických markerů i pro forenzní účely.

Mikrosatelity (viz kapitola Repetitivní DNA), také známé jako krátké tandemové repetice (short tandem repeats, STRs) jsou sekvence náchylné k mutacím, díky mechanismu nerovnoměrné rekombinace a "skluzu" polymerázy na repetici, které mění počet repetice. Nicméně tato nestabilita je za normálních okolností detekovatelná pouze v evolučním měřítku, pokud srovnáme rychlost vývoje mikrosatelitů s vývojem jiných sekvencí. To znamená, že v tomto časovém rámci došlo k akumulaci mnoha mutací mikrosatelitů v populaci, tyto mutace pak byly většinou neutrální (ani výhodné, ani nevýhodné). Výsledkem je přítomnost několika alel lišících se počtem opakování pro daný mikrosatelit v lidské populaci. To dělá z mikrosatelitů užitečné genetické markery, neboť délka repetice může být snadno zjišťována pomocí amplifikace metodou PCR, s oligonukleotidovými primery umístěnými v unikátní sekvenci obklopující repetici, a porovnání délky amplikonů gelovou elektroforézou. Vyšetření založené na mikrosatelitech je relativně jednoduché a levné, a proto se široce využívá ve vazebných, asociačních a populačních studiích, pro DNA diagnostiku i pro forenzní účely.

Výhodou mikrosatelitů je vysoká míra jejich informativnosti v rodokmenech. V rodokmenu je marker neinformativní tehdy, pokud je jeden rodič homozygotem, nebo pro 1/2 potomků, pokud jsou oba rodiče heterozygoti pro stejnou kombinaci alel. Jaká je tedy pravděpodobnost, že je marker informativní? Budiž $p(i)$ frekvence alely i v populaci, $p(i)^2$ je pravděpodobnost, že je 1 rodič homozygot pro alelu i , $4p(i)^2p(j)^2$ je pravděpodobnost, že oba rodiče budou heterozygoti pro alely i a j . Potom se míra informativnosti daného polymorfismu (polymorphism information content, PIC) bude rovnat:

(8.7)

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n \sum_{j=i+1}^n 2 p_i^2 p_j^2$$

Pro typický mikrosatelitní marker s 5 alelami, každá s frekvencí 0,2 je PIC 0,77 (nebo 77%), zatímco pro RFLP polymorfismus s dvěma alelami, s frekvencí 60% a 40% je PIC pouze 0,36 nebo 36%.

Podrobnosti ve Strachan a Read,

<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.1387>.

Hlavní nevýhoda mikrosatelitů je omezená možnost paralelního vyšetření více markerů v jedné zkumavce (multiplexování). Nejvíce užívaný multiplex je DNA vyšetření pro forenzní účely, které obsahuje standardně 13 mikrosatelitních markerů (databáze CODIS na FBI, více informací také v databázi mikrosatelitních markerů - STRbase, <http://www.cstl.nist.gov/div831/strbase/fbicore.htm>). Multiplexní zpracování je možné díky použití 4 různých fluorescenčních barviv na značení primerů pro PCR a různé délce ampliconů se stejnou barvou (různé alely mikrosatelitního markeru se obvykle liší jen o délku 1 nebo několika repetice, je tedy možné např. pro tetranukleotidovou repetici o 12 alelách navrhnout délku ampliconu PCR na 100-150 bp, a pro druhou podobnou repetici 200-250 atd., aby nedocházelo k překryvu. Příklady sekvencí mikrosatelitů a jejich podoby při elektroforéze je na obr. 8.8.

SNP (vyslovuje se, alespoň v anglickém prostředí "snip") = single nucleotide polymorphism, polymorfismus v jednom nukleotidu, je nejmenší možná genetická změna, týkající se pouze 1 páru bazí. Přesněji, SNP se objevuje tam, kde jeden nukleotid jako třeba adenin je nahrazen jakýmkoli jiným nukleotidem T, C nebo G, a to ve významné části populace (>1%). Je třeba mít na mysli, že jeden nukleotid znamená ve skutečnosti 1 pár bazí v DNA. Jestliže je A nahrazen C, musí být v komplementárním řetězci T nahrazen G. Odhaduje se, že v lidském genomu je průměrně 1 SNP na 1000 bp. Většina SNP pak leží mimo kódující sekvence, ty které leží uvnitř exonů jsou převážně bez vlivu na aminokyseliny, díky degeneraci genetického kódu. Některé SNP jsou v restričních místech - jedna alela tvoří restriční místo, druhá se liší od konsenzu pro štěpení restriční endonukleázou. Tyto SNP mohou být stanovovány jako RFLP (restriction fragment length polymorphism, polymorfismus v délce restričních fragmentů) pomocí Southernova blotu s DNA štěpenou příslušnou restriční endonukleázou a s detekcí značenou sondou, nebo pomocí PCR, amplifikací fragmentu s primery okolo restričního místa a následným rozštěpením produktu reakce restriční endonukleázou (Obr. 8.9A). Ne všechny SNP ale leží v restričních místech, a proto jsou třeba jiné metody. Ty jsou založeny na hybridizaci DNA s oligonukleotidy, sekvencně specifickými pro každou variantu, nebo na jedнокrokové polymerázové reakci s primerem těsně vedle SNP a značeným nukleotidem. Využitím DNA mikročipů lze tyto metody efektivně multiplexovat. Pomocí nejmodernějšího 500k GeneChip od firmy Affymetrix lze určit najednou genotyp 500 000 SNP, na dvou čipech o 250 000 SNP (princip viz obr. 8.9B, který popisuje starší, "jen" 100k set). Metoda je založena na měření síly hybridizace oligonucleotidů odpovídajících jednotlivým sekvencním variantám, jako kontrola se používají oligonucleotidy se záměrně vnesenou chybou.

Nevýhoda SNP je omezené množství alel v populaci. Teoreticky zde můžou být až 4 alely na SNP, ale obvykle jsou jen 2. Informativnost jednoho SNP je tedy relativně malá. Výťažnost se ovšem dá zlepšit paralelním zpracováním více SNP a analýzou haplotypů spíše než jednotlivých polymorfismů. Je tedy třeba více SNP k získání té samé informace jako poskytuje jeden mikrosatelit, ale paralelní zpracování SNP je daleko snazší. SNP představují významný příspěvek ke genetické variabilitě člověka. Část této variability může být funkční - SNP v kódujících sekvencích, které vedou ke změně aminokyseliny, a SNP v regulačních sekvencích. Probíhá nyní velký mezinárodní projekt HapMap (<http://www.hapmap.org/index.html>), s ambiciózním plánem postihnout většinu této diverzity pomocí typizace rozdílných populací na SNP. Pro ilustraci, nejnověji přispěla data získaná v

projektu HapMap k objasnění genetické změny, která je podkladem světlé barvy pleti u Evropanů (Lamason a spol., 2005).

Referenční genetické mapy pro člověka

Pro člověka jsou nyní k dispozici 3 celogenomové genetické (vazebné) mapy. Všechny jsou k dispozici na internetu, např. přes lidský genom v NCBI, <http://www.ncbi.nlm.nih.gov/genome/guide/human/>. 2 mapy (Genethon a Marshfield) jsou založeny na rodinách z CEPH (Centre d'Études du Polymorphisme Humaine in Paris), mapa od společnosti deCODE na 146 rodinách z Islandu.

Vztah genetické mapy a DNA sekvence

Pořadí lokusů určené pomocí vazby by mělo být vždy stejné jako pořadí těchto lokusů v DNA sekvenci podél chromozomu. Kdykoli tedy nesouhlasí pořadí v sekvenci s genetickou mapou, musí být chyba v genetické mapě, v sekvenci nebo v obou. Protože sestavování sekvence a genetická mapa jsou na sobě nezávislé, může být genetická mapa využita k hledání a opravě chyb v sekvenci.

Vztah mezi vazebnými vzdálenostmi v cM a DNA sekvencí je na druhou stranu složitý. Celogenomové porovnání genetické mapy z deCODE a veřejné sekvence lidského genomu vede k průměrnému odhadu 1,13 cM/Mb (1,13 centimorganu na 1 megabázi, tedy 1000000 párů bazí v DNA), ale rozpětí hodnot je více než jeden řád. Současný model předpokládá střídání horkých míst pro rekombinace a rekombinačních "pouští" podél chromozomu. Lze jmenovat určitá obecná pravidla: Četnost rekombinací i variabilita této četnosti je větší u žen. Typické rekombinační pouště (tedy místa kde dochází k nejméně crossing-overům) jsou centromery, frekvence rekombinací stoupá směrem k telomerám. Typická horká místa pro rekombinaci jsou pseudoautozomální oblasti na obou koncích chromozomu X a Y, kde vzniká při spermatogenezi obligátní chiazma, kvůli zajištění správné segregace chromozomu X a Y do spermatid. Díky projektu HapMap bychom měli získat náhled jemné rekombinační struktury chromozomů, zejména sdílení SNP haplotypů, které nebyly vůbec rozděleny rekombinací (haplotypy v úplné vazbě).

Klasické využití vazby v medicíně - nepřímá DNA diagnostika

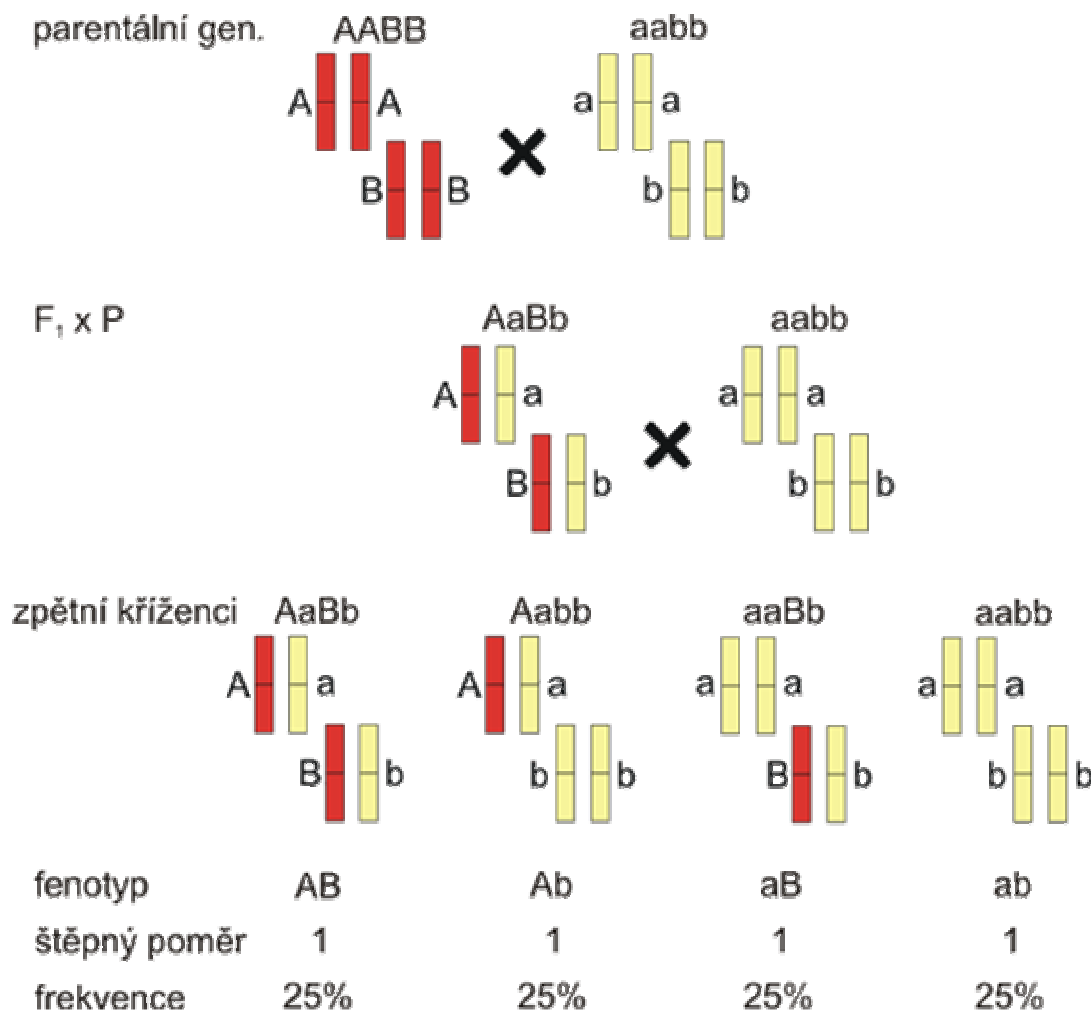
Vazba může být využita v DNA diagnostice. Pokud neznáme vlastní genetický defekt, který je podkladem monogenně podmíněného onemocnění, nebo pokud v daném genu existuje velké množství různých mutací, lze přesto konzultovat rodiny v riziku, pokud známe alespoň přibližně chromozomální lokalizaci vlastního genetického postižení. Můžeme jednoduše použít nějaký polymorfismus, který se nachází na daném místě genomu, stanovit genotyp zdravých i nemocných členů rodiny (pro tento polymorfismus) a snažit se dedukovat, která alela daného polymorfismu je ve vazbě s patogenní alelou a takto predikovat genotyp v lokusu pro onemocnění a stanovit riziko v prenatální nebo presymptomatické diagnostice, případně riziko přenašečství. Můžeme s výhodou využít i dosti vzdálené polymorfnní markery. Vezměme si například autosomálně dominantní onemocnění a polymorfismus ležící 1 Mb od patogenního genu, typicky s několika dalšími geny mezi těmito dvěma místy. Přesto bude pravděpodobnost rekombinace pouze kolem 1%, a tak bude prenatální diagnóza správná na 99% - významné zpřesnění oproti genealogickému riziku 50%. Nevýhodou této nepřímé metody je potřeba úplné rodiny, s již postiženými členy. Jiným komplikujícím faktorem je skutečnost, že v každé rodině bude onemocnění ve vazbě s jinou alelou polymorfismu (je to

pouze vazba, nikoli příčina onemocnění). Některé rodiny budou tedy neinformativní a bude třeba stanovit genotyp pro další polymorfní lokusy, než najdeme lokus informativní.

Obrázky

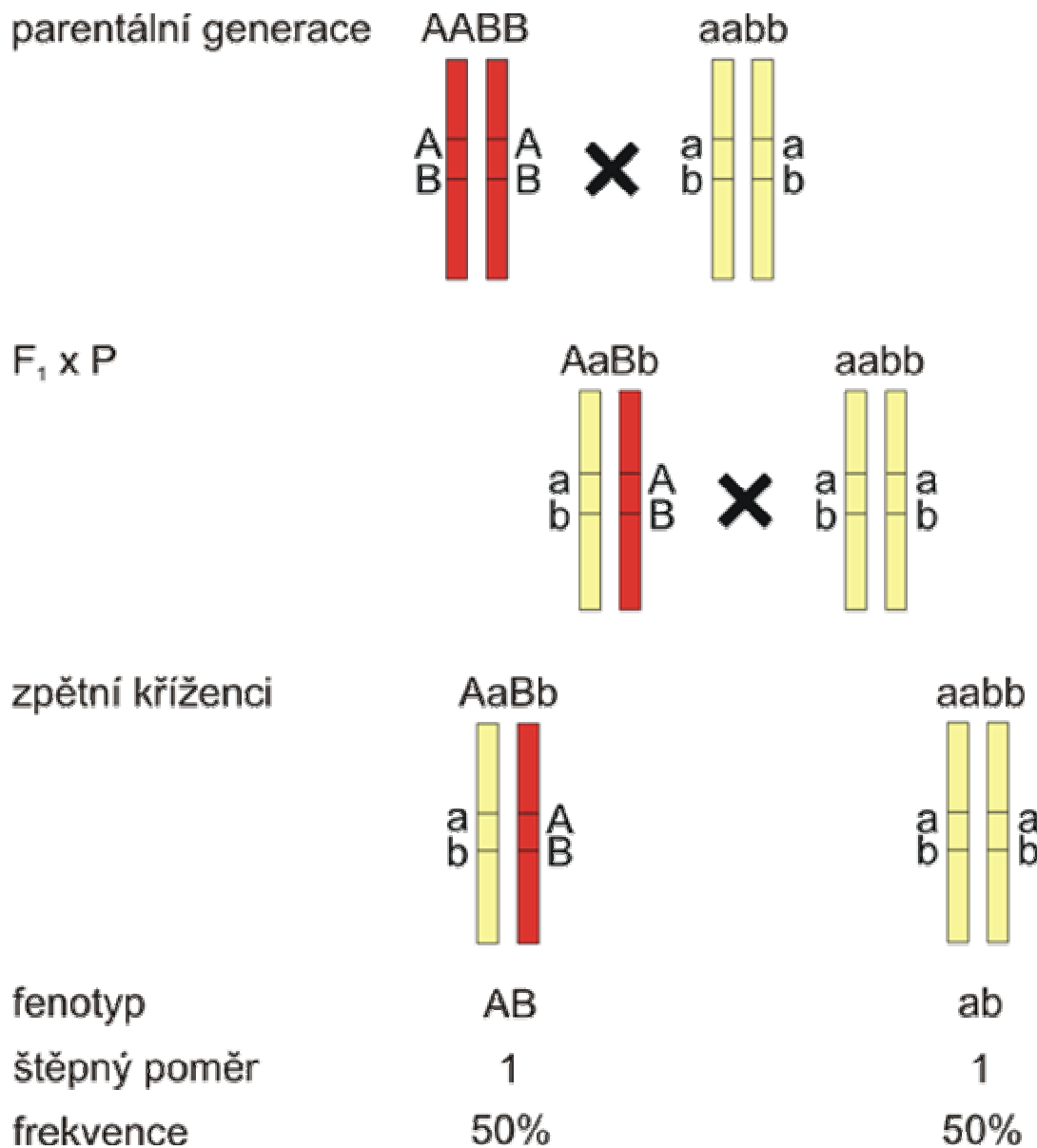
Obr. 8.1 Dihybridismus podle Johanna Gregora Mendela

Pro lokusy na různých chromozomech jsou v populaci zpětných kříženců $AaBb \times aabb$ čtyři různé genotypy, které mohou být pozorovány jako čtyři různé fenotypy jak pro kodominantní, tak pro dominantní a recesivní alely. Každý genotyp má stejnou pravděpodobnost 25%, štěpný poměr je 1:1:1:1.



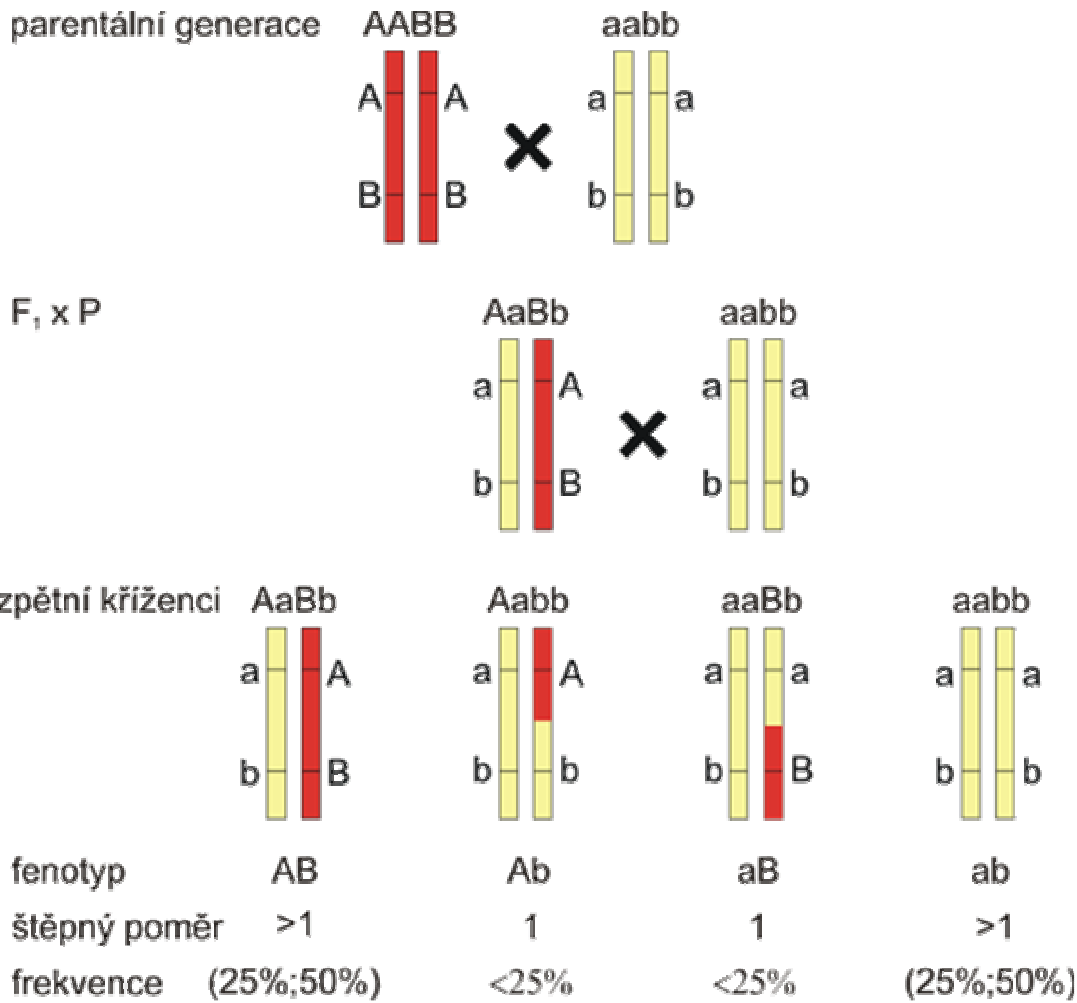
Obr. 8.2 Lokusy v těsné vazbě ve zpětném křížení - vazba úplná

Lokusy které jsou fyzicky velmi blízko u sebe se dědí vždy společně. V populaci zpětných kříženců jsou jen dvě genotypové i fenotypové třídy se stejnou pravděpodobností 50%.



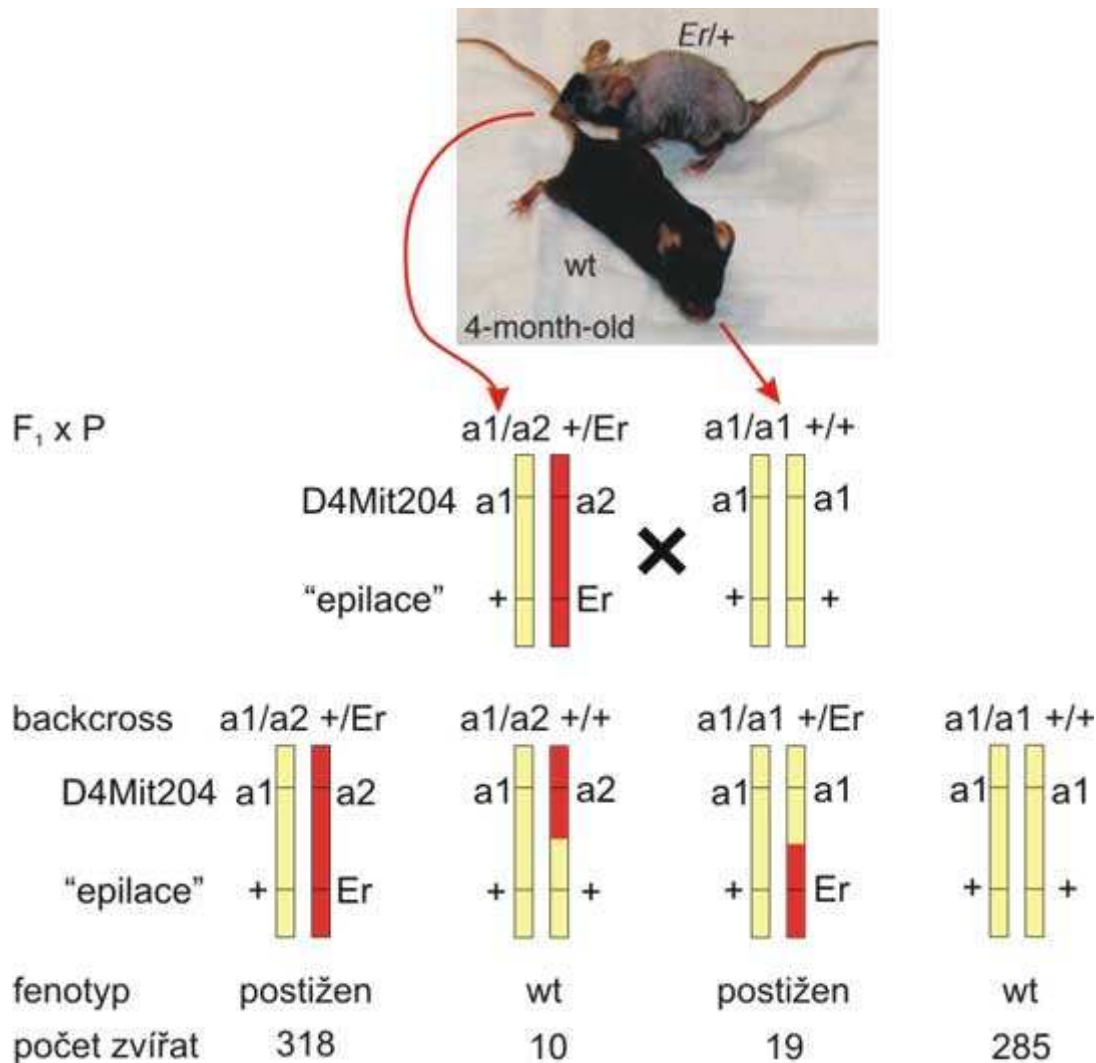
Obr. 8.3 Vazba ve zpětném křížení - obecný model

F1 hybrid má různé alely v každém lokusu na dvou homologních chromozomech, neboli každý chromozom má svůj haplotyp. Pokud křížíme F1 hybrida zpětně s parentálním kmenem, očekáváme, že v potomstvu někteří zdědí žlutý chromozom, jiní červený (od F1 hybrida, od parentální linie je to vždy žlutý chromozom). Jestliže však dojde ke crossing-overu mezi lokusy A a B, vzniknou další dvě genotypové třídy s novou kombinací alel (s rekombinantním chromozomem), tyto potomci se nazývají rekombinanti. Kolik bude rekombinantů? Jak je vidět na obrázku 8.2, pokud jsou lokusy velmi blízko, nemusí být žádný rekombinant. Na druhou stranu maximální množství rekombinantů je 50%, jak může být dedukováno ze skutečnosti, že rekombinace nastává v páru duplikovaných homologních chromozomů. Všimněte si, že situace s 50% rekombinantů je ekvivalentní a nemůže být rozlišena od případu volně rekombinovatelných lokusů na různých chromozomech (obr. 8.1).



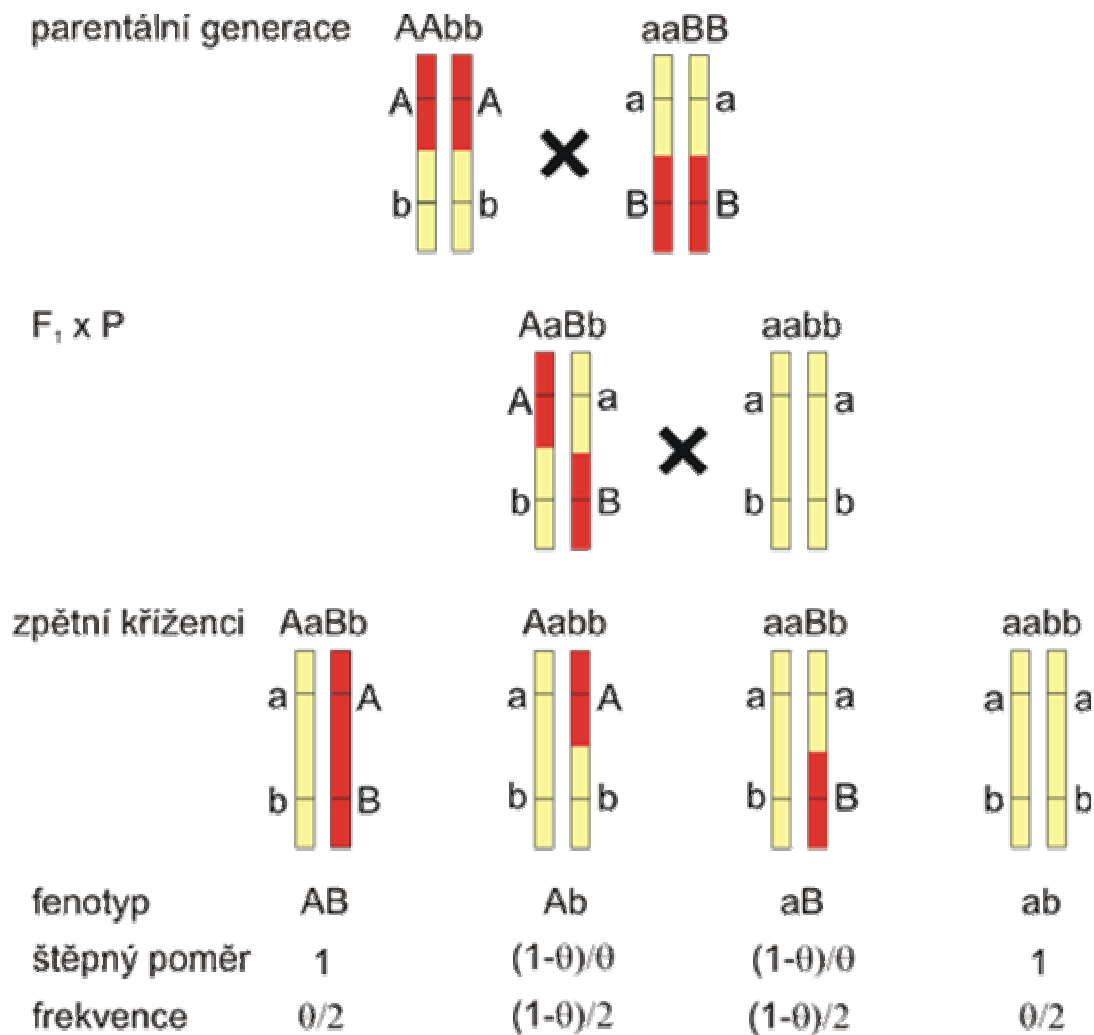
Obr. 8.4 Zpětné křížení u modelového onemocnění u myši

"opakovaná epilace" (repeated epilation, Er) je autozomálně dominantní mutace s abnormální proliferací a diferenciací keratinocytů u heterozygotů. Er/Er homozygoti umírají in utero. Aby byl zjištěn genetický defekt, který je podkladem tohoto onemocnění, byla testována vazba Er a genetických markerů v myším genomu. Na obrázku vidíte vazbu Er (mutantní alela Er, alela divokého typu [wildtype, wt] značena +). a markeru D4Mit204 (DNA segment na chromozomu 4, Massachusetts Institute of Technology číslo 204, se dvěma alelami a1 a a2). Tato studie (Herron a spol., 2005) vedla nakonec k identifikaci mutace v genu pro stratifin a ukázala na důležitost stratifinu jako regulátoru buněčného cyklu myších i lidských keratinocytů.



Obr. 8.5. Vazba ve fázi trans

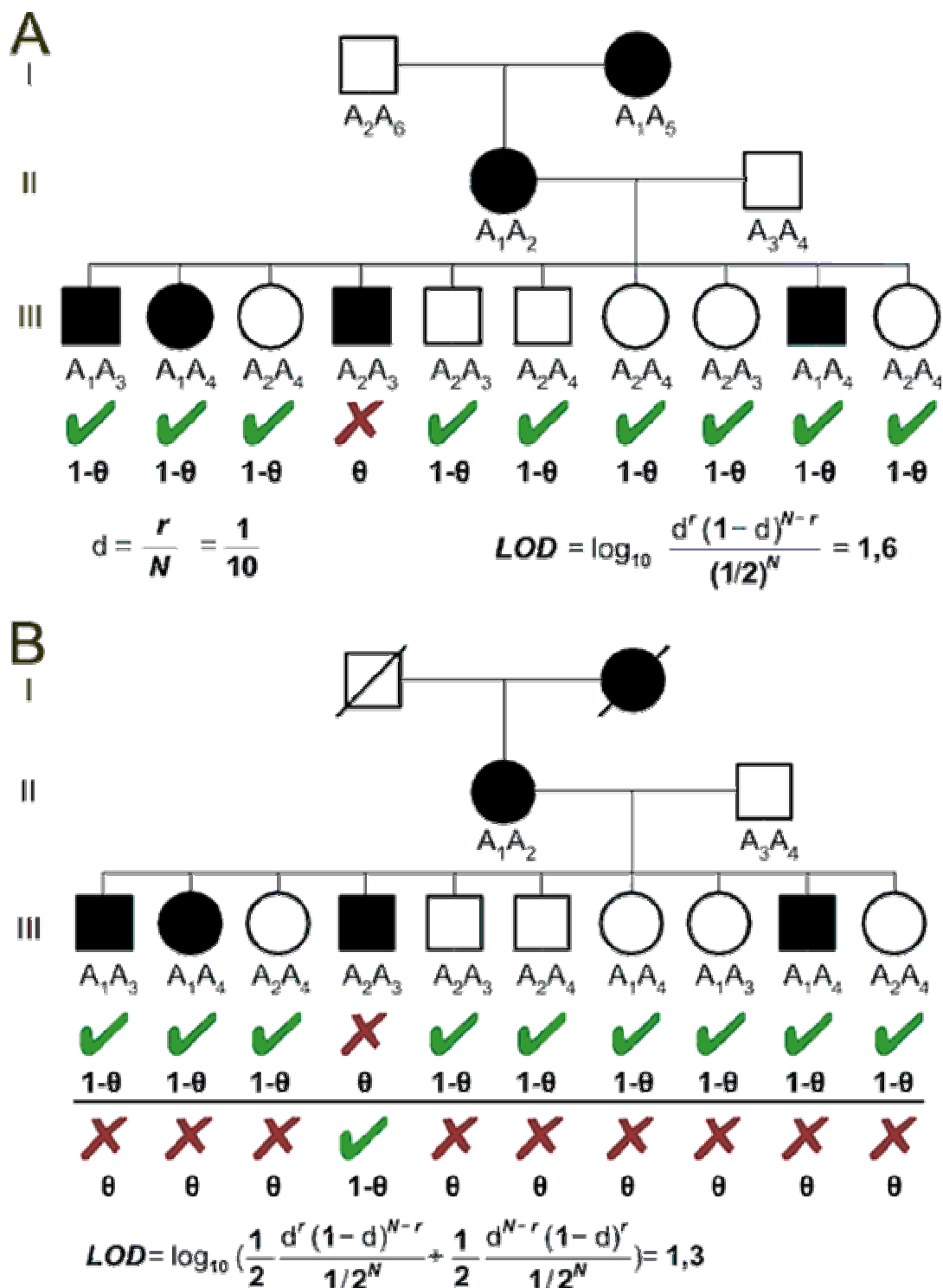
Situace je velmi podobná obrázku 8.3. Jen u F1 hybridu je recesivní alela a na stejném chromozomu s dominantní alelou B (haplotyp aB), obráceně na druhém chromozomu. F1 hybrid má tedy sestavu haplotypů aB/Ab . Všimněte si, že pokud napíšete genotyp klasicky $AaBb$, nerozlišíte F1 hybridu z obr. 8.3, s haplotypy AB/ab . Četnosti potomků jsou obrácené ve srovnání s obr. 8.3. - non-rekombinanti z obr. 8.3 jsou nyní rekombinanti a naopak. Na tomto obrázku jsou uvedeny obecné vrozce pro štěpný poměr a frekvenci jednotlivých genotypů.



Obr. 8.6. Vazba v lidském rodokmenu

A: Rodina s jedinci trpícími autosomálně dominantním onemocněním byla genotypována pro vysoce polymorfní marker s alelami A1-A6. Onemocnění je ve vazbě s alelou A1, s výjimkou jedince III/4, který je pravděpodobně rekombinant. Rekombinační zlomek je 0,1. LOD skóre 1,6 je však nedostatečné pro potvrzení vazby.

B: Stejná rodina jako v A, ale genotypy prarodičů nejsou známy. To komplikuje vazebnou analýzu, neboť není jasné, jestli je ve vazbě s onemocněním alela A1 nebo A2. I když výsledky třetí generace mluví pro alelu A1, LOD skóre musíme počítat tak, abychom vzali do úvahy a priori stejně pravděpodobné možnosti, a tak je LOD skóre pouze 1,3.



Obr. 8.7 Pořadí lokusů

A: gen h pro autosomálně recesivně dědičnou infertilitu samců (hh homozygoti neplodní, HH a Hh normální) byl "zamapován" mezi dva mikrosatelitní markery D10Rat34 (s alelami A a a) a D10Rat57 (s alelami B a b) ve zpětném křížení aabbhh x AaBbHh. aabbhh musela být kvůli neplodnosti samců samice. V tabulce je u potomků pouze jeden haplotyp, druhý (maternální) je vždy abh. Genotypy jsou seskupeny tak, že rekombinace mezi lokusem A a lokusy B+H vyústí v genotypy aBH/abh a Abh/abh. Dvojitý crossing-over má řádově nižší

pravděpodobnost než jednoduchý, identifikace této skupiny rekombinantů je tedy snadná.

B: genetická mapa vytvořená z dat v tabulce A, s vazebnými vzdálenostmi v centimorganech (cM), vypočítanými pomocí Haldanovy nebo Kosambiho mapové funkce.

C: Mapování myši autosomálně recesivní mutace luxoid (lu), která ovlivňuje vývoj končetin a spermatogenezi s použitím markerů na myším chromozomu 9. Porovnáním genotypů v markerech (bílé jsou alely z postiženého kmene, černé z kontrolního kmene) s fenotypem (+ je divoký fenotyp, lu je postižený fenotyp) nám ukazuje, že luxoid musí ležet mezi markery D9Mit256 a D9Mit99. Tato studie (Buaas a spol., 2004) odhalila mutaci v Plzf (promyelocytic leukemia zinc finger) a rozpoznala Plzf jako důležitý regulátor proliferace kmenových buněk.

A

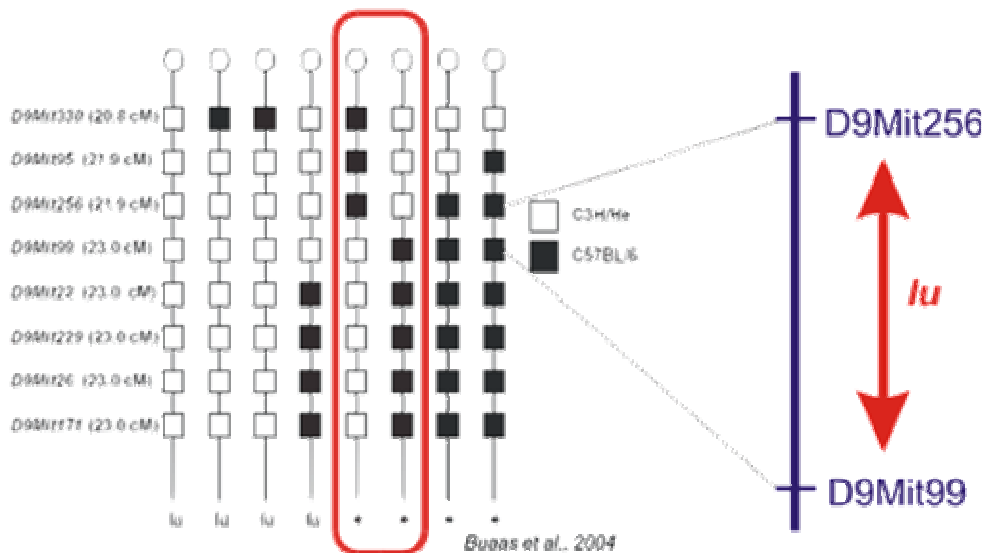
D10Rat34 = A, D10Rat57 = B, *infertilita* = *h*,
homozygoti hh postiženi

genotyp	počet	%	rekombinace?
ABH <i>abh</i>	266	78,24	parentální genotypy
AB <i>h</i> <i>abH</i>	1	0,29	dvojnásobný crossing over
<i>a</i> BH Ab <i>h</i>	43	12,65	jednoduchý crossing over
<i>a</i> B <i>h</i> AbH	30	8,82	jednoduchý crossing over
	340	100,00	

B



C



Obr. 8.8 Dinukleotidové a trinukleotidové repeticie jako příklady mikrosatelitních sekvencí.

Vlevo - mikrosatelity s okolní sekvencí DNA. Sekvence primeru na 5' straně je žlutě, primer na 3' straně je komplementární k sekvenci označené zeleně. Vlastní mikrosatelit je modře.

Všimněte si určité degenerace trinukleotidového mikrosatelitu. V sekvenci obklopující dinukleotidový mikrosatelit je ještě jeden tetranukleotidový.

Napravo - příklad elektroforézy v polyakrylamidovém gelu pro oba mikrosatelity. DNA byla

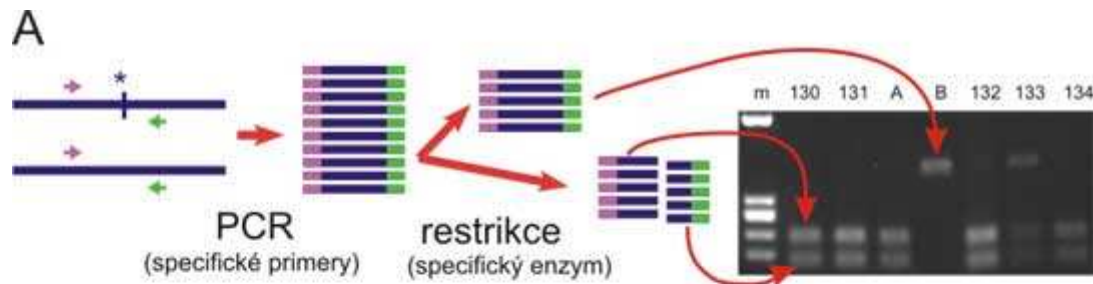
obarvena fluorescentním barvivem ethidium bromidem a vyfotografována při ultrafialovém osvětlení. Všimněte si komplikované struktury amplifikovaných fragmentů, zvláště u dinukleotidového mikrosatelitu. To se připisuje "skluzu" polymerázy (chyba při polymeraci vedoucí k přidání nebo méně často odebrání několika jednotek repetice, viz kapitola repetitivní sekvence) a vzniku heteroduplexů při PCR (heteroduplexy mají nižší pohyblivost, zejména v polyakrylamidovém gelu).



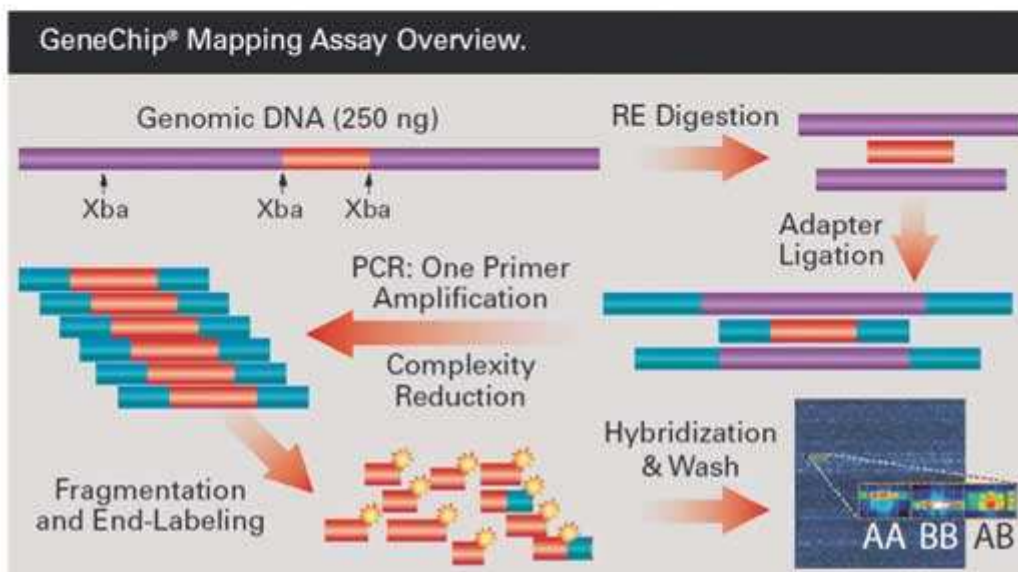
Obr. 8.9 Analýza polymorfismů typu SNP

A: SNP v restričním místě. DNA segment se amplifikuje pomocí PCR, s primery 5' a 3' od restričního místa a produkt PCR reakce se štěpí patřičnou restriční endonukleázou. Fragmentsy jsou separovány elektroforézou v agarózovém gelu. Vzorky A a B jsou homozygotní kontroly. Pokud označíme alelu s restričním místem + a alelu bez restričního místa -, bude A ++ kontrola, B kontrola -/- a genotypy v rodině budou 130-132 ++, 133 heterozygot +/- a 134 opět homozygot ++.

B: . Analýza SNP na mikročipu. Genomická DNA se štěpí restriční endonukleázou na konstantních místech. Takto vzniklé fragmenty jsou pak ligovány k syntetickým oligonukleotidovým adaptorům a amplifikovány pomocí PCR (primer odpovídá sekvenci adaptoru). Pomocí PCR nelze amplifikovat příliš dlouhé fragmenty, čímž se zredukuje komplexita vzorku. Tyto fragmenty se rozlámou na malé kousky, označí fluorescenčně a hybridizují s mikročipem. Pro každý SNP mikročip obsahuje řadu několika oligonukleotidů, které hybridizují s alelou A (obsahující např. guanin) a druhou řadu oligonukleotidů, které hybridizují s alelou B (obsahující např. adenin jako polymorfní nukleotid). Mikročip se vyfotografuje pod fluorescentním mikroskopem a genotypy se odečtou, jak je naznačeno na obrázku. Mikročip na obrázku je vyroben pro analýzu 50000 SNP.

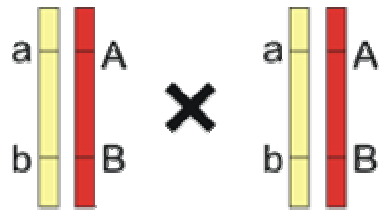


B



Tabulka 8.1. Vazba u interkrosu

Tabulka představuje interkros, ale vyobrazený jinak než je obvyklé u dihybridismu - zde je jeden lokus ve sloupcích a druhý v řádcích. nr = non-rekombinant (např. genotyp AABB), c.o. = crossing-over (např. genotyp AaBB), 2 x c.o. = crossing-over na obou chromozomech (nepleťte si s dvojitými rekombinanty u třibodového pokusu). Pro genotyp AaBb, nerekombinovaný haplotyp AB/ab nelze odlišit od dvojnásobně rekombinovaného Ab/aB. Pro lokusy vzdálené 10 cM, dvojnásobně rekombinovaných bude v genotypu AaBb asi 1,2%, a nemusíme tedy většinou tuto možnost brát v úvahu.



	lokus A		
lokus B	AA	aa	Aa
BB	nr	2 x c.o.	c.o.
bb	2 x c.o.	nr	c.o.
Bb	c.o.	c.o.	???

