

Genetické haraburdí - repetitivní DNA

Repetitivní DNA

DNA eukaryot a také člověka obsahuje značný podíl nekódujících sekvencí. Tak jako kódující DNA i nekódující může být unikátní anebo se může nacházet v genomu ve více identických nebo podobných kopiích. Sekvence DNA s vysokým množstvím kopií se nazývají repetitivní sekvence. Pokud jsou kopie sekvenčního motivu v blocích, v řadě za sebou, hovoříme o tandemových repetitcích, od nich odlišujeme repetitivní sekvence rozptýlené v genomu jako jednotlivé kopie (rozptýlené repetice - anglicky interspersed repeats).

Podstata rozptýlených repetit - transpozony

Většina rozptýlených repetit vzniká procesem transpozice, což je "skákáni" segmentu DNA na jiné místo genomu. Rozlišujeme v podstatě dva typy transpozibilních elementů DNA, neboli transpozonů: DNA transpozony a retrotranspozony. Hlavní skupiny rozptýlených repetit se schopností transpozice jsou schematicky znázorněny na obr. 1.

DNA transpozony

DNA transpozony jsou v lidském genomu považovány za inaktivní, díky akumulaci mutací v průběhu fylogeneze obratlovců, a tak můžeme najít pouze jejich evolučně staré zbytky, neboli "fosilie". Nicméně aktivní transpozon odvozený z lidských fosilních elementů může být "vyroben" s použitím informací získaných z lidského genomu i genomu ostatních obratlovců. Jedním z příkladů je transpozon "Sleeping Beauty" (Šípková Růženka), který by se mohl např. stát základem další generace genové terapie, díky více specifickému místu integrace, než je tomu např. u retrovirů. Jak funguje typický DNA transpozon? Jádrem je sekvence kódující enzym transpozázu. Tento enzym se váže k oběma koncům repetitivního elementu, které jsou tvořeny invertovanými repeticemi. Tyto invertované konce si tedy mohou "vyměnit" řetězce a stabilizovat tak strukturu stopka-klička, nezbytnou pro aktivitu transpozázy. Transpozáza pak vyštěpí transpozon a liguje takto vzniklé volné konce chromozomální DNA. [Téměř shodný mechanismus je činný během maturace genů pro imunoglobuliny (V-D-J rekombinace) a TCR (T-cell receptor, receptor T-lymfocyty) při vyštěpení mezilehlých sekvencí. Je zajímavé, že enzym katalyzující tuto reakci (skládá se z dvou podjednotek RAG1 a RAG2) se skutečně pravděpodobně vyvinul z transpozázy.]. Uvolněný komplex transpozon-transpozáza se váže na specifický sekvenční motiv jinde v genomu, transpozáza štěpí hostitelskou DNA a liguje transpozon na nové místo. Takto se transpozon pohybuje mechanismem vyjmout-vložit (cut and paste) a počet kopií zůstává stabilní.

Retrotranspozony

Retrotranspozony jsou v lidském genomu mnohem důležitějšími transpozibilními elementy. Zaprvé jsou daleko hojnější, přímo tvoří nejméně 45% lidského genomu (odhady se různí, ale mnoho výzkumníků věří, že by to mělo být více, neboť starobylé retrotranspozony které byly inaktivovány, divergovaly díky mutacím tak, že jsou již nerozeznatelné). Zadruhé jsou retrotranspozony v lidském genomu stále aktivní.

Pro "skákání" vyžadují buněčné RNA polymerázy (II nebo III), kterými jsou přepsány do RNA, zatímco původní kopie zůstává na svém místě. RNA kopie podléhá reverzní transkripci do DNA, která je vložena do genomu na nové místo. Tyto elementy tedy expandují (co do množství) mechanismem duplikace (kopírovat-vložit, copy and paste). Jak je dále popsáno pro L1 retrotranspozony, proces retrotranspozice je náchylný k různorodým chybám, a tak jsou nově vzniklé kopie většinou inaktivovány delekcemi nebo bodovými mutacemi. Protože je většina kopií inaktivní, další expanze dané rodiny retrotranspozonů je řízena několika aktivními úplnými elementy. Avšak i když by později během fylogeneze došlo ke ztrátě všech aktivních elementů, genom může být doslova přeplněn fosilními členy dané rodiny sekvencí.

Retrotranspozony mohou být dále klasifikovány jako autonomní nebo neautonomní. Autonomní retrotranspozony kódují proteiny nezbytné k jejich transpozici, ačkoli pro úspěšné "skákání" jsou také závislé na hostitelových RNA polymerázách a enzymech opravujících DNA. Neautonomní retrotranspozony nekódují proteiny a musí tak zneužít enzymy jiného transpozonu aby byly schopné transpozice.

LTR retrotranspozony - endogenní retroviry

Endogenní retroviry, také nazývané LTR retrotranspozony, připomínají svým složením proviry skutečných retrovirů - obsahují LTR (long terminal repeats, dlouhé terminální repetice) a geny *gag*, *pol*, *env* a *prt*, ale alespoň jeden z genů nezbytných pro sestavení infekčních virových částic je mutován nebo chybí, zvláště se to týká genu *env*. Proto se mohou endogenní retroviry pohybovat pouze uvnitř buněk, jinak je jejich životní cyklus podobný infekčním retrovirům, jako je HIV. Ačkoli jsou endogenní retroviry aktivní u mnoha savců, včetně šimpanze, lidský genom v současné době obsahuje pouze fosilie endogenních retrovirů (mutované a neschopné transpozice), které zaplňují asi 8% genomu. Intaktní endogenní retroviry jsou dlouhé 7-9 kb, ale stejně jako u L1 retrotranspozony (viz dále) mnoho z nich je zkrácených, zejména na 5' konci. Často také můžeme najít pouze samostatné LTR, jako výsledek integrace retroviru a následné intrachromozomální rekombinace mezi oběma LTR nebo nerovnoměrné rekombinace dvou homologních chromozomů vedoucí k deleci kódující části retroviru (viz obr. 5).

Non-LTR retrotranspozony

LINE

LINE (long interspersed nuclear elements = dlouhé rozptýlené jaderné elementy) jsou autonomní retrotranspozony. Tvoří asi 21% lidského genomu. Aktivní elementy patří k nejhojnější rodině LINE-1 neboli L1, která sama o sobě zahrnuje 17% genomu. Ze zhruba půl milionu kopií L1 v našem genomu, skoro 10 000 má úplnou velikost a asi 100 je stále schopno retrotranspozice. Aktivní L1 element je dlouhý asi 6 kb a obsahuje dva otevřené čtecí rámce (open reading frames), ORF1 a ORF2. 5' UTR (untranslated region, nepřekládaná oblast) funguje také jako promotor, 3' UTR obsahuje signál k polyadenylaci. Funkce ORF1 není jasná, známo je jen, že se váže na L1 mRNA. ORF2 obsahuje doménu s aktivitou reverzní transkriptázy a endonukleázovou doménu a je enzymem zodpovědným za integraci. Životní cyklus L1 začíná transkripcí L1 DNA buněčnou RNA polymerázou II a standardní maturací v mRNA molekulu. L1 mRNA je transportována do cytoplazmy, kde je syntetizován protein ORF1. Pak je translace reiniciována na "vnitřním místě pro vstup ribozomu" (internal ribosomal entry site, IRES) (proces to nekanonický a tím neefektivní u eukaryot, a tak jen část L1 mRNA molekul získá svůj protein ORF2). Oba proteiny se po své translaci neprodleně

váží na L1 mRNA. Tento komplex protein-mRNA je transportován do jádra. ORF2 štěpí chromosomální DNA v cílovém místě (cílové místo není úplně specifické jak je tomu např. v případě restričních endonukleáz, ale je zde určitá preference pro sekvence bohaté A a T, místo štěpení je přibližně TT/AAAA). Štěpení DNA je nerovnoměrné (vytváří se kohezní konce). Volná 3' OH skupina na jedné straně štěpené DNA molekuly je užita reverzní transkriptázou proteinu ORF2 k zahájení syntézy prvního řetězce cDNA (target primed reverse transcription, reverzní transkripce s cílovou sekvencí jako primerem). Detailní mechanismus syntézy druhého řetězce cDNA je stále předmětem diskuse, proces však končí stabilní integrací dvouvláknové L1 DNA na novém místě genomu. Díky stupňovitému zlomu cílové DNA vyrobenému endonukleázou transpozonu je integrovaný L1 element obklopen duplikací cílového místa o velikosti 7-20 párů bazí (bp = base pairs) (obr. 2). Reverzní transkriptáza je většinou neschopna ukončit syntézu prvního řetězce, což vede ke zkrácení nové kopie na 5' konci (obr. 3A). Reverzní transkriptáza také nemá 3' - 5' exonukleázovou aktivitu a tak často zavádí do nové kopie bodové mutace. Je zajímavé, že L1 mRNA je exprimována zejména v meiotických a postmeiotických spermatocytech, zvyšující tak potenciál L1 pro expanzi (kopie introdukované to zárodečné linie jsou na rozdíl od nových somatických integrací dědičné).

Neautonomní retrotranspozony - SINE

SINE (short interspersed nuclear elements = krátké rozptýlené jaderné elementy) jsou typicky kratší než 500 bp a nemají žádný kódující potenciál. Hlavní rodinou SINE u člověka jsou Alu elementy (jméno je odvozeno od jejich objevu spojenému s párem konzervovaných restričních míst pro endonukleázu AluI). Více než 1 milion Alu elementů tvoří asi 11% lidského genomu. Alu elementy sdílí konsenzus 282 bp který je příbuzný a byl patrně odvozen z RNA podjednotky SRP (zvané 7SL RNA). SRP (signal recognition particle = částice rozpoznávající signál) je ribonukleoproteinový komplex, který rozpoznává signální peptid, váže se na něj a přemístí komplex ribozom-mRNA-nascentní peptid ke kanálu endoplazmatického retikula (ER), skrz nějž je nascentní peptid translokován do lumen ER nebo integrován v membráně ER. Alu jsou, stejně jako gen pro 7SL RNA transkribovány RNA polymerázou III. Alu RNA váže dva proteiny SRP (9 a 14). Pravděpodobně se tak může Alu vázat na ribozom a díky svému "ocasů" bohatému na adenin také (pokud ribozom zrovna zpracovává LINE-1 mRNA) na nascentní protein ORF2 a zneužít ORF2 k reverzní transkripci a integraci vlastní RNA a nikoli LINE-1 (obr. 4).

Funkce transpozonů

Z bezprostředního pohledu nemají transpozony žádnou důležitou funkci v buňce - hovoří se o "starém harampádí" - odpadní DNA (junk DNA); nebo o sobecké DNA, neboť se transpozony propagují na úkor buněčných energetických zdrojů. Z širšího úhlu pohledu může být mobilita retrotranspozonů důležitá pro plasticitu genomu. Příležitostná inserce do genu může vyřadit gen z funkce a způsobit dědičné onemocnění (obr. 3C). LTR a LINE elementy mohou také měnit genovou expresi, pokud se inserují do blízkosti nějakého genu, neboť LTR a LINE 5'UTR mají silnou promotorovou aktivitu v obou směrech (obr. 3F).

Protože má LINE-1 retrotranspozon relativně slabý polyadenylační signál, stává se, že RNA polymeráza II se skrz něj proče, a tak připojí k L1 mRNA i následující sekvenci, která podlehne reverzní transkripci a přesunu na nové místo. Tak může být LINE-1 vektorem pro mobilitu samostatně nemobilních sekvencí. Navíc jsou retrotransponované kopie L1 často zkrácené na 5' konci, a tak se mobilizovaná DNA (která je na 3' konci) může dostat na nové

místo i beze zbytků L1 sekvence. To může mít význam hlavně pro mobilitu menších DNA fragmentů - např. k výměně exonů mezi geny (obr. 3D).

Retrotranspozice L1 může dokonce vyústit v delece a inverze, jako na obr. 3E.

Zřídka je normální buněčná mRNA předmětem reverzní transkripce a transpozice enzymem z L1 nebo z jiného retrotranspozonu. V tomto případě dochází k duplikaci genu. Nová kopie se nazývá "procesovaný pseudogen" (processed pseudogene), neboť je odvozena ze zralé "zpracované" mRNA bez intronů, a je obvykle nefunkční, díky chybějícímu promotoru (obr. 3B). Zřídka však může procesovaný pseudogen přijmout novou funkci pod selekčním tlakem. Velmi známý příklad je gen pro podjednotku E1 alfa pyruvátdehydrogenázy. Tento gen (PDHA1) leží u placentálních savců na chromozomu X. Ale exprese mnoha genů na chromozomu X je v průběhu spermatogeneze zastavena, včetně PDHA1, ačkoli jeho produkt je nezbytný pro funkci všech buněk. Tato chybějící funkce byla očividně zachráněna retrotranspozicí: na chromozomu 4 se nachází velmi podobný gen PDHA2, ale tento gen postrádá introny - a to je pro procesované pseudogeny typické. Vysoce exprimované "provozní" (housekeeping) geny mají samozřejmě větší pravděpodobnost retrotranspozice. Nacházíme tak mnoho procesovaných pseudogenů pro ribozomální proteiny, glykolytické enzymy, beta-aktin, a podobně. Procesované pseudogeny by neměly být zaměňovány za druhou kategorii "obyčejných" pseudogenů, které vznikají duplikací genomické DNA (např. pseudogeny ve skupině genů pro hemoglobin) a zachovávají proto původní strukturu (exony, introny, promotor...i když s porušenou funkcí).

Bylo objeveno několik genů přímo odvozených z retrotranspozonů. Poslední přídavek je gen Peg10 (paternally expressed 10, paternálně exprimovaný gen 10), odvozený z LTR retrotranspozonu z rodiny Ty3/gypsy (velmi podobný retrotranspozon byl nalezen v aktivní formě u ryby fugu {*Takifugu rubripes*}). Peg10 je nezbytný pro vývoj placenty u myši a stejnou funkci bude mít pravděpodobně u člověka. Jiné příklady zahrnují geny pro syncytin, odvozené z endogenních retrovirů z rodiny HERV-W. Produkty těchto genů jsou důležité pro vytvoření syncytia z buněk trofoblastu, mechanismus fúze membrán připomíná vstup retroviru do buňky.

I neaktivní repetitivní elementy zvětšují plasticitu genomu tím, že podporují mezichromozomový nerovnoměrný crossing-over nebo intrachromozomovou rekombinaci (obr. 5).

V neposlední řadě se spekuluje o tom, že by transpozony mohly mít nějakou reálnou fyziologickou funkci, např. proto, že jejich exprese je obecně zvýšena během stresové odpovědi. Ale různé hypotézy, které mohou být koncipovány na základě takových pozorování jsou v současné době zcela nepotvrzené.

Tandemové repetice

Tandemové repetice jsou tvořeny za sebou jsoícími identickými a nebo téměř identickými jednotkami. Tolik se však různí v délce jednotky repetice i celé repetice, že je jakákoli klasifikace neuspokojivá, a je nutno ji brát "cum grano salis". Největší repetice, které mají tendenci být složeny z relativně dlouhých jednotek se nazývají **satelity**. Jméno satelity je pochází z centrifugace DNA v hustotních gradientech. Nejprve, během konvenční izolace DNA, je tato předmětem namáhání smykem (shear stress), s výslednou fragmentací DNA (in vivo obsahuje jeden chromosom v G1 fázi 1 molekulu DNA). Tyto fragmenty mohou být

centrifugovány v hustotních gradientech tak, že molekuly DNA obsazují v gradientu místa se stejnou hustotou prostředí jako má molekula DNA. Většina DNA vytvoří jednotný "proužek". Ale fragmenty DNA se signifikantně odlišným obsahem CG/AT, způsobeným např. rozsáhlými monotónními repeticemi vytvoří méně intenzivní přídatné "satelitní" proužky. Označení satelitní DNA bylo později rozšířeno a zahrnuje i podobně repetitivní sekvence, které však nevytváří tyto satelitní proužky. Primární jednotky repetice u satelitů jsou různorodé, od GGAAT u satelitu 2 a 3 až po 171 bp u alfa satelitu. Ale tyto primární jednotky jsou často degenerované, s určitými nepravidelnostmi. Tyto nepravidelnosti se mohou periodicky opakovat a tak tvořit sekundární jednotky. Satelitní DNA je hojná v oblasti centromer a konstitutivního heterochromatinu. Přestože je lidský genom považován za úplně sestavený, oblasti centromer a heterochromatin obsahující satelitní sekvence nejsou zahrnuty, neboť sekvenování takových oblastí je z různých důvodů problematické (absence patřičných restričních míst, obtížné sekvenování, téměř nemožné sestavení jednotlivých sekvencí do tzv. kontigu apod.). Z mnoha satelitů nacházených v oblasti centromer, tvoří rodina alfa satelitu (s primární jednotkou dlouhou 171 bp) pravděpodobně funkční jádro centromery, neboť je důležitá pro "poskládání" kinetochory během buněčného dělení (některé proteiny kinetochory se váží na alfa satelit v centromeře a tím zahajují sestavování kinetochory). Funkce ostatních satelitů je neznámá, jsou považovány obvykle za odpadní (junk) DNA.

Minisatelity jsou kratší tandemové repetice, v rozsahu kilobází, které se více vyskytují v subtelomerických oblastech chromozomů. Jsou obvykle vysoce polymorfní co do počtu opakování jednotky repetice (mnoho alel v populaci) a mohou být použity jako genetické markery - VNTR (variable number of tandem repeats = variabilní množství tandemových repetice). VNTR jsou často příliš dlouhé pro amplifikaci pomocí PCR a jsou tudíž typicky stanovovány pomocí Southernova blotu (a jejich obliba tudíž klesá). Někdy se uvažuje o tom, že by některé minisatelity mohly mít regulační funkce, jako např. VNTR v promotoru inzulinového genu, kde byla různá délka VNTR asociována s různými typy diabetu. Jedna z alel insulínového VNTR je zobrazena na obr. 7. Telomery lidských chromozomů, tvořené několika kilobazemi hexamerové repetice TTAGGG patří rozsahem také k minisatelitům, i když vznikají specifickým mechanismem - pomocí enzymu telomerázy. Telomeráza je složena z bílkovinné podjednotky a z RNA podjednotky obsahující sekvenci komplementární k TTAGGG, která slouží jako templát pro elongaci telomery (bílkovinná podjednotka je příbuzná reverzním transkriptázám non-LTR retrotranspozonů). Nicméně se mohou telomery elongovat i pasivně, mechanismem nerovnoměrného crossing-overu (obr. 5D), např. v nádorových buňkách.

Možná by stálo za to na tomto místě znovu připomenout, že sekvence lidského genomu zahrnuje euchromatické úseky, ohraničené proximálně (ale nezahrnující) centromerou a pericentromerickým heterochromatinem a distálně telomery, které také, spolu se subtelomerickými oblastmi, nejsou obsaženy

Microsatelity jsou zpravidla tvořeny opakováním 1-5 bp, s množstvím opakování zřídka překračujícím stovky. Nejčastější jsou dinukleotidové repetice, ze kterých převažuje typ (CA)_n. Mikrosatelity jsou v genomu velice časté, vysoce polymorfní a jsou často používány jako genetické markery (příklady mikrosatelitů jako genetických markerů jsou v kapitole o genetickém mapování).

Nemoci způsobené expanzí trinukleotidových repetic

Pokud jsou uvnitř nebo v blízkosti genů, mohou mít mikrosatelity, resp. jejich různá délka, závažné důsledky, např. v heterogenní skupině monogenních nemocí podmíněných expanzí trinukleotidových repetic. Nejznámějším příkladem je Huntingtonova chorea, fatální neurologické onemocnění s nástupem v dospělosti, projevující se jako demence s extrapyramidovou poruchou motoriky. V genu pro huntingtin je repetitivní sekvence (CAG)_n, která kóduje úsek bílkoviny tvořený zbytky glutaminu (polyglutaminový úsek, polyglutamine tract). Za normálních okolností mají lidé méně než 20 trinukleotidů CAG a tedy i glutaminů v huntingtinu, kde tyto tvoří důležitou doménu pro interakce s jinými proteiny. Pokud se však mutací toto množství zvětší nad 30 glutaminů, protein nepracuje správně (jak přesně je předmětem rozsáhlého výzkumu a je mimo rozsah tohoto textu) s výsledným progresivním odumíráním neuronů v nucleus caudatus. U jiného onemocnění, myotonické dystrofie (svalová dystrofie se svalovou slabostí provázenou paradoxně zvýšeným svalovým tonem) se nachází patologická expanze trinukleotidu CTG v 3' nepřekládané oblasti genu DMPK (dystrophia myotonica protein kinase). Mutantní mRNA má sama o sobě patogenní potenciál, škodí pravděpodobně sekvestrací různých transkripčních faktorů. Další příklady "expanzních" onemocnění viz kapitola [Nemendelovská dědičnost](#).

Mechanismy expanze/zmenšení tandemových repetic

Prvním mechanismem, který přispívá k polymorfismu tandemových repetic je nerovnoměrný crossing-over. To je typické zvláště pro větší repetice (obr. 5D). malé mikrosatelitní repetice často mění svoji délku díky chybám při syntéze DNA, např. mechanismem "klouzání" polymerázy (obr. 8). Na čele replikace není dvojí šroubovice DNA ještě extrémně stabilní a podléhá významným termálním fluktuacím. Pokud právě DNA polymeráza replikuje mikrosatelit, DNA řetězce během fluktuace nemusí reasociovat exaktně, ale s posunem o několik jednotek repetice. Tento mechanismus je zesílen u některých typů repetic, které mohou stabilizovat metastabilní stavy tvorbou dvouřetězcové kličky, např. CAG/CTG trinukleotid.

Odkazy

Repetitivní sekvence jsou ukládány do centrální databáze, Repbase (bohužel přímé použití Repbase je vyhrazeno pouze akademickým institucím). <http://www.girinst.org/>

Existují také specializované databáze, pokrývající jen některé aspekty, jako je např. databáze lidských endogenních retrovirů. <http://herv.img.cas.cz/>

RepeatMasker je počítačový program provádějící identifikaci repetitivních sekvencí pomocí Repbase a jejich případné maskování v dané sekvenci např. k usnadnění objevování genů. <http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>

SRPDB (signal recognition particle database) poskytuje sekvence a strukturní data ve vztahu k funkci SRP. <http://psyche.uthct.edu/SRPDB/SRPDB.html>

AluGene je databáze Alu elementů které se včlenily do genů kódujících proteiny. <http://alugene.tau.ac.il/>

L1Xplorer je databáze věnovaná detekci a anotaci intaktních L1 elementů. <http://l1explorer.molgen.mpg.de>

Obrázky

Obr. 1 Různé třídy transpozonů objevujících se v lidském genomu

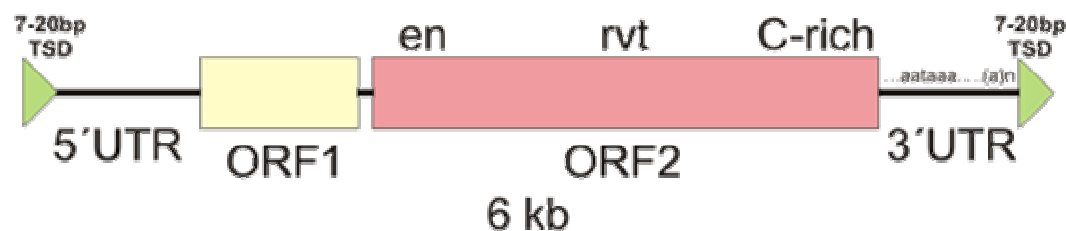
A: Non-LTR retrotranspozony. LINE (long interspersed repeats, dlouhé rozptýlené repetice) jsou zastoupeny LINE-1 (L1). Element dlouhý 6 kb obsahuje dva otevřené čtecí rámce. ORF2 obsahuje doménu s endonukleázovou aktivitou (en), reverzní transkriptázovou aktivitou (rvt) a také doménu bohatou cysteinem (C-rich). 5' nepřekládaná oblast (5' UTR) obsahuje interní promotor pro RNA polymerázu II (u obvyklého genu je promotor před 5' UTR). 3' nepřekládaná oblast obsahuje kanonický polyadenylační signál (AATAAA) a polyA konec (ten se také v genech normálně nevyskytuje, přidává se až k mRNA prostřednictvím polyA polymerázy). L1 je obklopen duplikací cílového místa která vzniká během reverzní transkripce (viz text a další obr.).

B: LTR-retrotranspozon - endogenní retrovirus. Vyobrazena je typická struktura retroviru, resp. proviru, tj. formy integrované do DNA. Endogenní retroviry mohou být na úrovni sekvence odlišeny od infekčních na základě bodových mutací nebo delecí v genech nezbytných pro sestavení infekčních částic - většinou je to gen env (envelope = obal). gag (group specific antigen = skupinově specifický antigen) je protein nukleokapsidy. pol (polymeráza) má aktivitu reverzní transkriptázy (rvt) pro syntézu prvního a druhého řetězce DNA, aktivitu RNázy H pro štěpení RNA v hybridu RNA/DNA po syntéze prvního řetězce a aktivitu integrázy (int) (štěpí cílovou DNA a liguje retrovirus do tohoto místa). prt (proteáza) je nezbytná pro sestavování virové částice tím že štěpí proteinové prekurzory (např. gag a pol jsou často syntetizovány jako jeden velký polyprotein). LTR (long terminal repeats, dlouhé terminální repetice) jsou identické sekvence na obou koncích retroviru. Každý LTR je složen z U3 (3' nepřekládaná oblast), R (oblast rekombinace) a U5 (5' nepřekládaná oblast). Názvy jsou odvozeny od struktury retrovirové mRNA, která je tvořena sekvencí pouze od 5' R do 3' R. Jak je z této mRNA odvozena úplná cDNA je mimo rozsah této kapitoly. I když se reverzní transkripce retrovirů odehrává v cytoplasmě a teoreticky tedy nepotřebuje duplikaci cílového místa, tato se obvykle také tvoří, i když kratší než u L1.

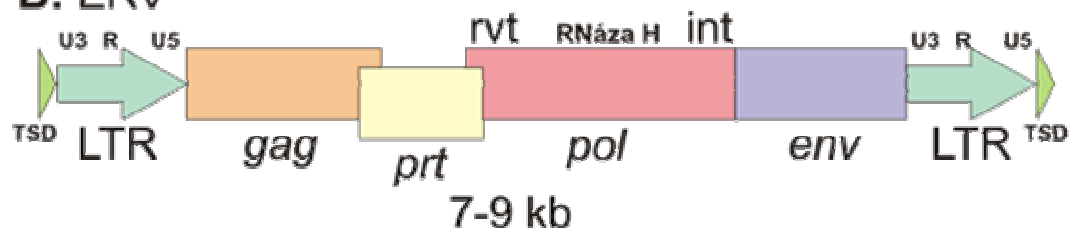
C: je reprezentován 1,2 kb dlouhou rodinou "mariner" (námořník). Synthetický DNA transpozon Sleeping Beauty (Šípková Růženka) také patří do této rodiny. Centrálně uložený gen pro transpozázu je obklopen invertovanými repeticemi. Při integraci se vytváří duplikace cílového místa. Ta je zanechána v genomu jako stopa po integraci, když transpozon přeskočí na jiné místo.

D: neautonomní nonLTR retrotranspozony patří k **SINE** (short interspersed repeat, krátká rozptýlená repetice). Rodina aktivní u lidí je reprezentována 282 bp dlouhým Alu elementem. Alu je dimerem složeným z dvou téměř shodných monomerů (světlá a střední šedá). Levý monomer má delecí tmavě šedého fragmentu. Monomer je odvozen od 7SL RNA, což je RNA podjednotka SRP (signal recognising particle = částice rozpoznávající signál). SRP je komplex rozpoznávající signální peptid proteinů, které mají být transportovány do lumen a nebo membrány endoplasmatického retikula. Pozor, 7SL gen je nakreslen v 50% měřítku oproti Alu. Oblast polyA není částí 7SL genu, ale je důležitá pro "úspěch" Alu jako retrotranspozonu.

A: LINE-1



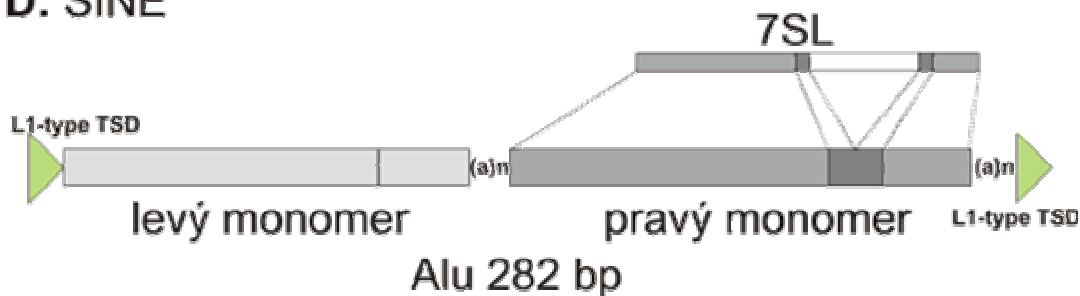
B: ERV



C: DNA transpozón

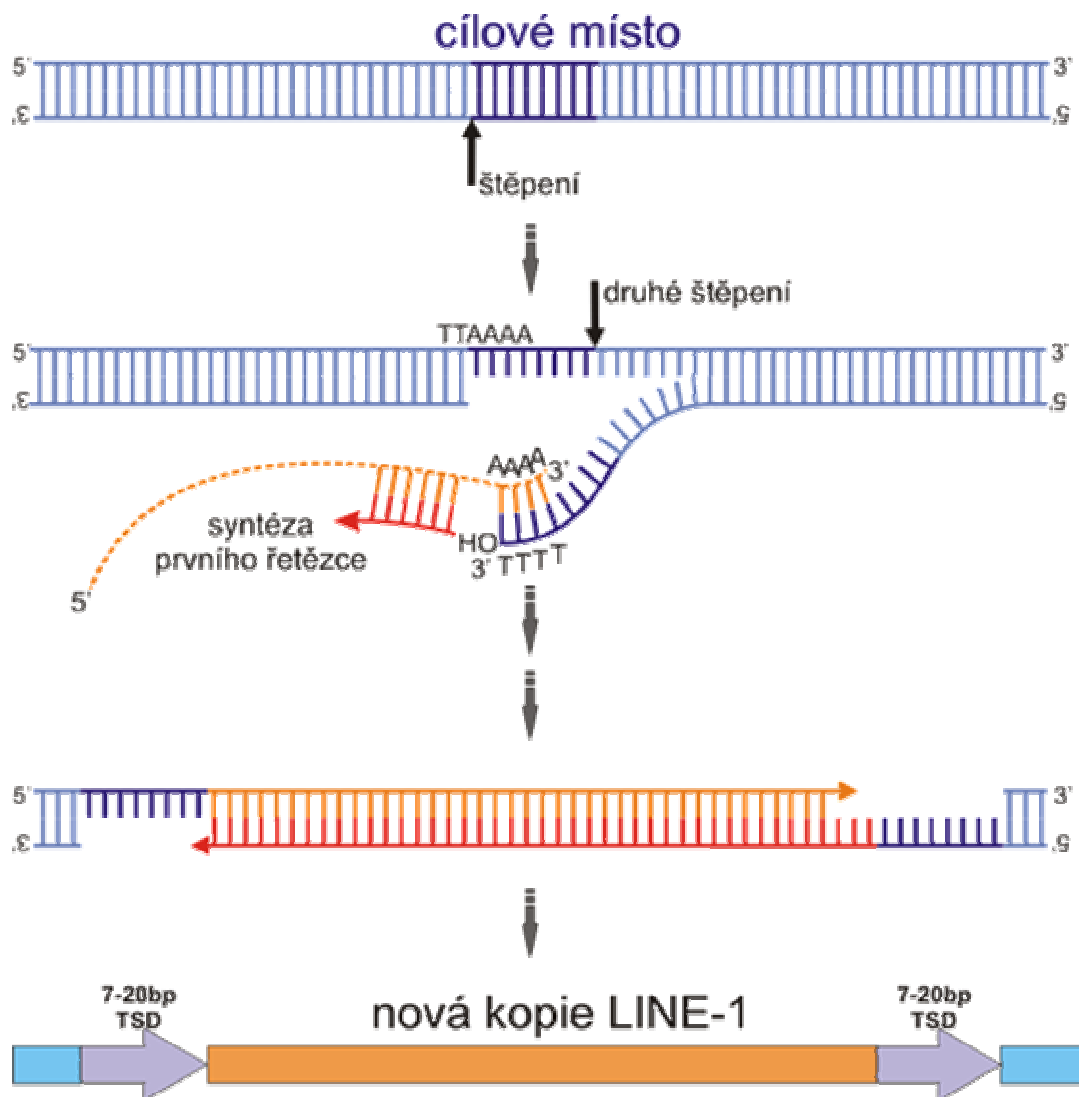


D: SINE



Obr. 2 Reverzní transkripce zahajovaná cílovou DNA (target primed reverse transcription).

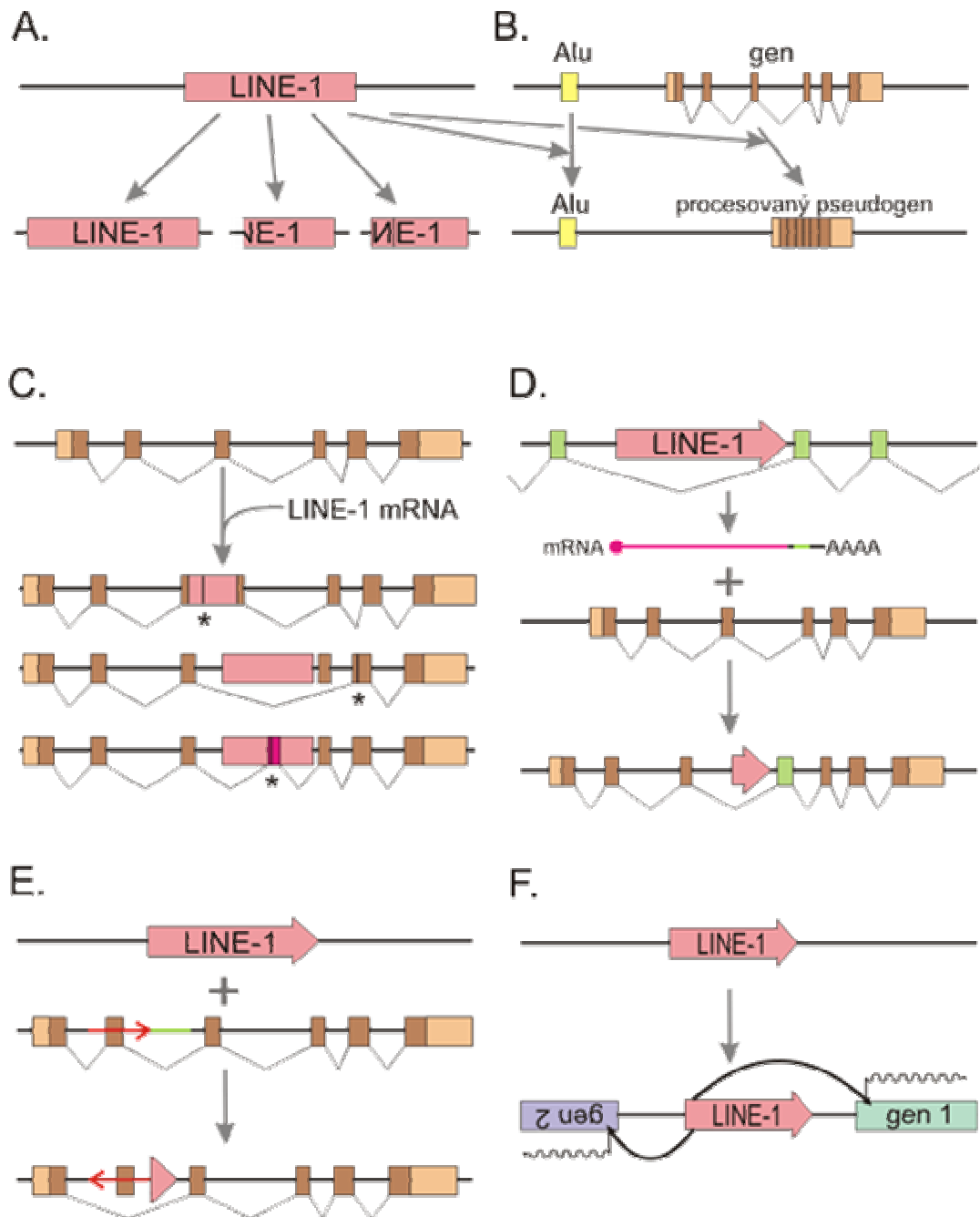
Protein ORF2 štěpí nejprve jeden DNA řetězec cílové sekvence (cílová sekvence je bohatá na A+T, a je obvykle podobná konsenzu TTAAAA, štěpení nastává mezi T a A v komplementárním řetězci). Rozštěpený řetězec se uvolní z dvojšroubovice a spojí se následně s polyA koncem L1 mRNA (přerušovaná oranžová čára). Volná 3' OH skupina řetězce DNA slouží jako primer pro syntézu prvního řetězce cDNA. Štěpení druhého řetězce DNA nastává 7-20 nt od prvního a takto vzniklá volná 3'OH skupina je použita jako primer pro syntézu druhého řetězce L1 cDNA. Mechanismus syntézy druhého řetězce není ještě zcela objasněn. Celý proces končí vytvořením nové DNA kopie L1 elementu, obklopené duplikovaným cílovým místem.



Obr. 3 LINE-1 pozměňuje genom různými způsoby

A: cis-retrotranspozice. L1 produkuje retrotransponované kopie sebe sama. Kopie mohou být úplné, ale daleko častěji jsou neúplné na 5' konci a nebo neúplné s inverzí na 5' konci. **B:** Protein ORF2 z L1 elementu může mobilizovat SINE (jako Alu) nebo jiné buněčné mRNA, čímž vytváří procesované pseudogeny (trans-retrotranspozice). Kódující exony představují hnědé obdélníky, 5' and 3' UTR (nepřekládané oblasti) jsou světlejší barvou, sestřih exonů do mRNA naznačen lomenými čarami. **C:** Retrotranspozon se může vmezeřit do genu. Umístí-li se do exonu, dochází obvykle k porušení čtecího rámce a ke zkrácení proteinu (hvězdička naznačuje, kde se může nacházet nový stop kodon). Ale i v intronu může retrotranspozon napáchat škody - například může dojít k přeskočení exonu při sestřihu nebo ke vzniku nového exonu, což obvykle také vede k přerušení kódující sekvence. Inzerce retrotranspozonu je dobře zdokumentovanou příčinou různých dědičných onemocnění. Nejčastěji se vmezeří Alu elementy, následované L1 elementy. **D:** 3' transdukce. L1 element má relativně slabý polyA signál, a tak se může RNA polymeráza při transkripci "pročíst" až do okolní DNA a připojit k 3' konci mRNA různě dlouhý segment okolní DNA. Tato hybridní mRNA je potom retrotransponována, čímž se přesune nejen L1 element (který je nicméně často zkrácený nebo může dokonce úplně chybět) ale i okolní DNA. To může být mechanismem výměny exonů mezi geny. **E:** Inzerce retrotranspozonu je často doprovázena přestavbou - zde delece

zeleného segmentu a inverze červeného segmentu, který obsahuje exon, s následným přeskočením tohoto exonu při sestřihu. **F:** L1 promotor může podporovat transkripci nejen svého elementu, ale i sousedících genů, a to na obě strany.

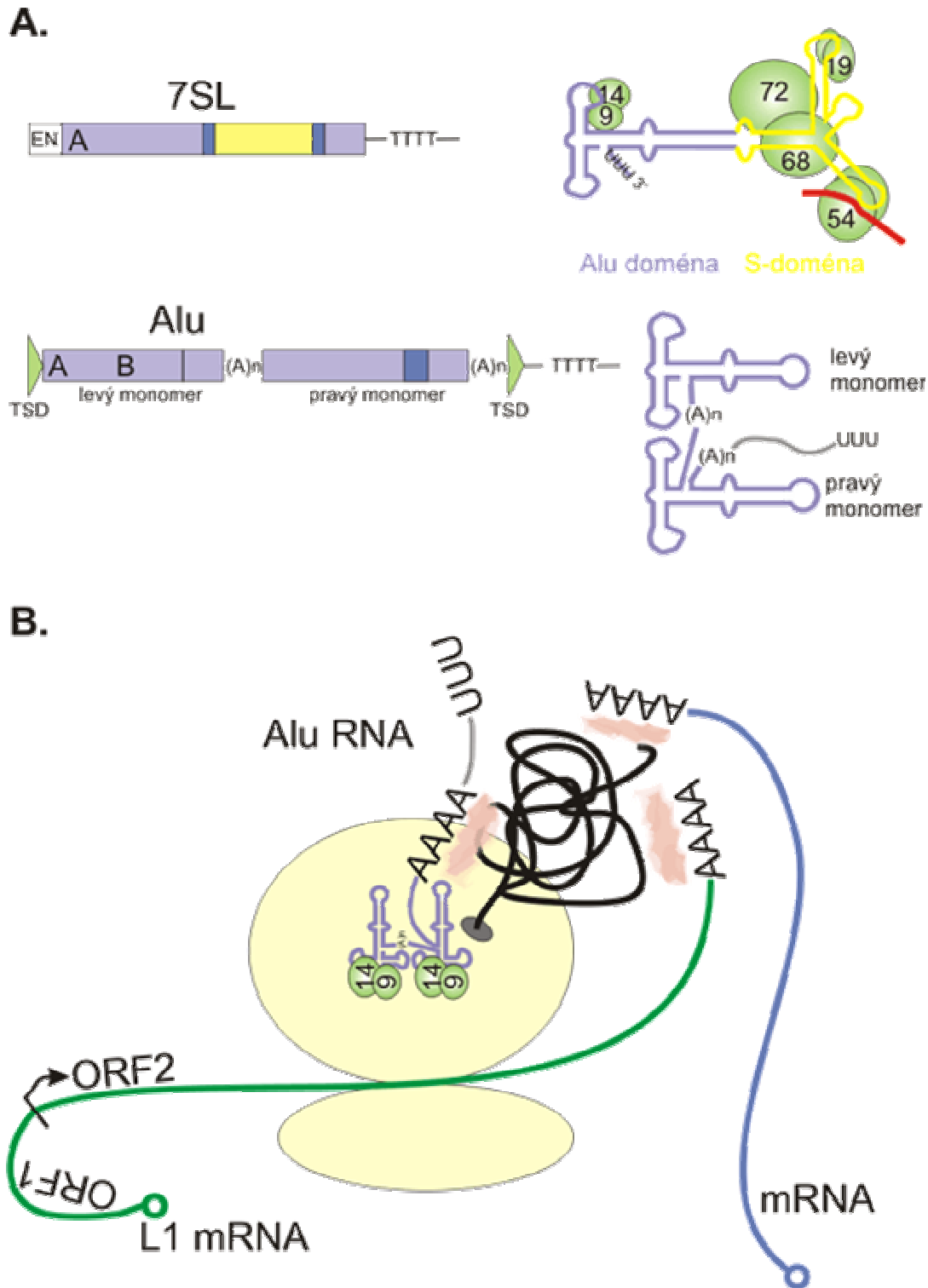


Obr. 4 Alu sekvence jsou hyperparaziti

A: Struktura genu 7SL RNA a Alu elementu (vlevo) a sekundární struktura odpovídajících RNA molekul (vpravo). Transkripce genu 7SL RNA je řízena vnitřním promotorem pro RNA polymerázu III (A) a enhancerem - zesilovačem transkripce (EN). Alu gen má složený vnitřní promotor (A+B). Přírozený terminátor pro RNA polymerázu III je tetranukleotid TTTT. Transkripce je přerušena po prvních třech T. 7SL RNA je složena z Alu domény (modrá) a S-domény (žlutá). SRP proteiny 9 a 14 se váží na Alu doménu, která slouží k navázání na ribozom. Ostatní proteiny se váží na S-doménu, včetně proteinu 54, který spolupracuje na

rozpoznání signálního peptidu. Alu RNA je tvořena v podstatě dvěma Alu doménami 7SL RNA, s přídatnou polyA sekvencí.

B: Alu RNA se váže na ribozom. Pokud ribozom právě překládá ORF2 z LINE-1 mRNA (zelená linie), polyA sekvence Alu elementu kompetuje s polyA sekvencí L1 mRNA o vazbu s nascentním ORF2. Interakce je zprostředkována proteiny vážící polyA sekvencí. Pokud se ORF2 váže na Alu, reverzně transkribuje a transponuje se Alu místo L1 a tím na něm parazituje. Pokud považujeme L1 transpozon za parazita genomu, je Alu hyperparazitem neboli parazitem parazita. Jiné buněčné mRNA (modrá linie) mohou také kompetovat s vazbou ORF2, avšak s mnohem nižší efektivitou (je odhadováno, že z 3000 L1 retrotranspozic, 300 případů bude "ukradeno" Alu elementem a cca 1 případ jinou mRNA).



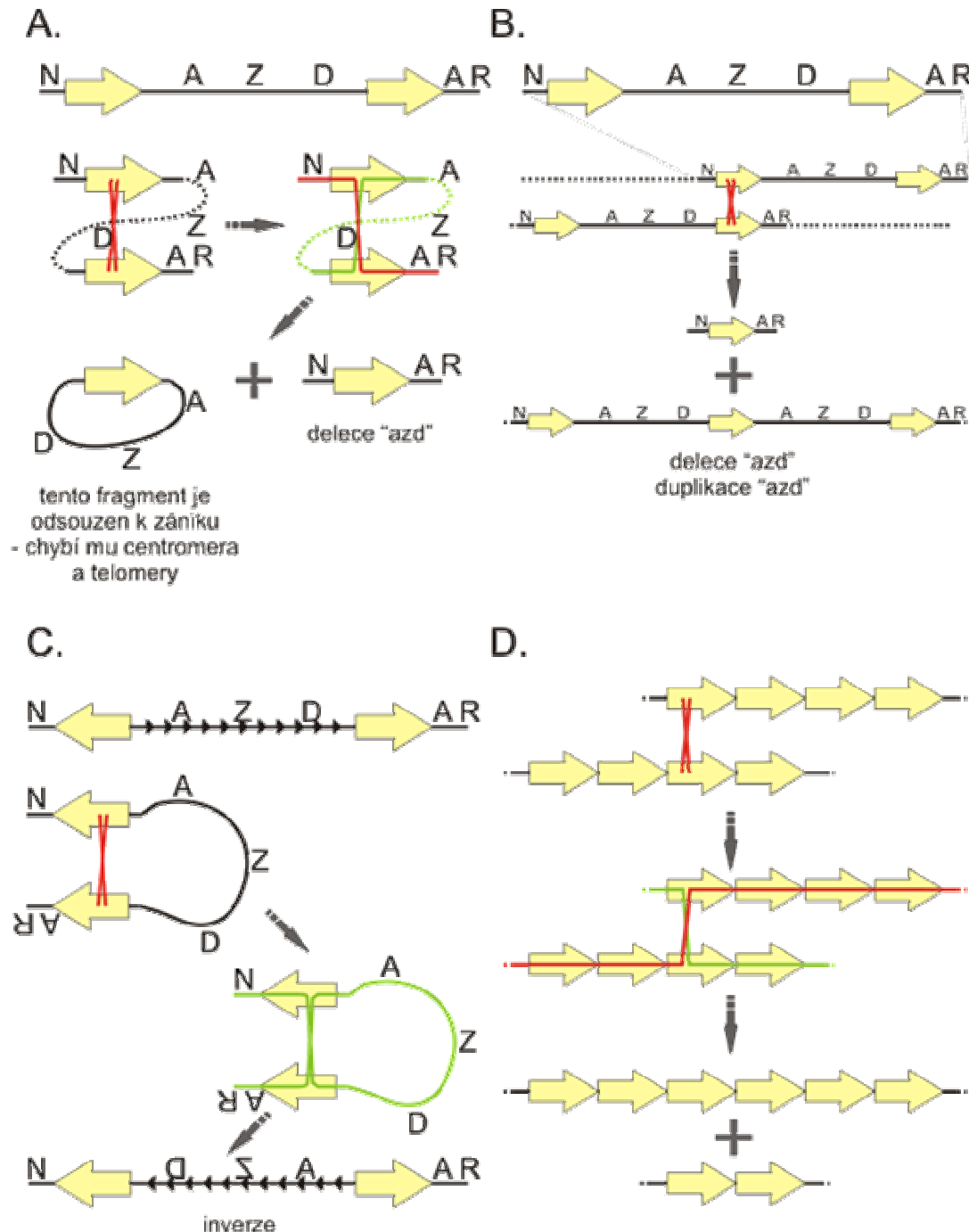
Obr. 5. Repetice podněcují přestavby genomu

A+B: oblast genomu obsahující přímé repetice (ve stejném směru, stejná sekvence je na stejném řetězci DNA). Tyto dvě repetice se mohou párovat a rekombinovat.

Intrachromozomální rekombinace (A) vede k delecí. Hypotetický kruhový fragment je ztracen - nemá centromeru. Nerovnoměrný crossing-over s interchromozomální rekombinací (B) způsobuje delecí a duplikaci.

C: Intrachromozomální rekombinace mezi dvěma invertovanými repeticemi (v opačném směru, stejná sekvence DNA je v opačných řetězcích DNA) vede k inverzi mezilehlé sekvence DNA. Funkční následky takových přestaveb mohou být různé, od němých změn až po letální, jak může být očekáváno.

D: Polymorfismy v tandemových repeticích mohou vznikat mechanismem nerovnoměrného crossing-overu.

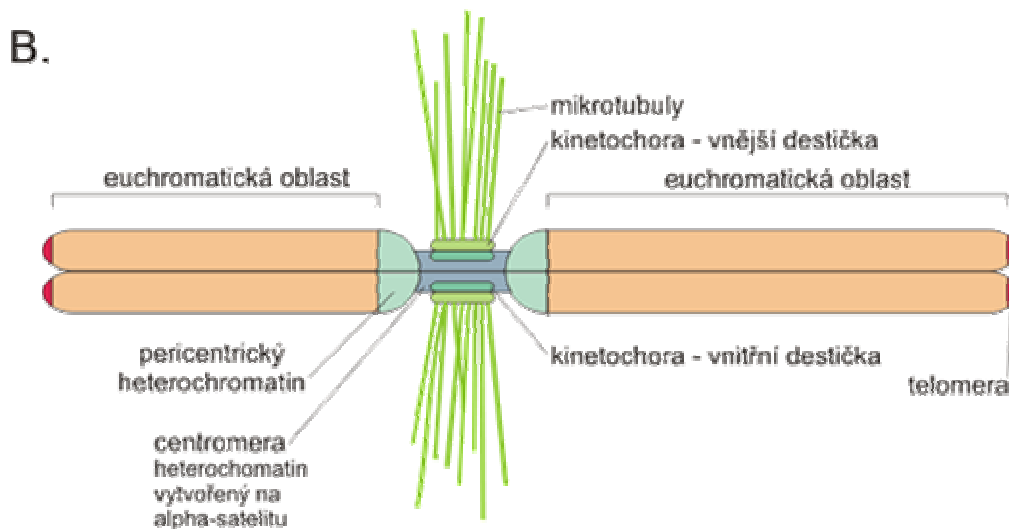
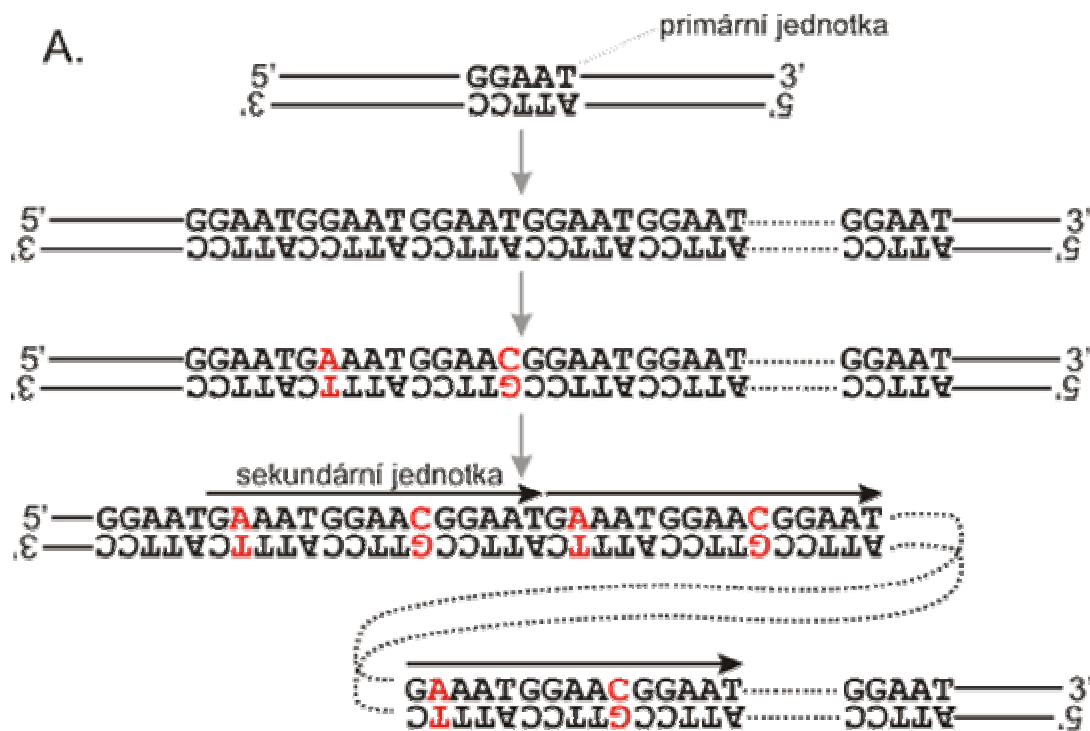


Obr. 6. Satelity

A: primární a sekundární jednotky tandemových repeticí. Pravděpodobná "evoluční historie" repetice na příkladu sekvence GGAAT. Tato sekvence se multiplikuje a vytvoří tak dokonalou monotónní repetici. Některé páry bazí později podlehnou mutaci (červeně) čímž vznikne

nedokonalá (degenerovaná) repetice. Následně dojde k další multiplikaci této repetice, nyní se však několik degenerovaných jednotek kopíruje jako jednotka jediná, a tak se vytváří dokonalá repetice této delší, sekundární jednotky (šipka). Sekvence GGAAT je základem lidských satelitů 2 a 3. Tyto satelity se liší právě sekundárními jednotkami.

B: Struktura lidského mitotického chromozomu s přihlédnutím k satelitním sekvencím. Alfa-satelit tvoří heterochromatin centrální části centromery. Mimo proteiny asociované s heterochromatinem se na centromeru váží "proteiny vážící se na alfa satelit" které vytvářejí "vnitřní kinetochorovou destičku". Některé z těchto proteinů jsou asociovány s centromerou v průběhu celého buněčného cyklu. Na "vnitřní kinetochorové destičce" se sestavuje "vnější kinetochorová destička" která interaguje s mikrotubuly dělicího vřeténka. Centromera je často obklopena pericentromerickým heterochromatinem, který je tvořen jinými typy satelitních sekvencí. Konce chromozomu (telomery) jsou tvořeny telomerickou repeticí TTAGGG, navazující úseky subtelomerických oblastí jsou také vysoce repetitivní.



Obr. 7. VNTR v genu pro inzulín

A: Úsek DNA (kódující řetězec, směr od 5' ke 3' konci) obsahující gen pro inzulín. Gen pro inzulín obsahuje tři exony (velká písmena) které tvoří zralou mRNA. Důležité regulační sekvence jsou červeně - TATA box před místem zahájení transkripce, ATG jako začátek translace (transkribován jako AUG do mRNA, kde slouží jako iniciační kodon, zařazuje první zbytek methioninu vznikajícího polypeptidového řetězce), konzervované dinukleotidy GT a AG v intronech v místech sestřihu, stop kodon TAG a polyadenylační signál AATAAA. Místa polymorfismů jednoho nukleotidu (SNP) jsou tučně (to znamená, že někteří jedinci mají na daném místě jiný nukleotid než vyznačený). Minisatelit je modře, pochopitelně je zobrazena jen jedna alela, ostatní se liší počtem opakování jednotky repetice. **B:** Tato alela

VNTR je složena z 29 opakování motivu GGGGTGTGGGACA, ačkoli všechna opakování nejsou zcela shodná s konsenzem (odlišné baze jsou černě). Všimněte si, že repetice obsahuje palindrom TGTnnnnACA, který může stabilizovat vlásenkovité struktury a tím podporovat nestabilitu množství opakování (viz obr. 8). Variabilní délka minisatelitu v oblasti promotoru může diferenciálně interagovat s transkripčními faktory vážícími se na promotor a vést tím k diferenciální expresi inzulinu. Skutečně byly některé alely VNTR asociovány s vývojem diabetu (je nicméně velmi obtížné odlišit přímý účinek od "pouhé" vazby - viz kapitola o vazbě pojednávající).

A.

```
ccaccctggggagctgagggcctcagctgggggtgctgtcctaaggcagggtgggaactag
gcagccagcaggaggggaccctcctcactcccactctcccaccccacacccttgccc
catccatggcggcatctctgggcatccgggactggggacaggggtcctggggacaggggtg
tggggacaggggtcctggggacaggggtctggggacaggggtctggggacaggggtgtg
ggacaggggtgtggggacaggggtgtggggacaggggtcctggggacaggggtctggggac
aggggtctgaggacaggggtgtggggacaggggtgtggggacaggggtgtggggacagggg
tgtggggacaggggtctggggacaggggtccggggacaggggtgtggggacaggggtgtg
ggggacaggggtgtggggacaggggtctggggacaggggtgtggggacaggggtctggggac
aggggtgtggggacaggggtgtggggacaggggtgtggggacaggggtgtggggacaggg
gtctggggacagcaggcgaagagcccccgcctgcagcctccagctctcctggtctaatgt
ggaaagtggccagggtgagggctttgctctcctggagacatttgcctccagctgtgagcag
ggacaggtctggccacccgggcccctggttaagactttaatgacccgctggtcctgaggaag
aggtgctgacgacccaaggagatcttcccacagaccagcaccagggaatgggtccggaaat
tgcagcctcagccccagccatctgcccagccccccccagggcctaatggccaggcg
gcaggggttgaggaggtaggggagatgggctctgagactatagggcagccagggggcccagca
gcccctcAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGGCTCTGTTCRAAGG
GCCTTTGCGCTCAGgtgggctcaggatctccagggtggctgggaccccagggccagctctgca
gcagggaggagctgggtgggctcgtgaagcctgtgggggtgagcccccaggggcccagga
gggacactggccttcagcctgcctcagccctgcctgtctccagATCACTGTCTTCTGCC
ATGGCCCTGTGGATGGCCCTCCTGCCCTGTGGGCGCTGTGGCCCTCTGGGGACCTGACC
CAGCCCGAGCCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAAGCTCTCTACCT
AGTGTGCGGGGAACGAGGCTTCTTCTACACACCCAAAGACCCGCCGGGAGGSCAGAGGACCTG
CAGGgtgagccaacctgcccactgtgctgcccctggccgccccagccaccccctgctcctggc
gctcccacccagcatgggcagaagggggcaggagggtgcccacccagcaggggggtcagggtgc
acttttttaaaaagaagtctcttgggtcacttcttaaaagtgcaccagctcctctggtgccc
gtcagaatctcagcctgaggacgggtgttggcttcggcagccccagagatacctcagaggggtg
ggcaccgctcctcctccactcggcccctcaaacaaaatgccccgagcccatttctccaccct
caatttgatgaccgcagatccaagtgttttggtaagtaagtcctgggtgacctgggggtcac
aggggtgccccacgctgcctgcctctggggcgaacaccccctcaecgcccggaggagggggtgg
ctgcccctgagtgaggccagaccccctgtgcccaggcctcagggcagctccatagtcagga
gatggggaaagatgctggggacagggccctggggagaagtactgggatcaactgttcaggtc
ccactgtgacgctgccccggggcgggggaagggaggtgggacatgtgggctgtggggcctgt
aggtccacacccaggtgtgggtgacccctcctctaacctgggtccagcccgggtggagatgg
gtgggagtgccgacctagggctggggggcaggcgggcaactgtgtctcctgactgtgtctctc
ctgtgtccctctgcctcggcctgtgttcgggaacctgtctgcggcagctcctggcagTG
GGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGCCCTGCAGCCCTTGGCCCTGGAGG
GGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGCTGTACCAGCATCTGCTCCTCTACCA
GCTGGAGAACTACTGCAACTAGACGCAGCCCGCAGGCAGCCCCACACCCGCCGCCCTCTGC
ACCGAGAGAGATGGAAATAAGCCCTTGAACCAAGCccctgctgtgcccctctgtgtctctggg
ggccctgggccaagcccaacttcccggcaactgtgtgagccctccagctctctccacgc
tctctgggtgcccacaggtgccaacgcccggccagggcccagcatgagtgagctctcccdaa
gcccacatgctctgggtgctctgcccccaactgtgggtcagggteaggtatggg
```

B.

```
ggg-actggggaca
ggggtcctggggaca
ggggtgtggggaca
ggggtcctggggaca
ggggtctggggaca
ggggtcctggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtcctggggaca
ggggtctggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtcctggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtcctggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtgtggggaca
ggggtctggggaca
```

Obr. 8. Polymorphismy mikrosatelitů mohou vzniknout "skluzem" DNA polymerázy
 Během polymerace může dojít díky termální fluktuaci k disociaci řetězců DNA. Následná reasociace řetězců je obvykle přesná a nedochází k žádné změně. Nicméně občas se mohou řetězce spárovat díky repetitivní sekvenci nepřesně. Buď pak vznikne klička na polymerovaném řetězci, což může vést k expanzi repetice (to je častější), nebo se prodlužující řetězec váže na templát více distálně (vznikne klička na templátovém řetězci) s následným zmenšením repetice. **Vložený obrázek:** Některé repetice mohou tento pochod podporovat stabilizací přechodového stavu tvorbou vlásenkovité struktury z neúplné dvojité šroubovice, což platí zejména pro CAG/CTG repetici, která je podkladem několika onemocnění z expanze trinukleotidových repetic. Čím delší je mikrosatelit, tím vyšší je pravděpodobnost skluzu, což

vytváří spolu s vyšší tendencí k prodloužení repetice pozitivní zpětnou vazbu.

