

Molekulární biologie člověka :

Mutace a nestability v lidském genomu

Dynamické mutace, repetitivní DNA

Multigenové rodiny

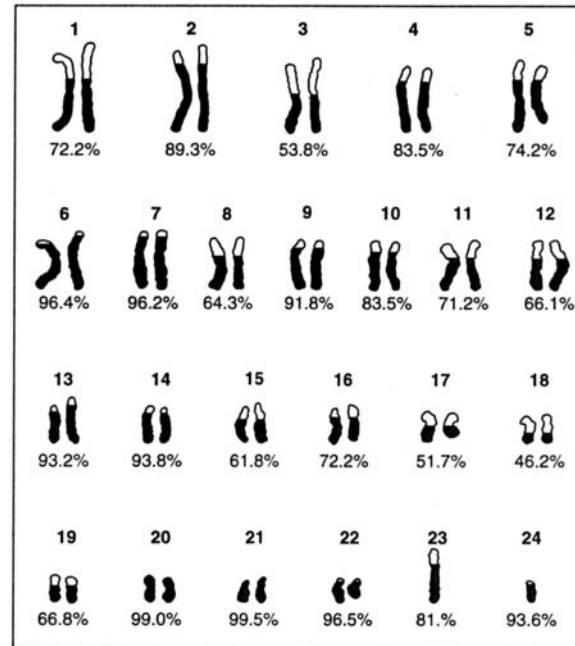
Marie Vojtíšková

OGMB, MU Brno 2007

Lidský genom

- kompletní sekvence 3miliard párů bází
- 20 000 – 30 000 genů jaderné DNA
- 37 genů mitochondriální DNA (10^3 kopií)
- poznání sekvence LG
- geny obsahují cca 3% celkové DNA
- (TESIMESENAVASISPOLUPRACI)
- Aplikace studia lidského genomu (LG)

% kompletní
sekvence LG
vzhledem k
jednotlivým
chromosomům



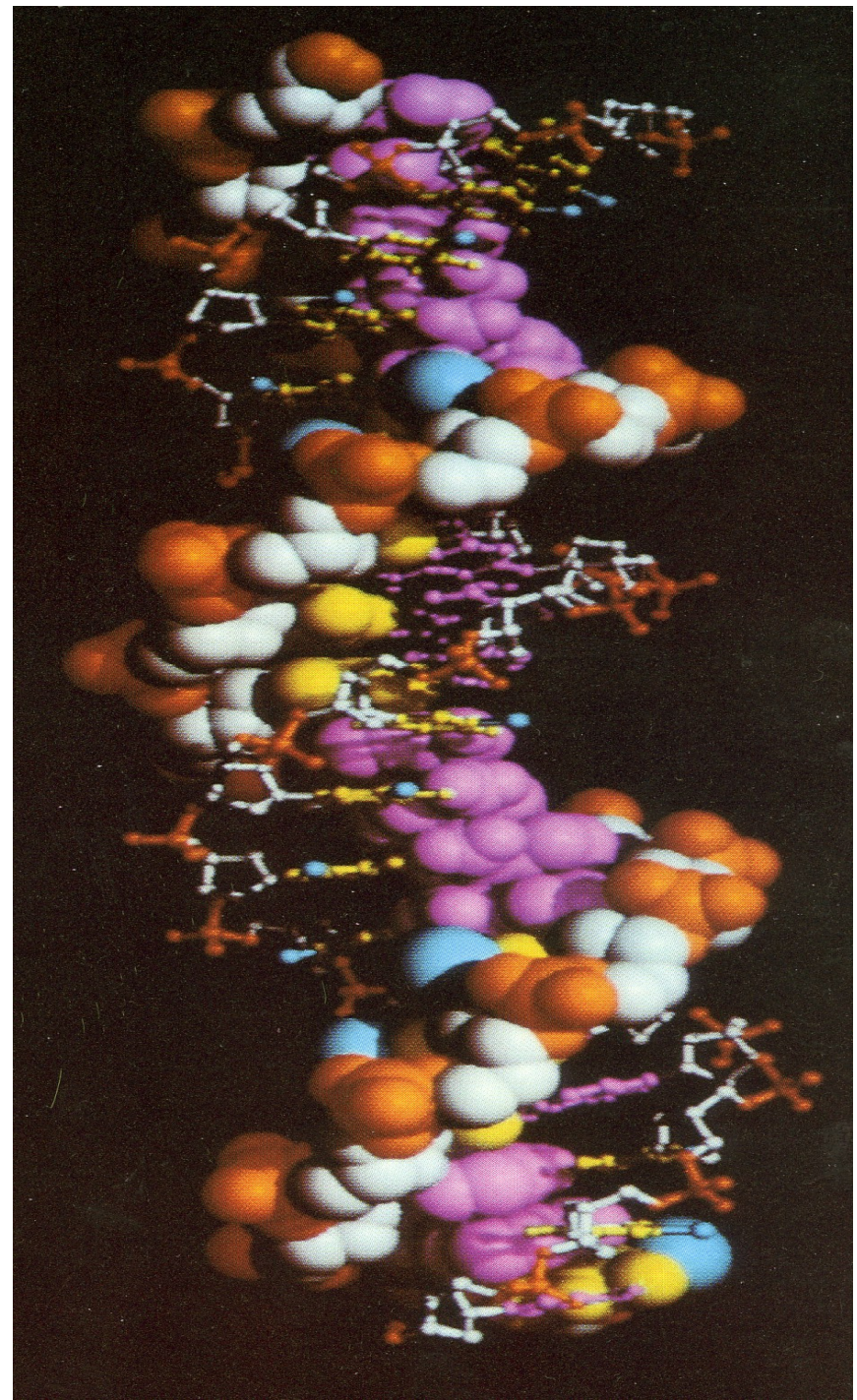
Science 296, 2002,
1600

DNA

uchování genetické informace

Posloupnost čtyř bází

A, G, T, C



Reasociační kinetiky lidské DNA prokázaly tři skupiny pakujících se sekvencí:

60 % velmi nízký počet kopií

30% středně repetitivní

10% vysoce repetitivní

Metody:

DNA sekvencování

DNA hybridizace

PCR

Základní třídy tandemových repeticí v lidské DNA

Satelitní DNA (bloky od 100 kb do několika Mb)

Satelity 2,3 – velikost jednotky repetice 5 pb. Lokalizace většinou na všech chromozomech

Satelity 1 (AT bohaté oblasti) o velikosti 25 – 48 bp, především lokalizovaných v centromerických a jiných heterochromatinových oblastech.

Alfa (alfoidní DNA) o velikosti repetitivní jednotky 171 pb s výskytem v centromerách všech chromozomů.

Beta (Sau3A rodina) s jednotkou 68 pb, zvláště se nacházející v centromerickém heterochromatinu chromosomů 1,9,13,14,15,21,22 a Y.

Minisatelitní DNA (bloky od 0.1 do 20kb)



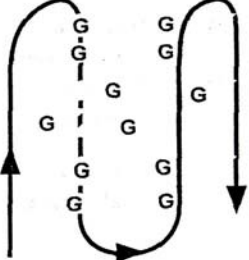

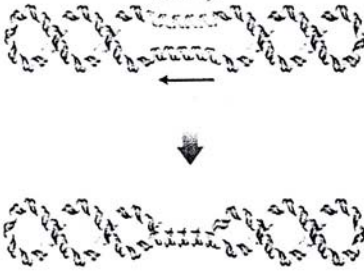
Telomerické sekvence T_2AG_3 , na koncích chromozomů
Hypervariabilní sekvence v blízkosti telomer

Microsatelitní DNA (bloky méně než 150 pb)

Krátké repetitivní sekvence např. typu $(CA)_n$, výskyt po celém genomu

non B
struktury
DNA

Závislost na
sekvenci
repetice

Structure	Conformation	General Seq. Requirements	Sequence
Slipped (Hairpin) Structure		Direct Repeats	<pre> →→→ CNGCNGCNG GNCGNCGNC →→→ </pre>
DNA Unwinding Element		A-T rich Regions	<pre> ATTCTATTCT TAAGATAAGA </pre>
Tetraplex		Single-Strand Oligo (G) _n Tracts	<pre> CGGCGGGCGG GCCGCCGCC </pre>
Triplex		(R•Y) _n Mirror Repeats	<pre> GAAGA AGAAG CTTCT TCTTC </pre>
Sticky DNA		2 G-A rich Tracts Direct Repeats	<pre> GAAGAAGAA CTTCTTCTT </pre>

Mutace LG- změny v genotypu

Četnost spontánních mutací nízká, pravděpodobnost 10^{-7}

- **Normální karyotyp člověka 46 chromosomů (XX/XY)/buňka (gamety 23 chromosomů)**
- **Mutace chromosomové (duplikace, delece, inverse, translokace, fragmentace, isochromosom, ring chromosom)**
- **Mutace genomové, polyploidie**
- **Mutace v genech, informace v genech, mutace na úrovni nukleotidů (inzerce, delece, substituce)**
- **dynamické mutace v závislosti na sekvenci**
- **buňky zárodečná linie - vrozené mutace**
- **maligní linie buněk – získané a neopravené mutace, vedoucí k neoplastické transformaci**
- **mRNA reflektuje exprese genů fyziol./patol.**
- **protein / funkce - fyziol. /patol.**

Poškození DNA na úrovni somatických buněk:

Intracelulární

- inkorporace chybného nukleotidu při replikaci
- náhodné chemické změny DNA v buňce
- reaktivní produkty metabolismu, kyslíkové radikály
- viry- inkorporace do genomu infikové buňky
- mobilní elementy (transpozony a retrotranspozony)

Extracelulární

- chemické látky, mutagenní vlivy vnějšího prostředí
- ionizační záření, UV záření

Cílené

- terapeutické působení cytostatik

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 repeticích, od nich odlišujeme repetitivní sekvence rozptýlené v genomu jako jednotlivé kopie (rozptýlené repetice - anglicky interspersed repeats).

Podstata rozptýlených repetic - transpozony

Většina rozptýlených repetic 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

repetic se schopností transpozice : LINE (long interspersed repeats), př. L1;

LTR-retrotranspozon; 1,2kb -gen pro transpozásu je obklopen ITR, duplikace cílového místa; SINE –(short interspread repeat) př. 282 bp dlouhý Alu element

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-lymfocytu) 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 **tvorí 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 retrotranspozon, **proces retrotranspozice je náchylný k různorodým chybám**, a tak jsou nově vzniklé kopie většinou **inaktivovány delecemi 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 semohou 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 retrotranspozonu (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 delecí kódující části retroviru .

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 kohezni 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 bází.

Reverzní transkriptáza je většinou neschopna ukončit syntézu prvního řetězce, což vede ke zkrácení nové kopie na 5' konci.

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, zvyšujíc 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 restrikčních míst pro endonukleázu AluI). **Více než 1 milión 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 "ocas" 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.

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.**

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 .

Protože má LINE-1 retrotranspozon relativně slabý polyadenylační signál, stává se, že RNA polymeráza II se skrz něj pročte, 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.

Retrotransposice L1 může dokonce vyústit v delece a inverze.

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.

Zřídka však může procesovaný pseudogen přijmout novou funkci pod selekčním tlakem. Velmi známý příklad je gen pro podjednotku E1alfa 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 crosing-over nebo intrachromozomovou rekombinaci

V neposlední řadě se uvažuje 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 nepotvrzené.**

Tandemové repetice

Tandemové repetice jsou tvořeny za sebou jdoucími 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 přídavné "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 restrikčních 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.

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 retrotranspozónů). Nicméně se mohou **telomery elongovat i pasivně, mechanismem nerovnoměrného crossing-overu, např. v nádorových buňkách**. (Výzkum struktury telomer - tetraplexy)

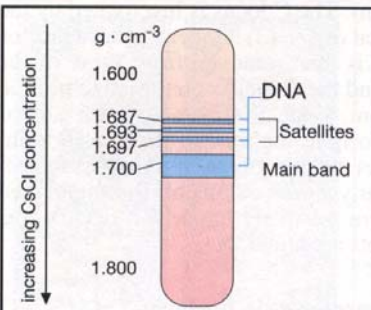
Možná by stálo za to na tomto místě znovu připomenout, že sekvence lidského genomu zahrnuje euchromatické úseky, ohraničené proximálně (ale nezahrnující) centromerou a pericentromerickým heterochromatinem a distálně telomery, které také, spolu se subtelomerickými oblastmi, nejsou obsaženy

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 pro genetickém mapování)



1. Cesium chloride density gradient



2. Classic satellite DNA (100-6500 bp repeats)

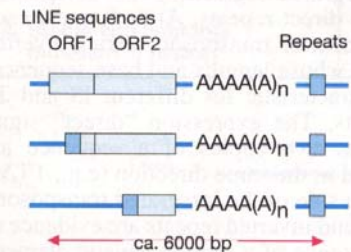


3. Minisatellite DNA (20-100 bp repeats)

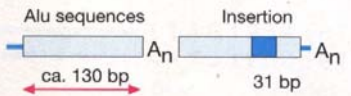
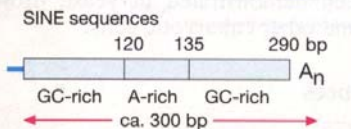


4. Microsatellite DNA (CA)_n repeats (n = 2-10 bp)

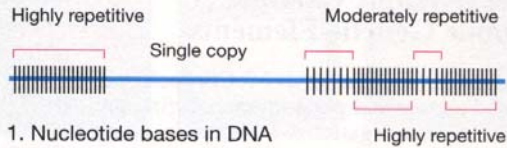
A. Satellite DNA



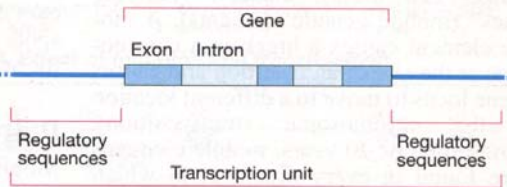
B. Long interspersed repeat sequences



C. Short interspersed repeat sequences



1. Nucleotide bases in DNA



2. Singular DNA sequences

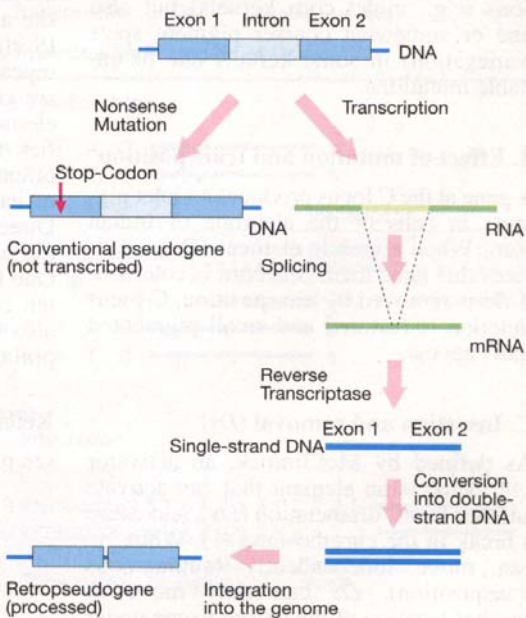
D. Repetitive and singular sequences



1. Gene cluster on the same chromosome

2. Multigene families on different chromosomes

E. Gene cluster and multigene families



F. Pseudogenes

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) s výsledným progresivním odumíráním neuronů v nucleus caudatus.

U jiného onemocnění, myotonické dytrofie (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.

**Nový mutační mechanismus - expanze
trinukleotidů -**

**Genomová nestabilita spojená s opakováním
trinukleotidů, (tetra -, penta.. ...dodekanukleotidů)**

**Počet dosud známých chorob spojený s expanzí
cca 15**

Rozdílná fyziologická a patologická hladina počtu opakování

Společný prvek :

na příslušném místě příslušného lokusu existuje tandemová repetice tripletu (tetra .. penta ---) bází, přičemž zvýšení aktuálního počtu repeticí vede ke vzniku choroby

patologie spojené s expanzí trinukleotidů:

a) *dědičné*

b) genetická *nestabilita v somatických nádorových buňkách* (poruchy reparačních mechanismů - NER, reparace chybného párování

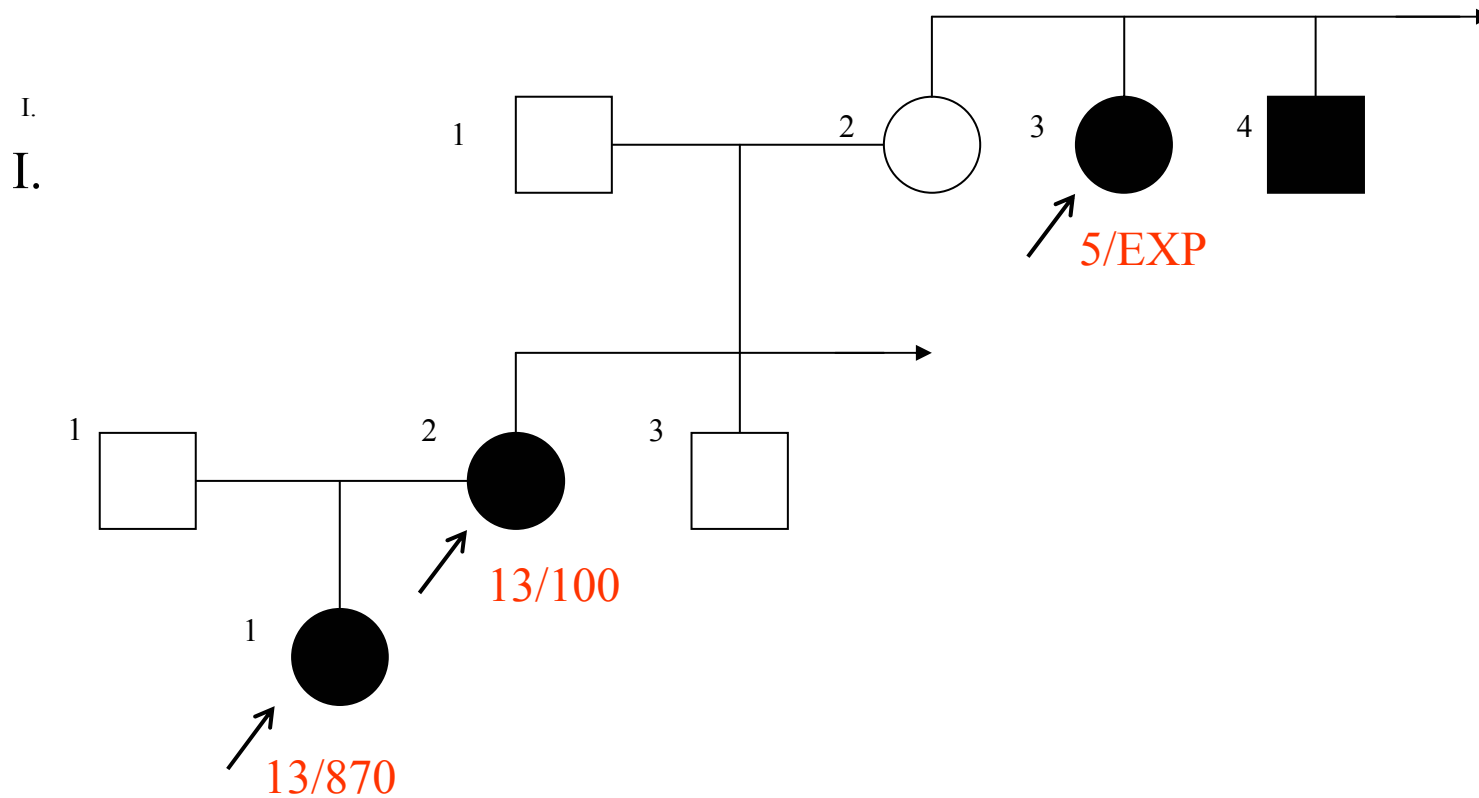
c) *genetická nestabilita v somatických buňkách v závislosti na věku*

Charakteristický rys dynamických mutací -

prodlužování nestabilní repetice během mezigeneračních transmisí

Exprese patologie dynamických mutací pouze u člověka

Unstable CTG repeats in the DM family



Family pedigree. Numbers of CTG repeats on both alleles are shown. EXP denoted long range pathological CTG repeats determined by TP-PCR. Individuals III/1, I/3,4 are with DM phenotype. Mother II/2 is healthy. Arrows represented investigated member of the family.

Molekulární příčiny genetické nestability dynamických mutací:

Trinukleotidové repetice - specifická podskupina STR-mikrosatelitů, vysoká frekvence (1/300- 500 kb) v **lidském genomu**

Sekvenční homogenost, symetrie repetitivních úseků - tandemový motiv $[NNN]_n$, n = počet opakování

tvorba non B struktur dvoušroubovice DNA:

např. vlásenky (hairpins),

lokální posun (klouzání) řetězců

Hoogsteenovo párování - triplexy, tetraplexy

Mutace způsobené expanzí trinukleotidových repeticí

Expanze trinukleotidových repeticí představuje nový mutační mechanismus, jehož kauzální role v oblasti výlučně lidských chorob se dotýká stále se rozšiřujícího počtu onemocnění, jejichž společným znakem je primární zasažení nervové tkáně .

Podle typu sekvence trinukleotidů na patologické alele mohou být choroby rozděleny na skupiny s expanzí v kodónu $(CAG)_n$ pro glutamin, choroby polyglutaminového traktu a na skupinu chorob s velkými expanzemi v nekódujících oblastech genů (SyFraX, Myotonická dystrofie, Friedreichova ataxie).

Původ mutací expandujících trinukleotidů dosud není spolehlivě vysvětlen. Bylo prokázáno, že nepřerušovaná sekvence CAG repetic je více náchylná k expanzi, než sekvence se vsuvkou CAT.

Faktory ovlivňující molekulární podstatu nestability dynamických mutací trinukleotidového opakování

- 1) typ sekvence
- 2) počet opakování
- 3) přerušení
- 4) orientace sekvence ve směru k počátku replikace

Navržený a diskutovaný model expanze/delece trinukleotidových repetací během replikace v závislosti na výskytu pseudosekundárních vlásenkových struktur ve fragmentech DNA bohatých na CGG, CTG a CAG sekvence (R.D. Wells, Nature Gen. 10 (1995) 213)

a, delece je způsobena přeskočením vytvořené vlásenkové struktury na opožděujícím se DNA řetězci DNA polymerázou

b, expanze je způsobena opakující se replikací vytvořené vlásenkové struktury v Okazaki fragmentu

Model chybného párování sklouznutím řetězce (SSM)

během DNA replikace v repetitivních sekvencích může způsobit zejména v dlouhých sekvencích inzerci až expanzi alely vedoucí k patologii nebo delecii v dceřiném řetězci v závislosti na tom, na kterém řetězci se „bublina“ chybného párování nachází.

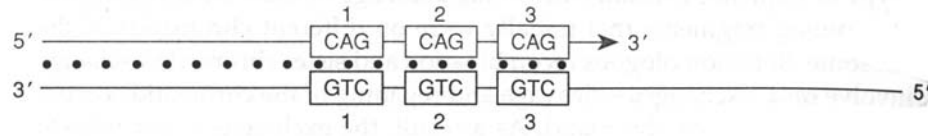
Je-li „bublina“ na dceřiném řetězci, tak dojde k inzerci, jeli na rodičovském, tak dochází k delecii.

Normální alely vykazují určitou variabilitu v počtu opakování repetice ve fyziologickém rozhraní.

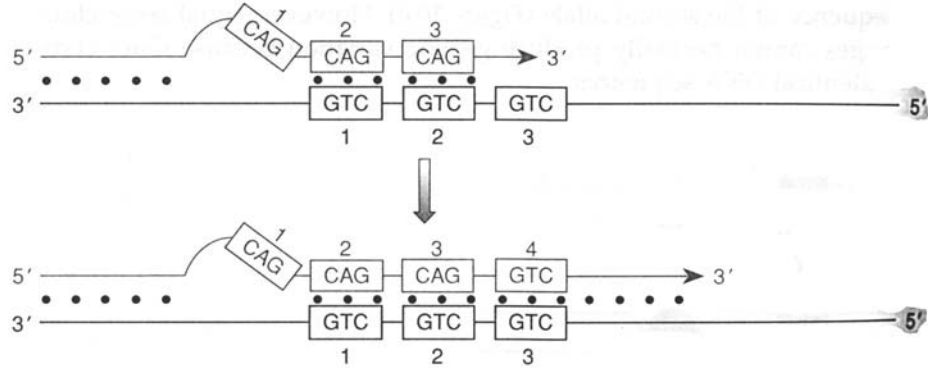
Mutation and instability of human DNA

(A)

Normal replication



Backward slippage causes insertion



Forward slippage causes deletion

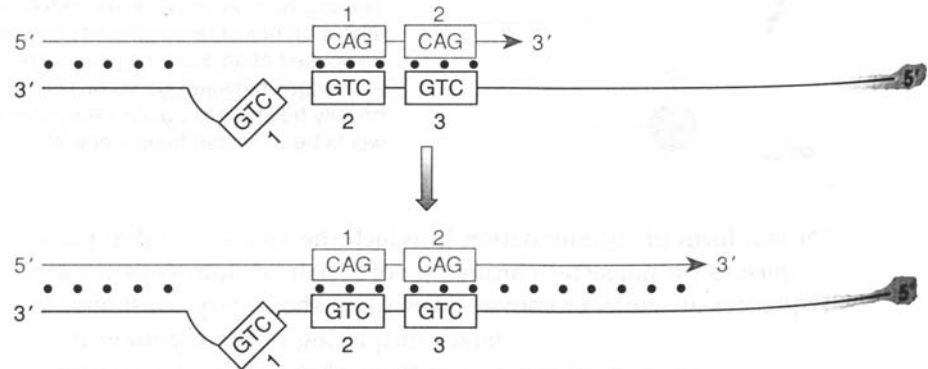
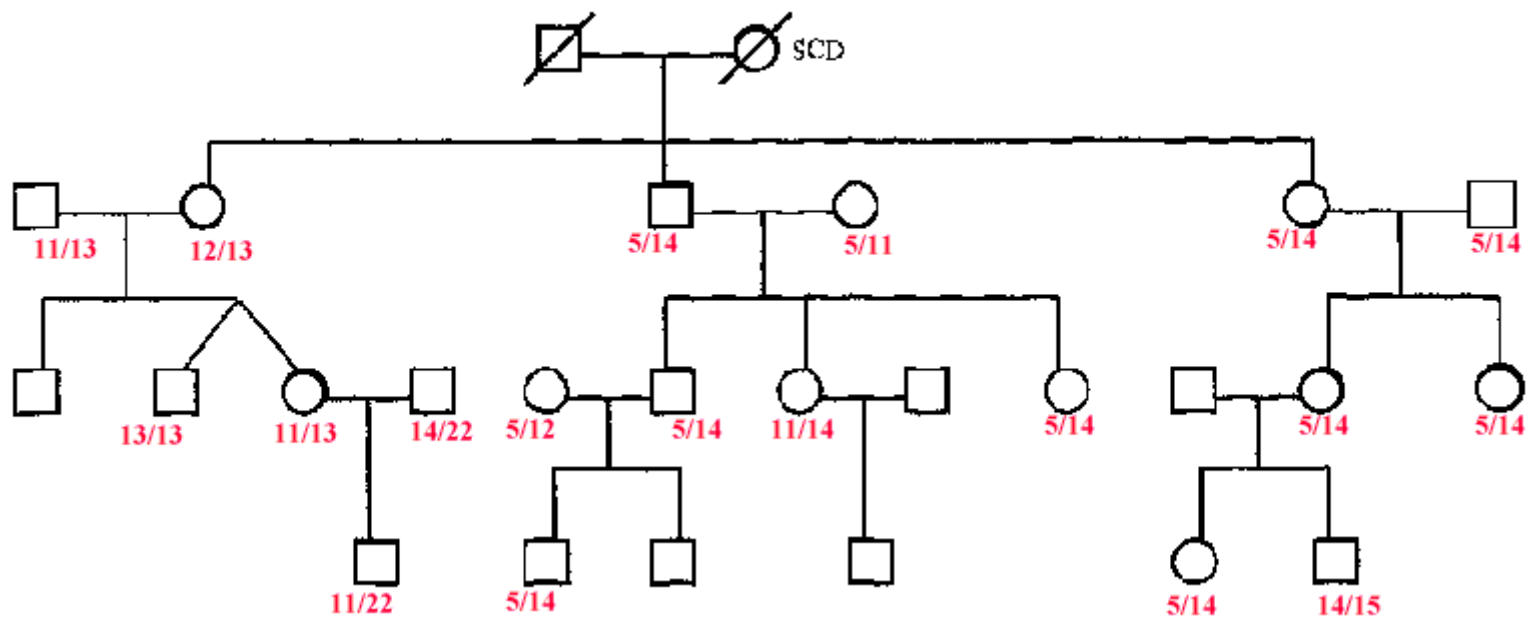


Figure 10.5: Slipped strand mispairing during DNA replication can cause insertions or deletions.

(where $n = \text{age in years} - 15$)

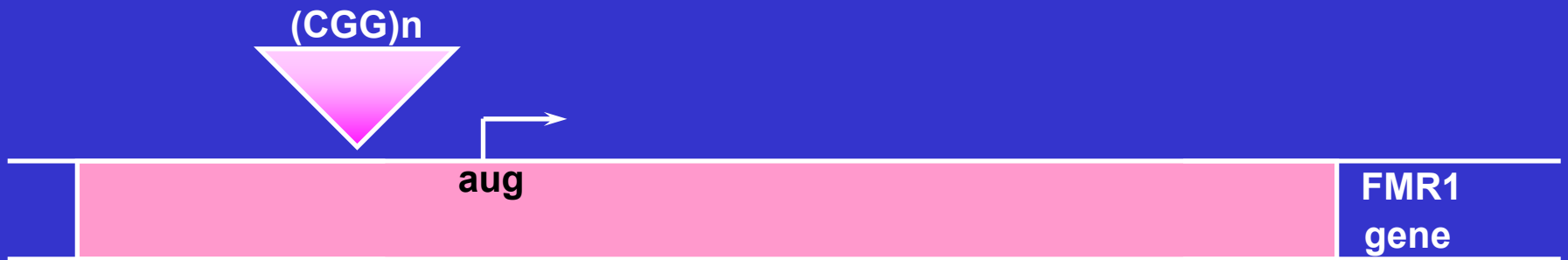


Fyziologické zastoupení alel $(CTG)_n$ v genu pro myotonickou dystrofií (19q13.3)

Table 1. Neurological disorders due to unstable trinucleotide repeats.

<i>Disorders</i>	<i>Transmission mode</i>	<i>Chromosome locus</i>	<i>Implicated gene</i>	<i>Trinucleotide repeat</i>		<i>Normal range</i>	<i>Pre-mutation range</i>	<i>Disease range</i>	<i>Instability</i>
				<i>Type</i>	<i>Localisation</i>				
Fragile X syndrome	Dominant	Xq27.3	FMR-1	CGG	Non coding region	6-54	54-200	200-4000	Maternal
Spinal and bulbar muscular atrophy	Recessive	Xq21.3	Androgen receptor	CAG	Coding region	11-33	Unknown	40-62	Paternal
Myotonic dystrophy	Dominant	19q13.3	Myotonin protein kinase	CTG	Non coding region	5-30	Unknown	45-3000	Paternal maternal for congenital DM
Huntington's disease	Dominant	4p16.3	IT15	CAG	Coding region	11-34	?	37-121	Paternal
Spinocerebellar ataxia type 1	Dominant	6p24	SCA1	CAG	Coding region?	25-36	Unknown	43-81	Paternal
Dentatorubral-pallidolusian atrophy	Dominant	12p12-ter	DRPLA	CAG	Coding region?	7-25	Unknown	49-68	Paternal

FraX



transcription

Absence of FMR 1 protein



Disruption of RNA processing
in brain and testes

DM



↓

Altered expression of RNA
CUG-binding proteins

↓

Altered processing of CUG
containing mRNAs in skeletal
muscle, heart, brain, and testes

Metody pro detekci expandujících tripletů (CTG)_n Algoritmus DNA diagnostiky

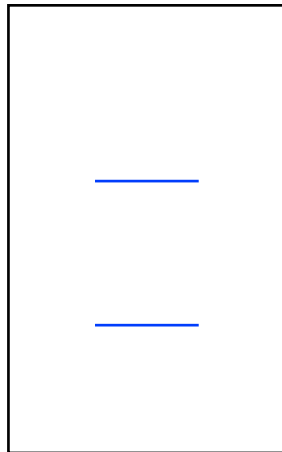
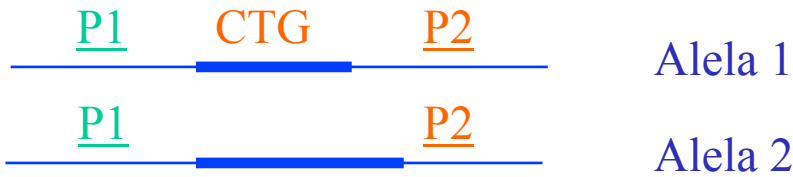
- 1. krok: PCR (P1,P2) alely pro $n < 100$
- 2. krok: TP PCR (P1, P3,P4CTG/P2,P3,P4CAG
• pro $n > 100$, bez omezení
- 3. krok: XL PCR (+ 7-deaza-dGTP) pro $n > 100$ $n = (300-800)$
- Časový faktor: 1 den
- 4. krok: Southern blott/p5B1.4 EcoRI/BglI
• pro $n > 300 - 800$
- Časový faktor: 1 týden

TP PCR

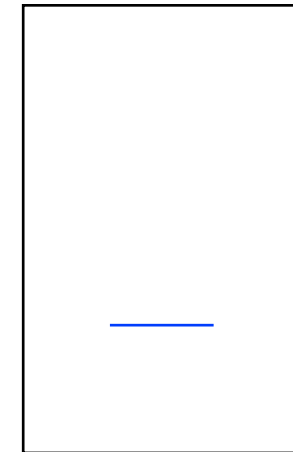
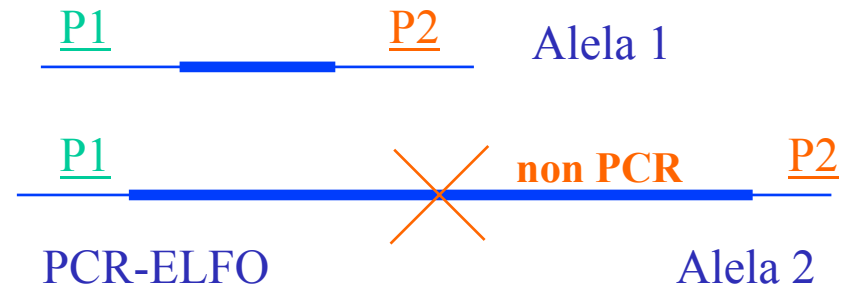
- PCR metoda pro detekci dlouhých opakujících se CAG (CTG) sekvencí
- metoda využívá současně jak specifický značkový primer pro CAG (CTG) repetici tak dvojici primerů syntetizujících z několikanásobných míst v mezích rozsahu opakování
- Výsledkem TP syntézy je žebříček různě dlouhých fragmentů, lišících se o jeden trinukleotid
- Separace a detekce : Genomový analyzátor na principu kapilární elektroforézy a detekce LIF ,
- PAGE a barvení AgNO_3

PCR P1,P2 (fragment 75pb pro n=5)

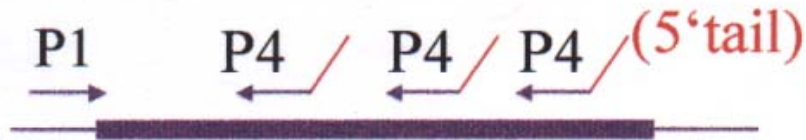
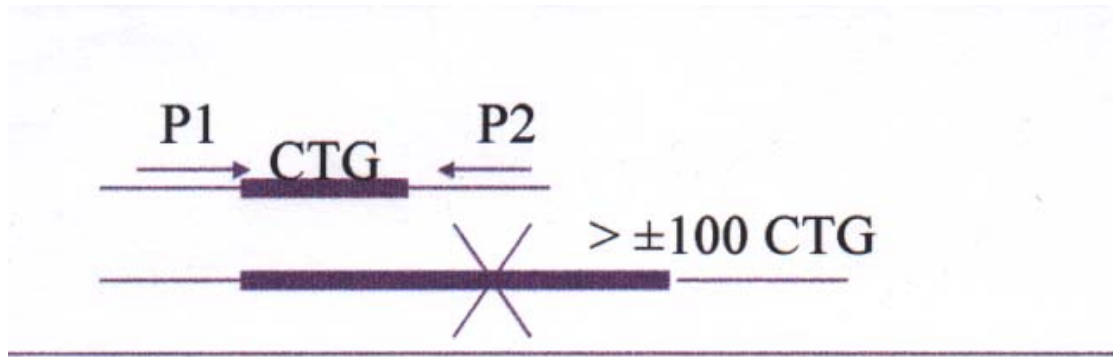
Alela cca 100 CTG



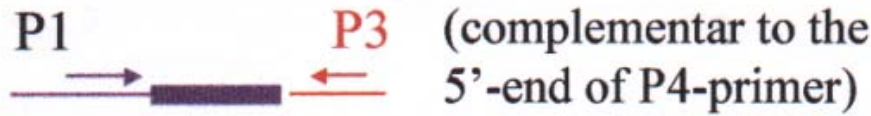
Alela > cca 100 CTG



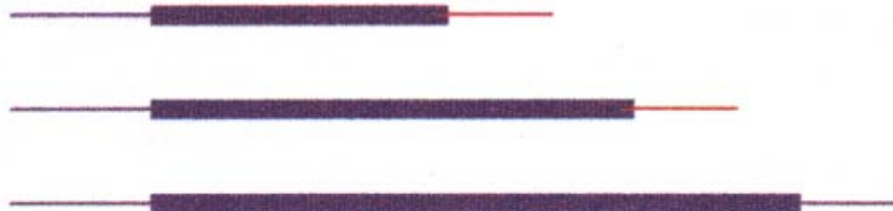
Princip TP PCR

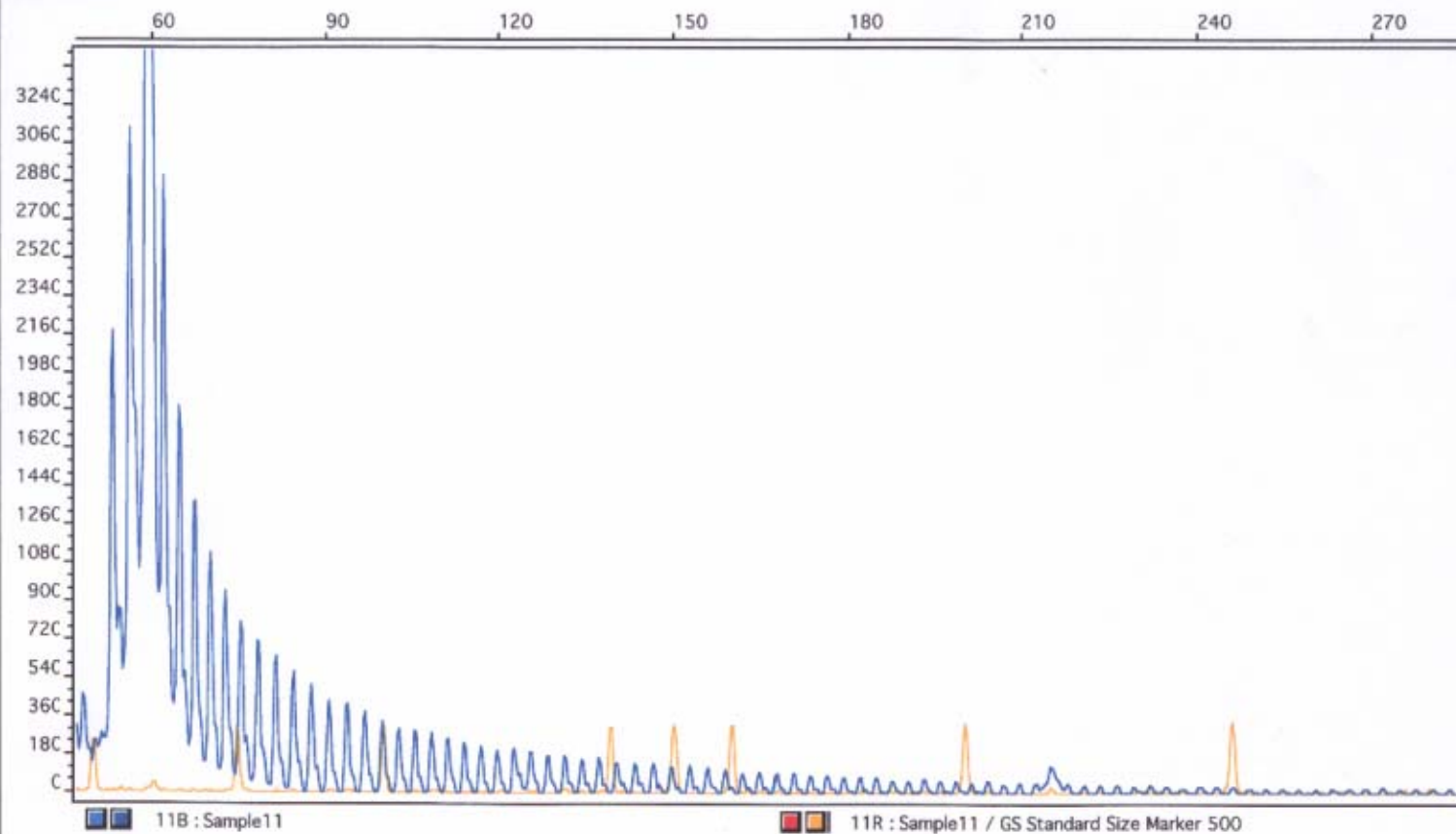


Počáteční hybridizace P4 CTG

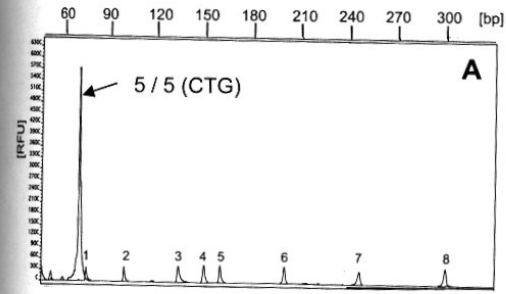


P3 není homologní k LG

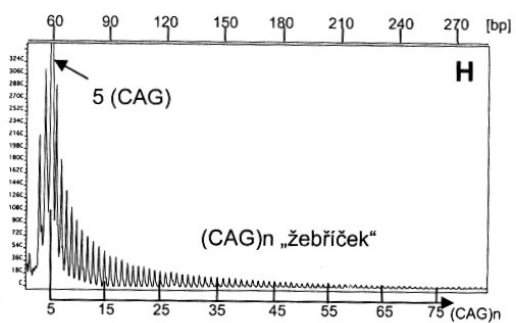
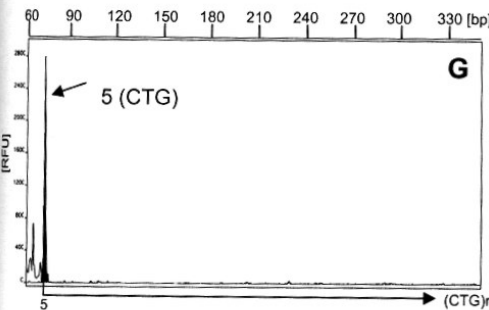
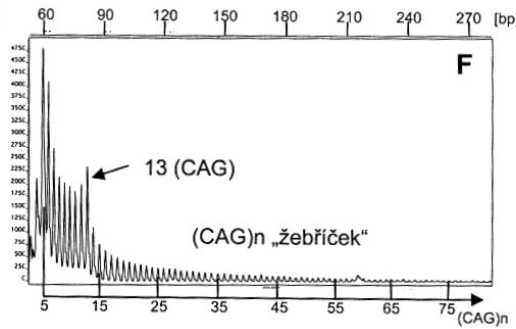
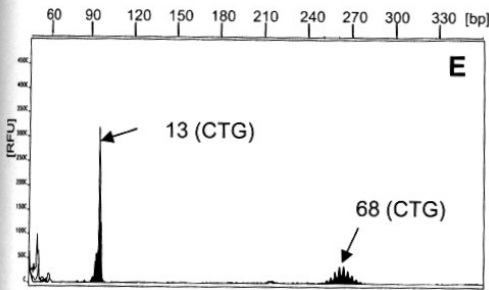
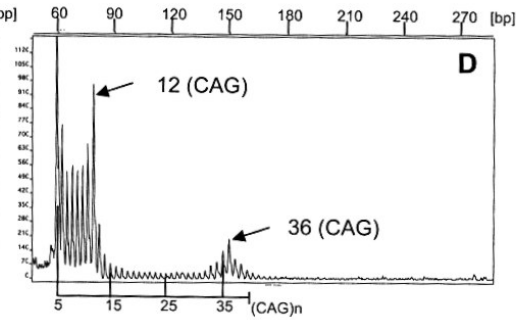
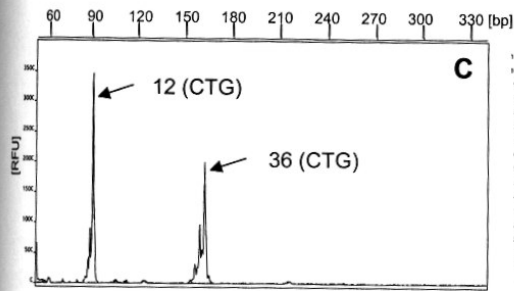
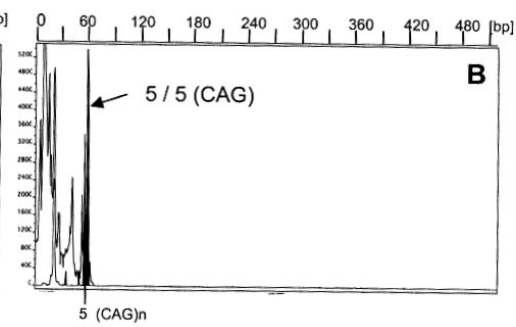




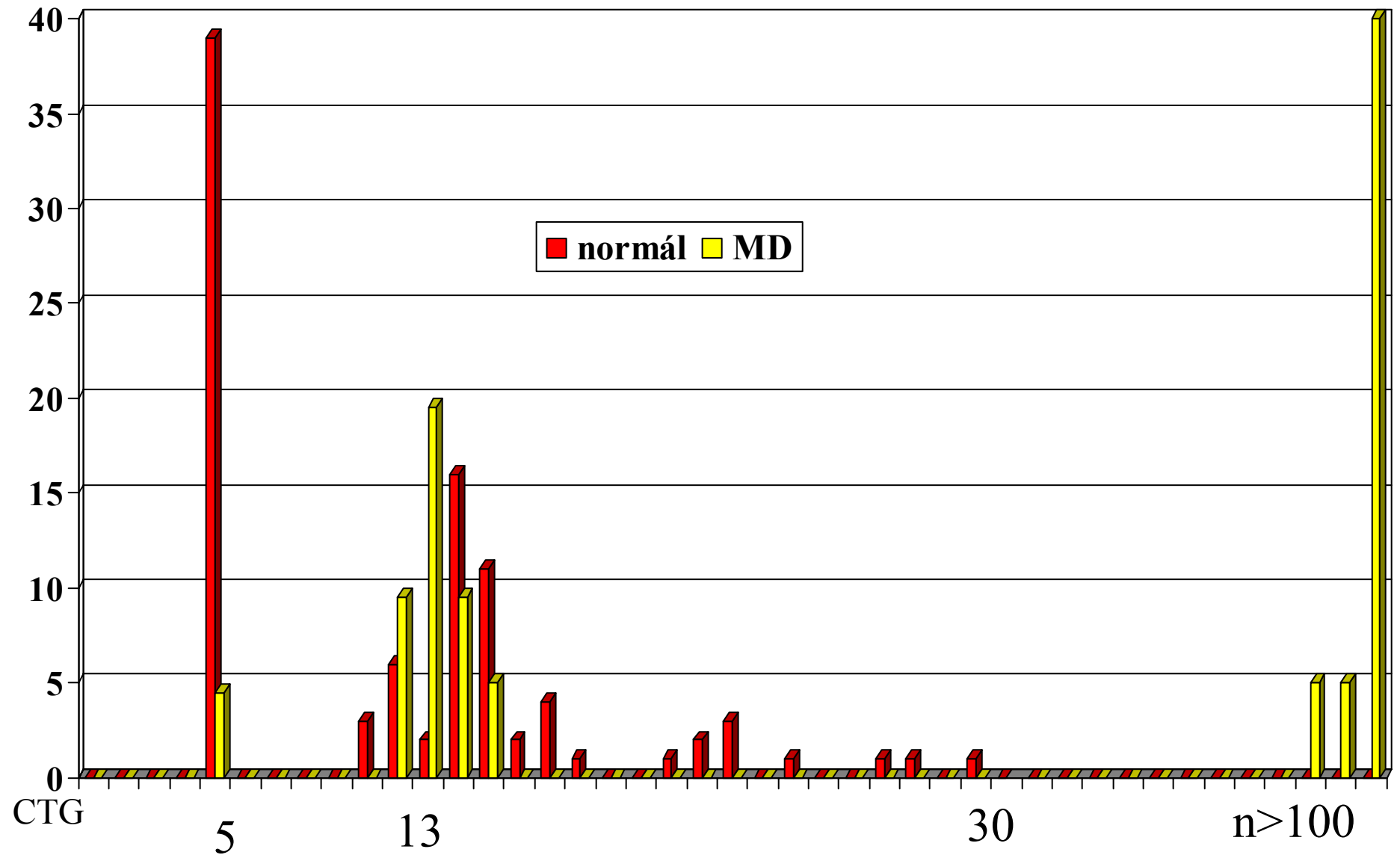
P1/P2-PCR



TP-PCR



Zastoupení alel v souboru vyšetřených jedinců



Myotonická dystrofie typu 2 – DM2

– (3q21), expanze **(CCTG)_n** v intronu 1 (ZNF9)
genu –

protein zinkového prstu

Normální alela – počet opakování tetranukleotidu
do $n = 30$

Patologická alela - expanze $n = (75 - 11\ 000)$
CCTG

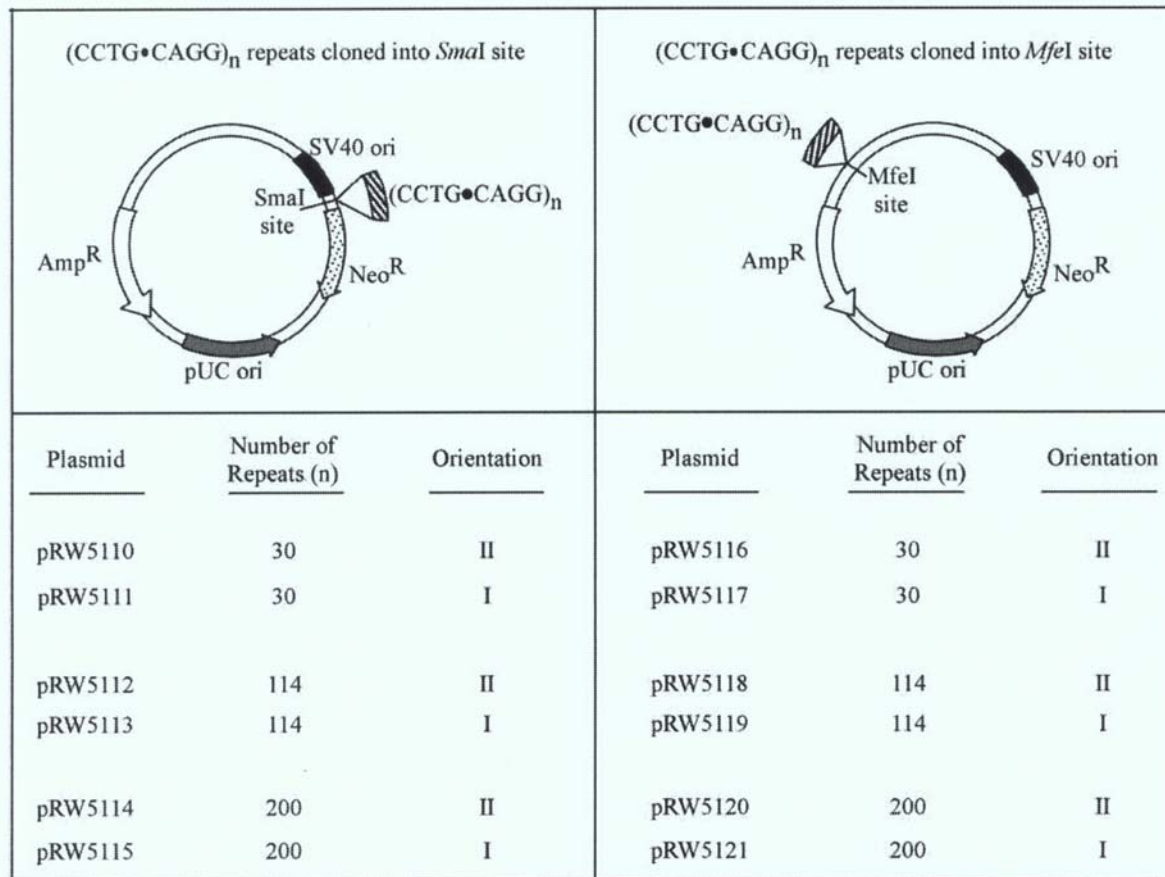
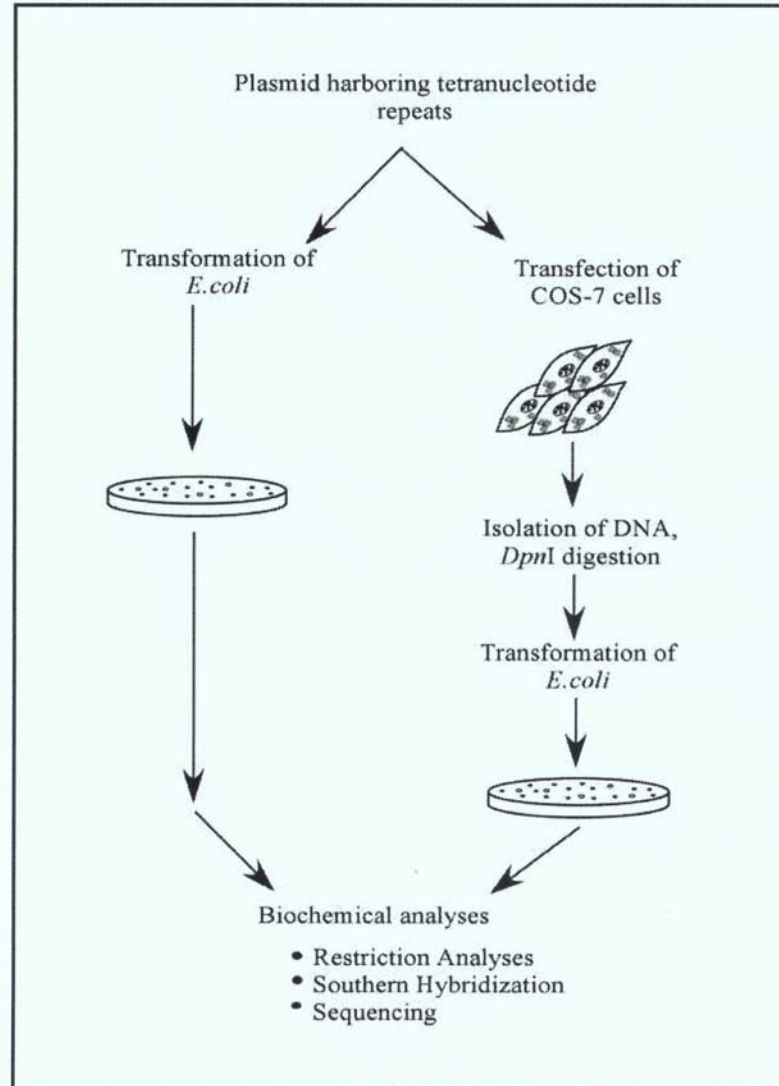


FIG. 1. **Plasmids used in the study.** The (CCTG•CAGG)_n repeats (where n = 30, 114, or 200) were cloned into the SmaI site (proximal to the SV40 origin of replication) or the MfeI site (distal to the SV40 origin of replication) of pcDNA3.1 in both orientations relative to the bidirectional SV40 origin of replication. Orientations I and II are defined under "Experimental Procedures."

FIG. 2. Experimental strategy using the mammalian cell culture assay. The (CCTG-CAGG)_n tracts (where *n* = 30, 114, or 200) were cloned either proximal to the origin of replication in the SmaI site (map position 2078) or distal to the origin of replication in the MfeI site (map position 162) (Fig. 1). These plasmids were then transfected into COS-7 cells, cultured for both 48 h and 2 weeks, and the episomal DNA was isolated using the alkaline lysis method. The episomal DNA was digested with DpnI to fragment the unreplicated DNA. The episomal DNA was transformed into *E. coli* HB101 and individual colonies were analyzed using biochemical analyses. Simultaneously, plasmids that were not replicated in COS-7 cells were also transformed into *E. coli* HB101 and individual colonies were subjected to similar biochemical analyses (see "Experimental Procedures").



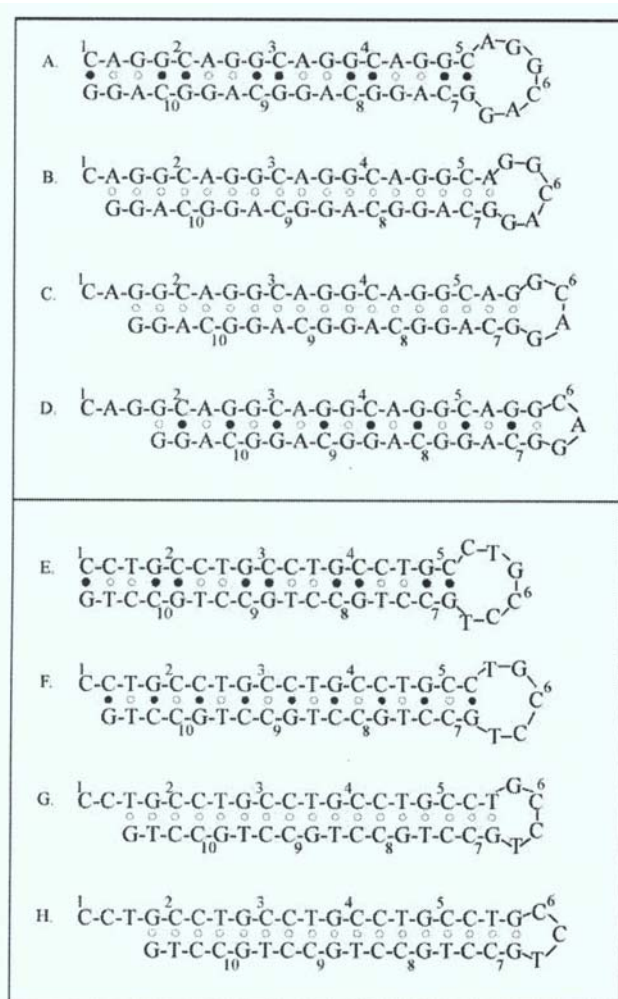


FIG. 7. Theoretical hairpin structures formed by d(CAGG) and d(CCTG) oligonucleotides. A portion of the various folded-back structures that can be formed by the d(CAGG) and d(CCTG) oligomers are shown. Hairpin structures with 6, 5, 4, or 3 residues in the terminal loop formed by slippage and misalignment of 0, 1, 2, or 3 nucleotides are shown for the d(CAGG) (A-D) and d(CCTG) (E-H) oligonucleotides. The filled circles between the two DNA strands indicate Watson-Crick pairing and the open circles denote non-Watson-Crick pairing. Although only 10 CAGG and CCTG repeats are shown for simplicity, the same types of loops and pairing arrangements would apply to oligonucleotides of any length. All oligomers are numbered from their 5'-ends.

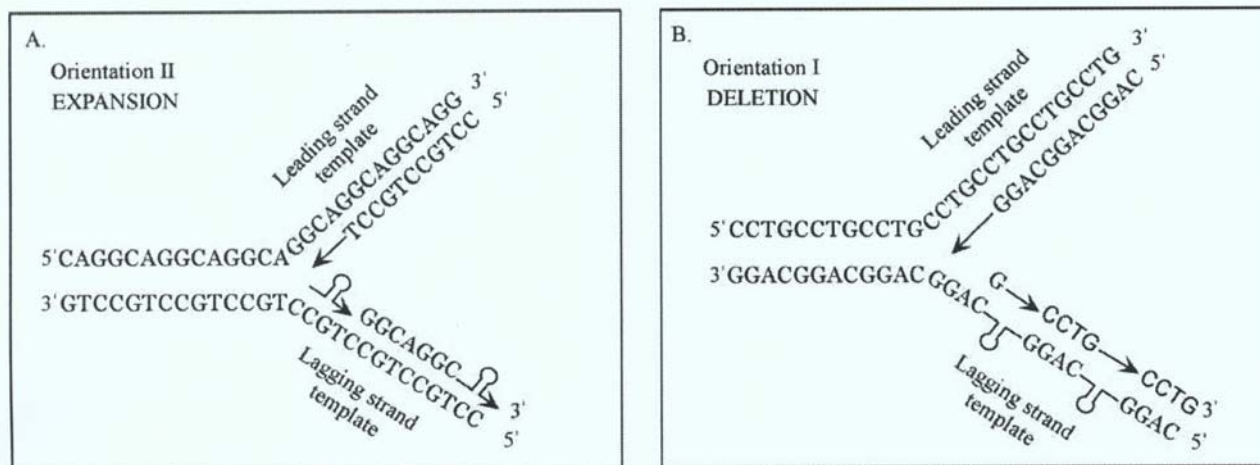


FIG. 8. Model for the orientation dependent instability of $(CCTG-CAGG)_n$ repeats in mammalian cells. A, the presence of the CAGG repeats on the leading strand template and on the newly synthesized products of the lagging strand template (orientation II) can give rise to expansions because the CAGG repeats on the nascent strand can form folded-back secondary structures by strand slippage and thus generate expansions. B, the presence of the CCTG repeats on the leading strand template and on the newly synthesized product of the lagging strand template (orientation I) preferentially give rise to deletions because the CAGG repeats on the lagging strand template can form slipped structures that may be bypassed during synthesis (see "Discussion").

Replikace v tetranukleotidových opakujících se sekvencích (CCTG) :

disociace řetězců - následná syntéza opožďujícího řetězce pomocí Ókazakiho fragmentů.

Sekundární pseudostruktury - hlavní příčina genetické nestability TR

vznik expanzí nebo delecí

**Příčiny nestability dynamických mutací v
několikanásobně se opakujících sekvencích
tri, tetra, penta nukleotidů :**

replikace, rekombinace, reparace

**Klouzání (slippage) v repetících vytváří non B-
DNA struktury**

Multigenové rodiny:

Skupina sekvenčně příbuzných genů, které mají společný evoluční původ a stejnou biologickou funkci

př. Geny kodující polypeptidové řetězce hemoglobinu

Pseudogeny: nepřesné kopie strukturních genů, které vznikly asi nepřesnou duplikací nebo mutací, kterou se původně aktivní gen v genové rodině inaktivoval

Tandemové genové repetice , jsou geny nebo skupiny genů opakující se bezprostředně za sebou, tedy v tandemu
Geny uvnitř skupiny mají vždy stejné pořadí

Tandemovým způsobem se opakují:

Geny přepisované do 5S rRNA (četnost repetice u člověka 250)

Geny přepisované do tRNA (četnost repetice 10 až 100 pro jednotlivý druh tRNA u člověka)

Geny kodující histony (četnost repetice 20 pro skupinu genů kodujících všech 5 histonů)

Biologický význam těchto tandemových genových repeticí spočívá v tom, že jedna kopie každého z uvedených genů by nestačila krýt spotřebu jeho produktů buňkou. V buňce musí být dostatečná zásoba 5S rRNA a tRNA pro realizaci translace a histonů pro syntézu histonového oktameru při replikaci.

Př. Multigenových rodin

Hlavní histokompatibilní komplex (HLA – Human Leucocyte Antigen) – na chromozomu 6

Biologickou funkcí HLA – proteinů je prezentace antigenních peptidů T – lymfocytům – spuštění komplexní imunitní odpovědi proti patogenu.

Obecná koncepce imunitní odpovědi – připravenost na jakýkoliv typ infekce.

HLA geny 1. třídy: HLA-A, HLA-B, HLA-C

HLA geny 2. třídy: HLA-DR, HLA-DQ, HLA-DP

Alelická bohatost HLA genů,

heterozygotní jedinec – schopnost čelit většímu počtu infekcí,

heterozygotní výhodnost.

Struktura HLA genové rodiny ,
počet cca 20 genů

(lokus 6p21.3). velikost 2 Mb.

Exprese **6 genů** (viz výše),

další 4 jsou konvenční **pseudogeny** a

7 reprezentuje **nefunkční geny** nebo
fragmenty genů

Genové rodiny kódující produkty s velkými konzervovanými doménami.

Výskyt homologie v silně konzervovaných oblastech genů, často jsou kódovány proteiny s významnou funkcí ve vývoji. Konzervovaný sekvenční motiv obsahují geny **PAX** (390 pb) nebo **homeoboxy** (180 pb).

Homeotické proteiny se váží na specifické sekvence DNA a regulují expresi genů na úrovni transkripce. Řada genů, které kódují transkripční faktory, obsahuje homeobox. HMG proteiny.

Genové rodiny kódující produkty s krátkými konzervovanými motivy aminokyselin jsou charakterizovány obecnou funkcí.

Např. Genová rodina s DEAD boxem obsahuje několik rozdílných genů s RNA helikázovou funkcí s krátkým charakteristickým motivem aminokyselin – DEAD box =

Asp-Glu-Ala-Asp.

Dalším typem je WD repetitivní genová rodina (tandemově se opakující tryptophan – aspartate) s pravěpodobnou regulační funkcí.

Genové superrodiny

Charakteristickým rysem je nízká sekvenční homologie velkých segmentů, ale v obecném smyslu funkční příbuznost a podobnost ve struktuře domén.

Př. Členové Ig imunoglobulinové superrodiny jsou povrchové proteiny s podobným typem struktury domén (S-S disulfidové můstky)

Rozptýlené genové repetice,

gen nebo genová rodina, jejichž kopie se vyskytují na různých místech haploidního genomu. Např. některé geny přepisované do tRNA

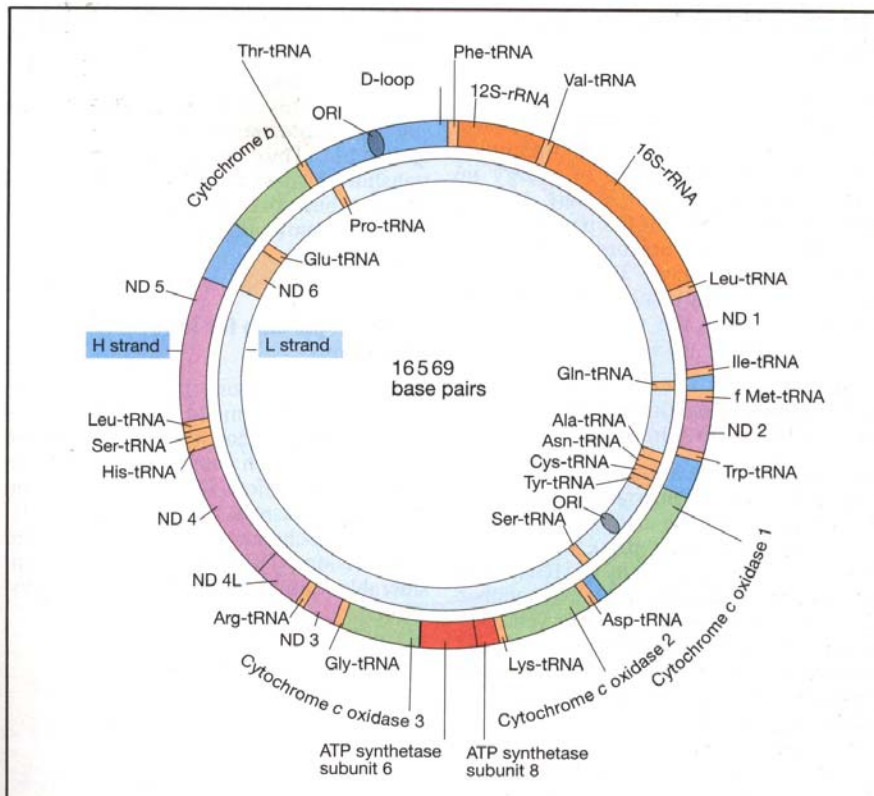
Nemendelovská dědičnost

Některá onemocnění a znaky, za které zodpovídají variace jednotlivých genů, nesledují do značné míry pravidla přenosu do dalších generací, která platí pro dědičnost "klasických,, monogenních onemocnění, ať už autozomálně či gonozomálně determinovaných.

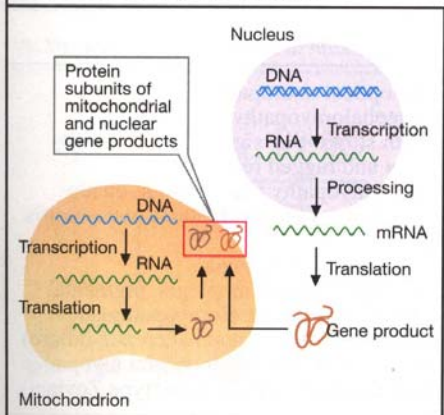
Proto se tato různorodá skupina někdy vyděluje pod pojmem tzv. nemendelovské dědičnosti (angl.non-Mendelian inheritance).

Řadíme sem především **mitochondriální dědičnost**, nestabilitu **repetitivních sekvencí** a **genomický imprinting**. Zcela zvláštní kapitolou je epigenetika .

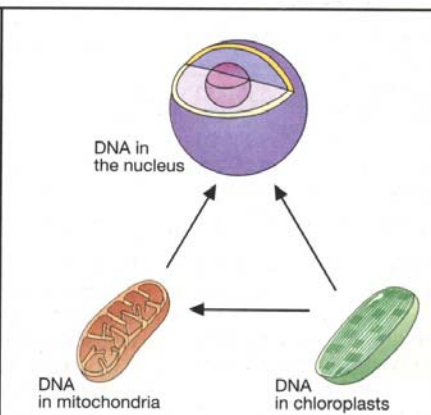
Naopak se do této skupiny obvykle nepočítají jevy, které "komplikují" monogenní dědičnost, např. **penetrance**, **expresivita** znaku atd.



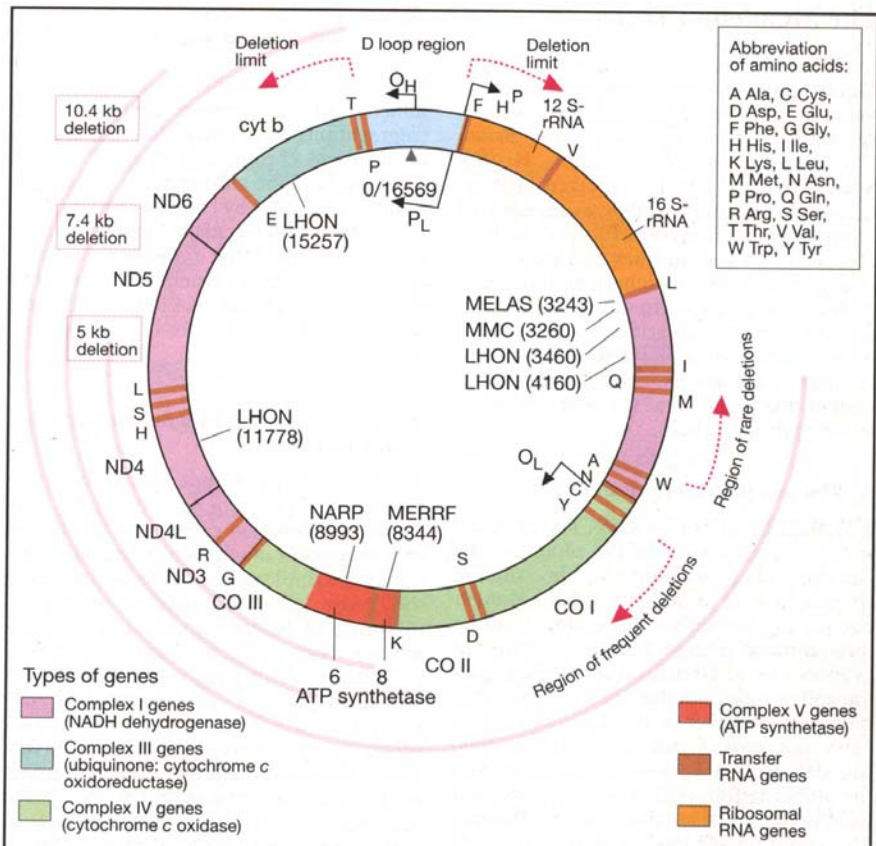
A. Mitochondrial genes in man



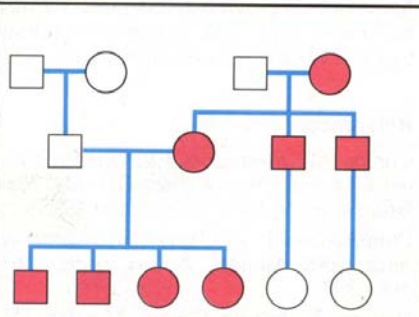
B. Cooperation between mitochondrial and nuclear genome



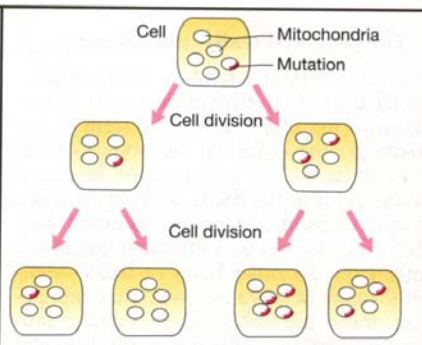
C. Evolutionary relationship of mitochondrial genomes



A. Mutations and deletions in human mitochondrial DNA



B. Maternal inheritance of a mitochondrial disease



C. Heteroplasmy for mitochondrial mutations

1. Mitochondrie a mitochondriální dědičnost

a) Mitochondriální DNA

Ačkoli je podstatná část lidského genomu uložena v jaderné DNA, **37 strukturních genů** je kódováno v cirkulárním mitochondriálním genomu, který je tvořen **16569 páry bází DNA** (označované jako mtDNA). **Třetina těchto genů kóduje podjednotky komplexů respiračního řetězce zodpovědného za produkci ATP v rámci tzv. oxidačně-fosforylačního (OXPHOS) systému** (13 polypeptidů, zbylých 74 je kódováno jaderným genomem), zbylé pak zodpovídají za tvorbu specifických molekul RNA potřebných pro syntézu proteinů (22 transferových RNA a dvě ribosomální RNA). **V cytoplasmě je přítomno několik set až tisíc kopií mtDNA**, když v každé mitochondrii je přítomna řada molekul mtDNA. O **homoplasmii** mluvíme, když je přítomna jen jedna homogenní populace mtDNA (odpovídá stavu u drtivé většiny "normálních" jedinců). Frekvence mutací je u mitochondrií 10-20x vyšší než u jaderné DNA, což vede k **heteroplasmii**, tedy stavu, kdy je přítomna smíšená populace s více variantami mtDNA. Mnohé vlastnosti mitochondriální DNA podporují teorii o původu mitochondrií coby **endosymbiontických bakterií** proto-eukaryotických buněk před cca 1,5 miliardami let. Mezi nejmarkantnější z nich určitě patří **cirkulární organizace genomu, absence histonů, absence intronů** či diskrétní počátky replikace. Navíc se mitochondrie obratlovců i dalších organismů se vyznačují odchylkami od univerzálního genetického kódu - kodón UGA není terminační, ale kóduje aminokyselinu tryptofan, naopak kodóny AGA a AGG nekódují leucin, ale ukončují translaci a nakonec kodón AUA místo leucinu kóduje metionin. Biologicky jsou asi nejbližšími příbuznými Rickettsie (mj. původci skvrnitého tyfu), což jsou obligátní intracelulární parazité. Předpokládá se, že Rickettsie a mitochondrie mají společného předka, který prodělal přechod od autonomní existence k endosymbióze. Mitochondriální DNA skutečně obsahuje jen jednu významnou "nekódující" sekvenci, takzvanou D-smyčku (z angl. D-loop), což je krátký úsek mtDNA, ve kterém je těžký řetězec vytěsňován fragmentem DNA (500-700 nukleotidů), komplementárním k řetězci lehkému (a v tomto úseku má tedy mitochondrie třířetězcovou DNA). Zde je počátek replikace tzv. těžkého řetězce (označovaného jako H z angl. heavy), zmiňovaný fragment funguje jako primer pro začátek replikace. Počátek replikace lehkého řetězce (L z angl. light) je umístěn mimo Dkličku, asi ve 2/3 mt DNA. Transkripty obou řetězců musí být rozštěpeny, aby se mohly uvolnit funkční RNA (rRNA, tRNA a mRNA).

b) Mitochondriální dědičnost

Z hlediska genetiky je zásadní fakt, že **mtDNA je předávána další generaci výhradně matkou (matroklinní dědičnost), když po oplodnění jsou zachovány pouze mitochondrie lidského vajíčka**. To patrně není pouhým důsledkem nepoměru počtu mitochondrií lidského oocyty (cca 100 000) a spermie (50-70), ale předpokládá se aktivní proces, který po oplození zlikviduje mitochondrie paternálního původu. Tomu odpovídá i typický maternální přenos chorob způsobených mutacemi mtDNA v rodokmenu (Obr.1.). Pokud je heteroplazmická mutace zděděna nebo k ní dojde v časných fázích embryogeneze, normální i mutovaná varianta jsou náhodně předávány při buněčném dělení dceřiným buňkám (mitotická i meiotická segregace). Distribuce a zastoupení mutované mtDNA v jednotlivých orgánech jsou proto patrně závislé na čase a vzniku mutace a rovněž na typu postižené buňky.

c) Onemocnění s mitochondriální dědičností

Je třeba si uvědomit, že zdaleka ne veškerá onemocnění způsobená dysfunkcí mitochondrií mají mitochondriální typ dědičnosti. Pokud je defektní protein, jehož některé podjednotky kódují geny jaderného genomu a jiné zase přímo mtDNA, může být dědičnost zcela medndelistická. Čtyři základní charakteristiky dědičných onemocnění, u kterých je třeba diferenciálně diagnosticky uvažovat o mitochondriální dědičnosti:

1. Maternální typ dědičnosti (viz výše)
2. Díky mitotické a meiotické segregaci nacházíme různý stupeň postižení v různých tkáních a variabilní projevy u potomků v jedné mateřské linii, což souvisí s existencí tzv. **prahového efektu**, což je pozorování, že pro manifestaci dysfunkce je nutné, aby byl překročen určitý poměr mitochondrií s mutantní a normální mtDNA (ve většině případů 60-90%).
3. Projevy poruchy OXPHOS systému. Dědičná mitochondriální onemocnění většinou postihují tkáně s vysokými nároky na přísun energie, jako jsou neuromuskulární či centrální nervový systém. Mezi běžně pozorované symptomy chorob s mitochondriální dědičností patří encefalopatie, ataxie, spasticita, (kardio)myopatie, hluchota nebo diabetes mellitus.
4. Rozdíl oproti gonozomálně dominantní dědičnosti vázané na chromozóm X je v absenci přenosu mitochondriálně vázaného onemocnění na potomstvo a signifikantní převahu postižených žen u X-vázané dominantní dědičnosti.