

# TEXTO 9

## ASPECTOS DA COMPLEXIDADE DOS GENOMAS

### Aspectos da Complexidade dos Genomas

#### I - O Paradoxo da Quantidade de DNA

Quantidade de Material Genético

Quantidade de DNA e complexidade do organismo

#### II - A Organização e Distribuição dos Genes nos Cromossomos

A densidade dos genes

Os genes que codificam proteínas podem ser solitários ou pertencer a famílias gênicas

Genes repetidos em tandem codificam histonas, RNAs ribossômicos e RNAs transportadores

#### III - A Natureza dos DNAs Não Codificadores

##### 1. DNA Repetitivo em Tandem

O DNA satélite

Centrômeros

Telômeros

VNTRs: minissatélites e microssatélites

##### 2. DNAs Repetitivos Dispersos: Os Elementos Genéticos Móveis

###### 2.1. Elementos genéticos móveis que se movem por meio de DNA.

###### 2.1.1. Elementos de Inserção (IS)

###### 2.1.2. Transposons

###### 2.1.3. Elementos genéticos móveis de DNA em eucariotos

###### 2.2 Elementos que se mobilizam por RNA (*retrotransposons*)

###### 2.2.1. Elementos semelhantes a retrovírus

###### 2.2.2. "Long Interspersed Elements" (LINEs)

###### 2.2.3. "Short Interspersed Elements" (SINEs)

###### 2.2.4. Pseudogenes e pseudogenes processados.

#### IV - As Sequências Repetitivas e o Projeto do Genoma Humano

#### V - O DNA das Organelas

O tamanho e a capacidade de codificação do DNA mitocondrial varia muito entre os diferentes organismos.

Os produtos dos genes mitocondriais não são exportados

O código genético mitocondrial é diferente do código genético padrão

O DNA mitocondrial humano

Mutações no DNA mitocondrial causam várias doenças genéticas nos humanos

Questões para orientar a leitura: Parte I

Questões para orientar a leitura: Parte II

## Apresentação

Sequenciar o genoma humano no encerramento do século da genética foi, sem dúvida, um feito notável. Mas, como sempre ocorre com grandes realizações científicas, mais novas perguntas são geradas do que respondidas. Descobrir que um pequena porcentagem do nosso genoma (no máximo 3%) corresponde aos exons dos genes codificadores de peptídeos e que temos muito menos genes do que imaginávamos nos desafia a tentar compreender o que nos tornou seres tão complexos. Cada vez mais pesquisadores acreditam que grande parte das respostas não está somente nos genes.

Essa apostila foi organizada pelos docentes do Instituto de Biociências da USP que ministraram e ministram as disciplinas “Biologia Molecular e de Microrganismos”, “Biologia Molecular” e “Fundamentos de Biologia Molecular” para os cursos de Bacharelado e Licenciatura em Ciências Biológicas. O texto foi adaptado e compilado de diversos livros didáticos e matérias instrucionais.

# Aspectos da Complexidade dos Genomas

## I - O Paradoxo da Quantidade de DNA

### Quantidade de Material Genético

Historicamente, a quantidade de DNA existente nos núcleos das células de eucariotos foi uma das primeiras características estudadas do material genético destes organismos. Esta quantidade era medida por uma técnica conhecida por citofotometria. Essa é uma técnica que emprega princípios de espectrofotometria e microscopia de varredura de luz transmitida. Lâminas com células são coradas com um reagente que se liga especificamente ao DNA (coloração de Feulgen). As lâminas são colocadas em um microscópio que emite um feixe de luz monocromático, o qual varre o campo. Uma célula fotossensível registra as variações da intensidade da luz transmitida fornecendo a quantidade de DNA circunscrita à região do núcleo das células. Através dessa técnica foi possível verificar, primeiramente, que o tamanho do genoma é constante para cada espécie em particular. A quantidade de DNA no genoma é conhecida como valor C (de Constante ou Característico). Este valor pode ser expresso em picogramas (pg,  $1\text{pg} = 1 \times 10^{-12}\text{g}$ ), que contém aproximadamente um bilhão de pares de bases de DNA. Hoje em dia é mais comum descrever o valor C com valores expressos em números de pares de bases. A constância do valor C pode ser violada, quando há, por exemplo, alguns eventos excepcionais de amplificação gênica, em certas partes do genoma.

### Quantidade de DNA e complexidade do organismo

O valor C não exprime uma relação direta com o grau de complexidade do organismo. Em geral, os eucariontes têm um valor C muito maior que os procariontes, mas há exceções: a levedura (*Saccharomyces cerevisiae*), por exemplo, tem um valor C menor que certas bactérias Gram-positivas. Nos eucariontes, a variação do valor C é muito grande, como mostrado na Tabela 1. A análise da Tabela 1 mostra claramente que não existe motivo para relacionarmos diretamente a complexidade de um organismo com a quantidade de DNA nuclear que possui. Por exemplo: dentre os protistas, há uma variação de quase 20.000 vezes! Essa discrepância entre a complexidade de organismos e o valor C ficou conhecida como o paradoxo do valor C.

Face a estes resultados, surge inevitavelmente a pergunta: será que todo o DNA se constitui em genes? Uma das primeiras respostas a esta questão surgiu de experimentos que envolveram a desnaturação e a renaturação de DNA. Se aquecermos e resfriarmos o DNA, a cadeia dupla desnatura e renatura em seguida. O DNA de fita simples pode ser distinguido do de fita dupla por algumas propriedades, tais como a afinidade com o mineral hidróxiapatita ou a

absorbância a 260 nm. Com essas técnicas foi possível estudar a cinética da reassociação das moléculas de DNA de vários organismos. O que se observou foi que uma fração de DNA desnaturado se reassocia rapidamente, uma outra fração demora um pouco mais e uma terceira fração demora muito para se reassociar. O que faz com que uma fração de DNA se reassocie mais rapidamente que outra? A única resposta possível é que as sequências que se reassociam rapidamente são muito reiteradas, ou seja, existem muitas cópias iguais da mesma sequência no genoma. Aquelas que demoram muito para se reassociar, por outro lado, existem em poucas cópias ou mesmo em cópias únicas.

Tabela 1. Valores C de DNA de alguns eucariotos, organizados em ordem de valor.

Espécie	Nome popular	X1000 Kb
<i>Navicola pelliculosa</i>	diatomácea	35
<i>Drosophila melanogaster</i>	drosófila	180
<i>Paramecium aurélia</i>	paramécio	190
<i>Gallus domesticus</i>	galinha	1200
<i>Cyprinus carpio</i>	carpa	1700
<i>Boa constrictor</i>	jibóia	2100
<i>Parascaris equorum</i>	lombriga de cavalo	2500
<i>Charcarias obscurus</i>	tubarão	2700
<i>Rattus norvegicus</i>	rato	2900
<i>Xenopus laevis</i>	sapo	3100
<i>Homo sapiens</i>	humanos	3400
<i>Nicotiana tabacum</i>	tabaco	3800
<i>Paramecium caudatum</i>	paramécio	8600
<i>Schistocerca gregária</i>	gafanhoto	9300
<i>Allium cepa</i>	cebola	18000
<i>Lilium formosanum</i>	lírio	36000
<i>Pinus resinosa</i>	pinheiro	68000
<i>Protopterus aethiopicus</i>	peixe pulmonado	140000
<i>Ophioglossum petiolatum</i>	samambaia	160000
<i>Amoeba proteus</i>	ameba	290000
<i>Amoeba dúbia</i>	ameba	670000

Em média, as sequências de DNA presentes na fração considerada medianamente repetitiva estão reiteradas 350 vezes. Já na fração de DNA altamente repetitivo dos genomas, as sequências estão em média repetidas 500.000 vezes.

Dentre os organismos, existe uma grande variação na proporção de DNA de “cópia única” com relação à totalidade do DNA. Em procariontes inferiores, a maioria do DNA é única. Apenas 20% dele pode ser considerado dentro da categoria de moderadamente repetitivo. Em células animais, em média, até 50% do DNA situa-se na categoria de DNA altamente ou moderadamente repetitivo; em anfíbios e plantas, essa categoria engloba cerca de 80% do genoma.

Em muitos organismos, como *Drosophila* e vertebrados, grande parte das sequências altamente repetitivas ocorre em enormes blocos. Já outros tipos de sequências repetitivas aparecem dispersas, ou seja, espalhadas pelo genoma.

Existe uma tendência geral de que espécies com valor C muito alto tenham alta proporção de DNA repetitivo. Isto parece resolver em parte o paradoxo do valor C, mas não é a resposta completa para explicar porque existe tanto DNA em certas espécies, pois tais espécies parecem ter também um aumento da quantidade de DNA de cópia única.

## II - A Organização e Distribuição dos Genes nos Cromossomos

Estudamos em módulos anteriores a relação entre as unidades de transcrição e os genes nos eucariontes e procariontes. Estudaremos agora a organização dos genes nos cromossomos eucariontes e, a seguir, a organização das sequências não-codificadoras nos genomas desses organismos.

## A densidade dos genes

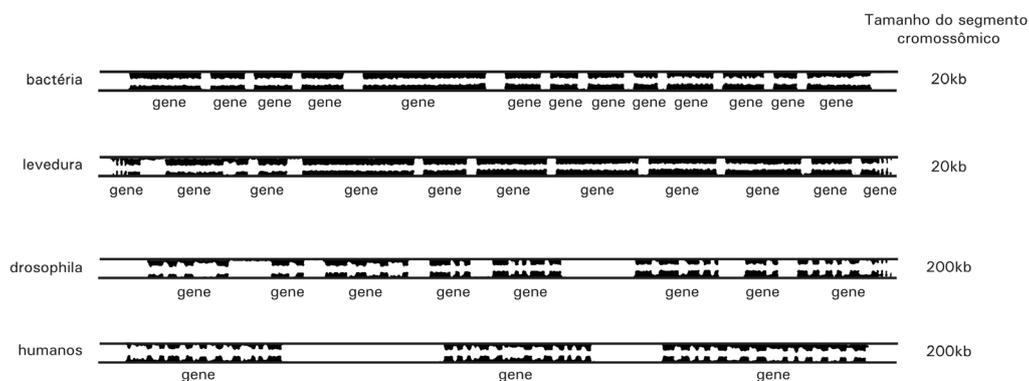
Ao observarmos a Figura 1, vemos que há importantes diferenças na organização dos cromossomos apresentados. Observa-se que nos cromossomos da bactéria e da levedura há pouco espaço entre os genes, enquanto uma fração pequena do cromossomo humano codifica proteína. Observe também que o número e o tamanho dos íntrons aumenta da levedura ao homem. Os projetos de sequenciamento de genomas já concluídos confirmam que, de modo geral, existe muito DNA nos organismos superiores que não codifica proteína, ou seja, a densidade de genes é baixa em relação ao total de DNA, nos organismos mais complexos. Não se conhece ainda a função de muitas categorias desse DNA não codificador.

## Os genes que codificam proteínas podem ser solitários ou pertencer a famílias gênicas

Nos organismos multicelulares, entre 25–50% dos genes que codificam proteínas ocorrem uma única vez no genoma e por isso são chamados genes solitários ou de cópia única. Os demais genes pertencem a famílias, que reúnem dois ou mais genes similares.

Frequentemente, nos 5–10 kb de DNA que ficam ao redor de um certo gene pode estar presente uma cópia incompleta desse mesmo gene. Essa sequência pode surgir a partir da duplicação do gene ancestral. Um conjunto de genes que surge a partir da duplicação de um gene ancestral e que produz proteínas com sequências de aminoácidos parecidas (mas não obrigatoriamente idênticas) constitui uma **família gênica**.

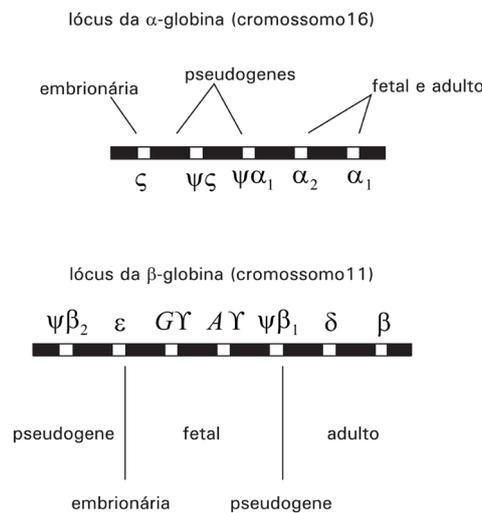
Os diferentes genes de beta-globina surgiram provavelmente a partir da duplicação de um gene ancestral, mais provavelmente por meio de uma permutação desigual durante a recombinação meiótica. Com o passar do tempo evolucionário, as duas cópias do gene ancestral devem ter acumulado mutações e aquelas que conferiram algum refinamento ao processo de transporte de oxigênio devem ter sido selecionadas. Mas, às vezes as mutações tornam uma das cópias de um gene ancestral inativa do ponto de vista genético, caracterizando então um **pseudogene**, como estudaremos a seguir.



**Figura 1.** Tamanho aproximado e espaçamento dos genes em vários organismos representativos. As regiões codificadoras são as escuras e os íntrons e as regiões espaçadoras são as claras. Observe que os segmentos de *Drosophila* e humanos representados na figura são dez vezes mais longos do que os de bactéria e de levedura.

## Genes repetidos em tandem codificam histonas, RNAs ribossômicos e RNAs transportadores

Nos invertebrados e nos vertebrados, os genes que codificam histonas, RNAs ribossômicos e RNAs transportadores ocorrem em arranjos de sequências repetidas em tandem, ou seja, as cópias repetidas são adjacentes e em mesma orientação. Esses arranjos são diferentes das famílias gênicas exemplificadas no item anterior, porque codificam RNAs ou proteínas praticamente idênticos. Essas cópias múltiplas de genes são necessárias para atender a enorme demanda celular de seus produtos. A maior parte do RNA de uma célula é representada pelos RNAs transportadores e ribossômicos. Por exemplo, uma célula embrionária humana precisa se duplicar a cada 24 horas e tem milhões de ribossomos.



**Figura 2.** Esquema da distribuição dos genes das globinas nos cromossomos humanos 16 e 11. Os pseudogenes têm como prefixo a letra grega ψ (psi).

Por isso, uma célula embrionária humana precisa de pelo menos 100 cópias dos genes ribossomais. Todos os eucariontes, incluindo as leveduras, possuem 100 ou mais cópias dos genes que codificam o precursor do RNA ribossômico (a molécula precursora de RNAr 45S) e o RNA ribossômico 5S. Mais de 20.000 cópias desse último podem ocorrer em uma célula de anfíbio.

Os genes que codificam o RNAr estão tipicamente concentrados em muitas cópias em tandem, em um ou alguns locais dos cromossomos. Esses locais podem ser identificados em microscopia como as regiões organizadoras do nucléolo. O nucléolo contém uma massa de material ribossômico recém-sintetizado, muito proeminente em núcleos interfásicos. As regiões organizadoras de nucléolo podem ser identificadas pela sua associação ao nucléolo na prófase e apresentam-se como regiões pouco condensadas na metáfase: Os organizadores nucleolares hibridam intensamente com sondas de DNAr.

A maioria das espécies tem centenas de cópias dos genes ribossomais. As estimativas para humanos, *Drosophila* e levedura são da ordem de 200, 200 e 140 cópias por genoma, respectivamente. Isto significa que o DNAr é um componente quantitativamente importante (cerca de 7%) do genoma total da levedura, onde é a mais abundante sequência repetitiva, enquanto no homem corresponde a apenas 0,3% do genoma. Além das sequências que transcrevem a grande molécula precursora de RNAr 45S, a partir da qual originam-se RNAs 18S, -5,8S e 28S, todos os eucariontes têm múltiplos genes que transcrevem a outra pequena molécula de RNAr, o 5S.

As histonas, importantes proteínas cromossômicas, também são necessárias em grandes quantidades e os genes codificadores dos cinco tipos principais de histonas também são reiterados nos genomas: na espécie humana, os genes das cinco histonas formam um agrupamento (“cluster”), repetido cerca de 20 vezes. Os genes do RNAr, das histonas e outros genes repetidos que codificam proteínas correspondem a uma porção apreciável da chamada fração de DNA medianamente repetitivo dos genomas eucariontes.

### III - A Natureza dos DNAs não Codificadores

#### 1. DNA Repetitivo em Tandem

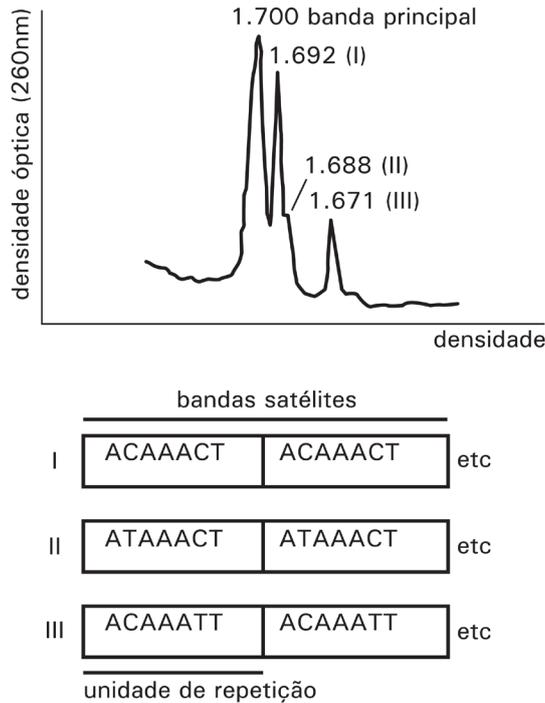
##### O DNA satélite

O termo DNA satélite foi originalmente aplicado a frações do DNA dos vertebrados que são separáveis do restante do DNA através de gradientes de densidade, como os gradientes de cloreto de céso, por exemplo. Uma fração de DNA satélite é identificada em um gradiente como uma banda separada do restante do DNA por diferença de densidade (Fig. 3).

O DNA satélite consiste de segmentos de DNA relativamente simples, curtos e altamente repetitivos, que diferem do resto do DNA pela composição de bases. Atualmente o uso termo satélite foi ampliado para descrever

o DNA altamente repetitivo, cujas sequências dispõem-se lado a lado (em tandem), mesmo que ele não possa ser separado do restante em função de sua densidade nos gradientes.

O DNA satélite foi extensivamente estudado em *Drosophila*, camundongos, ratos e humanos. Em todos esses organismos, o DNA satélite pode corresponder de 5 a 10% do genoma total. Ele é um dos componentes principais da heterocromatina, definida como a parte dos cromossomos que fica condensada e corada durante todo o ciclo celular, em contraste com o resto da cromatina, a eucromatina, que está estendida e difusa nos núcleos interfásicos. O DNA satélite está tipicamente organizado em blocos de milhares ou milhões de repetições em tandem, geralmente concentrados ao redor das regiões centroméricas dos cromossomos. Mas quantidades enormes de DNA repetitivo satélite também podem estar presentes em blocos de heterocromatina que estão distantes das regiões centroméricas dos cromossomos.



**Figura 3.** DNA satélite de *Drosophila virilis*. (a) DNA de embrião foi submetido à centrifugação em gradiente de cloreto de céσιο. O DNA contido em diferentes zonas do gradiente foi monitorado por medida de absorção a 260 nm (densidade óptica). A banda principal, ou seja, a maior parte do DNA, tem uma densidade de 1,700. As três bandas satélites (I, II e III) são menos densas. (b) A análise da sequência de bases mostrou que cada banda satélite é composta de DNA constituído por uma longa série de repetições de sete nucleotídeos. As sequências repetidas nas bandas II e III diferem da sequência da banda I apenas por uma base.

Até o momento, sabe-se muito pouco sobre a função do DNA satélite; o tipo e a quantidade de cópias das sequências que o constituem variam muito entre as diferentes espécies, mesmo dentro de um mesmo gênero, e a quantidade pode variar até mesmo entre indivíduos de uma mesma espécie.

### Centrômeros

O centrômero é uma estrutura do cromossomo, constituída por DNA e proteínas, altamente diferenciada e que exerce várias funções relacionadas à mitose e à meiose, como o emparelhamento das cromátides irmãs, ligação ao fuso e movimentação cromossômica. Nos centrômeros, uma estrutura conhecida como cinetócoro se organiza e promove a ligação do cromossomo às fibras do fuso.

Estudos em vários organismos indicaram que os centrômeros são de dois tipos principais: aqueles compostos por uma região bastante curta (cerca de 200pb) de DNA cromossomal, que ocorre nas leveduras, e aqueles que se estendem por várias centenas de milhares de pares de bases (de 40 kb até várias megabases). O último tipo é frequentemente caracterizado por grandes quantidades de sequências repetidas, como os DNAs satélites mencionados acima. As sequências centroméricas são flanqueadas por unidades de repetição mais heterogêneas,

algumas delas relacionadas a elementos de transposição. As repetições do flanco têm a função de mediar a adesão de cromátides irmãs durante a divisão celular.

O DNA centromérico está arranjado em nucleossomos nos quais a versão usual da histona H3 é substituída por uma variante, a proteína CENP-A, que é fundamental para formação do centrômero e sua função. As sequências flanqueadoras desse núcleo centromérico estão empacotadas na forma de heterocromatina, que é distinguível bioquimicamente da eucromatina pela presença de grupos metil ligados à lisina na posição 9 da H3.

O DNA repetitivo centromérico não é considerado imprescindível para o funcionamento dos centrômeros, mas há uma correlação muito alta entre a presença de certas famílias de DNA repetitivo centromérico (por exemplo, o DNA satélite alfa dos primatas) e a localização do centrômero.

## Telômeros

Telômeros são estruturas formadas por DNA repetitivo e proteínas. Estão presentes nas extremidades dos cromossomos de organismos eucarióticos. Eles são considerados parte essencial dos cromossomos, pois protegem suas extremidades de degradação, permitem que a célula possa distinguir os cromossomos íntegros daqueles quebrados, e são substrato para a replicação do DNA na extremidade dos cromossomos.

O DNA dos telômeros é formado por uma série de sequências simples repetidas que não codificam proteínas, mas que possuem uma função definida. Estas sequências estão repetidas centenas ou milhares de vezes, estendendo-se por várias kilobases de DNA.

As extremidades das moléculas de DNA que estão presentes nos telômeros oferecem uma dificuldade de ordem estrutural para a replicação do DNA. Pela impossibilidade de *primers* se hibridarem exatamente às extremidades dos cromossomos, trechos da molécula de DNA nas extremidades deixam de ser replicados a cada ciclo de replicação e divisão celular.

Disso decorre o encurtamento do DNA nas extremidades dos cromossomos a cada ciclo mitótico. Esse encurtamento pode ser compensado pela atividade da telomerase, uma transcriptase reversa específica de telômero, capaz de adicionar cópias adicionais de DNA repetitivo telomérico às extremidades dos cromossomos, utilizando como molde uma das suas subunidades que é constituída por uma molécula de RNA.

O comprimento médio do telômero pode também aumentar ou diminuir em resposta a alterações nutricionais, como acontece na levedura, por exemplo. Em células somáticas dos organismos superiores, os telômeros diminuem lentamente durante divisões celulares sucessivas, com o aumento da idade das células. Porém, a perda de sequências também pode ocorrer na ausência de divisões celulares, provavelmente como resultado de degradação.

Nos organismos multicelulares, a telomerase fica geralmente inativa nas células dos tecidos somáticos, mas ela está ativa na linhagem germinativa, em células-tronco e alguns tipos celulares com elevada capacidade de proliferação. No câncer, algum tipo de mecanismo que permita a extensão do DNA repetitivo telomérico deve estar presente; do contrário, a proliferação celular não seria possível. Em alguns tipos de células cancerosas, o telômero é estendido por causa da reativação da atividade da telomerase, mas há outros mecanismos complexos independentes dessa enzima para a manutenção da estrutura telomérica.

O encurtamento dos telômeros causa mudanças nas células que estão correlacionadas ao envelhecimento e ao risco de câncer; por isso eles são estruturas de relevância médica e têm sido muito estudados nos últimos anos.

*Drosophila* tem uma estrutura telomérica completamente diferente dos demais organismos: ao invés de repetições simples, o DNA das extremidades dos cromossomos é composto por elementos de transposição, que ao se transporem, aumentam o comprimento do DNA telomérico. Uma vez que há múltiplas possibilidades de sequências nucleotídicas nas unidades de repetição nos telômeros, a função do telômero não está associada a um único tipo de sequência. Além disso, em muitos organismos, DNA telomérico é também encontrado em sítios internos dos cromossomos, o que demonstra que a sequência de DNA sozinha não é suficiente para desempenhar a função de telômero.

Há numerosos estudos investigando as proteínas teloméricas e seu papel na regulação do tamanho dos telômeros. Muitas delas tiveram sua função correlacionada a outras funções bioquimicamente importantes para as células, como por exemplo, a regulação do reparo do DNA e do ciclo celular, o que certamente é muito relevante para o estudo de doenças e do envelhecimento celular.

## VNTRs: minissatélites e microsatélites

Essa é uma classe muito interessante de DNA repetitivo em tandem. A sigla quer dizer *Variable Number of Tandem Repeats*, pois o número de cópias de cada unidade de repetição varia muito entre os indivíduos de uma população, caracterizando polimorfismos genéticos. Quando a unidade de repetição tem entre 15 e 100 nucleotídeos, utiliza-se o termo **minissatélite**. Os minissatélites estão localizados em várias posições e em vários cromossomos. Se cortarmos o DNA de um indivíduo com uma enzima de restrição e tivermos uma sonda correspondente a um minissatélite, o resultado da hibridação serão fragmentos de tamanhos diferentes nos diferentes indivíduos. Se a sonda hibridar com diversos minissatélites distintos, teremos um *fingerprinting* de DNA. Por essas propriedades, os minissatélites foram muito utilizados em identificação individual, testes de paternidade e estudos de mapeamento genético.

Os **microsatélites** são semelhantes aos minissatélites no que diz respeito à variabilidade no número de cópias, mas nessas sequências as unidades que se repetem são mais curtas (de um até cerca de 14 nucleotídeos). Um termo também utilizado para descrever os microsatélites é STRs (*Short Tandem Repeats*).

Embora muito úteis aos geneticistas, ainda não foi atribuída nenhuma função genômica para a maioria das sequências de microsatélites e minissatélites. As aplicações dos microsatélites são semelhantes às descritas para os minissatélites. Os microsatélites são atualmente a ferramenta favorita para estudos de identificação individual e vínculo biológico, pois são facilmente amplificáveis por meio da reação em cadeia da polimerase e os genótipos são determinados em analisadores genéticos automáticos (sequenciadores de DNA).

## 2. DNAs Repetitivos Dispersos: Os Elementos Genéticos Móveis

Uma das mais interessantes descobertas da genética, nas últimas décadas, é a presença de elementos infiltrados no genoma, que parecem ocorrer simplesmente em decorrência da sua capacidade de proliferação independente em relação ao resto do genoma. Alguns desses “DNA parasitas” ou móveis codificam pelo menos parte do aparato necessário à sua própria proliferação, enquanto que outros parecem depender do aparato de proliferação de outros elementos.

Esses elementos podem ser divididos em dois grupos principais. Alguns deles têm a capacidade de se deslocar no genoma por um processo de excisão e reinserção ou por processo de cópia de moléculas de DNA. Outros se movem através de intermediários de RNA, por um processo de transcrição reversa. Nós vamos estudar a seguir algumas das categorias desses diferentes tipos de elementos móveis que podem ocorrer nos genomas de procariontes e eucariontes.

### 2.1. Elementos genéticos móveis que se movem por meio de DNA.

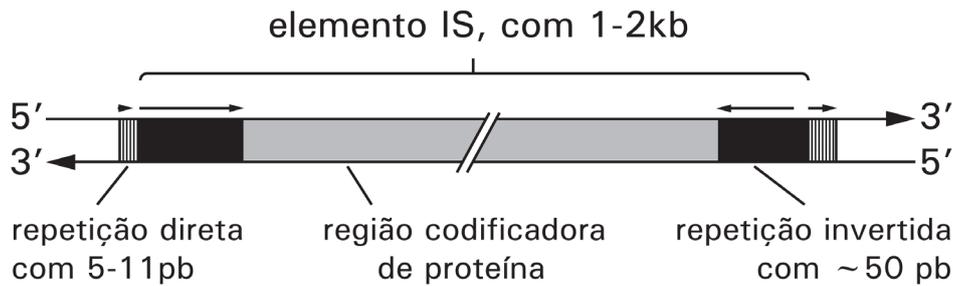
Os elementos transponíveis foram primeiramente identificados molecularmente em plasmídeos bacterianos, como elementos móveis capazes de carregar genes que conferem resistência a antibióticos. O uso da palavra *transposon*, que corresponde a uma das classes desses elementos transponíveis em bactérias, estendeu-se muito, passando a incluir quaisquer elementos móveis presentes tanto em procariontes como em eucariontes, que se movem por meio de moléculas de DNA.

Para compreender melhor o papel dos elementos transponíveis nos genomas de diversas espécies, vamos estudar primeiramente os mecanismos de transposição encontrados em procariontes, por serem os mais bem estudados molecularmente.

#### 2.1.1. Elementos de Inserção (IS)

Esses elementos foram detectados pela primeira vez em *E. coli*, mais especificamente inseridos no operon *gal*. Como são sequências que têm capacidade de mudar de posição no cromossomo, muitas vezes eles interrompem a sequência codificadora de um gene e inativam a expressão do seu produto. Hoje sabemos que o genoma da *E. coli* é muito rico em sequências IS (IS1, IS2, IS3, etc.) constituindo talvez mais de vinte tipos. A estrutura geral de um elemento IS está apresentada na Figura 4.

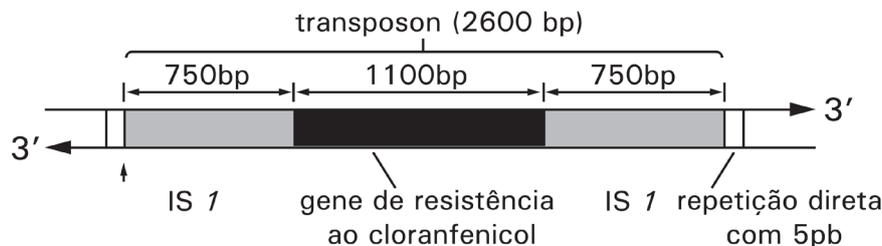
Os elementos IS codificam uma ou duas enzimas, dependendo do mecanismo que usam para a transposição, o qual pode ser de dois tipos básicos: não-replicativo e replicativo. Na transposição não-replicativa, os elementos excisam-se de um local do cromossomo da bactéria e inserem-se em nova posição. Já na transposição replicativa, os IS inserem uma cópia de si mesmos em uma nova posição.



**Figura 4.** Representação esquemática de um elemento IS. A região central (segmento negro) codifica enzimas necessárias à transposição. Flanqueando um elemento IS, existem repetições invertidas e curtas repetições diretas (segmentos listados) originadas no sítio-alvo, durante a mobilização do elemento. Como as repetições diretas dependem do local de inserção, sua seqüência varia, mas seu comprimento é constante para cada elemento IS. As setas indicam a orientação das seqüências.

### 2.1.2. Transposons

Essa classe de elementos de transposição de bactérias, os transposons, são maiores que os elementos de inserção e contêm um ou mais genes que codificam para proteínas, além dos que são indispensáveis para a mobilização do elemento. Muitos transposons são compostos por genes cujos produtos conferem resistência a antibióticos flanqueados por duas cópias dos elementos IS. Os mecanismos de mobilização dos transposons são semelhantes aos dos IS. (Fig 5)



**Figura 5.** Estrutura do transposon Tn9 de *Escherichia coli*, que contém um gene de resistência ao cloranfenicol, flanqueado por duas cópias do IS1.

### 2.1.3. Elementos genéticos móveis de DNA em eucariotos

Alguns elementos móveis de eucariotes também se mobilizam por transferência direta de DNA. Entre eles encontram-se o elemento P de *Drosophila* e os elementos Ac e Ds de milho.

O primeiro elemento transponível a ser identificado geneticamente foi estudado por Barbara McClintock em milho, na década de 50. Nesse material, eles eram responsáveis por mutações instáveis, caracterizadas por perdas e restaurações da função de certos genes. McClintock identificou duas famílias de elementos transponíveis em milho, e várias outras foram caracterizadas mais tarde. Dentro de cada família, alguns dos membros, como os elementos Ac, são capazes de se deslocar autonomamente e outros, como os Ds, deslocam-se apenas em resposta à atividade de um outro elemento autônomo da mesma família, situado em outro local do genoma.

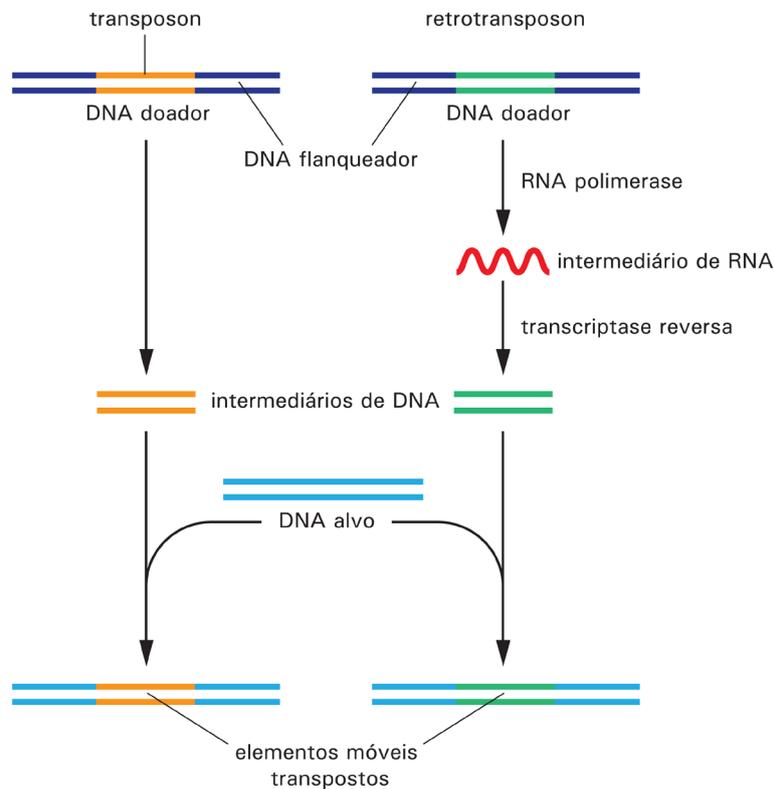
A capacidade de um elemento transponível se deslocar autonomamente depende da existência, em suas seqüências, de um quadro de leitura capaz de codificar polipeptídeo com função enzimática de transposase, necessária para sua excisão e reinserção.

A capacidade de realizar a transposição, autonomamente ou não, depende ainda de uma seqüência de bases específica, presente na extremidade das repetições invertidas terminais, características dos elementos transponíveis em geral, e que constitui um sítio de corte reconhecido pela transposase. Os elementos de uma família que não são capazes de transposição autônoma devem ter se originado por deleções em elementos íntegros: como eles mantêm as repetições invertidas, eles também podem ser alvo da ação das transposases codificadas por outros elementos.

Transposons típicos, que se mobilizam via DNA, não representam frações apreciáveis dos genomas humano ou de mamíferos. Mesmo os elementos P, importantes transposons de *Drosophila*, semelhantes aos elementos Ac e Ds do milho, correspondem no máximo a 0,1% do DNA total dessas espécies. A sua importância nas espécies em que ocorrem reside na sua capacidade de induzir mutações e rearranjos cromossômicos.

## 2.2 Elementos que se mobilizam por RNA (*retrotransposons*)

São elementos que não se mobilizam por deslocamento de sequências de DNA, mas sim através da transcrição reversa do RNA transcrito a partir delas, seguida da integração cromossômica do produto dessa transcrição reversa. (Fig 6). Há três tipos de elementos que se mobilizam por RNA: os semelhantes a retrovírus (retrovírus endógenos), LINES e SINEs. Além disso, há uma quarta categoria de elementos, os pseudogenes processados, cuja origem também está relacionada a eventos de retrotransposição.



**Figura 6.** Acima, esquema comparativo entre o mecanismo de transposição dos transposons de DNA e dos retrotransposons. Observe que na retrotransposição, a mobilização ocorre por meio de um intermediário que se trata de um RNA que é transformado em um DNA de dupla-fita pela ação da transcriptase reversa. Esse DNA é então inserido no DNA alvo, completando a retrotransposição.

### 2.2.1. Elementos semelhantes a retrovírus

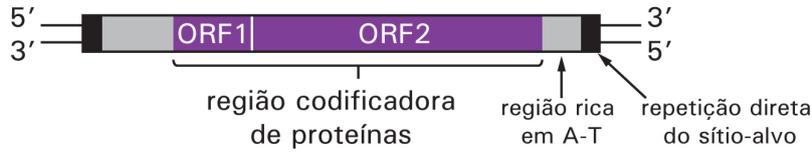
Esses elementos podem ser identificados no genoma por serem flanqueados por longas repetições terminais diretas, as LTRs, de *long terminal repeats*, que estão também presentes nos genomas dos retrovírus. Eles se mobilizam por retrotransposição e são capazes de codificar uma transcriptase reversa. Por essas características, eles se assemelham muito aos genomas de retrovírus integrados (provírus). No caso dos retrovírus, entretanto, além das sequências que codificam a transcriptase reversa, há sequências que codificam as proteínas que constituem a partícula viral infecciosa.

Elementos de transposição semelhantes aos retrovírus foram detectados nos genomas de vários organismos, incluindo os humanos. Em *Drosophila*, por exemplo, esse tipo de retrotransposon é muito frequente, podendo representar de 2 a 10% do genoma total: muitos tipos diferentes já foram caracterizados, havendo de 10 a 80 cópias de cada tipo, no genoma.

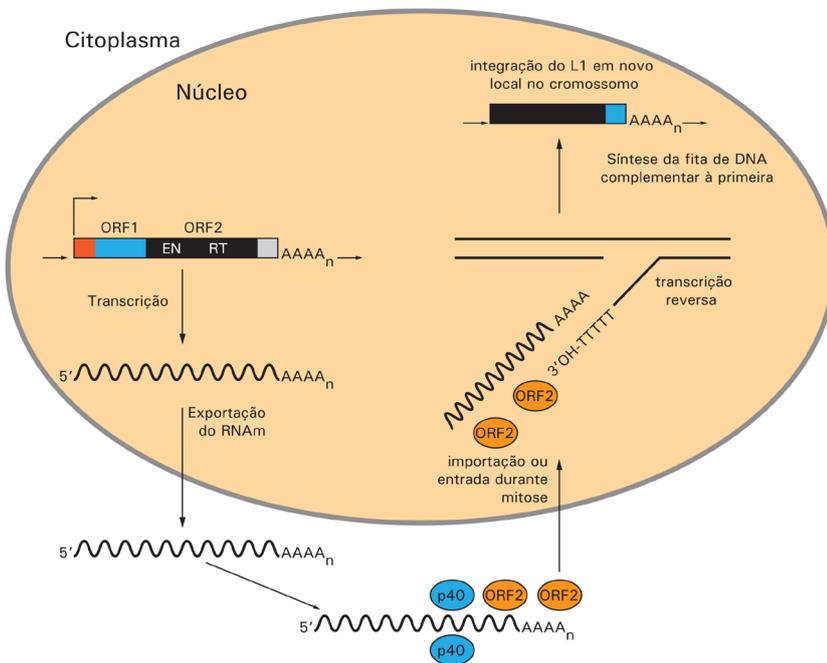
### 2.2.2. "Long Interspersed Elements" (LINEs)

Alguns dos elementos que se movem por retrotransposição são desprovidos de LTRs, mas apresentam sequências codificadoras de proteínas com alta homologia com transcriptases reversas, indicando que esses elementos também podem realizar a retrotransposição. Um exemplo importante dessa classe são os elementos LINE presentes em eucariontes, principalmente em mamíferos. No homem e no camundongo o elemento LINE mais comum é o L1, que nos camundongos pode corresponder a mais de 8% do genoma. (Fig. 7)

No genoma humano há cerca de 850 mil cópias do elemento L1 e há evidências de que a mobilização recente desse elemento seria a responsável por algumas mutações relacionadas a doenças hereditárias, como alguns casos de distrofia muscular e de hemofilia.



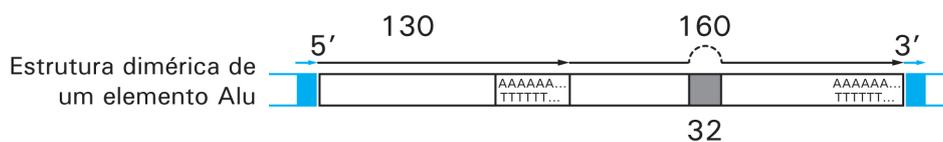
**Figura 7.** Estrutura geral de um elemento LINE L1. O elemento L1 é flanqueado por repetições diretas que diferem em diferentes cópias do elemento. Essas repetições parecem se originar dos sítios-alvos durante a inserção. Próximo à extremidade 5' da região de código tem início um quadro de leitura aberto (ORF1) que se estende por cerca de 1kb, finalizando com muitos códons de parada. A função da proteína codificada pela ORF1 é desconhecida. Uma ORF2, com aproximadamente 4kb, codifica uma proteína semelhante à transcriptase reversa de retrovírus e de elementos de retrotransposição. A extremidade 3' da fita superior direita do L1 é rica em resíduos de adenina, o que sugere que estes elementos tenham se originado a partir de um transcrito da RNA polimerase II (um RNA mensageiro com cauda poli-A). Os L1 não contêm LTRs.



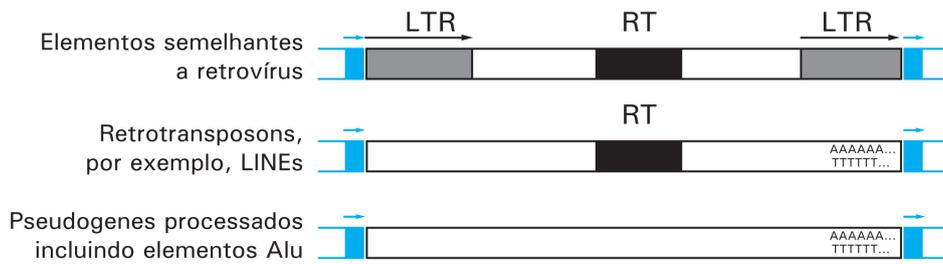
**Figura 8.** Mecanismo proposto para explicar a mobilização de retrotransposons do tipo LINE (por exemplo, L1). A sequência nucleotídica do elemento é transcrita no núcleo. A tradução do RNAm do LINE resulta em uma proteína com atividade de endonuclease e de transcriptase reversa. Essa enzima cliva o DNA em região rica em T, que pode se hibridar com a cauda Poli-A do RNAm do elemento LINE. A extremidade rica em T permite o alongamento da cadeia de DNA pela transcriptase reversa, que copia a sequência do RNAm em sequência de DNA. O DNA de dupla-fita que resulta da atividade da transcriptase reversa é integrado no cromossomo por meio de quebra adicional introduzida do DNA (Figura modificada de Kazazian & Moran, *Nature Reviews in Genetics*, 19:19-24, 1998).

### 2.2.3. "Short Interspersed Elements" (SINEs)

O exemplo mais conhecido de elementos SINE é a sequência *Alu*, presente na espécie humana e nos primatas. Estima-se que estejam presentes em um milhão e 500 mil cópias no genoma humano, totalizando cerca de 10% do genoma. Se utilizarmos uma sonda específica para elementos *Alu* em um "blotting", a maioria dos fragmentos de DNA vai hibridar com essa sonda, já que a sequência *Alu* é muito abundante. Há casos bem documentados na espécie humana em que doenças hereditárias foram causadas pela inserção de um elemento *Alu* em genes funcionais, o que prova que alguns deles ainda são ativos e se mobilizaram recentemente. Genomas de outros mamíferos também possuem sequências do tipo SINE em altas frequências.



**Figura 9.** Estrutura do elemento *Alu*, o elemento SINE mais abundante nos humanos. As duas partes do elemento mostram alta homologia de sequências entre si (exceto por uma inserção de 32 pb no segmento da direita que é maior) e com o RNA 7SL. O segmento da esquerda tem sequências semelhantes ao promotor da polimerase III do RNA.



**Figura 10.** Esquema comparativo da estrutura dos elementos semelhantes a retrovírus, LINEs e elementos Alu. LTR = Long Terminal Repeats e RT = sequência que codifica transcriptase reversa.

Os elementos *Alu* apresentam muita semelhança, quanto à sequência de bases, com um tipo de RNA conhecido como 7SL. Este RNA é um componente da “partícula reconhecadora de sinal”, uma ribonucleoproteína que auxilia na secreção de polipeptídeos recém-formados, na sua exportação através do retículo endoplasmático. É possível que a origem dos elementos *Alu* esteja relacionada à retrotransposição do RNA 7SL. Assim como os elementos SINE, as sequências *Alu* também possuem uma extremidade 3' rica em resíduos de A, o que indica sua origem provável por retrotransposição (Fig. 9). Com base nessa hipótese, podemos afirmar que *Alu* é um tipo de pseudogene processado, pois seria um pseudogene modificado da sequência que transcree o RNA 7SL.

É muito intrigante a abundância de elementos *Alu* no genoma de primatas, em especial nos humanos. O número estimado de cópias de elementos *Alu* no genoma humano é superior ao número estimado de elementos do tipo L1. Esse achado é mais intrigante se levarmos em conta que um elemento *Alu* não tem sequência nucleotídica que codifica uma transcriptase reversa. Sua abundância só pode ser explicada se admitirmos que o elemento *Alu* se retrotranspõe utilizando a transcriptase reversa sintetizada a partir do RNAm transcrito de outro elemento retrotransponível, como por exemplo, os elementos do tipo LINEs.

A variabilidade das populações humanas tem sido estudada em relação aos polimorfismos de inserção de sequências *Alu*. Como várias dessas inserções são recentes na história dos humanos, nem todas estão presentes em uma dada posição em todos os seres humanos, caracterizando, portanto, polimorfismos na nossa espécie. A comparação entre diferentes populações atuais, com relação à presença ou ausência dessas inserções, tem contribuído para os estudos evolutivos dessas populações, especialmente quanto às suas migrações.

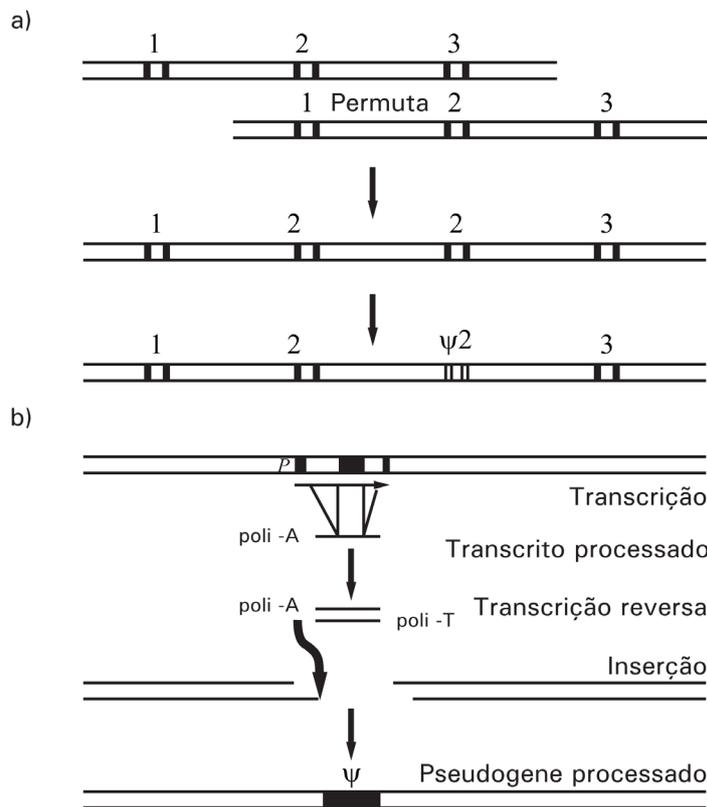
#### 2.2.4. Pseudogenes e pseudogenes processados.

É muito comum, ao se hibridar uma biblioteca genômica com uma sonda de cDNA, encontrar muitos sinais de hibridação, além do esperado para um gene de cópia única. O sequenciamento desses clones genômicos revelam a presença de sequências que, embora muito parecidas com a do gene utilizado como sonda, estão inativas ou pouco ativas em relação à transcrição e à tradução, em decorrência de várias mutações. Essas sequências inativas são chamadas de **pseudogenes** e foram classificadas basicamente em dois tipos.

No primeiro tipo, as sequências não-funcionais localizam-se próximas a agrupamentos de cópias funcionais do mesmo gene. O exemplo clássico é representado pelos conjuntos dos genes da alfa e beta globinas dos mamíferos, em que ocorrem vários pseudogenes, além dos genes funcionais. Acredita-se que estas sequências tenham surgido através de recombinações desiguais entre os membros de um arranjo de genes repetidos em tandem: um dos produtos do “*crossing over*” desigual daria origem a um cromossomo em que um gene funcional foi perdido, e que deverá, portanto, ser eliminado pela seleção natural. O produto recíproco da permuta vai originar um cromossomo com uma cópia a mais do gene em questão, que poderá então, ao acaso, sofrer mutações sem prejudicar o organismo. (Figura 11a).

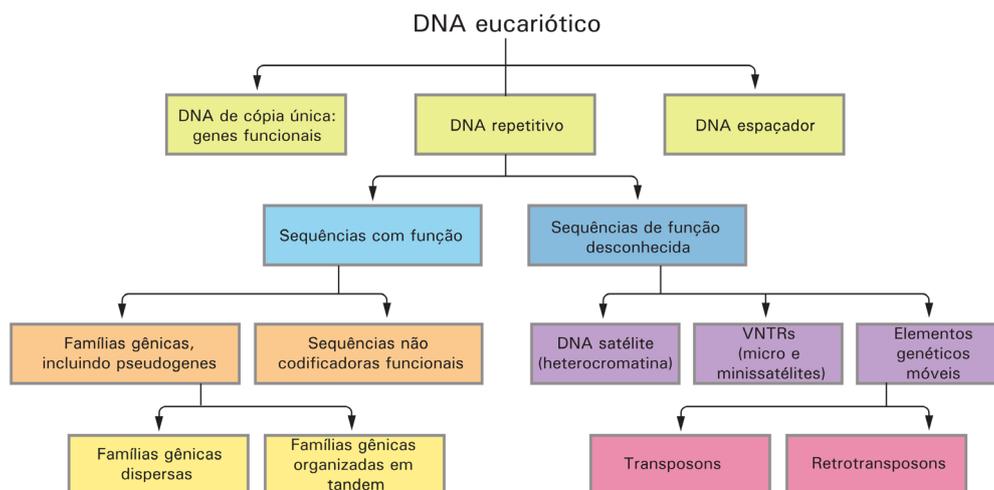
Uma segunda categoria de pseudogenes usualmente não pertence a arranjos de genes repetitivos e sua tendência é ocorrer em posições bem distantes dos seus equivalentes funcionais. A sua estrutura sugere a origem através de um mecanismo semelhante à retrotransposição: eles não possuem sequências promotoras da RNA pol II e, portanto, não são mais transcritos. Entretanto, possuem cauda de poli-A em sua extremidade 3', e não apresentam introns. Por estes motivos, eles são denominados **pseudogenes processados**, para distingui-los do primeiro tipo. Supõe-se que sua origem está relacionada à atividade da transcriptase reversa codificada por outros elementos móveis.

Os pseudogenes processados são muito numerosos e por serem desprovidos de introns, correspondem a quantidades menos importantes para o total de DNA da célula (Figura 11b).



**Figura 11.** Acima, duas possíveis origens para os pseudogenes: (a) recombinação desigual entre conjuntos de genes repetitivos, resultando em cópia adicional que, por acúmulo de mutações, transforma-se em um pseudogene ( $\psi$ ); (b) integração no cromossomo de DNA originado por transcrição reversa de um transcrito da polimerase II do RNA já processado, isto é, desprovido de introns e da cauda 3' de poli-A.

Finalmente, todo o DNA restante presente nos genomas dos organismos e que não pode ser classificado em nenhuma dessas categorias que estudamos é chamado por alguns autores de DNA espaçador. Na verdade, esse termo foi criado para congregiar nossa ignorância sobre a estrutura e a função de diversos tipos de sequências que não conseguimos ainda classificar. À medida que nosso conhecimento sobre a estrutura dos genomas aumenta, é bem provável que a quantidade de DNA considerado espaçador venha a se reduzir. A Figura 12 apresenta uma proposta de classificação da sequências de DNA que ocorrem em um genoma complexo.



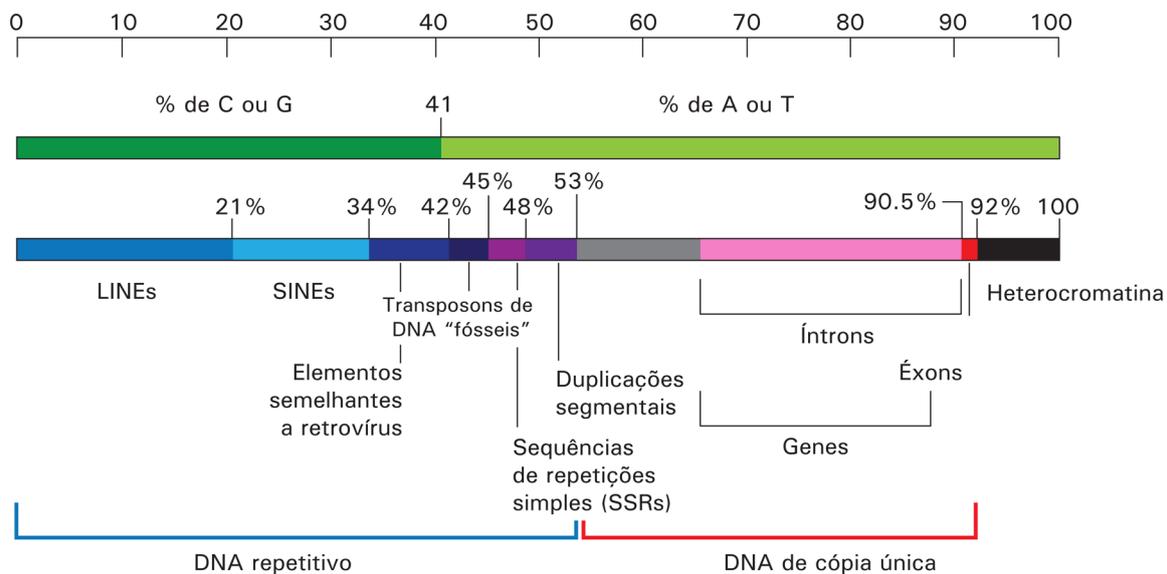
**Figura 12.** Classificação das sequências presentes nos genomas dos eucarióticos complexos.

## IV - As Sequências Repetitivas e o Projeto do Genoma Humano

Em 2001, as revistas científicas *Science* e *Nature* publicaram importantes artigos que procuraram resumir os principais achados relacionados à conclusão do chamado “rascunho” do Projeto do Genoma Humano. Na verdade, a conclusão do sequenciamento do genoma humano, mais completo e com melhor qualidade, veio alguns anos depois. A revista *Nature* publicou o resumo dos achados do projeto do genoma conhecido como “público”, coordenado por Francis Collins, e a revista *Science* publicou os achados do projeto genoma conduzido pela empresa privada, a Celera, liderada por Craig Venter.

Um caderno comemorativo desses fatos foi também publicado pela editora da revista *Nature*, ainda em 2001, e dessa edição especial foi retirado o esquema apresentado na Figura 13.

Esses achados confirmaram o que trabalhos mais antigos já apontavam: quase metade do genoma humano consiste de sequências repetitivas, e a maior parte delas, cerca de 45% do total, são derivadas de sequências relacionadas aos elementos de transposição. O número de elementos repetitivos é dos mais notáveis entre todos os genomas estudados: em uma bactéria, cerca de 1,5% do DNA é repetitivo e apenas 3% é desse tipo nas moscas. A maior parte do conteúdo repetitivo do nosso genoma é representada por elementos móveis que já perderam a capacidade de se mover. Apenas dois tipos ainda são capazes de mobilização: as sequências *Alu* e o elemento LINE 1. *Alu* e LINE 1 perfazem 60% do DNA repetitivo do nosso genoma. Surpreendentemente, cerca de um milhão de cópias de *Alu* devem estar presentes no genoma humano.



**Figura 13.** A composição do genoma humano: cerca de 41% é composto por bases C ou G. Cerca de metade do genoma é composto por sequências repetitivas. A maior parte das sequências repetitivas são elementos que resultam de transposição, como LINEs e SINEs, o mais abundante elemento é, sem dúvida, o *Alu*. Os genes incluindo seus íntrons ocupam um quarto do genoma, mas apenas 1,5% do genoma efetivamente codifica polipeptídeos, ou seja, é constituído por éxons. Modificada a partir do livro “*The Human Genome*”, editores Dennis e Gallagher, *Nature Publishing Group* New York, 2001.

## V - O DNA das Organelas

Embora a maior parte do DNA na maioria dos organismos eucariontes esteja presente no núcleo, uma parte do DNA celular está presente nas mitocôndrias dos animais, das plantas e dos fungos e também nos cloroplastos das plantas. Essas organelas são importantes para as células do ponto de vista energético, pois nas mitocôndrias ocorre a síntese de ATP e nos cloroplastos, a fotossíntese. Diversas linhas de evidência indicam que a mitocôndria e o cloroplasto evoluíram a partir de bactérias que sofreram endocitose e foram incorporadas em células ancestrais que continham um núcleo eucariótico, formando endosimbiontes. Com o passar do tempo evolutivo, a maior parte dos genes bacterianos que codificam componentes das organelas atuais foram transferidos para os cromossomos, no núcleo. Contudo, a mitocôndria e o cloroplasto ainda mantêm DNAs circulares que codificam proteínas essenciais ao funcionamento dessas organelas, assim como RNAs transportadores e ribossômicos necessários para a síntese protéica que ocorre no seu interior. Assim, as células eucariontes possuem sistemas genéticos múltiplos: um sistema predominantemente nuclear e sistemas secundários, com DNA próprio, como os que ocorrem nas mitocôndrias e cloroplastos.

Quando ocorre a mitose, cada célula filha recebe aproximadamente o mesmo número de mitocôndrias, mas como não há um mecanismo que garanta números exatamente iguais de mitocôndrias nas células filhas, algumas células podem conter mais DNA mitocondrial que outras. Todas as mitocôndrias nas células eucariontes contêm várias cópias das moléculas do DNA mitocondrial. Logo, a quantidade de DNA mitocondrial de uma célula depende do número de mitocôndrias, do tamanho do DNA mitocondrial da espécie e do número de moléculas de DNAm<sub>t</sub> por mitocôndria.

A herança do DNA mitocondrial em leveduras é biparental: durante a fusão de células haplóides, ambos os pais contribuem igualmente com o citoplasma da célula diplóide resultante. Nos mamíferos e na maior parte dos outros animais até agora estudados, o espermatozóide contribui com pouquíssimo ou nenhum DNA mitocondrial ao zigoto e, na prática, todas as mitocôndrias do embrião são derivadas do óvulo, e não do espermatozóide. Estudos realizados em camundongos demonstraram que 99,99% do DNA mitocondrial tem origem materna, mas uma pequena parte (0,01%) é herdada do pai. Em plantas superiores, o DNA mitocondrial é herdado de maneira exclusivamente materna (óvulo) e não do grão de pólen.

O DNA mitocondrial (às vezes chamado de genoma mitocondrial) em diversos organismos já foi completamente sequenciado e o DNA mitocondrial obtido de todas essas fontes geralmente codifica RNAs ribossômicos, RNAs transportadores e algumas proteínas mitocondriais essenciais. Todas as proteínas codificadas por genes mitocondriais são traduzidas nos ribossomos mitocondriais. A maioria dos polipeptídeos sintetizados nas mitocôndrias não são enzimas completas, mas sim subunidades de complexos multiméricos associados à cadeia de transporte de elétrons ou à síntese de ATP. A maioria das proteínas localizadas na mitocôndria, como as DNAs polimerases e RNAs polimerases, no entanto, são sintetizadas nos ribossomos citoplasmáticos e são importadas para essa organela.

O tamanho e a capacidade de codificação do DNA mitocondrial varia muito entre os diferentes organismos.

O DNA mitocondrial dos mamíferos, ao contrário do DNA nuclear, não apresenta íntrons e nem sequências não-codificadoras longas.

O DNA mitocondrial dos invertebrados é aproximadamente do mesmo tamanho do DNA mitocondrial humano, mas o DNA mitocondrial da levedura é cerca de cinco vezes maior (cerca de 78.000 pb) quando comparado ao DNA dos mamíferos. O DNA mitocondrial das leveduras e de outros eucariontes inferiores codificam muitos produtos gênicos em comum, mas outros genes são encontrados no núcleo das células dos mamíferos.

Ao contrário dos outros eucariontes, que contêm um único tipo de DNAm<sub>t</sub>, as plantas têm diversos tipos de DNA mitocondrial que podem sofrer recombinação entre si. Os DNAs mitocondriais encontrados em plantas são muito maiores e muito mais variáveis em tamanho do que nos demais organismos. Mesmo dentro de uma única família de plantas, o DNAm<sub>t</sub> pode variar por um fator de cerca de oito vezes (330.000 pb na melancia e 2.500.000 pb em um tipo de melão). Em contraste com o que ocorre entre os animais, leveduras e fungos, o DNA mitocondrial das plantas contém genes que codificam o RNA ribossômico 5S, que só ocorre no ribossomo das plantas. Os RNAs ribossômicos das plantas também são maiores do que os encontrados nos demais eucariontes. O recente sequenciamento dos DNAm<sub>t</sub> obtidos de plantas indicaram que a presença de longas regiões não codificadoras e duplicações explicam em parte o seu maior tamanho.

As diferenças nos tamanhos e na capacidade de código dos DNAs mitocondriais dos vários organismos indicam que durante a evolução houve a provável movimentação do DNA entre a mitocôndria e o núcleo. Evidências diretas desse movimento vieram da observação de que várias proteínas codificadas pelo DNA mitocondrial em algumas espécies são codificadas por genes nucleares em outras espécies. Parece que genes inteiros moveram-se das mitocôndrias para o núcleo durante o processo evolutivo. O mais notável exemplo desse fenômeno ocorre com o gene *cox II*, que codifica a subunidade 2 da citocromo c oxidase. Esse gene é encontrado no DNAm<sub>t</sub> em quase todos os organismos, exceto em uma espécie de leguminosa. Muitos RNAs transcritos a partir de genes mitocondriais de plantas são editados, principalmente por uma conversão catalizada por enzimas de alguns resíduos de C em U, e, ocasionalmente, de U para C. O gene nuclear *cox II* da leguminosa se parece mais com a versão editada do gene *cox II* encontrado nas outras espécies de leguminosas. Esses fatos são fortes evidências de que o gene *cox II* se moveu para o núcleo por meio de um processo que envolveu um intermediário de RNA. Possivelmente, essa transferência ocorreu por um mecanismo de transcrição reversa, similar ao processo pelo qual se originam os pseudogenes processados presentes no núcleo, a partir de RNAs mensageiros transcritos de genes nucleares.

## Os produtos dos genes mitocondriais não são exportados

Até o momento, todos os transcritos conhecidos do DNA mitocondrial e seus produtos traduzidos permanecem na mitocôndria e todas as proteínas codificadas pelos genes mitocondriais são sintetizadas nos ribossomos mitocondriais. O DNA mitocondrial codifica os RNAs ribossômicos que formam os ribossomos mitocondriais, embora a maioria das proteínas presentes no ribossomo sejam codificadas pelo DNA nuclear. Na maior parte dos eucariontes, todos os RNAs utilizados na síntese proteica mitocondrial são codificados pelo DNA mitocondrial. No entanto, no trigo e no *Trypanosoma brucei* (o causador da doença do sono), em protozoários ciliados, todos os RNAs transportadores são codificados pelo DNA nuclear e importados pela mitocôndria.

Como consequência de sua origem bacteriana, os ribossomos mitocondriais assemelham-se aos ribossomos bacterianos e são diferentes dos ribossomos citoplasmáticos na sua composição proteica e de RNAs, no tamanho e na sua sensibilidade a certos antibióticos. Por exemplo, o cloranfenicol bloqueia a síntese proteica bacteriana e também a síntese proteica dos ribossomos mitocondriais, mas não a dos ribossomos do citoplasma celular. Por outro lado, a cicloheximida bloqueia a síntese proteica nos ribossomos das células eucariontes, mas não afeta a síntese nos ribossomos mitocondriais e bacterianos. Em células de mamíferos, as únicas proteínas sintetizadas na presença da cicloheximida são as mitocondriais.

## O código genético mitocondrial é diferente do código genético padrão

O código genético utilizado nas mitocôndrias dos animais e dos fungos é diferente do código genético padrão empregado por todos os genes bacterianos ou eucarióticos nucleares. Surpreendentemente, esse código mitocondrial também é diferente entre diversas espécies já estudadas (Tabela 2). O porquê desse fenômeno ter ocorrido na história evolutiva ainda é um mistério. UGA, por exemplo, normalmente é um codon de parada, mas é interpretado como triptofano no sistema de síntese de proteínas nas mitocôndrias humanas e de fungos. No entanto, nas mitocôndrias de plantas, UGA permanece como um codon de parada; AGA e AGG que normalmente codificam a arginina, também significam arginina nas mitocôndrias de fungos e plantas, mas são codons de parada no DNAmt de mamíferos e significam serina no DNAmt de *Drosophila*.

Como demonstrado na Tabela 2, as mitocôndrias de plantas parecem usar o código genético padrão. Contudo, comparações entre as sequências de aminoácidos de proteínas mitocondriais das plantas e as sequências de DNAmt de plantas sugeriram que CGG poderia significar arginina ou triptofano. Essa aparente falta de especificidade do código no DNAmt das plantas é explicada pelo processo de edição dos RNAs transcritos a partir do DNAmt, que pode converter citosina em uracila. Se uma sequência CGG é editada em UGG, o codon vai especificar triptofano e não arginina.

## O DNA mitocondrial humano

O DNA mitocondrial humano é uma molécula de DNA circular de dupla-fita cuja sequência completa de nucleotídeos foi estabelecida em 1991. Tem 16569 pb de comprimento. As duas fitas do DNA têm composição de bases diferente e por isso receberam diferentes nomes: a cadeia H (de *heavy* ou pesada) é rica em guaninas e a cadeia L (de *light* ou leve) é mais rica em citosinas. Embora o DNA mitocondrial humano seja predominantemente de dupla-fita, em um pequeno segmento ocorre DNA com tripla-fita, estrutura que resulta da síntese adicional de um pequeno segmento do DNA mitocondrial, perto da região que chamamos de hipervariável. Células humanas contêm geralmente milhares de cópias do DNA mitocondrial. O DNA mitocondrial humano corresponde a cerca de 0,5% do DNA total de uma célula somática.

Na Figura 14 está representado o DNA mitocondrial humano que contém 37 genes: 28 são presentes na cadeia pesada e nove na cadeia leve. Um total de 13 dos 37 genes que codificam polipeptídeos que são sintetizados nos ribossomos mitocondriais. Eles codificam alguns elementos dos complexos respiratórios, enzimas relacionadas à fosforilação oxidativa. Todas as demais proteínas mitocondriais são codificadas pelo genoma nuclear e são sintetizadas pelos ribossomos do citoplasma celular antes de serem importadas pela mitocôndria. Os demais 24 genes mitocondriais codificam 22 tipos de RNAs transportadores e dois tipos de moléculas de RNA ribossômico, que constituem parte do aparato de síntese proteica mitocondrial. Esses genes ocupam a maior parte da extensão do DNA mitocondrial. No entanto, existe uma região não codificadora de aproximadamente 1200 nucleotídeos que é conhecida por nomes diferentes: alça D (*D loop*) ou região de controle. O termo região de controle se refere



DNA mitocondrial em teoria não afetam sua capacidade de replicação. Ao contrário, foram descritos casos em que grandes deleções do DNA mitocondrial acarretaram vantagem replicativa.

Todas as células contêm mitocôndrias, mas, curiosamente, as mutações mitocondriais afetam preferencialmente certos tecidos. Aqueles que são os mais comprometidos geralmente são aqueles com elevada demanda de ATP produzido pela cadeia de fosforilação oxidativa, ou aqueles com demandas rápidas e muito variáveis de ATP. A neuropatia óptica hereditária de Leber (degeneração no nervo óptico acompanhada de cegueira progressiva), por exemplo, é causada por uma mutação de sentido errado no gene mitocondrial que codifica a subunidade 4 da NADH-CoQ redutase. Qualquer uma das grandes deleções de DNA mitocondrial que já foram estudadas causam diversas outras doenças, como oftalmoplegia externa crônica progressiva e síndrome de Kearns-Sayre, caracterizadas por defeitos de visão e degenerações do sistema nervoso central. Um outro tipo de condição, que se caracteriza pela presença de fibras musculares vermelhas esgarçadas e movimentos atáxicos pode ser devida a mutações no gene do RNAt mitocondrial da lisina. Como resultado dessa mutação, a síntese de diversas proteínas mitocondriais parece estar comprometida. O interessante a respeito das doenças resultantes de mutações mitocondriais na espécie humana é que, salvo uma única exceção descrita, elas sempre têm herança materna. Em outras palavras, as mulheres transmitem a doença a seus filhos e filhas, mas homens afetados jamais transmitem a doença a seus descendentes.

## Questões para orientar a leitura: Parte I

1. No que consiste o chamado paradoxo do valor C? Discuta-o com base nos valores apresentados na Tabela 1.
2. Que características dos genes eucarióticos acabam por contribuir para o grande tamanho do genoma presente em muitos desses organismos?
3. O que os experimentos de reassociação do DNA revelaram sobre a organização dos genomas mais complexos?
4. De que modo podem estar organizadas as várias cópias de genes repetitivos no genoma eucariótico?
5. No que consiste o chamado DNA satélite, e o que se sabe a respeito do seu eventual papel no genoma dos organismos superiores?
6. Que tipos de sequências de DNA são importantes à manutenção, segregação e integridade dos cromossomos?
7. Quais as possíveis aplicações laboratoriais das sequências repetitivas variáveis?
8. A que se atribui a disseminação dos chamados “DNA parasitas” no genoma?
9. Quais as características dos retrotransposons que os tornam muito semelhantes aos retrovírus?
10. No que consistem as chamadas sequências *Alu*? Dado que elas não codificam transcriptase reversa, como atingiram elevado número de cópias?
11. O que há em comum entre sequências *Alu* (SINES), LINEs e sequências semelhantes a retrovírus (retrovírus endógenos)?
12. Defina pseudogenes. Que tipos de mecanismos podem estar envolvidos na sua origem? Discuta as evidências em favor de cada um deles.
13. Que tipos de sequências dos genomas poderiam ser chamados de DNA egoísta?
14. Com base no texto, podemos afirmar que os humanos produzem transcriptase reversa?

## Questões para orientar a leitura: Parte II

1. Quais os tipos de genes encontrados no DNA mitocondrial? Como estão organizados no genoma?
2. Qual é a relação entre o genoma da mitocôndria e o do núcleo?
3. Compare a organização gênica do genoma mitocondrial com a dos procariotos.
4. Discuta a afirmação: o código genético é universal.
5. Qual a hipótese mais aceita atualmente para explicar a origem das mitocôndrias?
6. Quais as características de herança do DNA mitocondrial?
7. Desenhe uma genealogia em que segrega uma doença decorrente de mutação no DNA mitocondrial em humanos.
8. Explique porque o quadro clínico das doenças de herança mitocondrial pode variar muito, mesmo entre indivíduos da mesma família.