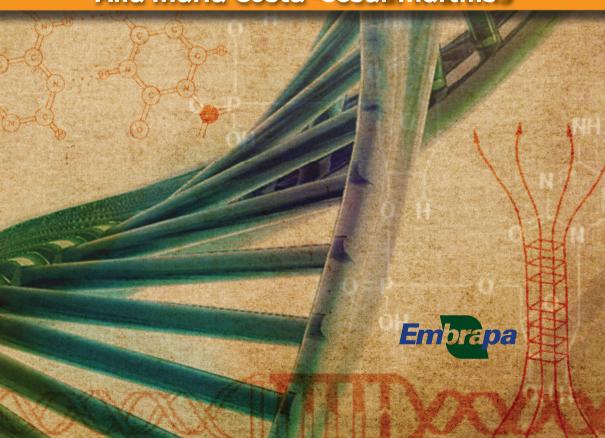


Ana Maria Costa César Martins



Genomas

Empresa Brasileira de Pesquisa Agropecuária Embrapa Cerrados Ministério da Agricultura, Pecuária e Abastecimento



Ana Maria Costa – César Martins

Embrapa Cerrados Planaltina, DF 2010 Exemplares desta publicação podem ser adquiridos na:

Embrapa Cerrados

BR 020, Km 18, Rodovia Brasília/Fortaleza Caixa Postal 08223 CEP 73310-970 – Planaltina-DF Fone (61) 3388-9898 – Fax (61) 3388-9879 www.cpac.embrapa.br sac@cpac.embrapa.br

Coordenação editorial Jussara Flores de Oliveira Arbués

Equipe de revisão Fernanda Vidigal Cabral de Miranda Francisca Elijani do Nascimento Jussara Flores de Oliveira Arbués

Normalização bibliográfica Shirley da Luz Soares de Araújo

Capa, projeto gráfico e diagramação Fabiano Bastos

1ª edição

1ª impressão (2010): 500 exemplares

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei n° 9.610).

Dados Internacionais de Catalogação na Publicação - CIP Embrapa Cerrados

C 937e Costa, Ana Maria.

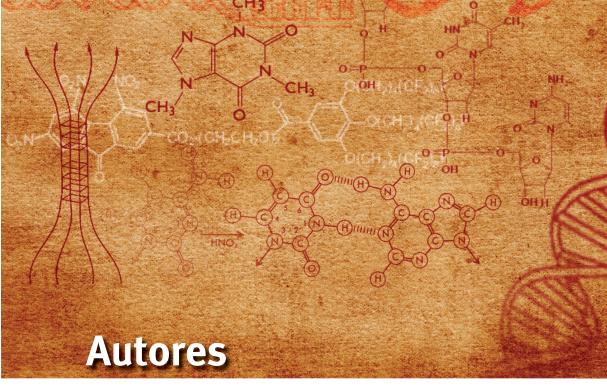
Estrutura e evolução dos genomas / Ana Maria Costa, Cesar Martins. – Planaltina, DF : Embrapa Cerrados, 2010.

110 p.: il.. color.

ISBN 978-85-7075-055-6

1. Genética. 2. Genoma. 3. Evolução. I. Martins, C.. II. Título.

572.86



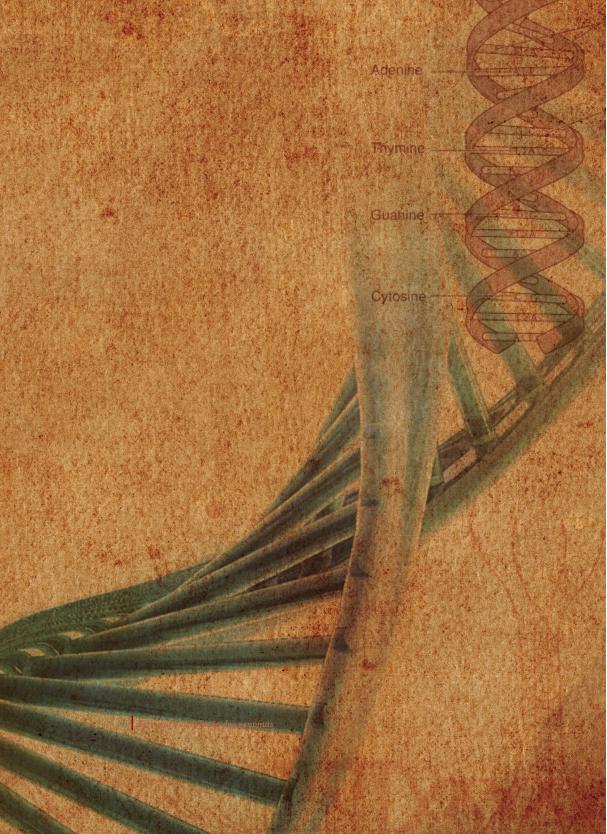
Ana Maria Costa

Engenheira agrônoma, D.Sc. Pesquisadora da Embrapa Cerrados abarros@cpac.embrapa.br

Cesar Martins

Professor Adjunto, D.Sc. Departamento de Morfologia Instituto de Biociências UNESP - Universidade Estadual Paulista







O retrato que se vê hoje dos seres vivos é o resultado das variações genéticas ocorridas ao longo dos milhões de anos. É comum a impressão de se estar diante do final evolutivo, em que toda diversidade e variabilidade existentes já foram alcançadas. Contudo, o tempo passa e as alterações genéticas continuam a acontecer tanto para o homem quanto para todos os seres vivos. Acreditar que o homem possa deter esse processo é um erro.

O ambiente contribui na seleção dos indivíduos mais aptos, mas também é um agente desencadeador de processos celulares que resultam na alteração genômica.

Os mecanismos que geram variabilidade são tão poderosos que nem sempre é fácil manter a uniformidade genética nas situações onde ela é desejável, como nos casos dos cultivos de microorganismos para fins de pesquisa ou industrial, alguns sistemas de produção agrícola e conservação de recursos genéticos. Em geral, essas

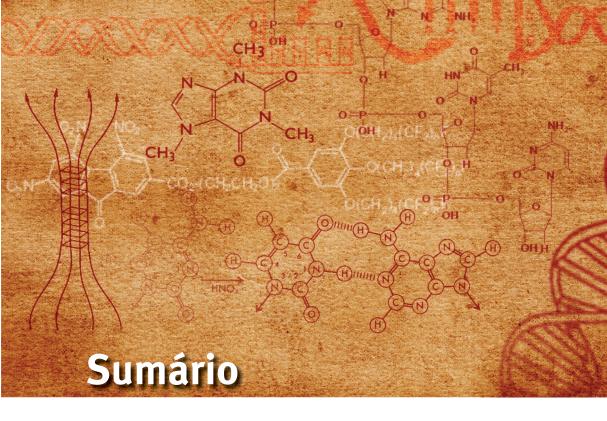


atividades exigem constantes esforços para selecionar e recuperar a linhagem original.

O livro *Estrutura e Evolução de Genomas* faz um apanhado geral dos esforços para se conhecer o material genético, apresenta conceitos básicos de biologia molecular e sintetiza os conhecimentos gerados nos projetos genomas quanto à estrutura e organização cromossômica das espécies. Finalizando, apresenta e discute os mecanismos moleculares que promovem variabilidade e diversidade biológica.

Os assuntos apresentados contribuem para o entendimento da dinamicidade dos processos evolutivos importantes aos programas de conservação e melhoramento genético.

José Robson Bezerra Sereno Chefe-Geral da Embrapa Cerrados



O Genoma e sua História
Uma Visão Geral dos Genomas: de Vírus a Humanos 19
Genoma viral
Vírus de DNA
Vírus de RNA fita simples positiva22
Vírus de RNA fita simples negativa
Vírus de RNA fita dupla
Vírus de transcrição reversa
Genoma bacteriano
Estrutura do genoma bacteriano
Evolução e dinâmica do genoma bacteriano
Organização dos genes bacterianos

Genoma eucarionte
Genoma das organelas
Genoma dos cloroplastos e mitocôndrias
Genoma mitocondrial de tripanossomatídeos 39
Organização Estrutural e Instabilidade do Material Genético 43
Elementos repetitivos
Elementos repetitivos móveis longos 50
Elementos repetitivos móveis curtos
Instabilidade de sequências
Repetições do tipo satélites
Instabilidade de insertos clonados
Instabilidade e estruturas não-B 62
Evolução do Genoma
Mutação intragênica
Embaralhamento de éxons
A transferência horizontal de genes
Transferência horizontal de genes e características estruturais
do genoma
Transferência horizontal de genes de sequências de
minicírculos de kDNA de Trypanosoma cruzi para o genoma
da célula hospedeira
Duplicação gênica: a maior força da evolução 80
Considerações Finais
Referências Bibliográficas



A palavra "genoma" foi cunhada por Hans Winkler, em 1920, como uma conjugação de gene e cromossomo. No entanto, o conceito geral de genoma pode ser atribuído ao século IV, antes da Era Cristã, quando Aristóteles apontou os primeiros conceitos em relação à hereditariedade. Embora os trabalhos de Mendel, no final do século XIV, tenham propiciado grandes contribuições no campo da hereditariedade, essa área mostrava-se, ainda, bastante abstrata. Com o avanço dos métodos científicos e das tecnologias, a hereditariedade passou a ser associada às estruturas presentes no núcleo das células chamadas cromossomos (final do século XIX e início do século XX) e finalmente com os polímeros de nucleotídeos duplafita chamados de DNA (meados do século XX), que formam os cromossomos.

Anteriormente ao advento do sequenciamento de moléculas de DNA, os cromossomos forneceram as primeiras noções detalhadas

sobre a organização física do genoma. Mesmo após avanços significativos obtidos por meio do sequenciamento completo de genomas, os cromossomos continuam sendo a base para a compreensão do genoma, já que possibilitam integrar diversos tipos de informações na geração de mapas genéticos físicos detalhados das espécies.

As primeiras evidências da existência do DNA, no núcleo das células, foram obtidas pelo jovem cientista suiço Friedrich Miescher, em 1869. Miescher, aos 25 anos de idade, identificou uma substância rica em fósforo que não se enquadrava como nenhuma das substâncias proteicas conhecidas até aquele momento. A essa substância ele deu o nome de nucleína. Embora alguns anos mais tarde Walther Flemming (1889) tenha cunhado a palavra cromatina e Wilhelm Waldeyer (1888) o termo *cromossomo*, nenhuma relação clara havia sido feita entre cromatina/cromossomo e nucleína. Em 1889, Richard Altman obteve amostras de nucleína livres de proteína e cunhou uma denominação mais apropriada para a substância: ácido nucleico. Após esse período, a composição das moléculas de ácido nucleico foi deduzida e, por volta de 1930, a nucleína se tornou o ácido nucleico desoxirribose e, posteriormente, ácido desoxirribonucleico. Até aquele momento, existiam questionamentos acerca da molécula que carregava a informação hereditária. As primeiras evidências de que o DNA era o material genético foram obtidas por Avery, MacLeod e MacCarty, em 1944. A partir desse momento, importantes descobertas ocorreram nos anos seguintes, incluindo as primeiras hipóteses sobre a estrutura do DNA e as descobertas que levaram à síntese de Watson e Crick, em 1953, sobre a estrutura do DNA e a subsequente decifração do código genético. A história dos estudos em genoma acompanham os avanços recentes em genética e biologia molecular. Ao contrário de outras áreas como a física, matemática e filosofia, essa área é muito recente e vem se desenvolvendo de forma muito dinâmica (Fig. 1). Dessa forma, a genômica como ciência existe desde os primeiros estudos básicos sobre estrutura cromossômica, chegando aos avanços recentes relacionados ao

sequenciamento de genomas inteiros. Assim, o genoma pode ser entendido como o complemento gênico completo (no sentido de genótipo) ou o DNA total presente no lote cromossômico haplóide.

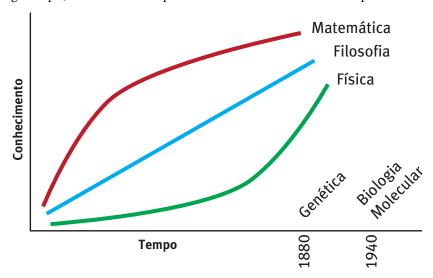


Fig. 1. Avanços no conhecimento, ao longo do tempo, nas diferentes áreas da ciência e da filosofia.

Por volta de 1950, diversos trabalhos começaram a utilizar a terminologia classes de DNA como referência ao conteúdo de DNA de células diplóides (Classe I) e haplóides (Classe IC). Uma forma derivada utilizada com base nesses trabalhos para definir a quantidade de DNA em um núcleo haplóide foi o "valor-C". Uma vez que o DNA é o material hereditário, poderíamos considerar que as variações detectadas no tamanho do genoma estariam refletindo variações no número de genes. No entanto, essa consideração passou a perder sua validade a partir das comparações que começaram a ser feitas entre espécies.

Em 1951, Mirsky e Ris publicaram um trabalho sobre o conteúdo de DNA das células de uma ampla gama de organismos animais, mostrando um verdadeiro paradoxo (paradoxo do valor-C), uma vez que o aumento da complexidade não estava correlacionado com aumento no genoma e do número de genes. No entanto, com a descoberta de que grande parte do DNA eucariótico é representada por sequências não-codificaroras, o paradoxo do valor-C parecia parcialmente resolvido. Entretanto, essa questão foi, ainda, tema de debates que persistiram durante a segunda metade do século XX, envolvendo diversos aspectos relacionados à evolução do genoma.

Diversos trabalhos de TR Gregory, publicados nos anos subsequentes a 2000, enfatizam que a evolução do tamanho do genoma é mais bem definida como um complexo quebra-cabeça denominado Enigma do Valor-C. Para compreender esse enigma, precisamos responder a diferentes e independentes questões que norteiam diversos aspectos da biologia das espécies, tais como: (1) variação do tamanho do genoma nas diferentes taxas; (2) quantidade, tipo e distribuição de DNA não-codificador nos diferentes organismos; (3) mecanismos de perda e ganho de DNA ao longo do tempo; e (4) função do DNA não-codificador.

As últimas décadas do século XX representaram um período particularmente excitante para a biologia. Novos métodos de análises de proteínas, DNA e RNA levaram a uma explosão de informações que permitiram aos cientistas estudar os organismos, as células e suas moléculas de forma mais detalhada. Essas novas tecnologias têm permitido não somente decifrar as informações, mas, também, a forma com que os componentes moleculares interagem, formando células e organismos complexos e funcionais.

Até o início dos anos de 1970, o DNA era uma molécula difícil de ser analisada bioquimicamente. Hoje, graças ao desenvolvimento das tecnologias de manipulação do DNA, ela é uma das moléculas mais fáceis de ser estudada. Regiões específicas do DNA podem ser facilmente isoladas, propagadas em número inimaginável de vezes e sequências nucleotídicas de milhões de pares de bases podem ser conhecidas em questão de horas. O desenvolvimento de três tecnologias - em particular, a restrição enzimática e a reação em cadeia da

polimerase (*Polymerase Chain Reaction* – PCR) e o sequenciamento do DNA - possibilitou avanços significativos na manipulação e no conhecimento do genoma.

Em 1962, W. Arber forneceu as primeiras evidências da existência das nucleases de restrição, que são enzimas bacterianas de clivagem de DNA que reconhecem sequências curtas de bases e fazem um corte na dupla hélice do DNA. Sua função na célula bacteriana é cortar e destruir DNAs virais estranhos que possam ter entrado na célula. Um aspecto importante é que as enzimas de restrição não clivam aleatoriamente o DNA e sim em sequências-alvo-específicas, o que faz delas uma ferramenta poderosa na manipulação do DNA. Qualquer molécula de DNA, de vírus a humanos, contém sítios-alvo de enzimas de restrição, distribuídos aleatoriamente de forma que fragmentos de DNA de tamanhos adequados possam ser produzidos para clonagem. Os sítios de restrição não possuem relevância funcional, uma vez que a maioria dos organismos não possui enzimas de restrição.

Em 1985, K. B. Mullis e colaboradores inventaram a PCR, tecnologia que permite uma rápida e eficiente amplificação de regiões específicas do genoma. Precisamente, essa reação possibilita que uma região selecionada do genoma seja amplificada milhões de vezes, permitindo seu isolamento do resto do genoma. Dessa forma, os conhecimentos gerados, nos últimos 50 anos, permitiram avanços significativos no esclarecimento da organização estrutural e funcional dos genomas.

Grande parte dos genomas que foram completamente sequenciados utilizou o método de Sanger (SANGER et al., 1977), que possibilita a determinação da sequência nucleotídica ordenada de um fragmento de DNA. Embora esse método tenha sido melhorado ao longo das últimas décadas, seus princípios continuam os mesmos. Nessa técnica, o DNA a ser sequenciado é desnaturado inicialmente em fitas simples por aquecimento, com posterior anelamento de uma sequência de iniciadores marcado com fósforo radioativo (P32) no final da extremidade 3' de uma das fitas simples. Essa solução é dividida em quatro amostras correspondendo às quatro bases do DNA. DNA polimerase e desoxinucleotídeos (dATP, dCTP, dGTP, dTTP) são adicionados a cada amostra, sendo posteriormente adicionada a cada uma das amostras dideoxinucleotídeos correspondendo às quatro bases (ddATP, ddCTP, ddGTP, ddTTP). Dessa forma, são montadas quatro reações idênticas, exceto pela presença de dideoxinucleotídeos de Adeninas na primeira reação, dideoxinucleotídeos de Timinas, na segunda reação, e assim sucessivamente.

Os dideoxinucleotídeos são estruturalmente similares aos nucleotídeos normais, exceto pela presença de um hidrogênio no carbono 3' do açúcar, em vez de um grupo hidroxila. Esse nucleotídeo modificado, quando incorporado na cadeia de DNA nascente, impede a incorporação de um próximo nucleotídeo e consequentemente interrompe a síntese da fita de DNA. Dessa forma, ao final de cada reação se forma uma coleção de milhares de moléculas terminando no mesmo nucleotídeo. Quando essas moléculas são separadas por eletroforese, é possível identificar a posição na fita de DNA dos dideoxinucleotídeos específicos daquela reação. A análise conjunta das reações contendo dideoxinucleotídeos para as quatro bases permite a leitura ordenada da sequência de bases do DNA.

Na década de 1980, iniciativas começaram a ser discutidas para o sequenciamento de diversos genomas de espécies modelo, incluindo humanos. No início da década de 1990, o desenvolvimento de dideoxinucleotídeos marcados com corantes, assim como o desenvolvimento das máquinas de sequenciamento por capilaridade, permitiram um aumento considerável da capacidade de sequenciamento.

O primeiro genoma completamente sequenciado foi do bacteriófago phiX174, com 5.375 pares de base (pb) contendo 9 genes (SANGER et al., 1977). O sequenciamento desse genoma viral representou um passo significativo, pois indicava que genomas mais complexos poderiam também ser sequenciados.

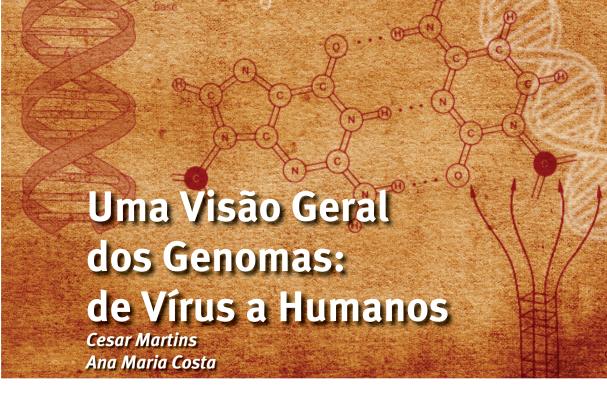
O primeiro eucarionte sequenciado foi o fungo Saccharomyces cerevisiae, com 13 milhões de pb, cujo genoma foi somente determinado quase vinte anos após o primeiro sequenciamento realizado (GOFFEAU et al., 1996). Nos anos seguintes, os avanços continuaram e o primeiro animal a ter seu genoma sequenciado foi o nematódeo Caenorhabditis elegans com 100 milhões pb (THE C. ELEGANS SEQUENCING CONSORTIUM, 1998); a primeira planta sequenciada foi *Arabidopsis thaliana* com 157 milhões de pb (THE ARABIDOPSIS GENOME INITIATIVE, 2000).

A idéia de sequenciamento do genoma humano começou a ser discutida por volta de 1985 e o primeiro rascunho da sequência nucleotídica do genoma humano saiu em 2001 (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001; VENTER et al., 2001). Embora o "sequenciamento completo" de diversos genomas tenha sido divulgado, incluindo o genoma humano, muitas incorreções ainda existem, principalmente relacionadas ao correto posicionamento no genoma de determinados segmentos sequenciados. Além disso, o correto arranjo e o posicionamento da grande quantidade de sequências repetitivas de DNA presentes no genoma representam, ainda, um desafio para os sistemas atuais de "montagem" da sequência nucleotídica ordenada do genoma.

Após 30 anos de avanços, existiam, em outubro de 2007, mais de 600 genomas sequenciados que cobrem mais de 180 espécies. A maior parte dos dados desses organismos sequenciados pode ser acessada on-line por meio dos seguintes sítios na internet:

- http://www.genomenewsnetwork.org/resources/sequenced_ genomes/genome_guide_p1.shtml
- http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi
- http://www.ncbi.nlm.nih.gov/sites/entrez?db=Genome
- http://www.genomesonline.org/

tet palacra "genoma" for sunhara por Hans Prinkler en 120, como una confusação le gene e cremotsomo. No entanto o concetto geral de genoma pode sen atributio ao seculo Itantes da era xista quanto e Axistoteles apontou os princiros concettot em relaçõe à hereditariedade. Emboracol travalhos de Mendel, no final do século XIII lenhan propisialo grandes contribuições no campo da hereditariedade, esta area mostrava-se, ainda, bastante althata. Com o avanço dos métotos científicos e das ternologias, a herekitariesake passou a ser associada as estruturas presentes no núcleo das células chamadas monossamos final no século III é início do . século X e finalmente som est polímeros de nucleothers aupla lita hanados de DHE, A meados do Seulo II) que forman es cromossemes.



Genoma viral

Parasitas intracelulares obrigatórios, os vírus são agentes infecciosos que utilizam uma célula hospedeira para se replicar. Os vírus infectam células dos três domínios de vida existentes na terra (Arqueobactéria, Bactéria e Eucariotos) e são de grande interesse de diversas áreas da ciência porque causam muitas doenças em humanos, em animais domésticos e em plantas utilizadas em culturas. Embora muito esforço tenha sido realizado e elguns vírus erradicados, ao menos em alguns continentes, muitos vírus ainda persistem aos tratamentos, em razão de sua habilidade de mudar rapidamente e se adaptar a uma nova condição. Além disso, os vírus têm sido utilizados como ferramentas para uma ampla variedade de estudos em biologia celular e molecular. O desenvolvimento de vetores virais utilizados para construções de bibliotecas genômicas de outros

organismos, ou mesmo como carreadores de genes a serem transferidos para células, como ocorre na sua aplicação na terapia gênica, representam algumas das aplicações atuais dos vírus.

Embora amplamente estudados em razão do seu envolvimento com doenças, principalmente em humanos e em animais de produção, e do seu potencial como ferramenta em biologia molecular, a complexidade organizacional do genoma viral é ainda pouco conhecida, frente à diversidade de famílias virais existentes. Entre algumas literaturas que poderiam ser recomendadas, a publicação recente de Strauss e Strauss (2008) apresenta uma rica fonte para consulta de diversos aspectos da biologia dos vírus, incluindo informações sobre seus genomas.

Os vírus contêm um genoma, que pode ser de DNA ou RNA, envolto em uma capa proteica chamada capsídeo ou nucleocapsídeo. Alguns vírus possuem ainda um envelope lipídico envolvendo o nucleocapsídeo. A partícula viral, extracelular e matura, é denominada de virion. De acordo com o Comitê Internacional de Taxonomia Viral (ICTV), os vírus são classificados em espécies, gêneros e famílias. A família pode ser considerada a unidade fundamental da nomenclatura viral e está baseada no tipo e no tamanho do seu genoma, na estrutura do virion e na estratégia de replicação. Unidades taxonômicas mais elevadas têm sido reconhecidas apenas para três ordens de vírus. Dessa forma, a essência da nomenclatura viral está baseada no agrupamento dos diferentes vírus em famílias.

A composição e estrutura dos genomas virais é mais diversa e complexa do que o observado em bactérias, plantas ou animais. O genoma dos vírus pode variar em tamanho de 3.200 nucleotídeos (e.g. hepadnavirus) a 1,2 milhões de pares de bases (mimivirus). A grande diversidade dos genomas virais pode ser facilmente visualizada nos tipos de moléculas de ácido nucleico e no seu arranjo nos genomas. O genoma viral pode conter DNA ou RNA, mas nunca os dois juntos. A molécula de ácido nucleico pode assumir diversas configurações incluindo fita simples, dupla-fita, linear, circular

ou segmentada, já que a base da nomenclatura viral está no tipo de genoma que possui. Em seguida, serão apresentados diferentes grupos virais com base na estrutura do seu ácido nucleico.

Vírus de DNA

Apresentam seu material genético na forma de fita dupla (dsDNA) ou fita simples (ssDNA).

Como exemplos de vírus de DNA fita dupla podem ser citados os vírus da varíola, herpes, hepatite B, citomegalovírus, adenovírus e alguns bacteriófagos. De acordo com o ICTV, existem 22 famílias de vírus que possuem dsDNA (Poxviridae, Iridoviridae, Herpesviridae, Adenoviridae, Polyomaviridae, Papillomaviridae, Asfarviridae, Myoviridae, Siphoviridae, Podoviridae, Tectioviridae, Corticoviridae, Plasmaviridae, Lipothrixviridae, Rudiviridae, Fuselloviridae, Guttaviridae, Phycodnaviridae, Polydnaviridae, Ascoviridae, Baculoviridae, Nimaviridae). Seus genomas variam de 5 quilos de base (kb) a 375 kb e infectam bactérias, arqueobactérias, amebas, micoplasmas, algas, fungos, invertebrados e vertebrados. Interessantemente, não se conhece vírus de dsDNA que infecta plantas.

Existem cinco famílias de vírus de ssDNA (Inoviridae, Microviridae, Geminiviridae, Circoviridae, Parvoviridae). Como exemplo de vírus de DNA fita simples, existem os bacteriófagos phiX-174 e o vírus AAV (Adeno-associated virus). Seus genomas variam de 1,7 kb a 8,5 kb e podem infectar bactérias, micoplasmas, spiroplasmas, plantas, invertebrados e vertebrados.

Durante a infecção, a maioria dos vírus dsDNA de vertebrados estimula a replicação do DNA da célula hospedeira, ou, pelo menos, os estágios iniciais de replicação do DNA, para preparar um ambiente adequado para a replicação do seu próprio DNA. Esse ambiente favorável inclui a presença de fatores celulares requeridos para a replicação do DNA, assim como, aumento celular da quantidade de

substrados requeridos para a síntese de DNA. Além disso, alguns vírus causam proliferação celular, ao menos no início da infecção. Por isso, muitos vírus são conhecidos como causadores de tumores em humanos e em outros animais.

Vírus de RNA fita simples positiva

Os vírus de RNA fita simples positiva [ssRNA(+)] compreendem um grupo extenso formado por muitas famílias virais. Entre esses vírus, estão os causadores de diversas doenças em humanos como a encefalite, a hepatite, a poliartrite, a febre-amarela, a dengue, a poliomielite e a gripe. Existem sete famílias de vírus ssRNA(+) que infectam humanos, dos quais alguns também infectam outros vertebrados, causando importantes patogenias em animais domésticos. A maioria dos vírus que infectam plantas possui genoma com ssRNA(+) e é classificada em nove famílias e alguns gêneros não definidos. Outras famílias de vírus ssRNA(+) incluem duas famílias que infectam bactérias.

O genoma dos vírus ssRNA(+) contém Open Read Frames - ORFs e codifica diversas classes de proteínas como helicase, protease papaína-*like* e RNA polimerase. Um exemplo de organização genômica dos vírus ssRNA(+) pode ser obtido por meio da análise de membros da família Picornaviridae. O genoma dessa família viral contém uma única molécula de RNA de cerca de 7.5 kb que possui uma ORF que é traduzida em uma única e longa poliproteína. Essa poliproteína é clivada por uma ou mais proteinases codificadas pelo vírus formando 25 diferentes polipeptídeos.

Vírus de RNA fita simples negativa

Sete famílias de vírus possuem seus genomas na forma de RNA fita simples negativa [ssRNA(-)] e provocam sérias epidemias em humanos como raiva, encefalite, infecções do trato respiratório e febre hemorrágica. Vírus pertencentes às famílias Paramyxoviridae, Rhabdoviridae, Filoviridae e Bornaviridae possuem seus genomas na forma de RNA não segmentados e organizados de forma similar. Essas famílias são agrupadas na ordem Mononegavirales. As outras três famílias de vírus ssRNA(-), que incluem Arenaviridae, Bunyaviridae e Orthhomyxoviridae, possuem genomas segmentados com dois, três e seis a oito segmentos, respectivamente. Independentemente de o genoma ser segmentado ou não, os vírus ssRNA(-) possuem um conjunto similar de genes. Nos vírus da ordem Mononegavirales, a ordem dos genes está conservada em todos os membros da família. Nos vírus com genomas segmentados, os genes podem ser ordenados por meio do alinhamento dos segmentos.

Para todos os vírus ssRNA(-), o primeiro evento na infecção é a síntese de RNAm a partir do genoma pela RNA polimerase presente no nucleocapsídeo, uma vez que essa polimerase é necessária para a produção do RNAm e as proteínas traduzidas requeridas para a replicação do genoma, o genoma nu dos vírus ssRNA(-), não são infecciosos. Múltiplos RNAs mensageiros são produzidos dos genomas ssRNA(-) e a maioria dos RNAm é traduzida em um único polipeptídeo. No entanto, alguns poucos genes produzem RNAs mensageiros que são traduzidos em mais de um produto. Produtos múltiplos podem ser produzidos com base em um mesmo gene por meio do uso de tradução alternativa dos códons da iniciação, durante a tradução de um RNAm. Ao contrário dos vírus ssRNA(+), os vírus ssRNA(-) não produzem poliproteínas que requerem processamento por proteases codificadas pelo próprio vírus. A replicação do genoma ssRNA(-) requer a produção de uma cópia complementar do genoma, chamada de antigenoma ou RNA vírus-complementar (vcRNA). Os genomas de todos os vírus ssRNA(-) possuem sequências terminais invertidas que podem ter papel na circularização do genoma de alguns desses vírus. A circularização do genoma pode ser requerida para a replicação do genoma ou a mesma para a síntese de RNAm.

Vírus de RNA fita dupla

A maioria dos vírus contendo RNA fita dupla (dsRNA) pertence à família *Reoviridae* e possue 10, 11 ou 12 segmentos de dsRNA no seu genoma, tatalizando 16 kb a 27 kb. Esse grupo de vírus infecta um amplo espectro de organismos incluindo vertebrados, invertebrados, plantas e fungos. Vários deles vírus são importantes agentes patogênicos em humanos. A família Reoviridae possui 12 gêneros, dos quais o *Orthoreovirus* é bastante estudado, já que infecta muitas espécies de mamíferos, incluindo humanos. O genoma dos *Orthoreovirus* consiste de dez segmentos de dsRNA que varia de tamanho de 1,2 kb a 3,9 kb e somam 23.5 kb. Os dez segmentos são agrupados em três classes chamadas L para grande, M para médio e S para pequeno, e são responsáveis pela produção de 11 ou 12 proteínas distintas.

Vírus de transcrição reversa

Duas famílias de vírus animais utilizam transcriptase reversa (RT) na replicação do seu genoma: *Retroviridae* e *Hepadnaviridae*. Além disso, dois gêneros de vírus de plantas, com ciclos de vida similares à família *Hepadnaviridae* também utilizam RT. Nesses vírus, a informação genética no genoma alterna entre o estado de RNA e DNA. A RT, codificada pelo genoma viral, converte o genoma de RNA dos retrovírus, ou uma cópia de RNA do genoma de DNA dos hepadnavírus, em DNA dupla-fita. No núcleo da célula infectada, a RNA polimerase celular transcreve o DNA do genoma do hepadnavírus ou a cópia de DNA do retrovírus produzindo RNA a ser transcrito pela RT. Os hepadnavírus e os vírus de plantas transcrevem, reversamente, o RNA em DNA durante o empacotamento. Dessa forma, a replicação do genoma dos retrovírus ocorre no sentido RNA→DNA→RNA, enquanto a replicação do genoma dos hepadnavírus ocorre no sentido DNA→RNA→DNA. Uma característica interessante da transcrição reversa é a destruição do molde de RNA

no processo de conversão para DNA. A RT possui, associada a ela, uma atividade RNase H que degrada a fita de RNA de uma molécula híbrida RNA-DNA. Essa atividade é essencial para a produção de uma cópia de DNA dupla-fita a partir do RNA viral pela RT.

Particularmente no que se refere aos retrovírus, após a infecção, a cópia de dsDNA é integrada no cromossomo da célula hospedeira e passa a ser chamada de provírus. Somente o DNA integrado é estável e e eficientemente transcrito pela maquinaria da célula hospedeira, sendo a integração requerida para produzir a infecção. Entretanto, durante a infecção de hepadnavírus, o DNA não é integrado no genoma hospedeiro. Ao contrário disso, o DNA é mantido no núcleo como epissomos não-replicativos estáveis e são eficientemente transcritos pela maquinaria da célula hospedeira.

Supõe-se que retrovírus tenham evoluído pela aquisição de genes, lhes permitindo deixarem a célula e se tornarem infecciosos (por exemplo, genes que coficam as proteínas que formam o envelope). É possível, também, que eventos contrários tenham ocorrido, em que retrovírus deram origem aos retrotransposons.

Genoma bacteriano

Tradicionalmente, a terminologia bactéria tem sido utilizada como referência a procariontes de forma geral. No entanto, recentemente, torna-se claro que os procariontes compreendem dois grupos distintos, denominados de Arqueobactéria e Eubactéria, que divergiram muito cedo na história evolutiva do planeta. Em nível molecular, as arqueobactérias são mais parecidas com os eucariotos, apresentando uma maquinaria de replicação, transcrição e tradução semelhante.

Estrutura do genoma bacteriano

Acredita-se que as bactérias mantenham características das primeiras formas de vida que existiram na terra e que representem os habitantes mais antigos do planeta. As dimensões micrométricas da célula bacteriana fornecem uma configuração estável para o metabolismo celular e replicação do DNA. Além disso, sua habilidade de crescer e se dividir sob uma ampla diversidade de condições ambientais revelam sua capacidade de adaptar-se e alterar a biosfera.

A principal característica que diferencia as células eucarióticas das células procariontes diz respeito à presença de um envelope ou envoltório que delimita o conteúdo de DNA no núcleo das células eucarióticas, que está ausente nos procariontes, e organelas membranosas. Além disso, as células procariontes são simples e pequenas e vivem, em sua maior parte, como indivíduos unicelulares. Essas células são protegidas por uma parede celular que reveste externamente a membrana plasmática, que, por sua vez, delimita um compartimento citoplasmático único contendo DNA, RNA, proteínas e muitas outras moléculas vitais.

A maioria dos genomas bacterianos consiste de uma única molécula circular de DNA, embora moléculas lineares de DNA tenham sido encontradas em algumas espécies como Agrobacterium tumefaciens, Rhodobacter capsulatus, Brucella melitensis e Rhizobium spp. Nos "cromossomos" bacterianos circulares, o DNA não existe em um círculo aberto relaxado. Os 3 a 4 milhões de pares de bases do DNA encontrados em um genoma bacteriano típico equivale a cerca de 1.000 vezes o tamanho da bactéria. O DNA bacteriano encontrase extremamente compactado, por meio da formação de complexos com várias proteínas, não está ligado a proteínas histonas como o DNA eucariótico (discutido mais adiante nesta obra). Na microscopia eletrônica, o DNA bacteriano frequentemente apresenta-se como um aglomerado distinto, o nucleoide, que é confinado a uma região definida do citoplasma. Se uma bactéria é rompida suavemente, seu DNA se espalha em uma série de alças torcidas. As alças

são presas umas às outras por proteínas. Além do DNA do nucleoide, muitas bactérias possuem DNA adicional sob a forma de pequenas moléculas circulares chamadas plasmídeos, que se replicam independentemente do cromossomo.

Evolução e dinâmica do genoma bacteriano

O genoma bacteriano possui capacidade significativa de sofrer mudanças, embora se mantenha relativamente constante. Esses estados opostos fornecem mecanismos essenciais para algumas alterações e também para a manutenção de certa constância no genoma bacteriano. O tamanho do genoma bacteriano, dentro de um gênero, permanece praticamente constante, com potencial para pequenas mudanças na ordem dos genes. No entanto, uma forma das células bacterianas para alterar o tamanho dos seus genomas é perder ou adquirir um ou mais plasmídeos. Os genomas bacterianos podem variar de 1 a 9 milhões de pares de bases (Tabela 1). Essa amplitude é por causa da diversidade evolutiva das diferentes espécies de bactérias. Os genomas de Escherichia coli e Salmonella typhimurium (ambos microorganismos Gram-negativos) possuem cerca de 4 milhões de pares de bases. Isso é equivalente ao peso molecular de 3 x 10⁹ Daltons e comprimento de cerca de 1,5 mm. Os "cromossomos" de E. coli e S. typhimurium, também, possuem propriedades comuns, uma molécula única, contínua, empacotada e helicoidizada negativamente, permitindo que sua estrutura nucleotídica forme curvaturas em sequências específicas de DNA.

Tabela 1. Espécies de bactérias e características dos seus genomas sequenciados.

Tamanho do genoma em Mb	Número de genes
0,58	468
0,82	687
1,11	834
1,14	1.041
1,55	1.544
1,67	1.590
1,75	1.013
1,83	1.740
1,87	1.877
3,57	3.168
4,21	4.099
4,45	4.402
4,64	4.289
9,20	7.380
1,66	1.750
2,17	2.493
0,49	552
	9enoma em Mb 0,58 0,82 1,11 1,14 1,55 1,67 1,75 1,83 1,87 3,57 4,21 4,45 4,64 9,20 1,66 2,17

Fonte: www.ncbi.nlm.nih.gov

Diversos trabalhos sobre evolução e tamanho do genoma bacteriano têm mostrado que o tamanho do genoma não está distribuído nas médias aritméticas de 3.812 kb para as bactérias Gram-negativas e 3.115 kb para as bactérias Gram-positivas. Outro aspecto evidente é que o tamanho do genoma pode variar grandemente entre diferentes cepas (LAWRENCE; HENDRICKSON, 2005). Isso ocorre, por exemplo, com Pseudomonas aeruginosa, cujos tamanhos dos genomas podem variar de 4.100 kb a 5.400 kb, 5.460 kb a 5.900 kb e de 2.200 kb a 2.800 kb, dependendo da cepa analisada (TREVORS,

1996). Adicionalmente, diferenças na média do tamanho dos genomas podem variar dependendo se forem utilizados tamanhos de genomas Gram-negativos ou Gram-positivos nos cálculos e a presença ou ausência de plasmídios, especialmente megaplasmídios como ocorrem em Rhizobium spp. Bactérias relacionadas possuem tamanhos similares do seu genoma com pequenas variações no conteúdo de G + C (guanina + citosina) em diferentes partes do genoma. Em adição, as bactérias apresentam uma ampla variação do conteúdo G + C de 23 % a 72 %.

É possível que bactérias com genomas grandes sejam resultado da duplicação do tamanho do cromossomo durante a evolução. Adicionalmente, a maior parte dos cromossomos bacterianos tem ficado estável durante milhões a bilhões de anos de evolução. As bactérias possuem sequências de DNA que, normalmente, não são expressas, mas estão sujeitas a efeitos mutacionais que podem torná-las funcionais. Os plasmídeos representam um pool significante de diversidade de informação genética. Os plasmídeos variam em tamanho de 1 kb a 400 kb, dos quais os de maior tamanho podem representar 10 % ou mais do genoma da bactéria, dependendo do tamanho do cromossomo da bactéria hospedeira. Por sua vez, a ocorrência de plasmídeos pode variar em espécies particulares. Populações naturais de Pseudomonas contêm plasmídeos de tamanhos diversos – o que ajuda na determinação de fenótipos similares ou diferentes. Os plasmídeos promovem transferência de informação genética entre bactérias pertencentes a uma mesma unidade taxonômica ou mesmo para bactérias pertencentes a diferentes grupos. Essa "engenharia genética" in vivo, promovida pelas bactérias, tem permitido mudanças acentuadas no tamanho do genoma. Outro aspecto significativo é o diferente número de cópias dos plasmídeos que podem existir nas bactérias.

As células bacterianas mantêm uma constância relativa do seu genoma ao mesmo tempo em que sofrem mudanças por mecanismos de recombinação genética, adquirindo e perdendo plasmídeos

e transposons (material genético capaz de promover sua mobilização dentro do genoma) e realizando transdução e deleção/inserção de DNA (DOBRINDT; HACKER, 2001). Esses mecanismos originam formas variantes e uma diversidade acentuada nas populações de bactérias. Exposições a condições severas de estresse levam a deleções e amplificações contíguas, conhecidas por tandem, muito comuns nos genomas bacterianos. A alta complexidade genômica das bactérias de vida livre, em comparação à complexidade genômica reduzida nas bactérias parasitas intracelulares, sugere que o conteúdo genômico reflete o estilo de vida das bactérias, como resultado de processos de otimização genômica (DOBRINDT; HACKER, 2001).

A análise dos genomas sequenciados de bactérias mostrou que quase todo seu DNA (85 % a 90 %) codifica para RNA estruturais ou proteínas. Dessa forma, o tamanho do genoma bacteriano é proporcional ao número de genes. Embora não se saiba a função de todos esses genes, cerca de 60 % deles apresentam homologia entre as diferentes espécies. Esses genes estão envolvidos com metabolismo, estrutura celular, transporte e regulação da expressão gênica, mecanismos esses essenciais na estrutura e função básica da célula. O número de genes pode variar de 500 a 1.200, entre as bactérias parasitas intracelulares obrigatórias, 1.500 a 7.500, entre as bactérias de vida livre, e 1.500 a 2.700 entre as arqueobactérias (Tabela 1).

Organização dos genes bacterianos

Nas bactérias, existe uma equivalência exata entre um trecho específico de DNA de um gene e o número de aminoácidos na proteína codificada por esse gene. De acordo com o código genético, três nucleotídeos do DNA gênico são necessários para codificar cada aminoácido do polipeptídeo da proteína. Dessa forma, um gene bacteriano consiste de uma sequência de DNA que é transcrita em uma molécula de RNA (RNA mensageiro - RNAm) que será traduzida na sequência de aminoácidos de um polipeptídeo (Fig. 2). Os genes bacterianos estão frequentemente organizados em blocos denominados de clusters que incluem genes que codificam proteínas cujas funções estão relacionadas. Isso ocorre, por exemplo, com os genes codificadores de enzimas que participam de uma via metabólica específica. Esses clusters gênicos representam uma unidade transcricional única, operon, responsável pela síntese de um RNAm policistrônico, que codificará as diferentes proteínas de funções relacionadas (LEWIN, 2004).

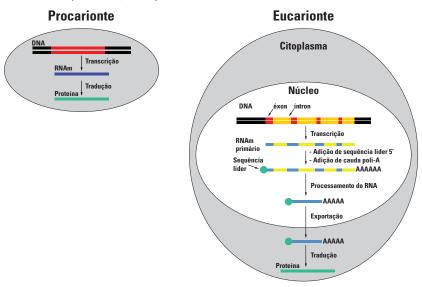


Fig. 2. Estrutura organizacional e funcional dos genes procariontes e eucariontes.

Genoma eucarionte

Ao contrário dos procariontes, a presença de um ambiente individualizado representado pelo núcleo permite a compartimentalização de diversas funções gênicas nos eucariontes.

No final da década de 1970, foi descoberto que o RNAm dos eucariontes e de seus vírus era menor que o DNA genômico do qual haviam sido transcritos (BERGER et al., 1977; CHOW et al., 1977). Os genomas eucariontes contêm genes interrompidos que consistem de sequências alternadas denominadas de éxons (que se mantém no RNAm final), interespaçadas por sequências denominadas de íntrons (que são removidas do RNAm imaturo). Os genes eucarióticos são transcritos individualmente, cada qual produzindo um RNAm com informação para apenas uma cadeia polipeptídica (RNAm monocistrônico). Existe apenas uma exceção geral a essa regra: no genoma do nematódeo *C. elegans*, ~15 % dos genes estão organizadas em unidades policistrônicas (um RNAm que codifica várias cadeias polipeptídicas).

Independente do tamanho do genoma, os procariontes em geral apresentam uma densidade média de 900 genes por Mb de DNA genômico, enquanto o genoma dos mamíferos apresenta uma densidade de 11 genes por Mb (PATTHY, 2003). Essas características representam consequências organizacionais da presença de genes interrompidos e de grande quantidade de sequências de DNA nãocodificadores (DNAs repetitivos) no genoma dos eucariontes.

A expressão dos genes interrompidos requer um passo adicional que não ocorre nos genes não interrompidos dos procariontes. O RNAm formado representa uma cópia exata da sequência genômica e apenas um precursor utilizado para a síntese da proteína. Primeiramente, os íntrons são removidos do transcrito primário, produzindo uma molécula de RNA composta apenas de uma série de éxons. Além dessa modificação denominada de "processamento", um grupo metil-guanosina é adicionado à extremidade 5' da molécula e uma cauda de aproximadamente 150 a 200 adenosinas (A) é adicionada na extremidade 3' do RNAm (Fig. 2). A proporção de genes interrompidos é baixa em fungos e aumenta consideravelmente nos eucariontes mais complexos. Em Saccharomices cerevisiae, a grande maioria dos genes (> 96 %) não é interrompida. Não existem genes em S. cerevisiae com mais de 4 éxons. Em insetos e mamíferos, a situação é inversa: poucos genes apresentam sequências codificadoras não interrompidas (6 % em mamíferos). Aproximadamente 50 % dos genes de mamíferos possuem mais de 10 íntrons. Uma consequência desse padrão de organização reflete no tamanho dos genes. A média de tamanho de um gene de fungo é de 1,4 kb e pouquíssimos genes são maiores do que 5 kb.

Grande parte dos genes dos organismos está presente em mais de uma cópia, o que faz com que o número de diferentes tipos de genes seja menor do que o número total de genes. A presença de genes duplicados é extremamente comum nos eucariontes e representa um aspecto importante para a evolução do genoma (discutido no item Duplicação Gênica: a maior força da evolução, nesta obra). A duplicação gênica permite que as novas cópias do gene assumam funções mais especializadas ou mesmo possibilita a origem de novos genes com funções totalmente distintas daquelas dos genes de origem. A duplicação gênica permitiu a origem de novos genes que passaram a dividir as funções celulares, permitindo, assim, um ganho de complexidade organizacional das células, tecidos e organismos. Em bactérias, a maior parte dos genes está presente na forma de cópia única, enquanto, nos eucariontes superiores, a maioria dos genes está organizada em famílias com um grande número de cópias (Tabela 2).

Tabela 2. Espécies e porcentagem de seus genes organizados em famílias*.

Espécies	Genes únicos	Famílias com 2 a 4 membros	Famílias com > 4 membros
Haemophilus influenzae	89 %	10 %	1 %
Sacharomyces cerevisiae	72 %	19 %	9 %
Caenorhabditis elegans	55 %	20 %	16 %
Drosophila melanogaster	72 %	14 %	14 %
Arabidopsis thaliana	35 %	24 %	41 %

^{*} Dados obtidos com base em Lewin (2004).

Outra característica marcante do genoma dos eucariontes é a presença de grande quantidade de segmentos repetitivos de DNA (discutido no item Estrutura e Instabilidade do Material Genético, nesta obra). Essas repetições representam duplicações de pequenas sequências, segmentos cromossômicos, cromossomos inteiros, ou até mesmo duplicações de todo o genoma, que ocorreram durante a história evolutiva dos eucariontes. A presença de sequências repetitivas de DNA não é exclusividade dos eucariontes, pois ocorre também em procariontes. No entanto, nesse último grupo, a redundância das repetições é bastante baixa.

O genoma humano, por exemplo, possui somente 3 % do seu conteúdo de DNA codificando para polipeptídeos (HORVATH et al., 2001). Sequências de DNA presentes na forma de cópia única no genoma correspondem a aproximadamente 2 % a 10 % do genoma dos eucariontes de uma forma geral, enquanto o restante do genoma é composto de sequências repetitivas. Os DNAs repetitivos incluem as sequências organizadas em tandem (DNAs satélites, minisatélites e microsatélites) e as repetições dispersas no genoma, como os elementos genéticos móveis (transposons e retrotransposons) (CHARLESWORTH et al., 1994). Os DNAs repetitivos, organizados em tandem, são encontrados principalmente nos centrômeros e telômeros dos cromossomos, enquanto os DNAs repetitivos dispersos são encontrados como unidades isoladas ou agrupamentos distribuídos ao longo do genoma. Embora extensivamente estudados nas últimas décadas, as forças moleculares que governam a evolução dos DNAs repetitivos no genoma são ainda muito discutidas (Fig. 3).

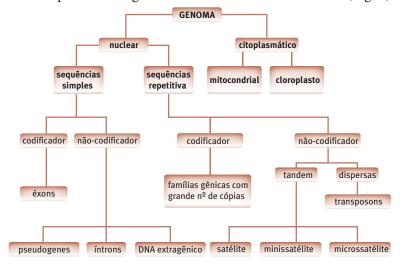


Fig. 3. Organização do genoma dos eucariontes.

O genoma dos eucariontes unicelulares apresenta a mesma escala de tamanho dos maiores genomas de bactérias (Tabela 3). Os eucariontes multicelulares apresentam um número maior de genes, mas não existe uma correlação exata com as variações no tamanho dos genomas. Os fungos S. cerevisiae e S. pombe possuem genomas de 13,4 Mb e 12,5 Mb, com aproximadamente 6.000 e 5.000 genes, respectivamente. O genoma do nematódeo C. elegans possui ~18.500 genes e 97 Mb. O genoma da mosca de frutas *Drosophila* melanogaster possui um genoma maior (165 Mb) do que C. elegans, mas um número menor de genes (13.600). A razão pela qual a mosca-de-frutas, organismo muito mais complexo, possuir apenas 70 % do número de genes do nematódeo não está bem compreendida. Isso mostra uma correlação entre o número de genes e a complexidade dos organismos. A planta A. thaliana possui cerca de 25.000 genes e um tamanho de genoma intermediário entre o nematódeo e a mosca das frutas. Mostrando também uma tendência das plantas possuírem um número maior de genes, provavelmente por causa dos eventos de duplicações genômicas ancestrais mais comuns nesse grupo do que entre os animais.

Tabela 3. Espécies eucariontes e características dos seus genomas sequenciados.

Espécies	Tamanho do genoma em Mb	Número de genes
Plasmodium falciparum	22,9	5.268
Sacharomyces cerevisiae	13,5	6.034
Sacharomyces pombe	12,5	4.929
Caenorhabditis elegans	97	18.424
Ciona intestinalis	160	16.000
Drosophila melanogaster	165	13.601
Anopheles gambiae	278	13.683
Arabidopsis thaliana	119	25.498
Rattus norvegicus	2.800	30.000
Takifugu rubripes	400	30.000
Homo sapiens	3.000	30.000

Fonte: www.ncbi.nlm.nih.gov

Um exemplo bastante interessante de complexidade e diversidade organizacional de genomas ocorre no peixe baiacu Takifugu rubripes, que possui o menor genoma dos vertebrados, com aproximadamente 390 Mb. Embora extremamente compacto, o genoma dessa espécie apresenta o mesmo número de genes quando comparado com o genoma humano, que é quase oito vezes maior (Tabela 3). O sequenciamento do genoma do baiacu mostrou uma alta densidade gênica e genes com íntrons bastante reduzidos em tamanho quando comparados com humanos. Enquanto o genoma humano possui apenas 3 % de sequências codificadoras de proteínas, o genoma do baiacu possui 17 %.

Análises dos genomas dos baiacus *Takifugu rubripes* e *Tetraodon* nigroviridis mostraram ainda uma quantidade reduzida de sequências repetitivas de DNA do tipo retrotransposons, quando comparados com humanos e camundongos. No entanto, esses peixes apresentam uma alta diversidade de classes de elementos transponíveis não encontrada nos mamíferos. Essa diversidade de retrotransposons foi também observada no peixe *Danio rerio* (paulistinha) e não foi perdida após a compactação do genoma nos baiacus (VOLFF et al., 2003). Acredita-se que a alta diversidade de sequências repetitivas de DNA presente nos peixes tenha representado um papel significativo na radiação das espécies, o que levou a alta diversidade de espécies presentes neste grupo.

Genoma das organelas

Genoma dos cloroplastos e mitocôndrias

Os cloroplastos e mitocôndias são organelas celulares portadoras de material genético próprio. Acredita-se que tenham se originado nos endossimbiontes ancestrais e que evoluíram em organelas. Os cloroplastos provavelmente se originaram de endossimbiontes do grupo das cianobactérias, enquanto as mitocôndrias, de parentes das alfa-protobactérias, grupo das riquetsias e semelhantes (EMELYANOV, 2001, 2003; GRAY, 2001; BARBROOK et al., 2006, KUROIWA, et al., 2006; KHACHANE et al., 2007).

Em virtude das suas origens, o material genético dessas organelas normalmente é composto por um único cromossomo circular com replicação semelhante aos seus ancestrais: do tipo anéis FtsZ (MIYAGISHIMA, 2006) no caso dos cloroplastos, e do tipo MD no caso das mitocôndrias (KUROIWA et al., 2006).

Nas últimas duas décadas, muitos genomas de mitocôndrias e cloroplastos foram sequenciados e analisados em termos de estrutura, regiões não-codificadoras, presença de intros, conteúdo de genes, RNA ribossomais (RNAr) e conteúdo de RNA transportadores (EMELYANOV, 2001; 2003; GRAY, 2001; KHACHANE et al., 2007).

Desses estudos, observou-se que os genomas mitocondriais variam significativamente no tamanho e na organização estrutural, principalmente em relação à extensão das regiões não codantes e presença de íntrons (EMELYANOV, 2001; 2003; GRAY, 2001; KHACHANE et al., 2007). Por exemplo, os genomas mitocondriais de angiospermas são os maiores e mais complexos quando comparados com as demais mitocôndrias. Contudo, o conteúdo gênico das mitocôndrias e dos cloroplastos parece não variar, significativamente, nas espécies analisadas (BARBROOK et al., 2006).

Nos cloroplastos, a estrutura e conteúdo gênico são altamente conservados, principalmente nos das plantas superiores. Já os plastídeos das algas são mais variáveis e contêm uma série de genes em cópia única que estão ausentes nos cloroplastos das plantas superiores. Os genomas dos plastídeos codificam proteínas envolvidas na transcrição e tradução, além daquelas responsáveis pelo aparato fotossintético (BARBROOK et al., 2006).

Em termos evolutivos, verifica-se uma nítida diminuição no tamanho e complexidade dos genomas de mitocôndrias e cloroplastos em relação aos seus prováveis ancestrais (GRAY et al., 2001; BARBROOK et al., 2006; KHACHANE et al., 2007). Além da redução no tamanho dos genomas, observa-se também uma redução nos percentuais de C + G nesses genomas (KHACHANE et al., 2007). Observa-se que muitos genes foram transferidos das organelas para o núcleo dos hospedeiros, ou mesmo excluídos do genoma (GRAY et al., 2001; BARBROOK et al., 2006; KHACHANE et al., 2007). Por exemplo, os genes relacionados à divisão dos cloroplastos e mitocôndria foram transferidos para o núcleo da célula (MIYAGISHIMA, 2006; KUROIWA, et al., 2006). O gene FtsZ foi duplicado e modificado, surgindo uma nova família de proteínas envolvidas na regulação da divisão cloroplastidial (MIYAGISHIMA, 2006).

Uma característica dos genomas mitocondriais de tripanossomatídeos é a presença de genes que realizam modificações nos RNAm mitocondriais num processo chamado de edição. Esse mecanismo, detalhado no tópico seguinte, gera variabilidade nos transcritos pela substituição de bases, em geral, C para U e, em menor frequência, de U para C, com o auxílio de RNA guias (RNAg) e enzimas específicas. Até recentemente, acreditava-se que a edição fosse um mecanismo restrito a essa família, contudo identificaram-se, em Arabidosis thaliana, aproximadamente 450 genes de enzimas participantes do processo, despertanto a atenção dos pesquisadores para o assunto (SHIKANAI, 2006).

Genoma mitocondrial de tripanossomatídeos

O cinetoplasto é uma mitocôndria dos protozoários da família kinetoplastidae. Está presente em cópia única e contém o DNA mitocondrial do cinetoplasto (kDNA). O kDNA representa 10 % a 25 % do DNA total dos kinetoplastidea (TEIXEIRA et al., 2006) e está peculiarmente arranjado, formando uma estrutura complexa, composta de pequenos segmentos de DNA circular catenados (minicírculos) e de segmentos círculares maiores (maxicírculos) (SHAPIRO; ENGLUND, 1995).

Os maxicírculos estão presentes na organela em número de 25 a 100, conforme o tripanossomatídeo. Em termos de sequência de bases do DNA, observa-se que cada molécula de maxicírculo apresenta um alto grau de homogeneidade entre si e carregam informações para a biogênese mitocondrial. Os genes presentes nessas moléculas, diferentemente de outros eucariotos, codificam RNA mensageiros que apresentam discrepâncias nas fases de leitura, quando comparados às sequências das suas proteínas correspondentes, sugerindo participação de mecanismos de processamentos pós-transcricionais característicos para se tornarem ativos (BENNE et al., 1986; FEAGIN et al., 1988).

Os minicírculos variam em quantidade e tamanho entre os cinetoplastideos. Por exemplo, em *Trypanosoma brucei*, essas moléculas variam entre 0,5 kb a 2,9 kb e estão presentes em mais de 10.000 cópias por organela. No *Trypanosoma cruzi*, a grande maioria dos minicírculos está na faixa de tamanho entre 1,4 kb a 1,5 kb, e é composta por quatro regiões variáveis semelhantes entre si, intercaladas, a cada 90° na molécula circular, por regiões conservadas com aproximadamente 120 pb (SIMPSON, 1997). A heterogeneidade entre minicírculos de tripanosomatídeos é observada tanto interespécie, como intraespécie. Em T. brucei, estima-se a existência de 300 classes de minicírculos diferentes, enquanto, em Trypanosoma equiperdum, todos os minicírculos apresentam alta homogeneidade na sequência de bases do DNA. Já o T. cruzi mostra um grau intermediário entre os dois citados (RIOU; YOT, 1977). Em decorrência de sua alta complexidade, a caracterização de esquizodemas permite identificar isolados do parasita (MOREL et al., 1980; STURM, 1989; SIMPSON, 1997).

O processo de edição foi descrito pela primeira vez por Blum et al. (1990), em *T. brucei*. A princípio, acreditava-se que os RNAgs eram codificados pelos maxicírculos, porém Sturm e Simpson (1990) mostraram que provinham da transcrição dos minicírculos. Os RNAgs são normalmente pequenos, em torno de 50 a 70 bases, e possuem uma cauda de uridina (poli-U) na extremidade 3'-OH da molécula.

O processo de edição promove modificações nas sequências de bases dos RNA mensageiros imaturos (pré-RNAms), por meio da inserção, ou, em alguns casos, pela remoção de resíduos uridinilados da molécula, como dito anteriormente. O mecanismo proposto para o processo prevê a existência de uma molécula de RNAg, com sequência complementar em sua extremidade 5'-fostado ao pré-RNAm. Inicialmente, haveria o pareamento de 4 a 15 bases da região

complementar, que teria a finalidade de manter as fitas unidas. As bases não pareadas da extremidade 3'-OH do RNAg teriam a função de direcionar a edição (BLUM et al., 1990; HAJDUK et al., 1993; SEIWERT, 1996). Vários estudos sugerem que a estrutura secundária do pré-RNAm a ser editado já seria suficiente para expor o sítio onde ocorreria a clivagem para a inserção dos resíduos uridinilados (PILLER et al., 1995a, 1995b, 1996). Porém, existem transcritos cuja exposição do sítio é mediada pelo RNAg (KABLE et al., 1996; SEIWERT et al., 1996; ALDER; HAJDUK, 1997). Após a clivagem do pré-RNAm, as enzimas uridilil transferase e (ou) terminal uridilil exonuclease fariam, respectivamente, a adição ou remoção das uridinas de acordo com as informações contidas no RNAg. Findo o processo, as fitas do RNA mensageiro maduro seriam religadas pela RNA ligase mitocondrial (CRUZ-REYES; SOLLNER-WEBB, 1996; KABLE et al., 1996; SEIWERT et al., 1996; ADLER; HAJDUK, 1997).

Por meio da edição, o RNAm pode receber códigos de iniciação e de terminação, podem ocorrer mudanças das fases de leitura ou aumento do RNAm, o que, supostamente, permitiria uma maior variabilidade de transcritos (ALDER; HAJDUK, 1994; SEIWERT, 1996). Esse processo está presente em vários tripanosomatídeos: Herpetomonas, Crithidia, Leishmania, e cinetoplastídeos correlacionados, tais como o Tripanoplasma (SCOTT, 1995), e, inclusive, no T. cruzi (ÁVILA; SIMPSON, 1997).

Thertulien et al. (1991; 1994) identificaram transcritos obtidos com base em um banco de DNA complementar (cDNA) de T. cruzi, que correspondiam a sequências de minicírculos, mas, sem as características típicas de um RNAg. A função desses transcritos ainda não está bem definida, e sugere que os minicírculos sejam potencialmente capazes de codificar pequenos peptídeos.

A quantidade de DNA, presente no cinetoplasto, varia em função do estágio de desenvolvimento do parasita. Essa peculiaridade é conhecida por descinetoplastidização, e se caracteriza pela perda parcial ou total do kDNA (VICKERMAN; PRESTON, 1976). O fenômeno foi detectado, em condições naturais, nas formas sanguíneas das espécies T. equiperdum T. evansi e T. brucei e recentemente nas formas sanguíneas de T. cruzi. Mas pode ocorrer em consequência ao tratamento com drogas, tais como brometo de etídio, acriflavina, diamidinas e compostos relacionados. Nesse caso, as drogas atuariam ao nível da enzima topoisomerase II, essencial para o processo de replicação do kDNA.

Em adição, Lee et al. (1992a, 1992b, 1993) relataram um novo fenômeno conhecido como transcinetoplastidização, em que variantes de Leishmania leishmania mexicana, resistentes a arseniato e tunicamicina, apresentaram uma alteração na sequência dos maxicírculos, bem como uma significativa mudança na população de minicírculos. Durante o fenômeno, classes, antes pouco representadas, passam a ser majoritárias, enquanto as inicialmente bem representadas, minoritárias. São necessárias 15 gerações do parasita para que o evento se complete e populações intermediárias são observadas no decorrer do processo (LEE et al., 1994). Embora não se conheçam os mecanismos de regulação que levam a expansão ou perda de determinados kDNAs, acredita-se que esse material genético é de extrema importância na adaptação do parasita às diferentes condições do meio ambiente.



Os mecanismos de recombinação são fundamentais para a geração da variabilidade e diversidade genética. Esses mecanismos podem ser agrupados genericamente em três grandes categorias: (1) Recombinação homóloga - recombinação que ocorre com a participação de grandes regiões de similaridade, normalmente entre cromátides irmãs de cromossomos homólogos; (2) recombinação sítio--específica - mediada por pequenas regiões de identidade no local da recombinação; (3) recombinação não-homóloga - sem similaridade de sequências no ponto da recombinação (LEWIN, 2004).

Basicamente, as ezimas que participam desses processos identificam sequências e (ou) estruturas características no DNA que, ao reconhecê-las, desencadeiam uma sequência de eventos que culminam nos rearranjos do material genético.

Grande parte das mutações e aberrações cromossômicas decorreu das recombinações "sítio-específicas" e "não-homólogas". Essas

formas de recombinações estão presentes na incorporação de DNAs exógenos no genoma (incorporação de vírus, retrovírus, DNAs obtidos por transformação, elementos de transposição) e no reparo de estruturas consideradas anormais pela célula, como as quebras da fita simples ou fita dupla do DNA, regiões com pareamento de bases incorreto provenientes da recombinação homóloga, estruturas do tipo alça, grampo, entre outras (LEWIN, 2004).

Ao longo do tempo, os eventos recombinativos promovem modificações significativas no conteúdo e distribuição dos genes nos cromossomos contribuindo para evolução dos genomas (Fig. 4). Na Fig. 5, está ilustrada a complexidade dos eventos de recombinação que contribuíram para a formação dos cromossomos de A. thaliana, fenômeno evidenciado após o sequenciamento e análise do genoma dessa planta (THE ARABIDOPSIS GENOME INIATIVE, 2000).

Apesar do assunto ter recebido muita atenção nas últimas décadas, ainda existem muitas lacunas do conhecimento para serem preenchidas por causa da complexidade do fenômeno.

Como será discutido adiante, atualmente observa-se que o foco principal de grande parte das pesquisas no tema está voltado para a identificação das enzimas participantes do processo. Acredita-se que o resultado da recombinação mudaria conforme o perfil enzimático no momento do evento, e, de acordo com o contexto celular, poderia ocorrer perdas (deleção), ampliações (duplicação ou inserção de DNA da célula ou de outro organismo) ou rearranjos diversos nas fitas de DNA (inversão ou translocação de segmentos dentro do próprio cromossomo).

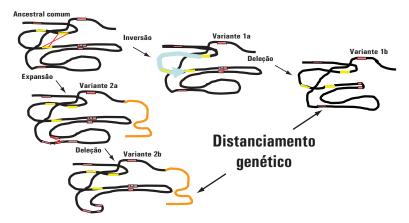


Fig. 4. Representação das mudanças na composição e na estrutura do DNA de um cromossomo, ao longo do tempo, em consequência de diferentes eventos recombinativos.

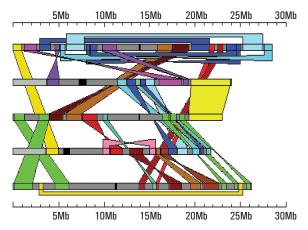


Fig. 5. Distribuição e arranjo das repetições de segmentos de DNA no genoma de Arabidospis thaliana. As cores semelhantes, destacadas do cinza, mostram repetições de segmentos cromossômicos presentes na forma direta (trapézios) ou invertidas (cones). Em preto, observam-se as regiões centroméricas de cada cromossomo (THE ARABIDOPSIS GENOME INITIATIVE, 2000).

Uma das questões mais intrigantes, que tem despertado a atenção da pesquisa, é o que torna uma determinada região genômica alvo preferencial de modificações na sequência de DNA em relação a outras. Regiões que são constantemente associadas a eventos de recombinação, ou que, quando clonadas, passam a ser denominadas de "instáveis".

Acredita-se que alterações na estrutura tipo B do DNA (dupla hélice, invertida e complementar passo direito), como predita por Watson e Crick (1953), poderiam ser reconhecidas pelas enzimas dos complexos de recombinação e desencadearem o fenômeno.

De concreto, verifica-se uma correlação direta entre os sítios de recombinação e a alteração da estrutura local do DNA provocada pela presença de elementos repetitivos e pela distribuição de bases numa ordem peculiar. Alguns trabalhos mostram que a clonagem de determinadas repetições gera recombinação sítio-específica. Esse tema será aprofundado no tópico Instabilidade de sequências.

Elementos repetitivos

Em consequência do "Projeto Genoma", sabe-se hoje que apenas 3 % do material genético haplóide de uma célula mamífera é composto por sequências que codificam para cadeias polipeptídicas (exons), o restante compreende as chamadas regiões não-codificadoras, onde estão distribuídas as diversas repetições, de origem viral ou não-viral, agrupadas genericamente em satélites, minissatélites microssatélites, telômeros, sequências repetitivas curtas, entre elas, Sines (short interspersed elements), ALU e MIR, e longas, tais como Line (long interspersed elements), L1 e L2 LTR (long terminal repeats) (LEWIN, 1994; BERNARDI, 1995; KAZAZIAN; MORAN, 1998, INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001; KIM, et al., 2004; HAN et al., 2007). Nos grupos de sequências repetitivas, curtas e longas, encontram-se os elementos transponíveis (elementos móveis), as sequências interruptoras de genes (íntrons) e os provírus tanto de origem viral como retroviral (BERNARDI, 1995; FAWCETT, et al., 2006; HAN et al., 2007).

Os elementos móveis, curtos e longos, são agrupados em duas categorias de acordo com a existência ou não de um intermediário de RNA durante o processo de mobilidade no genoma. Na primeira categoria, estão as sequências com capacidade de transposição direta, tais como elementos transponíveis de origem viral, sequências IS, Tn, e muitos íntrons, que se mobilizam por mecanismos de recombinação não-homóloga replicativa ou não-replicativa. A segunda categoria compreende os elementos de origem retroviral e transposons que se locomovem via mecanismo de retrotransposição (FINNEGAN, 1983; WEISS et al., 1985; VARMUS, 1988; LEWIN, 2004). De acordo com a similaridade de sequências os diferentes elementos móveis podem, ainda, ser agrupados em superfamílias, famílias e subfamílias dentro de cada categoria (HAN et al., 2007).

A grande maioria das partículas móveis apresenta, em sua estrutura, as informações necessárias para promover a sua expansão, ou transferência de um ponto do genoma para outro. Contudo, algumas famílias necessitam da expressão de enzimas de outros elementos móveis ativos para se transporem, como no caso dos Sines, que se movimentam por retrotransposição com o auxílio da maquinaria de outros retrotransposons (HAN et al., 2007; FAWCETT, et al., 2006; GU, et al., 2007).

Existe uma forte correlação entre as sequências de bases das repetições que interrompem sequências gênicas, os íntrons, e as sequências que compõe o grupo dos elementos móveis no genoma. Essa semelhança levou a especulação de que os íntrons e os elementos móveis poderiam estar evolutivamente correlacionados (LAMBOWITZ, 1989; CAVALIER-SMITH, 1991; HICKEY, 1992).

As inserções de elementos móveis no genoma, dependendo da natureza do elemento repetitivo, podem alterar a expressão gênica

(DEININGER et al., 2003); se rico em regiões CG, como no caso dos elementos Alu, podem introduzir novas ilhas CpG (LANDER et al., 2001), gerar deleções (HAN et al., 2005), criar novos genes ou famílias gênicas (XING et al., 2006).

As sequências repetitivas semelhantes, móveis ou não móveis, podem estar arranjadas próximas umas das outras formando estruturas simples ou complexas, genericamente classificadas quanto às suas orientações na fita de DNA. Desse modo, duas repetições próximas, estando na mesma orientação, formam uma estrutura simples denominada de repetição direta. Igualmente, duas repetições com sequências invertidas entre si, ou invertidas e complementares, são classificadas no grupo das repetições invertidas, ou no grupo das palindrômicas, respectivamente. Essas sequências podem estar ligadas umas às outras, formando estruturas complexas, também conhecidas por "tandens". Estruturas simples ou complexas são encontradas tanto dentro, como fora de regiões transcritas (LEWIN, 2004).

Conforme a organização das repetições e da natureza da sequência em termos de composição e distribuição das bases, o DNA pode assumir localmente estruturas diferentes das preditas por Watson e Crick (1953), ou seja, pode formar estruturas diferentes das do tipo B (dupla hélice invertida e complementar de passo direito) e formar alça, grampos, estruturas cruciformes, estruturas do tipo Z, curvas e dobramentos, estruturas triplas ou tipo H, quádruplas ou tetraplex (WANG et al., 2006; KMIEC et al., 1985; STEINMETZ, 1986, LINIAL; SHLOMAI, 1987; KIYAMA; KIYAMA 1996; GABRIELIAN et al., 1996; GABRIELIAN; PONGOR, 1996; VLAHOVICEK; PONGOR, 2000; KIYAMA et al., 1999; WANG; VASQUEZ, 2004; KAUSHIK et al., 2007; COSTA, 2008), que podem ser reconhecidas pelos mecanismos de recombinação celular e propiciar rearranjos diversos no material genético demonstradas em diferentes tipos de estudos (LEACH; STAHL, 1983; WYMAN, et al., 1985; 1986; MEIMA, et al., 1997; STRADER; HOWELL, 1997;

SONG et al., 2001; KOUPRINA et al., 2003; LAN; MUGUIRA, 2005; GRINDLEY et al., 2006; DU et al., 2007; WELLS, 2007; COSTA, 2008).

Estruturas repetidas invertidas e diretas não móveis são encontradas flanqueando muitos elementos transponíveis ativos e inativos. Os tamanhos das estruturas invertidas variam significativamente de elemento para elemento, podendo apresentar de 9 pb a mais de 200 pb (LEWIN, 2004; DOAK et al., 1997; BENITO; WALBOT, 1997). As repetições diretas são resultantes da duplicação do sítio de entrada da integração. São sequências normalmente pequenas entre 3 a 12 bases. Acredita-se que as enzimas envolvidas reconheçam as sequências do sítio de entrada e direcionem o evento de transposição, a exemplo do que acontece na integração do fago λ (LEWIN, 2004; BELFORT; PERLMANS, 1995; BENITO; WALBOT, 1997).

No genoma, estão presentes milhares de estruturas que possuem capacidade de transposição, porém a maioria encontra-se inativa. Muitas inserções novas, com base em estruturas transgênicas, são silenciadas pouco tempo depois de sua integração. Acredita-se que o silenciamento seja o resultado das discrepâncias entre a composição de CG do elemento inserido e o ambiente do *isochore* receptor. Observa-se também que elementos integrados em regiões pobres em CG são mais frequentemente inativados que os inseridos em regiões mais ricas em CG (BERNANRDI, 1995).

Como há uma coincidência entre os sítios inativos e o grau de metilação, vários pesquisadores sugeriram que a metilação poderia ser responsável pelo controle da transposição ao nível de transcrição (BESTOR; TYCKO, 1996; YOLDER et al., 1997). O'Neill et al. (1998) mostraram a associação entre a diminuição da metilação e ativação de retroelementos, correlacionando o silenciamento à estabilidade do genoma.

Elementos repetitivos móveis longos

Entre os elementos repetitivos longos de origem viral, os mais frequentes em mamíferos são os LTR (long terminal repeats) e Lines (long interspersed elements).

Os LTR caracterizam-se por apresentarem longas regiões repetidas nas extremidades de suas sequências integradas, com tamanho entre 250 pb a 1.400 pb, flanqueadas pela duplicação de 4 a 6 bases correspondentes ao sítio de entrada no genoma. Esses elementos, durante o processo de mobilização, são transcritos pela RNA polimerase II. A fita de RNAm gerada não apresenta cauda poliadenilada, e frequentemente possui íntrons, que são removidos antes de a fita ser traduzida. De um modo geral, os retroelementos possuem as informações necessárias para promover a sua retrotransposição (transcriptase reversa e (ou) integrase), mas, ao contrário dos retrovírus, os genes do envelope viral estão ausentes, ou defectivos (LEWIN, 2004). Dentro desse grupo estão os elementos Ty (em levedura), copia (em drosophila), IAP (em roedores) e THE1 (em humanos) (LEWIN, 2004).

O segundo grupo de elementos móveis longos em mamíferos compreende a família Line. Inicialmente, esses elementos foram classificados no grupo dos retroelementos de origem não-viral, em decorrência da presença de sequências poliadeniladas terminais e ausência de LTR. Entretanto, as análises das fases abertas de leitura mostraram alta identidade com enzimas retrovirais responsáveis pela retrotransposição (LEWIN, 2004). A estrutura dos Line compreende duas fases abertas de leitura. A primeira codifica uma proteína semelhante à gag com sítio de ligação com RNA e propriedade chaperônica para ácidos nucleicos (HOHJOH; SINGER, 1996; KOLOSHA; MARTIN, 1997; MARTIN; BUSHMAN, 2001; MARTIN et al., 2005), e a segunda fase aberta, uma proteína do tipo polimerase com atividade endodesoxirribonuclease (endonuclease) e transcriptase reversa (MATHIAS et al., 1991; FENG et al., 1996). As análises in silico dos sítios de integração de Line sugerem a participação do mecanismo de reparo de dupla-fita do DNA na transposição desses elementos (ICHIYANAGI; OKADA, 2006).

Trata-se de uma família com pelo menos 11 classes (MALIK; EICKBUSH, 1999). Line de diferentes categorias foram encontrados em um grande número de genomas eucariotos (ARKHIPOVA; MESELSON, 2000; compilado de RepBase em http://www.girinst. org/repbase/index.html).

Em mamíferos, o elemento Line-1 é o mais frequente. Aproximadamente 17 % do genoma humano são compostos por esse tipo de repetição, o que corresponde a 100 mil cópias, variando entre 50 mil a 100 mil cópias nos demais genomas (HWU, et al., 1986; SMITH, 1996; HAN et al., 2007). Dessas, apenas 3.500 cópias possuem a sequência Line-1 completa, das quais 95 % restantes encontramse truncadas na sua porção 5' fosfato ou rearranjados (FANNING; SINGER, 1987; DOMBROSKI et al., 1993).

Em razão do grande número de cópias presentes no genoma, os elementos repetitivos são tidos como sítios preferenciais para mediar recombinação homóloga desigual (FANNING, SINGER, 1987; BURWINKEL; KILIMANN, 1998).

Esses elementos não são completamente idênticos, variam em tamanho podendo atingir mais de 7 kb, sendo o consenso na ordem de 6 kb. Muitos elementos Line-1 completos acumularam códigos de terminação da tradução em suas sequências de DNA, o que impede a síntese das proteínas responsáveis pela sua locomoção, sendo raros os elementos capazes de se transpor a outros sítios no genoma. Acredita-se que apenas 20 a 40 elementos possam estar ativos no genoma humano (SASSAMAN, 1997). Desses, pelo menos sete já foram clonados e apresentaram atividade de retrotransposição em células mamíferas, o que tem permitido ampliar as informações sobre o mecanismo de transposição desse elemento (MORAN et al., 1996; SASSAMAN et al., 1997; HOHJOH; SINGER, 1997; MIKI, 1998).

De acordo com Miki (1998), parece existir uma divergência evolucionária entre o elemento Line-1 murino e o Line-1 humano.

Essas famílias apresentam identidade em torno de 60 %. O Line-1 murino compõe a família IAP. O elemento Line-1 murino é flanqueado por repetições diretas sugestivas do sítio de entrada no genoma, enquanto o Line-1 primata apresenta repetições do tipo CpG na região próxima a extremidade 5' fosfato. Na região 5' fostato, que antecede a fase aberta de leitura 1, também denominada de ORF-1 (open reading frame) do elemento murino, diferentemente do humano, observam-se repetições com aproximadamente 200 pb denominada de monômeros, que possui atividade promotora e ativadora da transcrição (GOODIER et al., 2001).

Nos elementos Line-1 completos, a extremidade 5' fosfato apresenta uma região de aproximadamente 1 kb, contendo vários códigos sinalizadores para final de tradução ("stop codons") em todas as fases de leituras ("frames"). Woodcock et al. (1996) mostraram que essa região codifica dois RNA pequenos, um de 305 nucleotídeos e outro de 275 nucleotídeos, cuja função ainda é desconhecida. Seguem-se a essa região duas fases abertas de leitura (ORF), a primeira, com 1kb (ORF 1) e a outra, de 4kb (ORF 2), intercaladas por uma pequena sequência que não codifica proteína. A região correspondente à extremidade 3' OH não é traduzida e possui uma sequência terminal rica em "A" (FANNING; SINGER, 1987; LANKINEN et al., 1996.; MCGLYNN et al., 2001).

O mecanismo de transposição dos elementos Line-1 ainda não está esclarecido. Sabe-se que são transcritos pela RNA polimerase II. Contudo, Kurose et al. (1995) mostraram a dependência da região promotora do Line-1 humano pela RNA polimerase III, com possível participação de fatores associados a RNA polimerase II. As ORF codificam para proteínas relacionadas a mobilidade do elemento (KOLOSHA; MARTIN, 1997; HOHJOH; SINGER, 1997), das quais a primeira codifica para uma proteína tipo gag e segunda carrega informações para transcriptase reversa, endonuclease AP, e uma possível RNAe H, como nos demais elementos dessa família. O produto da ORF 2 tem localização predominantemente citoplasmática, contudo a perda da região carboxiterminal direciona a proteína para o nucléolo (GOODIER et al., 2004).

A geração de uma nova cópia de Line-1 no genoma requer a ruptura e o reparo da fita de DNA alvo. Acredita-se que, à semelhança das retrotransposições de elementos LTR R2Bm (LUAN et al., 1993; COST et al., 2001), a ruptura no sítio-alvo geraria uma extremidade livre que atuaria como primer para a síntese da fita de DNA complementar ao RNA do elemento de transposição. No caso dos Line-1, identificou-se no genoma a sequência alvo 5'- TTTT/A - 3' como ponto preferencial de integração (GILBERT et al., 2002; SYMER et al., 2002; COST et al., 2002; GILBERT et al., 2005).

Farkash et al. (2006a e b), com base nas observações in silico de Zingler et al. (2005), que mostravam a existência de pequenos trechos de identidade entre os sítios de inserção e as extremidades 5' dos fragmentos de Line-1 de tamanho menores, não encontrada nas junções fragmentos maiores de Line-1, postularam uma segunda via de transposição com participação de endonucleases do mecanismo de reparo do DNA.

Inserções recentes de Line-1 foram associadas a doenças tanto em murinos, como em humanos. A integração da família IAP (Line-1 murino) foi identificada nas regiões codificadoras dos genes c-Fos, em plasmocitomas de camundongos, na extremidade 5' fosfato do gene c-myc, no gene Fu murino, e no gene reeler (TAKAHARA et al., 1996; MIKI, 1998). Em humanos, foi descrita pela primeira vez por Kazazian et al. (1988), interrompendo e exon 14 do fator de coagulação VIII, e posteriormente várias outras mutações foram associadas à integração "de novo" desse elemento (MIKI, 1998).

Belgnaoui et al. (2006) demonstraram que a ativação de Line-1 na célula gera múltiplas quebras na dupla-fita do DNA em células tumorais, resultando em apoptose.

Segundo Bernardi (1995), os Line estariam localizados preferencialmente em regiões ricas em AT e com baixa concentração gênica. Contrariamente, Ostertag e Kazazian (2005) sugerem que, apesar da aparente aleatoriedade na distribuição desses elementos no genoma, inserções recentes tenderiam a ocorrer em regiões próximas a genes que estariam sendo expressos. O fenômeno seria evidente em culturas de células neuronais e *in vivo*, sendo possível notar mudanças nos padrões de expressão gênica em função da entrada do elemento. Babushok et al. (2006), por meio de estudos com murinos transgênicos, também demonstraram uma tendência de direcionamento das novas inserções de Line-1 para as regiões intergênicas.

Fragmentos de Line-1 degenerados estão presentes nas regiões estruturais silenciosas, regulatórias, e em regiões transcritas de muitos genes, compondo principalmente íntrons. Esse elemento é encontrado fazendo parte da estrutura genômica do complexo de histocompatibilidade principal (MHC) (emb X87344), cadeia β dos receptores de linfócitos (gb L36092), fatores de coagulação (gb M22333 e gb M64554), receptores para interleucinas (emb X67285), fatores do complemento, e genes para βglobina (gb UO1317), entre outros (KAZAZIAN et al.,1988; MIKI, 1998; KAZAZIAN; MORAN, 1998).

Elementos repetitivos móveis curtos

As principais características dos retroelementos não virais são: ausência de LTRs; transcrição via RNA polimerase II ou III; duplicações de 7 pb a 21 pb nas extremidades; presença de poliadenilação na extremidade 3' OH; ausência ou poucos íntrons (LEWIN, 2004; BLINOV 1998). Considerados inicialmente como elementos estruturais sem função, sabe-se hoje que muitos participam de sítios regulatórios, a exemplo das ilhas de CpG e sítios recombinativos (CHRISTMANN et al., 1991; CRAIG; BICKMORE, 1994; RUBIN et al., 1994; BERNARDI, 1995; MIGHELL et al., 1997; GRAFF et al., 1997; BLINOV et al., 1998). Nesta categoria, encontram-se os Sines (short interspersed elements).

As sequências repetitivas do tipo Sine movimentam-se por retrotransposição com o auxílio de outros elementos móveis ativos, já que não codificam as enzimas necessárias para a sua própria mobilização, e são transcritos por intermédio de um promotor interno para RNA polimerase III (MATHIAS et al., 1991; WALLACE et al., 1991; BELFORT; PERLMANS, 1995; FAWCETT et al., 2006; GU et al., 2007).

Em geral, possuem tamanho entre 100 pb a 500 pb (YASUI et al., 2001; FAWCETT et al., 2006; GU et al., 2007). Em termos evolutivos, os Sines contribuíram significativamente para a história evolucionária da maioria das espécies eucarióticas (FAWCETT et al., 2006). Em vertebrados, compõe duas superfamílias Core-Sine e V Sine (KAZAZIAN et al., 2004, GILBERT; LABUDA, 1999; OGIWARA, 2002; FAWCETT et al., 2006). Em plantas, sua distribuição é mais limitada, sendo encontradas famílias de Sines do tipo AU (YASUI et al., 2001; FAWCETT et al., 2006).

Entre os Sines, as sequências do tipo Alu são as mais presentes no genoma primata. Esses elementos representam 5 % a 10 % do genoma primata – o que corresponde a pelo menos 1 cópia a cada 3 kb a 6 kb. Sua distribuição não é randômica e normalmente são encontrados no genoma, próximos uns aos outros (HWU et al., 1986; MOYZIZ et al., 1989; PITTMAN; SCHIMENTI, 1998; BLINOV et al., 1998). Por exemplo, o gene Bat2 do lócus HLA classe II possui 42 repetições de Alu, o que corresponde a 1,9 repetições a cada 1 kb (MIGHELL et al., 1997).

O elemento Alu, em murinos, é um monômero com aproximadamente 150 pb e é conhecido como elemento B (B1 e B2). Acredita-se que os monômeros tenham um ancestral comum, homólogo à partícula 7SL de RNA nuclear. A estrutura dimérica é exclusiva de primatas e estima-se que tenha surgido a 65 milhões de anos (BLINOV et al., 1998), mas a forma monomérica também está presente (ZUCMAN-ROSSI et al., 1997).

A família Alu é composta por várias subfamílias que apresentam identidade da ordem de 70 % a 75 % (DEININGER et al., 1992). Descrevem-se inserções antigas e recentes desse elemento, existindo subfamílias espécie-específicas, a exemplo de alguns tipos de Alu característicos de determinados grupos de primatas (DEININGER et al., 1992; FAWCETT et al., 2006).

O Alu dimérico possui aproximadamente 280 pb, é composto por dois monômeros ligados por uma região rica em A, sendo o direito 31 pb maior que o esquerdo. Apesar das diferenças estruturais entre a família B e Alu, ambos possuem dois sítios promotores (A e B) para RNA polimerase III e cauda poli A (BLINOV, 1998). Recombinação homóloga desigual entre elementos Alu é frequente e não randômica, ocorrendo preferencialmente entre as regiões A e B nos Alus diméricos (KASS et al., 1995).

No Alu primata, os sítios de ligação à polimerase estão no braço esquerdo e os monômeros estão ligados por uma sequência poli A (MIGHELL et al., 1997). Em elementos de origem recente, é possível observar pequenas regiões repetidas diretas nos flancos (WALLACE et al., 1991).

O conjunto das informações sugere que esses elementos tenham origem retrotransponível (MIGHELL et al., 1997; MIKI, 1998). Como esses elementos não possuem as enzimas necessárias para promoverem sua própria retrotransposição, contam com o auxílio de outros elementos ativos, tais como, Line-1 (MATHIAS et al., 1991, LANDER et al., 2001; DEWANNIEUX et al., 2003; HULME et al., 2007). Desse mecanismo, sabe-se que a proteína codificada pela ORF 2 dos elementos Line-1 é requerida no processo, e a transposição pode ser bloqueada pela proteína APOBEC3G, uma citidina deaminase que inibe a infecção pelo vírus HIV, de forma independente da expressão da proteína da ORF 1 do elemento Line -1 (HULME et al., 2007).

Cópias simples de Alu ou regiões ricas em Alu são normalmente encontradas em regiões com altos conteúdos de C/G, em regiões próximas a ilhas CpG, e flanqueando genes (GRAFF et al., 1997), íntrons e espaços intergênicos, mas não foi demonstrada a presença, até o momento, em regiões codificadoras de RNA (BLINOV et al., 1998). Existem controvérsias quanto aos motivos dessa distribuição, se essas regiões seriam sítios preferenciais de integração, ou se haveria uma seleção positiva para regiões CG ou negativa para as entradas em A/T (GROVER et al., 2003; GROVER et al., 2004; JURKA 2004; HACKENBERG et al., 2005; CORDAUX et al., 2006). Contudo, estudos com populações antigas e recentes de Alu sugerem que a entrada desses elementos no genoma seria randômica em termos de distribuição de C/G, havendo uma seleção positiva para os elementos inseridos em regiões ricas nessas bases (BELLE; EYRE-WALKER, 2002; CORDAUX et al., 2006; GU et al., 2007).

A princípio acreditava-se que a função desses elementos fosse apenas estrutural (MIGHELL et al., 1997). Porém, vários trabalhos indicam que os elementos Alu possam atuar na regulação da expressão gênica. Rubin et al. (1994) sugerem funcionalidade das sequências Alu em virtude da variação no seu padrão de metilação em diferentes tecidos. Graff et al. (1997) propõem a participação das sequências Alu no processo de metilação "de novo" das regiões CpG próximas. Lander et al. (2001) mostram que os elementos Alu podem introduzir a formação de novas ilhas CpG. Blinov et al. (1998), com base na potencialidade de os Alus formarem estruturas triplex, sugerem que as sequências Alu comporiam um código espacial formado pelas combinações das diferentes repetições completas e truncadas, diretas e invertidas.

Instabilidade de sequências

Repetições do tipo satélites

Todos os organismos eucariotos possuem pequenas sequências repetidas múltiplas vezes (TAUTZ; SCHLOTTERER, 1994).

São chamadas de microssatélites as regiões que possuem de 10 a 60 unidades de repetições seguidas, minisatélites e satélites quando apresentam um número de repetições entre 60 a 180 acima de desse valor, respectivamente. Estudos mostram que essas sequências nãomóveis são mais instáveis quando comparadas ao restante do genoma (LEVINSON; GUTMAN, 1987; HENDERSON; PETES, 1992; WEBER; WONG, 1993; WOJCIECHOWSKA et al., 2005), sendo a alteração mais comum a amplificação ou redução no número das repetições. O fenômeno foi observado em produtos clonados em E. coli (LEVINSON; GUTMAN, 1987; STRAUSS et al., 1997), em leveduras (STRAND et al., 1993, 1995) e em mamíferos (MODRICH; LAHUE, 1996; MARCADIER; PEARSON, 2003; PEARSON et al., 2005).

Foram identificadas pelo menos 40 doenças causadas pela variação no tamanho de satélites (PEARSON et al., 2005). Um exemplo é o aumento das repetições "CGG" no sítio X-Frágil. Nessa doença, observa-se uma herança estritamente materna, em que o alelo materno se expande nas células somáticas do zigoto. Quando a pré-mutação é de origem paterna, a expansão não ocorre (MCMURRAY, 1995). Acreditava-se que defeitos na meiose fossem os principais responsáveis pela expansão, contudo, verificou-se que o fenômeno ocorre normalmente na fase pós-zigótica da divisão celular, mais precisamente durante a replicação ou reparo do material genético (MARCADIER; PEARSON, 2003; PEARSON et al., 2005).

Conforme o tipo de sequências repetitivas, existe a formação de estruturas secundárias do tipo grampo ou alça. Baseado nessas informações, McMurray (1995) propôs mecanismos que poderiam explicar a expansão e (ou) redução desses segmentos. A dissociação transitória das fitas moldes e cópia, no momento da replicação, poderia ocasionar a formação dessas estruturas secundárias pela reassociação incorreta das repetições. Essas estruturas são normalmente reconhecidas e removidas pelos mecanismos de reparo do DNA. Caso a alça se dê na fita mãe, o reparo ocasionaria a perda de parte do segmento repetitivo, enquanto, na fita filha, a continuidade da replicação geraria a expansão das repetições na fita nascente (KOKOSKA et al., 1998; MARCADIER; PEARSON, 2003). Apesar de outras teorias serem plausíveis, em comum a maioria teria uma etapa onde se formam quebras na dupla-fita do DNA por endonucleases do reparo (MARCADIER; PEARSON, 2003; PEARSON et al., 2005).

Várias mutações que afetam o reparo de estruturas secundárias têm como consequência a elevação da frequência de mutações em satélites. Em leveduras, a proteína Msh2 juntamente com a Msh3 ou Msh6 e MutL, homóloga a Mlh1 e Pms1 corrigem a formação de alças com até 14 nucleotídios. Alças maiores são corrigidas por um mecanismo de reparo não usual, que envolve os genes Msh2 e Rad1 (KIRKPATRICK; PETES, 1997). Aparentemente, outras proteínas, além das que participam do reparo, estão envolvidas na estabilidade dessas regiões. Mutações no gene rad27 (FEN1 de mamíferos), responsável pela atividade 5'-3' exonucleásica, atuam na remoção da região ribonucleotídica dos fragmentos de Okaski e, na DNA polimerase δ (pol3-t), promovem aumento na instabilidade de satélites (KOKOSKA et al., 1998). Em humanos, haveria a participação do complexo Mre11-Rad50-NBS1 no reparo dessas alças (PAUL; GELLERT, 1999). Haveria, também, a participação de outras enzimas específicas (CONSTANTINOU et al., 2001) e de helicases (KAMATH-LOEB et al., 2001). Sequências repetitivas simples quando clonadas tendem a sofrer deleções ou inversão das repetições (MARCADIER; PEARSON, 2003). Próximo aos sítios de quebra e junção da região que sofreu a deleção ou inversão, observam-se, normalmente, estruturas do DNA do tipo não-B (WOJCIECHOWSKA et al., 2005).

Instabilidade de insertos clonados

Observa-se que alguns segmentos de DNAs quando clonados propiciam o rearranjo do vetor. As análises desses fragmentos mostraram que a instabilidade estaria correlacionada a presença de se-

quências repetidas diretas, invertidas e invertidas complementares, repetições do tipo microssatélite e minissatélite, todas potencialmente capazes de promover estruturas não-B no DNA (LEACH; STAHL, 1983; WYMAN, et al., 1985, 1986; MEIMA, et al., 1997; TRINH; SINDEN,1993; STRADER; HOWELL, 1997; SONG et al., 2001; KUPRINA et al., 2003; BACOLLA et al., 2004; LAN; MUGUIRA, 2005; WOJCIECHOWSKA et al., 2005; GRINDLEY et al., 2006; WANG; VASQUEZ, 2006; WOJCIECHOWSKA et al., 2006; DU et al., 2007; WELLS, 2007).

Os rearranjos podem ser complexos e resultar em perdas, duplicações, inversões, translocações de DNAs, com ou sem a presença de regiões de identidade nos sítios recombinativos (LEACH; STAHL, 1983; WYMAN et al., 1985; 1986; MEIMA et al., 1997; STRADER; HOWELL, 1997, LEEM et al., 2004). Em alguns casos, observa-se o comprometimento de sequências do vetor, inviabilizando a sua multiplicação (LEACH; STAHL, 1983; WYMAN et al., 1985; 1986; MEIMA et al., 1997; STRADER; HOWELL, 1997, LEEM et al., 2004).

Leem et al. (2004), em um trabalho que envolveu sobreposições de sequências clonadas em vetores bacterianos e de leveduras, identificaram que as regiões de instabilidade de quatro genes eram constituídas principalmente por grandes blocos de microssatélite e minissatélite e alta densidade de repetições do tipo Alu. Outras sequências, também identificadas como altamente instáveis em vetores bacterianos, foram as LineS repetições invertidas longas, sequências ricas em AT e sequências com estruturas do tipo Z-DNA (HAGAN; WARREN, 1982; SCHROTH; HO, 1995; KANG; COX, 1996; RAZIN et al., 2001). Em leveduras, tanto o tamanho das regiões repetitivas como a distância entre elas estimulam a recombinação (LOBACHEV et al., 1998). Já a presença de origem de replicação forte entre as repetições diminui a recombinação, sugerindo que a origem forte rompa a estrutura secundária que seria importante no favorecimento dos rearranjos.

Insertos ricos em CG clonados em plasmídios do tipo YACs podem ser instáveis em leveduras, indicando a participação de estruturas peculiares no mecanismo recombinativo (BERNANRDI, 1995).

Plasmídios que se mantêm em grande número de cópias na célula hospedeira, independente de presença de insertos, podem sofrer recombinação, mesmo em hospedeiras deficientes para esse tipo de mecanismo. Os produtos intermediários da recombinação formam estruturas do tipo cauda-cabeça, onde dois ou mais plasmídios ficam ligados entre si, podendo posteriormente ser resolvidos à forma monomérica. Observa-se que 95 % dos plasmídios mantidos em E. coli estão na forma monomérica, e 5 %, na forma de dímeros, trímeros ou tetrâmeros (BOE; TOLKER-NIELSEN, 1997).

Esses segmentos, eventualmente, são estabilizados quando a linhagem hospedeira é deficiente para enzimas do mecanismo de reparo e salvaguardas do DNA (LEACH; STAHL, 1983; WYMAN et al., 1985; 1986; MEIMA et al., 1997; STRADER; HOWELL, 1997; KOUPRINA et al., 2003; LEEM et al., 2004; INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2004; LAN; MUGUIRA, 2005).

Trinh e Sinden (1993) assinalaram a influência da estrutura primária e secundária de insertos com diferentes repetições diretas na estabilidade de plasmídios. Esses autores concluíram que as repetições são capazes de formar estruturas que podem ser removidas ou amplificadas, mesmo em bactérias hospedeiras deficientes em RecA, principal enzima responsável pela recombinação. Outra observação é a de que o número, o tamanho, a distância entre as repetições e a presença de palíndromos influenciam na frequência dos rearranjos que resultam em perdas. Duplicações e inversões ocorreriam a uma taxa elevada, quando não houvesse uma repetição direta ou invertida no seguimento contido entre duas repetições diretas.

Vetores contendo palíndromos superiores a 30 pb, instáveis em plasmídios, foram mantidos com sucesso em fago λ , utilizando hospedeiros deficientes para o gene recBC (exonuclease V) e sbcB (exonuclease I) (LEACH; STAHL, 1983).

Repetições superiores a 300 pb mantiveram-se estáveis quando clonadas em fago λ no hospedeiro *E. coli recA*⁺ deficientes para *recB*, recC, e sbcB ou recD (WYMAN et al., 1985; 1986). Verificou-se também que a utilização de vetores com sítios *chi*, e (ou) gene *gam* funcional pode, eventualmente, estabilizar sequências com problemas de clonagem (WYMAN et al., 1986).

Apesar do grande número de vetores e linhagens hospedeiras geradas para acomodar segmentos instáveis, a não-clonabilidade ainda representa obstáculo para o conhecimento pleno dos genomas tidos como completos (KOUPRINA et al., 2003; LEEM et al., 2004; INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2004; LAN; MUGUIRA, 2005). Estima-se que 1 % das sequências eucarióticas são de difícil clonagem e as regiões reconhecidas como instáveis permanecem nos rascunhos dos bancos genômicos até sua plena elucidação (LEEM et al., 2004).

Instabilidade e estruturas não-B

O material genético, segundo o modelo predito por Watson e Crick (1953), é constituído por duas fitas de DNA pareadas de formas invertidas e complementares, unidas por pontes de hidrogênio, que formam uma estrutura dupla hélice com giro passo direito, denominada B. Essa é a conformação predominante do material genético. Contudo, de acordo com distribuição de bases púricas e pirimídicas ao longo das fitas, a dupla hélice pode assumir estruturas diferentes das preditas por Watson e Crick e apresentar dobramentos (bend) e curvaturas (LINIAL; SHLOMAI, 1987; KIYAMA; KIYAMA 1996; GABRIELIAN et al., 1996; GABRIELIAN; PONGOR, 1996; VLAHOVICEK; PONGOR, 2000; KIYAMA et al., 1999), modificar o giro da hélice para passo esquerdo formando a estrutura tipo Z (WANG et al., 2006), parear em hélices triplas, denominadas estruturas triplex ou DNA tipo H (WANG; VASQUEZ, 2004) ou quádruplas (tetraplex) (KAUSHIK et al., 2007), ou não parear com a fita complementar e sim com segmentos na própria fita formando alças, grampos ou estruturas cruciformes (LEWIN, 2004). Essas estruturas alternativas foram demonstradas in vitro, e em muitos casos, se correlacionam a fenômenos biológicos in vivo, associados ou não a quadros clínicos (KIYAMA; KIYAMA 1996; SAMADASHWILY et al., 1997; WANG; VASQUEZ, 2004; WOJCIECHOWSKA et al., 2005; WANG, et al., 2006; WELLS, 2007; KAUSHIK et al., 2007). Na Fig. 6, pode-se ver a análise e a visualização de uma estrutura não-B do tipo curvatura típica de minicírculo de Leishmania sp.

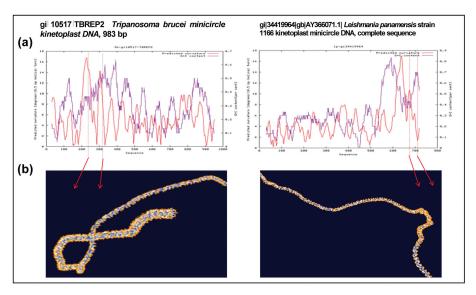


Fig. 6. Estrutura não-B, tipo dobra/curva típica, obtida da análise de minicírculos de tripanosomatídeos, por meio do programa bend.it (a) e model.it (b).

Fonte: Desenvolvido por Vlahovicek et al., (2003) (http://hydra.icgeb.trieste.it/~kristian/dna/).

Nas últimas duas décadas, muitos esforços procuraram elucidar os mecanismos que promovem a recombinação por meio da identificação de enzimas envolvidas e assinaturas genéticas nos pontos

de rearranjo (PITTMAN; SCHIMENTI, 1998; GRINDLEY et al., 2006; DU et al., 2007). E, apesar da complexidade dos mecanismos de recombinação, foi possível identificar a existência de um padrão. Em todos os casos relatados de segmentos instáveis, verificou-se que a região era propensa a formar estruturas do tipo não-B (DERE et al., 2004; BACOLLA et al., 2004; WOJCIECHOWSKA et al., 2005; WOJCIECHOWSKA et al., 2006; MARCADIER et al., 2003), sugerindo que a conformação do DNA seria um aspecto preponderante no processo (BACOLLA et al., 2004).

O efeito recombinogênico foi descrito para repetições tipo satélite (SAMADASHWILY et al., 1997; MARCADIER et al., 2003; DERE et al., 2004; WOJCIECHOWSKA et al., 2005), regiões com expectativa de estrutura Z e regiões R.Y do gene PKD1 (WANG et al., 2006; WANG; VASQUEZ, 2004), regiões do tipo curvatura e dobramento (LINIAL; SHLOMAI, 1987; KIYAMA; KIYAMA 1996; KIYAMA et al., 1999; BACOLLA et al., 2004), estrutura H, hélices triplas ou quádruplas (tetraplex) (KAUSHIK et al., 2007; WANG; VASQUEZ, 2004), repetições diretas, invertidas e invertidas complementares (LEACH; STAHL, 1983; WYMAN et al., 1985, 1986; MEIMA, et al., 1997; TRINH; SINDEN, 1993; STRADER; HOWELL, 1997; SONG et al., 2001; KUPRINA et al., 2003; LAN; MUGUIRA, 2005; GRINDLEY et al., 2006; WOJCIECHOWSKA et al., 2006; DU et al., 2007; WELLS, 2007).

Existe uma correlação entre os pontos de recombinação e bandeamento cromossômico. Bandas R e regiões de interseção R/G são sítios preferenciais nos processos de trocas (incluindo translocações espontâneas, trocas de cromátides irmãs, aberrações cromossômicas, quiasmas mitóticos) e sítios frágeis. A análise da composição de bases dos quiasmas mostra alta concentração de CG (EYRE-WALKER, 1993). Entre elas, sequências do tipo Alu, CpG, minisatélites, e teloméricas (CHRISTMANN et al., 1991; CRAIG; BICKMORE, 1994; FERRARO et al., 1993; KORENBERG; RIKOWISKI, 1988; ROYLE et al.,1988; KATINKA; BOURGAIN, 1992).

Acreditava-se, inicialmente, que a integração de DNAs exógenos não virais ocorresse aleatoriamente pelo genoma, porém são muitas as evidências de sítios e cromossomos preferenciais (revisto em RIJKERS et al., 1994). O DNA exógeno pode entrar em cópia simples, ou formar regiões com múltiplas cópias, sendo normal a orientação "cauda-cabeça". McFarlane e Wilson (1996) correlacionaram os consensos encontrados nos sítios de recombinação com os da sequência Chi de bactéria, e propuseram a participação da enzima topoisomerase I no mecanismo de integração. Da mesma forma, MacLeod et al. (1991) sugeriram que a topoisomerase I poderia favorecer a integração de uma marca de resistência clonada. Além do mais, a semelhança de uma pequena região nas extremidades da construção direcionaria a entrada para regiões inteRNA, ou adjacentes, a ilhas CpG presentes em promotores ativos. Bullock et al. (1985) também mostraram a associação entre sítios para topoisomerase I e a recombinação não-homóloga de fagos. Milot et al. (1994) observaram que nove dentre dez sequências provenientes de sítios de integração de retroelementos virais estavam correlacionados à presença de estruturas do tipo bent, conforme mostrado pela mobilidade anormal desses DNAs em gel de agarose e acrilamida.

Sequências repetitivas, como a microssatélite e a minissatélite, apresentam modificações na mobilidade eletroforética e são potencialmente capazes de formar estruturas do tipo não-B. Regiões que compreendem segmentos repetitivos como CTG/CAG, CCTG/ CAGG e GAA/TTC podem adotar estruturas do tipo desligamento transiente, grampo, alças, assumir estrutura do tipo triplex, tetraplex, cruciforme, zDNA e sticky DNA (WELLS, 2007).

Wahls et al. (1990) sugerem que microssatélites poderiam servir como sítio-específico de clivagem do DNA e promover a recombinação, após analisar a influência de uma região consenso de microssatélite clonado e o efeito na frequência da recombinação.

Sarkar et al. (1998) mostraram que a estabilização do microsatélite (CTG)_{35,50} em hospedeiras mutantes para SbcB, uma proteína que modula a clivagem da fita simples do DNA ou degradação de segmentos de DNA duplex com base em quebras na dupla-fita, sugerindo a formação de estrutura secundárias não-B.

Dere et al. (2004) mostraram que a estrutura repetitiva microssatélite (CAGG)₂₆ adotam estruturas não-B do tipo grampo, enquanto a fita complementar com a repetição (CCTG)₂₆ não possui essa propensão.

Sequências de bases potencialmente formadoras de estruturas como tipo Z-DNA in vitro são reconhecidos pela proteína rec1 (KMIEC et al., 1985; STEINMETZ, 1986). As repetições CAGG, presentes na região Eβ no sítio-recombinativo do complexo de histocompatibilidade principal, assemelham-se às estruturas Z formadas pelas repetições CGGG (STEINMETZ, 1986).

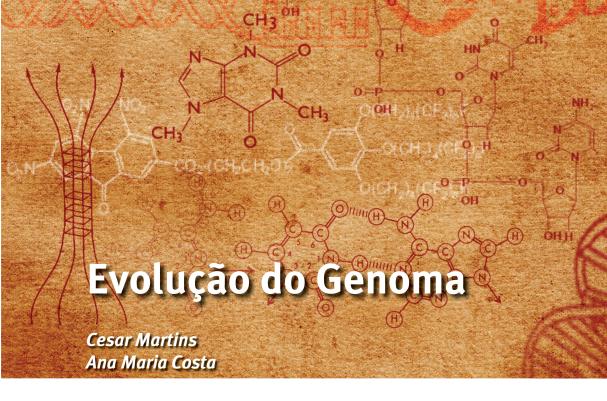
Sítios preferenciais de recombinação (hotspot) foram caracterizados em alguns genes. No cromossomo 17 murino, a região que compreende o complexo de histocompatibilidade principal apresenta quatro regiões com alta frequência de recombinação (STEINMETZ et al., 1982; 1986; SHIROISHI et al., 1995). Duas dessas regiões estão bem caracterizadas. Uma delas localiza-se na porção 3' do segundo íntron do gene Eb e compreende 1kb. O sítio-recombinativo apresenta algumas estruturas que podem atuar como ativadoras da recombinação. A primeira delas é uma sequência repetitiva que contém 10 a 18 unidades de AGGC denominadas MT, que aparece a uma frequência de 1 kb para cada 100 kb (KOBORI et al., 1986; BRYDA et al.,1992). Essa unidade de repetição lembra a sequência *Chi*, que em bactérias estimula a recombinação e integração de bacteriófagos (LEWIN, 2004; BRYDA et al., 1992; PITTMAN; SCHIMENTI, 1998).

Na mesma região, existiriam outras duas sequências frequentemente associadas a sítios recombinativos: um fragmento de um retroelemento de origem viral (LTR) e um minisatélite. Acreditase que a ligação de fatores de transcrição pode influenciar a atividade de recombinação em sítios específicos próximos (revisto por PITTMAN; SCHIMENTI, 1998). Nessa região, estão presentes também dois sítios de ligação de fatores de transcrição: a sequência B, que é reconhecida pelos fatores H2TF1/ kbF1, e NFkβ, e um domínio octâmero de ligação. A sensibilidade à DNAse também está associada a sítios recombinativos e é um indicativo da presença de struturas não-B do tipo bend (GABRIELIAN, et al., 1996).

Vizinhos ao *hotspot* existiriam dois pontos de hipersentibilidade a DNAse I, sendo um específico do paquíteno na meiose e o outro é uma região de 2 kb localizada adjacente ao gene Lmp2. Como na situação anterior, foram encontradas, no local, várias sequências que favorecem a recombinação, tais como uma repetição (CAGA)4-6, uma região semelhante a um LTR e uma sequência medianamente repetitiva; mas não foi encontrado sítio de hipersensibilidade a DNAse (PITTMAN; SCHIMENTI, 1998).

Observa-se, em regiões com repetições formadoras de estruturas não-B, a presença de quebras na dupla-fita nos pontos de formação dessas estruturas entre repetições. Acredita-se que as quebras estariam relacionadas à instabilidade da região (SARKAR et al., 1998; BACOLLA et al., 2004; WELLS; ASHIZAWA, 2006; WOJCIECHOWSKA et al., 2006). Supõe-se que a baixa estabilidade dessas conformações poderia dificultar ou impedir a replicação (SAMADASHWILY et al., 1997) e ou transcrição (BOWATER; WELLS, 2001; MIRKIN; MIRKIN, 2005).

the following genomal for sunhava por Hank Prinkler en 120, como una confusação le gene e cromottomo. No entanto o conceito jeral le genoma pode sen atributio ao seculo Itantes da era nista quando e Aristoteles apontou os princiros concettos em relação à hereatlariedade. Embarasol travalhos de Mendel, no final do século XIII lenhan propisialo grandes contribuições no campo da hereditariedade, esta area mostrava-se, ainda, bastante alstrata. Com o avanço dos métodos científicos e das ternologias, a herekitariesake passou a ser associada as estruturas presentes no núcleo das células chamadas monorsamos finalis século III é início do . século X e finalmente som est polímeros de nucleothers aupla lita hanados de DHE, A meados do seulo II) que forman os cromossomos.



A matéria-prima para a evolução é representada pelas moléculas de DNA e genes que já existem; não existem mecanismos naturais para produzir longos trechos de novas sequências de DNA. Dessa forma, novos genes e sequências de DNA surgem por modificação de sequências pré-existentes. Essas inovações podem surgir de diferentes maneiras (Fig. 7):

- 1) Mutação intragênica: um gene existente pode sofrer mutações na sua sequência de DNA, produzindo uma forma gênica modificada.
- 2) Embaralhamento de éxons: dois ou mais genes podem sofrer quebras e reorganização de forma a produzir genes quimeras consistindo de segmentos de DNA que, originalmente, pertenciam a genes separados.
- 3) Transferência horizontal: um segmento de DNA ou mesmo um genoma inteiro pode ser transferido de uma célula para

- outra da mesma espécie ou de uma espécie diferente. Esse processo é oposto à transferência vertical que ocorre dos pais para a progênie.
- 4) Duplicação gênica: um gene pode ser duplicado de forma a criar uma segunda cópia do gene dentro de uma mesma célula.

Embora esses mecanismos produzam novidades importantes para a evolução do genoma, eventos de deleções de genes e segmentos de DNA também atuam de forma eficiente produzindo mudanças no genoma. Alguns mecanismos representativos para a evolução do genoma serão discutidos a seguir.

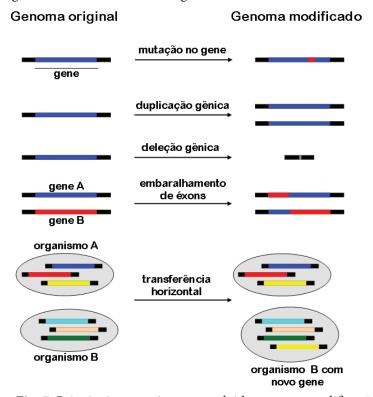


Fig. 7. Principais mecanismos envolvidos com as modificações que levam à evolução do genoma.

Mutação intragênica

Entende-se por mutação as alterações no material genético não detectáveis pelos métodos de citogenética convencional (bandeamento cromossômico). Normalmente, envolvem uma a algumas centenas de bases do DNA, sendo as modificações mais comuns:

- 1) Substiuição de base nitrogenada quando uma base do DNA é trocada ou convertida por outra.
- 2) Deleção quando ocorre a perda de uma ou mais bases nitrogenadas no segmento de DNA
- 3) Duplicação quando uma ou mais bases são duplicadas.
- 4) Inserção quando ocorre a entrada de um segmento de DNA no gene.
- 5) Inversão quando um segmento de DNA muda a sua orientação na dupla-fita.

Essas modificações podem ocorrer em regiões codificadoras ou não-codificadoras do genoma, mas são mais evidentes quando acontecem dentro de um lócus gênico pela possibilidade de mudar os níveis de expressão ou a proteína codificada (LEWIN, 2004; BALL et al., 2005).

As mutações que afetam os níveis de expressão, normalmente estão relacionadas à criação ou modificação de sítios regulatórios no lócus gênico. Um exemplo de mutação com surgimento de sítios que resultou na modificação diferencial da expressão gênica ocorreu durante o processo de formação da família gênica da globina. Essa família surgiu após sucessivas duplicações do gene mãe. As cópias ao longo da evolução acumularam mutações diversas. Algumas se tornaram inativas, enquanto outras se especializaram na expressão de duas isoformas de globina, uma específica para transportar o oxigênio na fase fetal e a outra na fase de respiração pulmonar. O gene fetal é inativado ao nível de transcrição, assim que o indivíduo nasce e passa a respirar com os pulmões (JOHNSON et al., 2006).

As mutações na região codificadora podem ser agrupadas em silenciosas, ou seja, que, apesar de presentes, não modificam a sequência da proteína, ou não silenciosas. A segunda categoria pode ser dividida em dois grupos: (a) a das mutações que modificam os aminoácidos da proteína, mas não a sua função, e (b) aquelas que modificam a proteína e sua função (LEWIN, 2004).

Embaralhamento de éxons

Logo após a descoberta da presença de éxons e íntrons nos genes de eucariontes, ficou claro que esse padrão organizacional poderia ter efeitos dramáticos na evolução das proteínas (GILBERT, 1978). A organização íntron-éxon sugere que novas combinações de éxons podem ser criadas por meio da recombinação dentro de um íntron, permitindo rearranjos de genes que podem assumir novas funções. Esse mecanismo de recombinação de éxons é chamado de embaralhamento de éxons (exon shuffling). Durante a evolução, segmentos de DNA codificadores de módulos ou domínios de proteínas sofreram duplicações e rearranjos. O embaralhamento desses domínios entre genes possibilitou a evolução das diferentes isoformas que compõem as famílias de muitas proteínas.

Uma vez que o tamanho dos íntrons é muito maior do que o tamanho dos éxons e a frequência de recombinação é proporcional ao tamanho do DNA, a grande maioria dos eventos de crossing (recombinação homóloga) ocorre nas sequências intrônicas nãocodificadoras. Além disso, íntrons contém um grande número de elementos transponíveis e sequências repetitivas, o que promove recombinação de genes não homólogos. Consequentemente, os íntrons representam hot spots (sítios preferências) para a recombinação genética, levando a um embaralhamento dos éxons com consequências significativas na evolução das proteínas. O grande número de exemplos de proteínas criadas por embaralhamento de éxons suporta o papel desse mecanismo genético na origem de novidades genéticas (KOLKMAN; STEMMER 2001).

Diversas evidências apontam que os éxons correspondam a domínios de proteínas e os íntrons servem para dividir o gene nesses pequenos segmentos funcionais. Dessa forma, os éxons funcionam como módulos ou segmentos gênicos úteis que podem ser embaralhados criando genes para novas proteínas com novas funções. Por exemplo, um éxon que codifica um domínio para inserção de uma proteína na membrana da célula poderia ser ligado a um éxon codificador de um domínio ligante de oxigênio. Isso permitiria que a nova combinação produzisse uma proteína de membrana com capacidade de estocar oxigênio (EICKBUSH, 1999).

As enzimas envolvidas no processo de coagulação sanguínea e fibrinólise representam um bom exemplo de uma família multigênica que evoluiu por embaralhamento de éxons. Análises comparativas entre essas proteases e arque-serino proteases, como as tripsinas e quimotripsinas, mostraram a presença de uma longa extensão N-terminal que consiste de uma série de domínios com funções bem definidas como reconhecimento do substrato, ligação a fosfolipídios de membrana ou interação com cofatores protéicos. A cauda N-terminal, inclui domínios EGF-like, domínios Gla, domínios fibronectina tipo I e tipo II, domínios kringle e domínios apple. A observação de que esses domínios são também encontrados em outras proteínas não relacionadas como apolipoproteína A (kringle domínios), fibronectina (domínios fibronectina tipo I e tipo II) e receptores de proteínas de baixo peso molecular (domínios EGF-like) abrem a possibilidade de que as regiões não catalíticas das proteases homeostáticas foram reunidas a partir de domínios individuais por meio de embaralhamento de éxons (KOLKMAN; STEMMER, 2001). De acordo com Patthy (1985), a diversidade de regiões N-terminal surgiu por inserção de domínios individuais entre o peptídeos sinal e o domínio serino-protease de um gene

ancestral tripsina-*like*, seguido de duplicação em *tandem* e troca de domínios.

A transferência horizontal de genes

Transferência horizontal de genes e características estruturais do genoma

A idéia clássica de evolução propõe que todos os organismos tenham um ancestral comum, e que a diversidade genética encontrada nas espécies seja decorrente das mutações, recombinações e seleção natural ocorridas ao longo dos milênios. Segundo essa teoria Darwiniana de evolução, uma espécie transfere, verticalmente, a herança genética contida em seu genoma para a sua progênie e a diversidade e a variabilidade decorreriam unicamente do acúmulo de mutações e rearranjos nos genomas. Até recentemente, acreditava-se que uma espécie assim diferenciada teria perdido a capacidade de transmitir, horizontalmente, sequências gênicas para outras espécies. Entretanto, com o advento da era genômica, verificouse que a diversidade genética é, em grande parte, promovida pela aquisição de DNAs exógenos, fusão de genomas seguida ou não de rearranjos no material genético híbrido (HACKER; KAPER, 2000; GOGARTEN, et al., 2002; KATZ, 2002; KURLAND et al., 2003; CHOI; KIM, 2007; GRINDLEY et al., 2006; MARRI et al., 2007; POPSTOVA; GOGARTEN, 2007).

Nos genomas dos mamíferos, grande parte do material genético compõe-se por segmentos de DNA repetitivos adquiridos pela integração de vírus, retrovírus e elementos transponíveis. Essas sequências, quando ativas, podem mobilizar segmentos do genoma tanto para outras células, como para outros organismos, não necessariamente da mesma espécie (LEWIN, 2004; KIM et al., 2004; FAWCETT et al., 2006; HAN et al., 2007).

Todos os segmentos de DNA são potencialmente capazes de serem transferidos de um genoma para outro, independente da participação de vírus, retrovírus ou outra partícula carreadora de material genético (LEWIN, 2004; LAWRENCE; HENDRICKSON, 2003). Desde que haja abundância de DNA no ambiente extracelular ou no citoplasma, é possível à absorção de segmentos e sua incorporação ao genoma. Esse fenômeno é comum às células procarióticas e eucarióticas e vem sendo explorado em muitos estudos da biologia molecular e engenharia genética (LEWIN, 2004). Apesar de ser um fenômeno natural, sabidamente muito comum nos estudos laboratoriais, o impacto da transferência horizontal na evolução dos genomas eucarióticos ainda é pouco estudado (RUJAN; MARTIN, 2001; STIBITZ et al., 2000; KATZ, 2002).

Sabe-se, hoje, que muitos genes nucleares tiveram origem nas organelas e em seus endossimbiontes (RUJAN; MARTIN, 2001; KATZ, 2002). A análise do genoma de Arabidopsis thaliana indica que 400 a 2.200 genes da planta foram herdados de cianobactérias (RUJAN; MARTIN, 2001), da mesma forma em Pyrenomonas helgolandii, uma alga criptofitas, a análise genealógica de dois grupos gênicos das actinas indicou que um desses grupos originou-se da transferência gênica de uma alga vermelha simbionte (STIBITZ et al., 2000; KATZ, 2002).

A transferência horizontal não se restringe apenas à transferência de fragmentos de material genético de uma célula para outra. Eventualmente genomas inteiros são fundidos, formando-se um organismo híbrido que pode ou não ser viável e gerar descendentes férteis. Esse fenômeno ocorre tanto em procariotos por meio de conjugação interespecífica, como em eucariotos (RAVEL et al., 2000). Em plantas, são transferidos genomas inteiros por fecundação cruzada. A atemóia é um exemplo desse fenômeno e resulta do cruzamento entre Annona cherimola Mill. x Annona squamosa L (NEVES; Yuhara, 2003; MELO et al., 2002). Em mamíferos, apesar de mais raro, existem relatos de cruzamentos interespecíficos, al-

guns aproveitados comercialmente, como no exemplo do cruzamento entre porco e javali (Sus scrofa X Sus scrofa scrofa), resultando no javaporco, animal servido como iguaria em muitos centros urbanos (LUI, 2000).

A incorporação do DNA exógeno pela célula receptora pode ser definitiva ou transiente. A estabilidade da transferência dependerá do tipo e da estrutura genética do material transferido, da capacidade de expressão de produtos tóxicos, dos locais da inserção no genoma, dos mecanismos de salvaguarda do DNA e da atuação de barreiras reprodutivas (LEWIN, 2004; KURLAND et al., 2003; MARRI et al., 2007; LEACH; STAHL, 1983; WYMAN, et al., 1985; 1986; MEIMA, et al., 1997; STRADER; HOWELL, 1997; SONG et al., 2001; KUPRINA et al., 2003; BACOLLA et al., 2004; LAN; MUGUIRA, 2005; WOJCIECHOWSKA et al., 2005; GRINDLEY et al., 2006; WANG; VASQUEZ, 2006; WOJCIECHOWSKA et al., 2006; DU et al., 2007; WELLS, 2007).

A transferência horizontal pode ocasionar instabilidade genética na célula receptora, gerando rearranjos variados. Após a integração, são frequentes os relatos de duplicações, translocações, ou deleções, envolvendo pequenos ou grandes segmentos cromossômicos formando novas combinações genômicas (WILKIE; PALMITER, 1987; COVARRUBIAS et al., 1986; KUPRINA et al., 2003; KIM, et al., 2004; GRINDLEY et al., 2006; DU et al., 2007; HAN et al., 2007). Isso contribuiu e, provavelmente, ainda vem contribuindo para o surgimento de novas combinações genômicas nos três grandes ramos filogenéticos Archea, Bacteria e Eukarya (HACKER; KAPER, 2000; KATZ, 2002; GOGARTEN et al., 2002; CHOI; KIM, 2007; MARRI et al., 2007; POPSTOVA; GOGARTEN, 2007).

Em termos evolutivos, a grande maioria das novas combinações resultantes desse fluxo gênico seria neutra (GOGARTEN et al., 2005; THOMPSON et al., 2005). Contudo, é inegável a importância do fenômeno na adaptação das espécies a nichos ecológicos específicos, conforme indicam os estudos filogenéticos de determinados grupamentos protéicos (HACKER; KAPER, 2000; KATZ, 2002; GOGARTEN, et al., 2002; MARRI et al., 2007).

Acredita-se que o impacto da transferência horizontal seja maior do que o estimado em função dos métodos atuais de detecção (KATZ, 2002; KURLAND et al., 2003; AZAD; LAWRENCE, 2005; POPSTOVA; GOGARTEN, 2007). Atualmente, os métodos utilizados para identificar a transferência horizontal podem ser divididos em dois grupos, os paramétricos e os filogenéticos (AZAD; LAWRENCE, 2005; POPSTOVA; GOGARTEN, 2007). O método paramétrico baseia-se na detecção de composições atípicas de sequências, quando a região é comparada com o restante do genoma. Enquanto o método filogenético procura por conflitos entre a filogenia do segmento gênico em relação ao restante do genoma. Essas duas técnicas identificam regiões de transferências que muitas vezes não se sobrepõe, sugerindo a necessidade de desenvolvimento de ferramentas mais adequadas para estudar o fenômeno (AZAD; LAWRENCE, 2005; POPSTOVA; GOGARTEN, 2007). Estudos recentes empregando modelagem in silico indicam que a taxa de transferência deva ser maior do que a estimada pelos métodos atuais (CHOI; KIM, 2007; POPSTOVA; GOGARTEN, 2007).

Transferência horizontal de genes de sequências de minicírculos de kDNA de Trypanosoma cruzi para o genoma da célula hospedeira

A progressão das lesões cardíacas em indivíduos chagásicos crônicos, a despeito da baixa parasitemia, motivou a hipótese elaborada por Teixeira et al. (1994) de que as formas amastigotas intracelulares de T. cruzi induziriam alterações no genoma hospedeiro, que poderiam estar, de alguma sorte, correlacionadas a sintomatologia característica dessa fase. As primeiras evidências de uma possível transferência de genes do parasita para o hospedeiro foram obtidas com base nos estudos citogenéticos e hibridização in situ. Esses autores analisaram macrófagos peritoniais isolados de camundongos chagásicos e mostraram a presença de estruturas anormais nas cromatinas de alguns cromossomos provenientes dos camundongos chagásicos, que não eram observadas nos normais. Os resultados das hibridações dos cromossomos com sondas provenientes do DNA do parasita foram sugestivos da associação ou uma possível integração de material genético do parasita no genoma hospedeiro, sendo o sinal mais frequente nos cromossomos 3, 6 e 11. No mesmo trabalho, os autores mostraram que o DNA encontrado provinha, exclusivamente, dos minicírculos do cinetoplasto. Com a finalidade de testar a hipótese de integração, os autores passaram a trabalhar com culturas de macrófagos murinos P388D1-IL1 infectadas com T. cruzi e, posteriormente, tratadas com benzonidazol. Erradicada a infecção, amostras de DNA foram colhidas para análise do perfil de restrição após a hibridização com sonda de minicírculos. O resultado indicou que o minicírculo estava presente em diferentes configurações no macrófago infectado-tratado, quando comparado com aquele do kDNA. Por sua vez, a hibridização com sonda de maxicírculo e de DNA nuclear, provenientes do parasita, revelou que essas sequências não estavam presente na preparação, sugerindo que o resultado positivo obtido com a sonda de minicírculo era em

consequência da integração e não, propriamente, da presença do parasita vivo. Esses achados foram confirmados pela técnica de PCR, quando foram amplificadas sequências específicas para minicírculo, sem que houvesse amplificação de segmentos de DNA nuclear do parasita (ARGAÑARAZ, 1996).

Para estudar esse fenômeno de integração, Argañaraz (1996) obteve linhagens de macrófago humano infectado-tratado. Por meio da amplificação por PCR, utilizando sequências de minicírculos como primer, esse autor obteve regiões com sequências híbridas minicírculo-DNA murino, cuja análise em banco de dados mostrou dois eventos contíguos, integração de minicírculos intercalados por um fragmento de elemento transponível da família Line-1.

Nitz et al. (2004) mostraram que a transferência de minicírculos de kDNA para as células hospedeiras ocorria também em coelhos e galinhas. Verificaram-se, nos clones provenientes das regiões de integração, alterações significativas na estrutura dos minicírculos, sugerindo sucessivos enventos de integração ou rearranjos complexos no material integrado.

Costa (2008) caracterizou os minicírculos integrados no genoma humano e sítios de integração e demonstrou a instabilidade genética das regiões quando clonadas em diferentes vetores e hospedeiras.

Simões-Barbosa et al. (1999) observaram que soros de pacientes chagásicos eram capazes de reconhecer alterações na expressão gênica das células murinas que continham minicírculos de kDNA integrados, sugerindo que a transferência poderia estar relacionada a patologia da Doença de Chagas.

Alterações na expressão gênica também foram observadas em células humanas contendo integrações de minicírculos. Simões-Barbosa et al. (2006) mostraram o bloqueio da expressão do RNAm pela entrada de um elemento Line-1 portando um minicírculo de kDNA no lócus p15, sugerindo, mais uma vez, que esse fenômeno poderia estar associado com a patologia.

Duplicação gênica: a maior força da evolução

Evidências sobre duplicação gênica e seu papel na origem de novidades genéticas vêm sendo descritas há muito tempo. Os primeiros relatos foram fornecidos por Bridges (1936), que mostrou a duplicação de uma banda cromossômica em um mutante de Drosophila melanogaster que exibia extrema redução do tamanho do olho. Nos anos subsequentes, diversos trabalhos apontavam a duplicação gênica como mecanismo potencial para a evolução (STEPHENS 1951; OHNO 1967; NEI 1969). A ideia da duplicação gênica como força motriz do processo evolutivo foi popularizada, posteriormente, pela publicação do livro Evolution by gene duplication de Susumu Ohno em 1970. No entanto, somente a partir da década de 1990, com a disponibilização de um grande número de genomas completamente sequenciados, é que se tornou possível um exame claro dos padrões de duplicação e como atuavam modificando a estruturação dos genomas.

Os mecanismos precisos dos eventos de duplicação são diversos e podem envolver crossing over desigual, retrotransposição, aneuploidia e poliploidia. Esses mecanismos podem levar às duplicações de pequenos segmentos de genes, genes inteiros, blocos de genes, cromossomos inteiros ou mesmo genomas inteiros. Acredita-se que as duplicações gênicas forneceram as condições necessárias para as principais transições evolutivas, incluindo a evolução da multicelularidade, da simetria bilateral e da evolução dos vertebrados (SIDOW 1996; HOLLAND, 1999; LUNDIN 1999; PATEL; PRINCE, 2000).

Uma vez duplicado, um gene pode estar sujeito a três destinos: (i) não-funcionalização, em que uma das duas cópias do gene degenera em um pseudogene que pode ser subsequentemente perdido do genoma; (ii) subfuncionalização, que consiste na divisão das funções originais do gene ancestral entre as duas cópias gênicas duplicadas;

(iii) neo-funcionalização, onde a cópia duplicada assume uma nova função (ZHANG, 2003).

A duplicação gênica permite que as novas cópias do gene assumam funções mais especializadas ou mesmo possibilita a origem de novos genes com funções totalmente distintas daquelas dos genes de origem. Os genes duplicados são denominados de parálogos e formam uma família multigênica. A maior família multigênica de D. melanogaster é a família codificadora das tripsinas, que possui 111 membros; enquanto a maior família de mamíferos codifica receptores olfatórios e possui 1.000 membros (GU et al., 2002; MOMBAERTS, 2001; ZHANG; FIRESTEIN 2002).

Um bom exemplo de duplicação de genes e de grupos de genes é representado pelas famílias de genes que estão envolvidos no controle do desenvolvimento. Uma dessas famílias é representada peles genes homeóticos que foram muito bem caracterizados em D. melanogaster e codificam fatores transcricionais que, possivelmente, estão envolvidos no controle do desenvolvimento de todos os animais. Alguns motifs se encontram conservados em muitos genes homeóticos. O grupo mais comum de *motifs* conservados nos genes envolvidos com o desenvolvimento é chamado de homeobox, que representa uma região de 180 pb localizada próximo do final 3' de diversos genes homeóticos. Existem cerca de 40 genes homeóticos em Drosophila que possuem homeobox. As sequências homeobox foram primeiramente identificadas e são predominantes em genes homeóticos, embora apareçam também em outros genes envolvidos no controle do desenvolvimento. A sequência proteica codificada pelo homeobox é chamada de homeodomínero e representa um motif de ligação ao DNA em fatores de transcrição.

Uma família gênica que compõe a superfamíla homeobox é representada pelos genes *Hox* que determinam, especificamente, o eixo antero-posterior dos metazoários animais. Na maioria das vezes, esses genes são expressos de acordo com sua ordem ao longo do cromossomo (isto é, genes localizados anteriormente no clus-

ter são expressos, inicialmente, nas partes anteriores do embrião) (AMORES et al., 1998). Existem 13 cópias parálogas de genes Hox nos invertebrados, organizadas em um único cluster (SANTINI; BERNARDI, 2005). Os invertebrados apresentam um único cluster de genes Hox e pouca diversidade axial, enquanto os tetrápodas (exceto peixes teleósteos) possuem quatro clusters e uma complexidade axial considerável. Os diferentes clusters de tetrápoda surgiram por duplicação do *cluster* ancestral contendo 13 genes. A origem dos vertebrados parece associada com o aumento no número de genes Hox, possivelmente como uma consequência de duas duplicações genômicas que ocorreram durante a história evolutiva das linhagens basais que deram origem aos vertebrados (HOEGG; MEYER 2005). Dessa forma, a origem de várias cópias parálogas desses genes, produtos de prováveis duplicações genômicas, permitiu ganho de complexidade na estruturação do plano do corpo e consequentemente vantagens evolutivas.

Dados recentes, envolvendo resultados da análise de genomas completamente sequenciados de peixes Actinopterigii (zebrafish, medaka e duas espécies de pufferfish) mostraram que os peixes apresentam oito clusters de genes Hox. Esses clusters adicionais, presentes em peixes, resultam de um terceiro evento de duplicação genômica que ocorreu exclusivamente nos ancestrais dos actinopterigii (HOEGG; MEYER 2005). Dessa forma, a análise da família multigênica *Hox* em vertebrados sugere a ocorrência de dois eventos de duplicação cromossômica (provavelmente duplicação genômica total) antes da divergência entre os peixes teleósteos (Actinopterigii) e tetrápoda (Sarcopterigii), e um terceiro evento de duplicação antes da radiação dos peixes teleósteos (AMORES et al., 1998). Uma vez que a duplicação gênica produz genes que podem assumir novas funções, acredita-se que os eventos de duplicação dos genes Hox tiveram papel fundamental na radiação que levou a ampla diversidade de peixes teleósteos encontrados atualmente.

O exemplo apresentado anteriormente reflete o modelo clássico de origem de novos membros de uma família multigênica por duplicação. No entanto, novos genes podem surgir com base na duplicação de genes ou sequências de DNA e assumir uma função específica e completamente distinta daquela do gene parental. Um exemplo disso está relacionado aos genes codificadores das proteínas anticongelantes presentes em peixes antárticos da subordem Nothonioidei. Essas proteínas, chamadas de AFGP (AntiFreeze GlycoProteins), impedem que esses peixes congelem a baixas temperaturas. O primeiro gene AFGP foi caracterizado na espécie Notothenia coriiceps e possui 46 genes codificadores de AFGP ligadas em tandem. Na espécie Dissostichus mawsoni, foi verificado que o gene AFGP deriva de um gene codificador do tripsinogênio pancreático (CHEN et al., 1997a, 1997b). A relação entre esses dois genes (AFGP e tripsinogênio) não pode ser explicada simplesmente por duplicação e divergência ou embaralhamento de éxons, processos comuns na origem de muitos genes. Em vez disso, a porção do gene AFGP, que codifica um domínio ice-ligante, deriva do recrutamento e interação de uma pequena região entre o primeiro íntron e o segundo éxon do gene do tripsinogênio. Esse segmento sofreu expansão e várias duplicações (talvez por *slippage* ou *crossing-over* designal), produzindo um fragmento de DNA com 41 repetições em tandem. Dessa forma, o gene AFGP atual retém, ainda, como evidências do seu surgimento, sequências em ambas as extremidades que são quase idênticas ao gene do tripsinogênio. A retenção do final 5' do gene do tripsinogênio pode ter sido significativa, pois essa região codifica um sinal peptídico utilizado para a secreção do pâncreas para o trato digestivo. Chen et al. (1997a; 1997b) hipotetizam que uma versão ancestral do gene AFGP desses peixes pode ter apresentado uma primeira função evitando o congelamento do fluído intestinal, que posteriormente expandiu para o sistema circulatório por meio da sua expressão no fígado.

the following genomal for sunhava por Hank Prinkler en 120, como una confusação le gene e cromottomo. No entanto o conceito jeral le genoma pode sen atributio ao seculo Itantes da era nista quando e Aristoteles apontou os princiros concettos em relação à hereatlariedade. Embarasol travalhos de Mendel, no final do século XIII lenhan propisialo grandes contribuições no campo da hereditariedade, esta area mostrava-se, ainda, bastante alstrata. Com o avanço dos métodos científicos e das ternologias, a herekitariesake passou a ser associada as estruturas presentes no núcleo das células chamadas monorsamos finalis século III é início do . século X e finalmente som est polímeros de nucleothers aupla lita hanados de DHE, A meados do seulo II) que forman os cromossomos.



Conforme apresentado, existem muitos mecanismos moleculares que geram variabilidade e diversidade genômica. Graças a eles, novas formas de vida surgem constantemente na natureza. Algumas se mantêm, outras se perdem de acordo com a pressão de seleção ou ausência dela. O ambiente contribui não só selecionando os indivíduos mais aptos, mas em alguns casos desencadeando processos celulares que resultarão na alteração de genomas e, consequentemente no aumento da variabilidade e diversidade.

Esses mecanismos são tão poderosos que nem sempre é fácil manter a uniformidade genética nas situações onde ela é desejável, como nos casos dos cultivos de microorganismos para fins de pesquisa ou industriais, alguns sistemas de produção agrícola e conservação de recursos genéticos. Em geral, essas atividades exigem constantes esforços para selecionar e recuperar a linhagem original.

O retrato que se vê, hoje, é o da vida como resultado de milênios a bilhões de anos de evolução. Acreditar que esse é o resultado final é um erro. O tempo não pára, o processo continua, tanto para o homem, quanto para todos os seres vivos, pois a variabilidade e diversidade genética são as bases para a manutenção da vida no planeta.

Há menos de 25 anos, os biologistas moleculares consideravam o genoma uma estrutura estável mantenedora do patrimônio genético. Os avanços, nestes últimos anos, vêm mostrando, de forma notável, o intenso dinamismo com que as sequências de DNA e genes que compõe o genoma sofrem alterações. A plasticidade do genoma em responder ao ambiente celular e externo é fantástica, pois possibilita a origem das variações tão importantes para a evolução e manutenção das formas de vida. Não podemos ver o genoma como uma estrutura governada unicamente por padrões Mendelianos de herança, mas sim uma estrutura complexa que sofre mudanças a todo o momento, permitindo que suas diferentes classes de sequências de DNA estejam em permanente transformação.



ALDER, B. K.; HAJDUK, S. L. Guide RNA requirement for editing-sitespecific endonucleolytic cleavage of preedited mRNA by mitochondrial ribonucleoprotein particles in Trypanosoma brucei. Molecular and cellular biology, v. 17, p. 5377-5385, 1997.

ALDER, B. K.; HAJDUK, S. L. Mechanisms and origins of RNA editing. Current Opinion Genetics & Development, v. 4, p. 316-322, 1994.

AMORES, A.; FORCE, A.; YAN, Y. L.; WANG, -L.; WESTERFIELD, M.; POSTLETHWAITET, J. H. Zebrafish hox clusters and vertebrate genome evolution. Science, v. 282, p. 1711-1714, 1998.

ARGAÑARAZ, E. R. Integração de sequências de minicírcilo de kDNA de *Trypanosoma cruzi* em elemento L1 no Genoma da Célula hospedeira. 1996. Tese (Doutorado em imunologia e genética aplicadas) - Faculdade de Ciências da Saúde, Universidade de Brasília, 1996.

ARKHIPOVA, I.; MESELSON, M. Transposable elements in sexual and ancient asexual taxa. Proceedings of the National Academy Science of the USA (PNAS), v. 97, p. 14473-14477, 2000.

ÁVILA, H.; SIMPSON, L. Organization and complexity of minicircle-encoded guide RNAs from Trypanosoma cruzi. RNA, v. 1, p. 939-947, 1997.

AZAD, R. K.; LAWRENCE, J. G. Use of Artificial Genomes in Assessing Methods for Atypical Gene Detection. PLoS Computational Biology, v. 1, n. 6, p. e56, 2005.

BABUSHOK, D. V.; OSTERTAG, E. M.; COURTNEY, C. E.; CHOI, J. M.; KAZAZIAN, H. H. Jr. L1 integration in a transgenic mouse model. Genome Research, v. 16, p. 240-250, 2006.

BACOLLA, A.; JAWORSKI, A.; LARSON, J. E.; JAKUPCIAK, J. P.; CHUZHANOVA, N.; ABEYSINGHE, S. S.; O'CONNELL, C. D.; COOPER, D. N.; WELLS, R. D. Breakpoints of gross deletions coincide with non-B DNA conformations. Proceedings of the National Academy Science of the USA (PNAS), v. 101, p. 14162-14167, 2004.

BALL, E. V.; STENSON, P. D.; ABEYSINGHE, S. S.; KRAWCZAK, M.; COOPER, D. N.; CHUZHANOVA, N. A. Microdeletions and Microinsertions Causing Human Genetic Disease: Common Mechanisms of Mutagenesis and the Role of Local DNA Sequence Complexity. Human Mutation, v. 26, n. 3, p. 205-213, 2005.

BARBROOK, A. C.; HOWE, C. J.; PURTON, S. Why are plastid genomes retained in non-photosynthetic organisms? Trends Plants Science, v. 11, n. 2, p. 101-108, 2006.

BELFORT, M.; PERLMAN, P. S. Mechanism of intron mobility. The Journal of Biological Chemistry, v. 270, n. 51, p. 30237-30240, 1995.

BELGNAOUI, S. M.; GOSDEN, R. G.; SEMMES, O. J.; HAOUDI, A. Human Line-1 retrotransposon induces DNA damageand apoptosis in cancer cells. Cancer Cell International, v. 6, p. 13, 2006.

BELLE, E. M. S.; EYRE-WALKER, A. A test of whether selection maintains isochores using sites polymorphic for Alu and 11 element insertions. Genetics, v. 160, p. 815-817, 2002.

BENITO, M. I; WALBOT, V. Characterization of the maize *mutator* transposable element MURA transposase as a DNA bindig protein. Molecular and Cellular Biology, v. 17, n. 9, p. 5165-5175, 1997.

BENNE, R.; VAN DEN BURG, J.; BRAKENHOFF, J. P. J.; SLOOF, P.; VAN BOOM, J. H. E TROMP, M. C. Major transcript of the frameshifted cox III gene from Trypanosoma mitoncondria contains four nucleotides that are not encoded in DNA. Cell, v. 46, p. 819-826, 1986.

BERGER, S. M.; MOORE, C.; SHARP, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proceedings of the National Academy Science of the USA (PNAS), v. 74, p. 3171-3175, 1977.

BERNARDI, G. The human genome: organization and evolutionary history. Annual Review Genetics, v. 29, p. 445-476, 1995.

BERNARDI, G. The neoselectionist theory of genome evolution. Proceedings of the National Academy Science of the USA (PNAS), v. 20, n. 104, p. 8385-8390, 2007.

BESTOR, T. H.; TYCKO, B. Creation of genomic methylation patterns. Nature Genetics, v. 12, p. 636-367, 1996.

BLINOV, V. M.; RESENCHUK, S. M.; UVAROV, D. L.; CHIRKOVA, G. B.; DENISOV, S. I.; KISSELEV, L. L. Alu elements in human genome: invariant secondary structure of Left and Right monomers. Molecular Biology, v. 32, n. 1, p. 70-77, 1998.

BLUM, B.; BAKALARA, N.; SIMPSON, L. A model for RNA editing kinetoplastid mithocondria "Guide" RNA molecular transcribed from maxicircle DNA provided the edited information. Cell, v. 60, p. 189-198, 1990.

BOE, L.; TOLKER-NIELSEN, T. Plasmid stability: comments on the dimer catastrophe hypothesis. Molecular Microbiology, v. 23, n. 2, p. 247-253, 1997.

BOWATER, R. P.; WELLS, R. D. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. Progress in Nucleic Acids Research Molecular Biology, v. 66, p. 159-202, 2001.

BRIDGES, C. B. The Bar 'gene' a duplication. Science, v. 83, p. 210-211, 1936.

BRYDA, E. C.; DEPARI, J. A.; SANT' ANGELO, D. B.; MURPHY, D. B.; PASSMORE, H. C. Multiple sites of crossing over within the Eb recombinational hotspot in the mouse. Mamalina Genome, v. 2, p. 123-129, 1992.

BULLOCK, P.; CHAMPOUX, J. J.; BOTCHAN, M. Association of crossover poinnts with topoisomerase I cleavage sites: a model for nonhomologous recombination. Science, v. 230, p. 954-958, 1985.

BURWINKEL B.; KILIMANN, M. Unequal homologous recombination between Line – 1 elements as a mutational mechanism in human genetic disease. Journal Molecular Biology, v. 277, p. 513-517, 1998.

CANN, A. J. Principles of molecular virology. 2. ed. San Diego: Academic Press, 1997. 310 p.

CAVALIER-SMITH, T. Intron phylogeny: a new hypotesis. Trends in Genetics, v. 7, n. 5, p. 145-148, 1991.

CHARLESWORTH, B.; SNLEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature, v. 371, p. 215-220, 1994.

CHEN, L.; DEVRIES, A. L.; CHENG, C. H. C. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proceedings of the National Academy Science of the USA (PNAS), v. 94, p. 3811-3816, 1997a.

CHEN, L.; DEVRIES, A. L.; CHENG, C. -H. C. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proceedings of the National Academy Science of the USA (PNAS), v. 94, n. 8, p. 3817-3822, 1997b.

CHOI, I. -G.; KIM, S. -H. Global extent of horizontal gene transfer. Proceedings of the National Academy Science of the USA (PNAS), v. 104, n. 11, p. 4489-4494, 2007.

CHOW, L. T.; GELINAS, R. E.; BROKER, T. R.; ROBERTS, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell, v. 12, p. 1-8, 1977.

CHRISTMANN, A.; LAGODA, P. J. L.; ZANG, K. D. Non-radioactive in situ hybridization pattern of the M13 minisatellite sequences on human methaphase chromosomes. Human Genetic, v. 86, p. 487-490, 1991.

CONSTANTINOU, A.; DAVIES, A. A.; WEST, S. C. Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. Cell, v. 104, p. 259-268, 2001.

CORDAUX, R.; LEE, J.; DINOSO, L.; BATZER, M.A. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene, v. 373, p. 138-144, 2006.

COST, G. J.; FENG, Q.; JACQUIER, A.; BOEKE, J. D. HumanL1element targetprimed reverse transcription in vitro. EMBO Journal, v. 21, p. 5899–5910, 2002.

COST, G. J.; GOLDING, A.; SCHLISSEL, M. S.; BOEKE, J. D. Target DNA chromatinization modulates nicking by L1 endonuclease. Nucleic Acids Research, v. 29, p. 573-577, 2001.

COSTA, A. M. Caracterização de regiões de integração de minicírculos de kDNA de *Trypanosoma cruzi* no genoma humano. 2008. 165 f. Tese (Doudorado) - Universidade de Brasília, Brasília, DF.

COSTANTINI, M.; CLAY, O.; FEDERICO, C.; SACCONE, S.; AULETTA, F.; BERNARDI, B. Human chromosomal bands: nested structure, high-definition map and molecular basis. Chromosoma, v. 116, p. 29-40, 2007.

COSTANTINI, M.; CLAY, O; AULETTA, F.; BERNARDI, G. An isochore map of human chromosomes. Genome Research, v.16, p. 536-541, 2006.

COVARRUBIAS, L.; NISHIDA, Y.; E MINTZ, B. Early postimplantation embryos lethality due to DNA rearrangments in a transgenic mouse strain. Proceedings of the National Academy Science of the USA (PNAS), v. 83, p. 6020-6024, 1986.

CRAIG, J. M.; BICKMORE, W. A. The distribution of CpG island in mammalian chromosomes. Nature Genetics, v. 7, p. 376-381, 1994.

CRUZ-REYES, J.; SOLLNER-WEBB. Trypanosome U-deletion RNA editing involves guide RNA directed endonuclease cleavage, terminal U exonuclease, e RNA activities. Proceedings of the National Academy Science of the USA (PNAS), v. 93, p. 8901-8906, 1996.

DEININGER, P. L.; BATSER, M. A.; HUTCHISON, III C. A; EDGEL, M. H. Mastar genes in mammalian repetitive DNA amplification. TIG, v. 8, n. 9, p. 307-311, 1992.

DEININGER, P. L.; MORAN, J. V.; BATZER, M. A.; KAZAZIAN H. H. JR. Mobile elements and mammalian genome evolution. Current Opinion in Genetic & Development, v. 13, p. 651, 2003.

DERE, R.; NAPIERALA M.; RANUM, L. P. W.; WELLS, R. D. Hairpin Structure-forming Propensity of the (CCTG-CAGG) Tetranucleotide Repeats Contributes to the Genetic Instability Associated with Myotonic Dystrophy Type 2. The Journal of Biological Chemistry, v. 279, n. 40, p. 41715-41726, 2004.

DEWANNIEUX, M.; ESNAULT, C.; HEIDMANN, T.Line-mediated retrotransposition of marked Alu sequences. Nature Genetics, v. 35, p. 41-48, 2.003.

DOAK, T. G.; WITHERSPOON, D. J.; DOERDER, F. P.; WILLIAMS, K.; HERRICK. G. Conserved features of TBE1 transposons in ciliated protozoa. Genetica, v. 101, p. 75-86, 1997.

DOBRINDT, U.; HACKER, J. Whole genome plasticity in pathogenic bacteria. Current Opinion in Microbiology, v. 4, p. 550-557, 2001.

DOMBROSKI, B. A.; SCOTT, A. F.; KAZAZIAN JR., H. H. Two additional potential retrotransposons isolated from human L1 subfamily that contain na active retrotransposable element. Proceedings of the National Academy Science of the USA (PNAS), v. 90, p. 6513-6517, 1993.

DU, Q.; LIVSHITS, A.; KWIATEK, A.; JAYARAM, M.; VOLOGODSKII, A. Protein-induced Local DNA Bends Regulate Global Topology of Recombination Products Journal of Molecular Biology, v. 368, p. 170-82, 2007.

EICKBUSH, T. Exon Shuffling in Retrospect. Science, v. 283, p. 1465-1467, 1999.

EMELYANOV, V. V. Mitochondrial connection to the origin of the eukaryotic cell. European Journal Biochemistry, v. 270, p. 1599-1618, 2003.

EMELYANOV, V. V. Rickettsiaceae, rickettsia-like endosymbiont and the origin of mitochondira. Bioscience Reports, v. 21, n. 1, p. 1-17, 2001.

EYRE-WALKER, A. Recombinationand mammalian genome evolution. Proceedings of the Royal Society B: Biological Sciences, Ser. B252, p. 237-243, 1993.

FANNING, T. G.; SINGER, M. F. Line 1: A mammalian transposable element. Biochimica et Biophysica Acta, v. 910, p. 203-212, 1987.

FARKASH, E. A.; KAO, G. D.; HORMAN, S. R.; PRAK, E. T. L. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. Nucleic Acids Research, v. 34, n. 4, p. 1196-1204, 2006a.

FARKASH, E. A.; PRAK, E. T. L. DNA Damage and L1 Retrotransposition. Journal of Biomedicine and Biotechnology, v. 2006, p. 1-8, 2006b.

FAWCETT, J. A.; KAWAHARA, T.; WATANABE H.; YASUI Y. A Sine Family Widely Distributed in the Plant Kingdom and its Evolutionary History. Plant Molecular Biology, v. 61, n. 3, p. 505-14, 2006.

FEAGIN, J. E.; ABRAHAM, J. M.; STUART, K. Extensive editing of the citocromo c oxidase III transcript in Trypanosoma brucei. Cell, v. 53, p. 413-422, 1988.

FENG, Q.; MORAN, J. V.; KAZAZIAN, H. H. Jr.; BOEKE, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell, v. 87, p. 905-916, 1996.

FERRARO, M.; PREDAZZI, V.; PRANTERA G. In human chromosomes telomeric regions are enriched in CpG relative to R-bands. Chromosoma, v. 102, p. 712-717, 1993.

FINNEGAN, D. J. Retrovirus transposable elements- which come first? Nature, v. 302, p. 105-106,1983.

GABRIELIAN, A.; PONGOR, S. Correlation of intrinsic DNA curvature with DNA property periodicity. FEBS Letters, v. 393, p. 65-68, 1996.

GABRIELIAN, A.; SIMONCSITS, S.; PONGOR S. Distribution of bending propensity in DNA sequences. FEBS Letters, v. 393, p. 124-130, 1996.

GILBERT, W. Why genes in pieces? Nature, v. 271, p. 501, 1978.

GILBERT, N.; LABUDA, D. CORE-Sines: eukaryotic short interspersed retroposing elements with common sequence motifs. Proceedings of the National Academy Science of the USA (PNAS), v. 96, n. 6, p. 2869, 1999.

GILBERT, N.; LUTZ, S.; MORRISH, T. A.; MORAN, J. V. Multiple fates of 11 retrotransposition intermediates in cultured human cells. Molecular Cell Biology, v. 25, p. 7780-7795, 2005.

GILBERT, N.; LUTZ-PRIGGE, S.; MORAN, J. V. Genomic deletions created upon Line-1 retrotransposition. Cell, v. 110, p. 315-325, 2002.

GOFFEAU, A.; BARRELL, B. G.; BUSSEY, H.; DAVIS, R. W.; DUJON, B.; FELDMANN, H.; GALIBERT, F.; HOHEISEL, J. D.; JACQ, C.; JOHNSTON, M.; LOUIS, E. J.; MEWES, H. W.; MURAKAMI, Y.; PHILIPPSEN, P.; TETTELIN, H.; OLIVER, S. G. Life with 6000 genes. Science, v. 274, p. 546, 563-567, 1996.

GOGARTEN, J. P.; TOWNSEND, J. P. Horizontal gene transfer, genome innovation and evolution. Nature Reviews Microbiology, v. 3, n. 9, p. 679-687, 2005.

GOGARTEN, J. P.; DOOLITTLE, W. F.; LAWRENCE, J. G. Prokaryotic evolution in light of gene transfer. Molecular Biology Evolution, v. 19, n. 12, p. 2226-2238, 2002.

GOODIER, J. L.; OSTERTAG, E. M.; DU, K.; KAZAZIAN, H. H. JR. A Novel Active L1 Retrotransposon Subfamily in the Mouse. Genome Research, v. 11, p. 1677-1685; 2001.

GOODIER, J. L.; OSTERTAG, E. M.; ENGLEKA, K. A.; SELEME, A. S.; KAZAZIAN, H. H. Jr. A potential role for the nucleolus in L1 retrotransposition. Human Molecular Genetics, v. 13, n. 10, 2004.

GRAFF, J. R.; HERMAN, J. G.; MYÖHÄNEN, S.; BAYLIN, S. B. E.; VERTINO, P. M. The Journal of Biological Chemistry, v. 272, n. 35, p. 22322-22329, 1997.

GRAY, M. W.; BURGER, G.; LANG, B. F. The origin and early evolution of mitochondria. Genome Biology, v. 2, n. 6, p. 1018.1-1018.5, 2001.

GRINDLEY, N. D. F; WHITESON, K. L.; RICE, P. A. Mechanisms of site-specific recombination Annual Review Biochemistry, v. 75, p. 567-605, 2006.

GROVER, D.; MAJUMDER, P. P. C. B. R.; BRAHMACHARI, S. K.; MUKERJI, M. Nonrandom distribution of Alu elements in genes of various functional categories: insight from analysis of human chromosomes 21 and 22. Molecular Biological Evolution, v. 20, p. 1420-1424, 2003.

GROVER, D.; MUKERJI, M.; BHATNAGAR, P.; KANNAN, K.; BRAHMACHARI, S. K. Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. Bioinformatics, v. 20, p. 813-817, 2004.

GU, W.; DAVID, A.; RAY, D. A.; WALKER, J. A.; BARNES, E. W.; GENTLES, A. J.; SAMOLLOW, P. B.; JURKA, J.; BATZER, M. A.; POLLOCK, D. D. Sines, evolution and genome structure in the opossum. Gene, v. 396, n. 1, p. 46-58, 2007.

GU, Z.; CAVALCANTI, A.; CHEN, F. C.; BOUMAN, P.; LI, W. H. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. **Molecular** Biology and Evolution, v. 19, p. 256-262, 2002.

HACKENBERG, M.; BERNAOLA-GALVAN, P.; CARPENA, P.; OLIVER, J. L. The biased distribution of Alus in human isochores might be driven by recombination. Journal of Molecular and Evolution, v. 60, p. 365-377, 2005.

HACKER, J.; KAPER, J. B. Pathogenicity islands and the evolution of microbes. Annual Review Microbiology, v. 54, p. 641-679, 2000.

HAGAN, C. E.; WARREN, G. J. Lethality of palindromic DNA and its use in selection of recombinant plasmids. Gene, v. 19, p. 147-151, 1982.

HAJDUK, S. L.; HARRIS, M. E.; POLLARD, V. W. RNA editing in kinetoplastid mitochondria. The FAESB Journal, v. 7, p. 54-63, 1993.

HAN, K.; KONKEL, M. K.; XING, J.; WANG, H.; LEE, J.; MEYER, T. J.; HUANG, C. T.; SANDIFER, E.; HEBERT, K.; BARNES, E. W.; HUBLEY, R.; MILLER, W.; ARIAN F. A. ULLMER, S. B.; BATZER M. A. Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome. Science, v. 316, p. 238-240, 2007.

HAN, K.; SEN, S. K.; WANG, J.; CALLINAN, P. A.; LEE, J.; CORDAUX, R.; LIANG, P.; BATZER, M. A. Genomic rearrangements by Line-1 insertionmediated deletion in the human and chimpanzee lineages. Nucleic Acids Research, v. 33, p. 4040-4052, 2005.

HANDLER, A. M.; GOMEZ, S. P. P element excision in Drosophila is stimulated by gamma-irradiation in transient embryonic assays. Genome Research, v. 70, n. 1, p.75-78, 1997.

HENDERSON, S. T.; PETES, T. D. Instability of simple sequence DNA in Saccharomyces cerevisiae. Molecular Cell Biology, v. 12, p. 2749-2757, 1992.

HICKEY, D. A. Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. Genetica, v. 86, p. 69-274, 1992.

HOEGG, S.; MEYER, A. Hox *clusters* as models for vertebrate genome evolution. Trends in Genetics, v. 21, p. 421-424, 2005

HOHJOH, H.; SINGER, M. J. Sequence-specific single-strand RNA binding protein encoded by the human Line-1 retrotransposon. EMBO Journal, v. 16, n. 19, p. 6034-6043, 1997.

HOHJOH, H.; SINGER, M. F. Cytoplasmic ribonucleoprotein complexes containing human Line-1 protein and RNA. EMBO Journal, v. 15, p. 630-639, 1996.

HOLLAND, P. W. H. The effect of gene duplication on homology. In: BOCK, G. R.; CARDEW, G. (Ed.). Homology. New York: John Wiley, 1999. p. 226-242.

HORVATH, J. E.; BAILEY, J. A.; LOCKE, D. P.; EICHLER, E. E. Lessons from the human genome: transitions between euchromatinand heterochromatin. Human Molecular Genetics, v. 10, p. 2215-2223, 2001.

HULME, A. E.; BOGERD, H. P.; CULLEN, B. R.; MORAN, J. V. Selective inhibition of Alu retrotransposition by APOBEC3G. Gene, v. 390, p. 199-205, 2007.

HWU, H. R.; ROBERTS, J. W.; DAVIDSON, E. H.; BRITTEN, R. J. Insertion and/ or deletion of many repeated DNA sequences in human and higher ape evolution. Proceedings of the National Academy Science of the USA (PNAS), v. 83, p. 3875-3879, 1986.

ICHIYANAGI, K.; OKADA. N. Genomic alterations upon integration of zebrafish L1 elements revealed by the TANT method. Gene, v. 383, p. 108-116, 2006.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. Nature, v. 409, p. 860-921, 2001.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Finishing the euchromatic sequence of the human genome. Nature, v. 431, p. 931-945, 2004.

JOHN, L.; GOODIER, E. M.; OSTERTAG, K. D.; KAZAZIAN, H. H. JR. A Novel Active L1 Retrotransposon Subfamily in the Mouse. Genome Research, v.11, p. 1677-1685, 2001.

JOHNSON, R. M.; PRYCHITKO, T.; GUMUCIO, D.; WILDMAN, D. E.; UDDIN, M.; GOODMAN, M. Phylogenetic comparisons suggest that distance from the locus control region guides developmental expression of primate betatype globin genes. Proceedings of the National Academy Science of the USA (PNAS), v.103, n. 9, p. 3186-3191, 2006.

JURKA, J. Evolutionary impact of human Alu repetitive elements. Current Opinion Genetics and Development, v. 14, p. 603-608, 2004.

KABLE, M. L.; SEIWERT, S. D.; HEIDMANN, S.; AND STUART, K. RNA editing: a mechanism for gRNA-specific uridylate insertion into precursor mRNA; Science, v. 273, p. 1189-1195, 1996.

KAMATH-LOEB, A. S.; LOEB, L. A.; JOHANSSON, E.; BURGERS, P. M.; FRY, M. Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. The Journal of Biological Chemistry, v. 276, p. 16439-16446, 2001.

KANG, H. K.; COX, D. W. Tandem repeats 3 of the IGHA genes in the human immunoglobulin heavy chain gene cluster. Genomics, v. 35, p. 189-195, 1996.

KASS, D. H.; BATZER, M. A.; DEININGR, P. LGene conversion as a secondary mechanism of short interspersed element (Sine) evolution. Molecular Cell Biology, v. 15, n. 1, p. 19-25, 1995.

KATINKA M. D.; BOURGAIN, F. M. Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of Paramecium primaurelia. EMBO Journal, v. 11, n. 2, p. 725-732, 1992.

KATZ, L. A.; Lateral gene transfers and the evolution of eukaryotes: theories and data. International Journal of Systematic and Evolutionary Microbiology, v. 52, p. 1893-1900, 2002.

KAUSHIK, M.; BANSAL, A.; SAXENA, S.; KUKRETI. S. Possibility of an Antiparallel (Tetramer) Quadruplex Exhibited by the Double Repeat of the Human Telomere. Biochemistry, v. 46, p. 7119-7131, 2007.

KAZAZIAN, H. H. JR.; MORAN, J. V. The impact of L1 retrotransposons on the human genome. Nature Genetics, v. 19, p. 19-24, 1998.

KAZAZIAN, H. H. JR.; WONG, C.; YOUSSOUFIAN, H.; SCOTT, A. F.; PHILLIPS, D. G.; ANTONARAKIS, S. E. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature, v. 332, p. 164-166, 1988.

KAZAZIAN, H. H. JR. Mobile elements: drivers of genome evolution. Science, v. 12, p. 1626-1632, 2004.

KHACHANE, A. N.; TIMMIS, K. N.; MARTINS DOS SANTOS, V. A. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. Molecular Biology Evolution, v. 24, n. 2, p. 449-456, 2007.

KIM, T.-M.; HONG, S.-J.; RHYU, M.-G. Periodic Explosive Expansion of Human Retroelements Associated with the Evolution of the Hominoid Primate. Journal Korean Med Science, v. 19, p. 177-85, 2004.

KIRKPATRICK, D. T; PETES, T. D. Repair of DNA loops involves DNAmismatch and nucleotide-excision repair proteins. Nature, v. 387, p. 929-931, 1997.

KIYAMA, W. Y.; SUZUKI, K.; KIYAMA, K. DNA bend sites in the human b-globin locus: evidence for a basic and universal structural component of genomic DNA. Molecular Biology and Evolution, v. 16, n. 7, p. 922-930, 1999.

KIYAMA, W. Y.; KIYAMA, K. An Intrachromosomal Repeating Unit Based on DNA Bending. Molecular and Cellular Biology, v. 16, p. 5664-5673, 1996.

KMIEC, E. B.; ANGELIDES, K. J.; HOLLOMAN, W. K. Left-handded DNA and the synaptic pairin reaction promoted by ustilago rec1 protein. Cell, v. 40, p. 139-145, 1985.

KOBORI, J. A.; STRAUSS, C. A.; MINARD, K.; HOOD, L. Molecular analysis of hotspot recombination in the murine histocompatibility complex. Science, v. 234, p. 173-179, 1986.

KOKOSKA, R. J.; STEFANOVIC, L.; TRAN, H. T. T.; RESNICK, M. A.; GORNENIN, D. A.; PETES, T. D. Destabilization of yeast micro- and minisatellite DNA sequences by mutation affecting a nuclease involved in okasaki fragment processing (rad27) and DNA polimerase δ (pol3-t). Molecular and cellular biology, v. 18, n. 5, p. 2779-2788, 1998.

KOLKMAN J. A.; STEMMER, W. P. C. Directed evolution of proteins by exon shuffling. Nature Biotechnology, v. 19, p. 423-428, 2001.

KOLOSHA, V. O.; MARTIN, S. L. In vitro properties of the first ORF protein from mouse Line-1 support its role in ribonucleoprotein particle formation during retrotransposition. Proceedings of the National Academy Science of the USA (PNAS), v. 94, p. 10155-10160, 1997.

KORENBERG, J. RIKOWISKI, M. Human molecular organization: alu, lines and molecular structure of metaphase chromosome bands. Cell, v. 53, p. 391-400, 1988.

KOUPRINA, N.; LEEM, S.-H; SOLOMON, G. L. Y. A.; KORIABINE, M.; OTSTOT, J.; PAK, E.; DUTRA, A.; ZHAO S.; BARRET, J. C.; LARIONOV, V. Segments missing from the draft human genome sequence can be isolated by transformationassociated recombination cloning in yeast. European Molecular Biology Organization (EMBO Reports), v. 4, n. 3, p. 257-262, 2003.

KURLAND, C. G.; CANBACK, B.; BERG, O. G. Horizontal gene transfer: A critical view. Proceedings of the National Academy Science of the USA (PNAS), v. 100, n. 17, p. 9658-9662, 2003.

KUROIWA, T.; NISHIDA, K.; YOSHIDA, Y.; FUJIWARA, T.; MORI, T.; KUROIWA, H.; MISUMI, O. Structure, function and evolution of the mitochondrial division apparatus. Biochim Biophys Acta, v. 1763, n. 5-6, p. 510-521, 2006.

KUROSE, K.; HATA, K.; HATTORI, M.; SAKAKI, Y. RNA polimerase III dependence of the human L1 promoter na possible participationn of the RNA polimerase II factor YY1 in the RNA polimerase III transcription system. Nucleic Acids Research, v. 23, n. 18, p. 3704-3709, 1995.

LAM, K. W. G.; JEFFREYS, A. J. Processes of de novo duplication of human -globin genes. Proceedings of the National Academy Science of the USA (PNAS), v. 104 n. 26, p. 10950-10955, 2007.

LAMBOWITZ, A. M. Infectious introns. Cell, v. 56, p. 323-326, 1989.

LAN, M. S.; MUGUIRA, M. An effective alternate cloning strategy for unstable mouse genomic sequences. Biochemical and Biophysical Research Communications, v. 330, p. 641-644, 2005.

LANDER, E. S.; WATERSTON, R. H.; SULSTON, J.; COLLINS, F. S. (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM). Initial sequencing and analysis of the human genome. Nature, 409, 860-921, 2001.

LANKINEN, M. H.; VILPO, L. M.; VILPO, J. A. UV- and γ-irradiation induced DNA single-strand breaks and their repair in human blood granulocytes and lymphocytes. Mutation Research/ Fundamental and Molecular Mechanisms of Mutagenesis, v. 352, n. 1-2, p. 31-38, 1996.

- LAWRENCE J. G.; HENDRICKSON, H. Lateral gene transfer: When will adolescence end? Molecular Microbiology, v. 50, p. 739-749, 2003.
- LAWRENCE, J. G.; HENDRICKSON H. Genome evolution in bacteria: order beneath chaos. Current Opinion in Microbiology, v. 8, p. 572-578, 2005.
- LEACH, R. F.; STAHL, F. W. Viability of λ phages carrying a perfect palindrome in the absence of recombination nucleases. Nature, v. 305, p. 448-451, 1983.
- LEE, S. T.; LIU, H. Y.; LEE, S. P.; TARN, C. Selection for arseniate resistance causes reversible changes in minicircle composition and kinetoplast organization in Leishmania mexicana. Molecular and Cellular Biology, v. 14, p. 587-596, 1994.
- LEE, S. T.; TARN, C. CHANG, K. P. Characterization of the switch of kinetoplast DNA minicircle dominances during development and reversion od druf resistance in Leishmania. Molecular and Biochemical Parasitology, v. 58, p. 187-204, 1993.
- LEE, S. T.; TARN, C.; WANG, C. Y. Characterization of sequence changes in kinetoplast DNA maxicircles of drug resistant Leishmania. Molecular and Biochemical Parasitology, v. 56, p. 196-208, 1992a.
- LEE, S. Y.; LEE, S. T.; CHANG, R. P. Transkinetoplastidy A novel phenome non involving bulk alteration of mitochondrion-kinetoplast DNA of trypanosomatid protozoa. The Journal of Protozoology, v. 39, p. 190-196, 1992b.
- LEEM, S. -H.; KOUPRINA, K.; GRIMWOOD, J.; KIM, J. -K; MULLOKANDOV, M.; YOON, Y.-H.; CHAE, J.-Y.; MORGAN, J.; LUCAS, S.; RICHARDSON, P.; DETTER, C.; GLAVINA, T.; RUBIN, E.; BARRETT, J. C.; LARIONOV, V. Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. Genome Research, v. 14, p. 239-246, 2004.
- LEVINSON, G.; GUTAMN, G. A. Slipped-strad mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution, v. 4, p. 203-221, 1987.
- LEWIN, B. Genes VIII. New Jersey: Prentice Hall, 2004. 1056 p.
- LINIAL, M.; SHLOMA, J. The Sequence-directed Bent Structure in kinetoplast DNA Is recognized by an enzyme from Crithidia fasciculate. The Journal of Biological Chemistry, v. 262, n. 31, p. 15194-5201, 1987.
- LOBACHEV, K. S.; SHOR, B. M.; TRAN, H. T.; TAYLOR, W.; KEEN, J. D.; RESNICK, M. R.; GORDENIN, D. A. Factor affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. Genetics, v. 148, p. 1507-1524, 1998.

- LUAN, D. D.; KORMAN, M. H.; JAKUBCZAK, J. L.; EICKBUSH, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell, v. 72, p. 595-605, 1993.
- LUI, J. -F. Cytogenetic study of pure breed and hybrid wild boar (Sus scrofa scrofa) in Southeast and South Brazil. Revista de Educação Continuada do CRMV-SP, v. 3, n. 1, p. 43-48, 2000.
- LUNDIN, L. -G. Gene duplications in early metazoan evolution. Cell and Developmental Biology, v. 10, p. 523-530, 1999.
- MACLEOD, D.; LOVELL-BADGE, R.; JONES, S.; JACKSON, I. A promoter trap in embrionic stem (ES) cells selects for integration of DNA into CPG islands. Nucleic Acids Research, v. 19, n. 1, p. 1-17, 1991.
- MALIK, H. S.; EICKBUSH; T. H. Retrotransposable elements R1 and R2 in the rDNA units of Drosophila mercatorum: abnormal abdomen revisited. Genetics, v. 151, p. 653-665, 1999.
- MARCADIER, J. L.; PEARSON, C. E. Fidelity of Primate Cell Repair of a Double-strand Break within a (CTG).(CAG) Tract. The Journal of Biological Chemistry, v. 278, n. 36, p. 33848–33856, 2003.
- MARRI, P. R.; HAO, W.; GOLDING, G. B. The role of laterally transferred genes in adaptive evolution. BioMed Central Evolutionary Biology, v. 7, (Suppl 1): S8, p. 1-14, 2007.
- MARTIN, S. L.; BUSHMAN, F. D. Nucleic acid chaperone activity of the ORF1 protein fromthemouse Line-1 retrotransposon. Molecular Cell Biology, v. 21, p. 467-475, 2001.
- MARTIN, S. L.; CRUCEANU, M.; BRANCIFORTE, D.; WAI-LUN, LI P.; KWOK, S. C.; HODGES, R. S.; WILLIAMS, M. C. Line-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. Journal of Molecular Biology, v. 348, p. 549-561, 2005.
- MATHIAS, S. L.; SCOTT, A. F.; KAZAZIAN, H. H. Jr.; BOEKE, J. D.; GABRIEL, A. Reverse transcriptase encoded by a human transposable element. Science, v. 254, p. 1808-1810, 1991.
- MCFARLANE, M.; WILSON, J. A model for the mechanism of precise integration of microinjected trasngene. Transgenic Research, v. 5, p. 171-177, 1996.
- MCGLYNN, P.; LLOYD, R. G. Recombinational repairand restart of damaged replication forks. Nature Reviews Molecular Cell Biology, v. 3, n. 11, p. 859-870, 2002.

MCMURRAY, C. T. Mechanisms of DNA expansion. Chromosoma, v. 104, p. 2-13, 1995.

MEIMA, R.; HAIJEMA B. J.; DIJKSTRA, H.; HAAN, G. -T.; VENEMA, G.; BRON, S. Role of enzymes of homologous recombination in illegitimate plasmid recombination in Bacillus subtilis. Journal Bacteriology, v. 179, n. 4, p. 1219-1229, 1997.

MELO, M. R.; POMMER, C. V.; KAVATI, R. Polinização artificial da atemoia com diversas fontes de polen comparada com a natural. Bragantia, v. 61, n. 3, p. 231-236, 2002.

MESERVY, J. M.; SARGENT, R. G.; IYER, R. R.; CHAN, F.; MCKENZIE, G. J.; WELLS, R. D.; WILSON, J. H. Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells. Molecular and Cellular Biology, v. 23, p. 3152-3162, 2003.

MIGHELL, A. J.; MARKHAM, A. F.; ROBONSIN, P. A. Alu sequences. FEBS Letters, v. 417, p. 1-5,1997.

MIKI, Y. Retrotransposal integration of mobile genetics in human disease. Journal of Human Genetecis, v. 43, p. 77-84, 1998.

MILOT, E.; BELMAAZA, A.; RASSART, E.; CHARTRAND, P. Association of a host DNA structure with retroviral integration site in chromosomal DNA. Virology, v. 201, p. 408-412, 1994.

MIRKIN, E. V.; MIRKIN, S. M. Mechanisms of transcription-replication collisions in bacteria Molecular Cell Biology, v. 25, p. 888-895, 2005.

MIYAGISHIMA, S. Y. Origin and evolution of the chloroplast division machinery. Journal of Plant Research, v. 118, n. 5, p. 295-306, 2005.

MODRICH, P.; LAHUE, R. Mismatch repair in replication fidelity, genetic recombination, and cancer. Annual Review of Biochemistry, v. 65, p.101-133, 1996.

MOMBAERTS, P. The human repertoire of odorant receptor genes and pseudogenes. Annual review of Genomics Human Genetics, v. 2, p. 493-510, 2001.

MORAN, J. V.; HOLMES, S. E.; NASS, T. P.; DEBERARDINIS, R. J.; BOEKE, J. D.; KAZAZIAN, H. H. JR. High frequency retrotransposition in cultured mammalian cells. Cell, v. 87, p. 917- 927, 1996.

MOREL, C.; CHIARI, E.; CAMARGO, E.; MATTEI, D.; ROMANHA, A.; SIMPSON, L. Strains and clones of Trypanosoma cruzi can be characterized by restriction endonuclease fingerprinting of kinetoplast DNA minicircles. Proceedings of the National Academy Science of the USA (PNAS), v. 77, p. 6810-6914, 1980.

MOYZIZ, R. K.; TORNEY, D. C.; MEYNE, J.; BUCKINGHAM, J. D.; WU, J. R.; BURKS, C.; SIROTKIN, K. M.; GOAD, W. B. The distribution of interspersed repetitive sequences in the human genome. Genomics, v. 4, p. 273-279, 1989.

NEI, M. Gene duplication and nucleotide substitution in evolution. Nature, v. 221, p. 40-42, 1969.

NEVES, C. S. V. J; YUHARA, E. N. Caracterização dos frutos de cultivares de atemoia produzidos no norte do Paraná. Semana de Ciências Agrárias de Londrina, v. 24, n. 2, p. 311-313, 2003.

NITZ, N.; GOMES, C.; ROSA, A. C.; SOUZA-AULT, M. R.; MORENO, F.; LAURIA-PIRES, L.; NASCIMENTO, R. J.; TEIXEIRA, A. R. Heritable integration of kDNA minicircle sequences from Trypanosoma cruzi into the avian genome: insights into human Chagas disease. Cell, v. 118, n. 2, p. 174-186, 2004.

O'NEILL, R. J. W.; O'NEILL, M. J. O.; GRAVES, J. A. M. Undermethylation associated with retroelement activation and chromosome remodelling in na interspecific mammalian hybrid. Nature, v. 393, p. 68-72, 1998.

OGIWARA, I.; MIYA, M.; OHSHIMA, K.; OKADA, N. V-Sines: a new superfamily of vertebrate Sines that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Research, v. 12, n. 2, p. 316- 324, 2002.

OHNO, S. Sex Chromosomes and Sex-Linked Genes. Berlin: Springer-Verlag, 1967. 192 p.

OHNO, S. Evolution by Gene Duplication. Berlin: Springer, 1970. 160 p.

OSTERTAG, E. M.; KAZAZIAN, H. H. JR. Lines in mind. Nature, v. 435, p. 16, 2005.

PATEL, N. H; PRINCE, V. E. Beyond the Hox complex. Genome Biology, v. 1, p. 1027.1-1027.4, 2000.

PATTHY, L. Modular assembly of genes and the evolution of new functions. Genetica, v. 1, n. 18, p. 217-231, 2003.

- PATTHY, L. Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. Cell, v. 41, p. 657-663, 1985.
- PAULL, T. T.; GELLERT, M. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes & Development, v. 13, p. 1276-1288, 1999.
- PEARSON, C. E.; EDAMURA, K. N.; CLEARY, J. D. Repeat instability: mechanisms of dynamic mutations. Nature Reviews Genetics, v. 6, n. 10, p. 729-42, 2005.
- PILLER, K. J.; DECKER, C. J.; RUSCHE, L. N.; HARRIS, E. M.; HAJDUK, S. L.; SOLLNER-WEBB B. Editing domains of Trypanosoma brucei mitochondrial RNA identified by secondary structure. Molecular and Cellular Biology, v. 15, p. 2916-2924, 1995a.
- PILLER, K. J.; DECKER, C. J.; RUSCHE, S. L.; SOLLNER-WEBB B. Trypanosoma brucei mitochondrial guide RNA - mRNA chimera-forming activity cofrationates with an editing-domain-specific endonucleaseand RNA ligase and is mimicked by heterologous nuclease and RNA ligase. Molecular and Cellular Biology, v. 15, p. 2925-2932, 1995b.
- PILLER, K. J.; RUSCHE, S. L.; SOLLNER-WEBB B. Trypanosoma brucei RNA editing. The Journal of Biological Chemistry, v. 27, p. 4613-4619, 1996.
- PITTMAN, D. L.; SCHIMENTI, C. Recombination in the mammalian germ line. Current Topics. in Developmental Biology, v. 37, p. 1-35, 1998.
- POPTSOVA, M. S.; GOGARTEN, J. P, The power of phylogenetic approaches to detect horizontally transferred genes. BioMed Central Evolutionary Biology, v. 7, n. 45, p. 1-17, 2007.
- RAVEL, J.; WELLINGTON, E. M. H.; HILL, R. T. Interspecific Transfer of Streptomyces Giant Linear Plasmids in Sterile Amended Soil Microcosms. Applied and Environmental Microbiology, v. 66, n. 2, p. 529-534, 2000.
- RAZIN, S. V.; IOUDINKOVA, E. S.; TRIFONOV. E.; SCHERRER, K. Nonclonability correlates with genome instability: A case of unique DNA region. Journal of Molecular Biology, v. 307, p. 481-486, 2001.
- RIJKERS, T.; PEETZ, A.; RUTHER, U. Insertional mutagenesis in transgenic mice. Transgenic research, v. 3, p. 203-15, 1994.
- RIOU, G. F.; YOT, P. Heterogeneity of kinetoplast DNA molecules of Trypanosoma cruzi. Biochemistry, v. 16, n. 11, p. 2390-2396, 1977.

ROYLE, N. J.; CLARKSON, R. E; WONG, Z.; JAFFREYS A. J. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics, v. 3, p. 352-360, 1988.

RUBIN, C. M.; VANDEVOORT, C. A.; TEPLITZ, R. L.; SCHMID, C. W. Alu repeat DNAs are differentially methylated in primate germ cells. . Nucleic Acids Research, v. 22, n. 23, p. 5121-5127, 1994.

RUJAN, T.; MARTIN, W. How many genes in Arabidopsis come from cyanobacteria? An estimate from 386 protein phylogenies. Trends in Genetics, v. 17, n. 3, p. 131-120, 2001.

SAMADASHWILY, G. M.; RACA, G.; MIRKIN, S. M. Tricucletide repeats affext DNA replication in vivo Nature. Genetics, v. 17, p. 298-304, 1997.

SANGER, F.; AIR, G. M.; BARRELL, B. G.; BROWN, N. L.; COULSON, A. R.; FIDDES, C. A.; HUTCHISON, C. A.; SLOCOMBE, P. M.; SMITH, M. Nucleotide sequence of bacteriophage phiX174 DNA. Nature, v. 265, p. 687-695, 1977.

SANTINI, S.; BERNARDI, G. Organization and base composition of tilapia Hox genes: implications for the evolution of Hox clusters in fish. Gene, v. 346, p. 51-61, 2005.

SARKAR, P. S.; CHANG, H. -C.; BOUDI, F. B.; REDDY, S. cgt Repeatas Shouw bimodal amplification in E. coli. Cell, v. 95, n. 13, p. 531-540, 1998.

SASSAMAN, D. M.; DOMBROSKI, B. S.; MORAN, J. V.; KIMBERLAND, M. L.; NASS, T. P.; DEBERARDINIS, R. J.; GABRIEL, A.; SWERGOLD, G. D.; KAZAZIAN, H. H. Jr. Many human L1 elements are capable of retrotransposition. Nature Genetics, v. 16, p. 37-43, 1997.

SCHROTH, G. P.; HO, P. S. Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Research, v. 23, p. 1977-983, 1995.

SCOTT, J. A place in the world for RNA editing, Cell, v. 81, n. 6, p. 833-836, 1995.

SEIWERT, S. D. RNA editing hints of remarkable diversity in gene expression pathways . Science, v. 274, p. 1636-1637, 1996.

SEIWERT, S. D.; HEIDMANN, S.; STUART, K. Direct visualization of uridylate deletion in vitro suggest a mechanism for kinetoplastid RNA editing. Cell, v. 84, p. 831-841, 1996.

SHAPIRO, T. A.; ENGLUND, P. The structure and replication of kinetoplast DNA. Annual Review of Microbiology, v. 49, p. 117-143, 1995.

SHIKANAI, T. RNA editing in plant organelles: machinery, physiological function and evolution. Cellular and molecular life sciences: CMLS, v. 63, n. 6, p. 698-708, 2006.

SHIROISHI, T.; KOIDE, T.; YOSHINO, M; SAGAI, T.; MORIWAKI, K. Hotspots of homologous recombination in mouse meiosis. Advances in biophysics, v. 31, p. 119-132, 1995.

SIDOW, A. Gen(om)e duplications in the evolution of early vertebrates. Current opinion in genetics & development, v. 6, p. 715-22,1996.

SIMÕES-BARBOSA, A.; BARROS, A. M.; NITZ, N., Argañaraz E.R, Teixeira A. R. L Integration of *Trypanosoma cruzi* kDNA Minicircle Sequence in the Host Genome May Be Associated with Autoimmune Serum Factors in Chagas Disease Patients. Memorial do Instituto Oswaldo Cruz, v. 94, Suppl. 1, p. 249-252, 1999.

SIMÕES-BARBOSA, A.; ARGAÑARAZ, E. R.; BARROS, A. M.; ROSA, A. C.; ALVES, N. P., LOUVANDINI, P.; SOUZA-AULT, M. R. D.; NITZ, N.; STURM, N. R.; NASCIMENTO, R. J.; TEIXEIRA, A. R. L. Hitchhiking Trypanosoma cruzi minicircle DNA affects gene expression in human host cells via Line-1 retrotransposon Memorial do Instituto Oswaldo Cruz, v. 101, n. 8, p. 833-843, 2006.

SIMPSON, L. The genomic organization of guide RNA in kinetoplastid protozoa: several conundrums and their solutions. Molecular and Biochemistry Parasitology, v. 86, p. 133-141, 1997.

SMITH, A. F. A. The origin of interspersed repeats in the human genome. Current opinion in genetics & development, v. 6, p. 743-748, 1996.

SONG, J.; DONG, F.; LILLY J. W.; STUPAR, R. M.; JIANG, J. Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA Sequences. Genome, v. 44, p. 463-469, 2001.

STEINMETZ, M.; STEPHAN, S.; LINDAHL, F. Gene organization and recombinational hotspots in the murine major histocompatibility complex. Cell, v. 44, p. 895-904, 1986.

STEINMETZ, M.; WINOTO, A.; MINARD, K.; HOOD, L. Cluster of genes enconding mouse transplantation antigens. Cell, v. 28, p. 489-498, 1982.

STEPHENS, S. G. Possible significance of duplication in evolution. Advances in Genetics, v. 4, p. 247-265, 1951.

- STIBITZ, T. B.; KEELING, P. J.; BHATTACHARYA, D. Symbiotic origin of a novel actin gene in the cryptophyte Pyrenomonas helgolandii. Molecular Biolology Evololution, v. 17, p. 1731-1738, 2000.
- STRADER, E.; HOWELL. Stable maintenance of tandem array of four R67 dihydrofolate reductase genes. Focus, v. 19, n. 2, p. 24-25, 1997.
- STRAND, M.; EARLY, M. C., EARLEY, M. C.; CROUSE, G. F.; PETES, T. D. Mutations in the MSH3 gene preferentially leads to deletion within tracts of simple repetitive DNA in Saccharomyces cerevisiae. Proceedings of the National Academy Science of the USA (PNAS), v. 92, p. 10418-10421, 1995.
- STRAND, M.; PROLLA, T. A.; LISKAY, R.M.; PETES, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutation affecting DNA mismatch repair. Nature, v. 365, p. 274-276, 1993.
- STRAUSS, B. S.; SAGHER, D.; ACHARYA, S. Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA. Nucleic Acids Research, v. 25, p. 806-813, 1997.
- STRAUSS, J. H.; STRAUS, E. G. Viruses and human disease. San Diego: Academic Press, 2008. 468 p.
- STURM, N. R.; SIMPSON, L. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome c oxidase subunit III mRNA. Cell, v. 61, p. 879-884, 1990.
- STURM, N.; DEGRAVE, W., MOREL, C.; SIMPSON, L. Sensitive detection and schizodeme classification of Trypanosoma cruzi cells by amplification of kinetoplast minicircle DNA sequences: use in diagnosis of Chagas' disease. Mol. Molecular and biochemical parasitology, v. 33, p. 205-214, 1989.
- SYMER, D. E.; CONNELLY, C.; SZAK, S. T.; CAPUTO, E. M.; COST, G. J.; PARMIGIANI, G.; BOEKE, J. D. Human 11 retrotransposition is associated with genetic instability in vivo. Cell, v. 110, p. 327-338, 2002.
- TAE, -M. K.; SEUNG, -J. H.; MUN, -G. R. Periodic Explosive Expansion of Human Retroelements Associated with the Evolution of the Hominoid Primate. Journal Korean Med Science, v. 19, p. 177-85, 2004.
- TAKAHARA, T.; OHSUMI, T.; KUROMITSU, J.; SHIBATA, K.; SASAKI, N.; OKASAKI, Y.; SHIBATA, H.; SATO, S.; YOSHIKI, A.; KUSAKABE, M.; MURAMATSU, M.; UEKI, M.; OKUDA, K.; HAYASHIZAKI, Y. Dysfunction of the *Orleans* reeler gene arising from exon skipping due to transposition of a fulllength copy of na active L1 sequence into the skipped exon. Human molecular genetics, v. 5, n. 7, p. 989-993, 1996.

TAUTZ, D.; SCHLOTTERER, C. Simples sequences. Current opinion in genetics & development, v. 4, p. 832-837, 1994.

TEIXEIRA, A. R. L.; ARGAÑARAZ, E. R.; FREITAS JR, L. H.; LACAVA, Z. G. M.; SANTANA, J. M.; LUNA, H. Possible Integration of Trypanosoma cruzi kDNA minicircles into the host cell genome by infection. Mutation Research, v. 305, p. 197-209, 1994.

TEIXEIRA, A. R. L.; NASCIMENTO, R. J; STURM, N. R. Evolution and pathology in Chagas Disease - a review. Memórias do Instituto Oswaldo Cruz, v. 101, n. 5, p. 463-491, 2006.

THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the fowering plant Arabidopsis thaliana. Nature, v. 408, p. 796-815, 2000.

THE C. ELEGANS SEQUENCING CONSORTIUM. Sequence and analysis of the genome of C. elegans. Science, v. 282, p. 2012-2018, 1998.

THERTULIEN, R.; AHIDARIS, C. G. Evidence that the entire length of a kinetoplast DNA minicircle is transcribed in Trypanosoma cruzi. Molecular microbiology, v. 5, p. 207 - 215, 1991.

THERTULIEN, R.; SIMPSON-HAIDARIS, P. J.; HAIDARIS, C. G. Intracellular localization of a Trypanozoma cruzi kDNA minicircle transcript using RNA:RNA in situ hybridization. J. Euk. v. 41, n. 4, p. 402-407, 1994.

THOMPSON, J. R.; PACOCHA, S.; PHARINO, C.; KLEPAC-CERAJ, V.; HUNT, D. E.; BENOIT, J.; SARMA-RUPAVTARM, R.; DISTEL, D. L.; POLZ, M. F. Genotypic diversity within a natural coastal bacterioplankton population. Science, v. 307, n. 5713, p. 1311-1313, 2005.

TREVORS, J. T. Genome size in bacteria. Antonie Leeuwenhoek, v. 69, p. 293-303, 1996.

TRINH, T. Q.; SINDEN, R. R. The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in Escherichia coli. Genetics, v. 134, p. 409-422, 1993.

VARMUS, M. Retroviruses. Science, v. 240, p. 1427-1435, 1988.

VENTER, J. C.; ADAMS, M. D.; MYERS, E. W. The sequence of the human genome. Science, v. 291, p. 1304-1351, 2001.

VICKERMAN, D.; PRESTON, T. M. Comparative cell biology of the kinetoplastid flagellates. In: LUMSDEM, W. H. R.; EVANS, D. A. Biology of the kinetoplastidae. New York, 1976.

VLAHOVICEK, K.; KAJAN, L.; PONGOR, S. DNA analysis servers: plot.it, bend.it, model.it and IS. Nucleic Acids Research, v. 31, n. 13, p. 3686-7, 2003.

VLAHOVICEK, K.; PONGOR, S. Model.it: building three dimensional DNA models from sequence data. Bioinformatics Applications Note, v. 16, n. 11, p. 1044-1045, 2000.

VOLFF, J. -N.; BOUNEAU, L.; OZOUF-COSTAZC, C.; FISCHER, C. Diversity of retrotransposable elements in compact pufferfish genomes. Trends in Genetics, v. 19, p. 674-678, 2003.

WAHLS, W. P.; WALLACE, L. J.; MOORE, P. D. Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. Cell. v. 60, n. 1, p. 95-103, 1990.

WALLACE, M. R.; ANDERSEN, L. B.; SAULINO, A. M.; GREGORY, P. E.; GLOVER, T. W; COLLINS, F. S. A de novo Alu insertion results in neurofibromatosis type 1. Nature, v. 353, p. 864-866, 1991.

WANG, G.; CHRISTENSEN, L. A.; VASQUEZ, K. M. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. Proceedings of the National Academy Science of the USA (PNAS), v. 103, p. 2677-2682, 2006.

WANG, G.; VASQUEZ, K. M. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proceedings of the National Academy Science of the USA (PNAS), v. 101, p. 13448-13453, 2004.

WATSON, J. D.; CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, v. 171, n. 4356, p. 737-8, 1953.

WEBER, J. L.; WONG, C. Mutation of human short tandem repeats. Human molecular genetics, v. 2, p. 1123-1128, 1993.

WEISS, R. A.; TEICH, N.; VARMUS, H.; COFFIN, J. (Ed.). Molecular biology of tumor viruses: RNA tumor virus. Cold Spring Harbor: Cold Spring Harbor Laboratory, 1985.

WELLS, R. D. Non-B DNA conformations, mutagenesis and disease. TRENDS in Biochemical Sciences, v. 32, n. 6, p. 271-278, 2007.

WELLS, R. D.; ASHIZAWA, T. Genetic Instabilities and Neurological Diseases, 2nd San Diego: Elsevier-Academic Press, 2006.

WILKIE, T. J.; PALMITER, R. D. Análysis of the integrant in MyK-1103 transgenic mice in with males fail to transmit the integrant. Molecularand cellular biology, v. 7, p. 1946-1655, 1987.

WINKLER, H. Verbreitung Und Ursache Der Parthenogenesis Im Pflanzen-Und Tierreiche. JENA: VERLAG FISCHER, 1920.

WOODCOCK, D. M.; WILLIAMSON, M. R.; DOHERTY, J. P. A sensitive RNAse protection assay to detect transcripts from potentially functional human endogenous L1 retrotransposons. Biochemical Biophysical Research Communications, v. 222, p. 460-465, 1996.

WOJCIECHOWSKA, M.; BACOLLA, A.; LARSON, J. E.; WELLS, R. D. The Myotonic Dystrophy Type 1 Triplet Repeat Sequence Induces Gross Deletions and Inversions. The Journal of Biological Chemistry, v. 280, p. 941-952, 2005.

WOJCIECHOWSKA, M.; NAPIERALA, M.; LARSON, J. E.; WELLS, R. D. Non-B DNA Conformations Formed by Long Repeating Tracts of Myotonic Dystrophy Type 1, Myotonic Dystrophy Type 2, and Friedreich's Ataxia Genes, Not the Sequences per se, Promote Mutagenesis in Flanking Regions The Journal of Biological Chemistry, v. 281, n. 34, p. 24531-24543, 2006.

WYMAN, A. R.; WERTMAN, K. F.; BARKER, D.; HELMS, C.; PETRI, W. H. Factors wich equalize the representation of genome segments in recombinant libraries. Gene, v. 49, p. 263-271, 1986.

WYMAN, A. R.; WOLFE, L. B.; BOTSTEIN, D. Propagation of some human DNA sequences in vector requires mutant Escherichia coli host. Proceedings of the National Academy Science of the USA (PNAS), v. 82, p. 2880-2884, 1985.

XING, J.; WANG, H.; BELANCIO, V. P.; CORDAUX, R.; DEININGER, P. L.; BATZER, M. A. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proceedings of the National Academy of Science USA, v. 103, p. 17608-17613, 2006.

YANT, S. R.; KAY, M. A. Nonhomologous-end-joining factors regulate DNA repair fidelity during sleeping beauty element transposition in mammalian cells. Molecular and Cellular Biology, v. 23, p. 8505-8518, 2003.

YASUI, Y.; NASUDA, S.; MATSUOKA, Y.; KAWAHARA, T. The Au family, a novel short interspersed element (Sine) from Aegilops umbellulata. Theoretical and Applied Genetics (TAG), v. 102, p. 463-470, 2001.

YOLDER, J. A.; WALSH, C. P.; BESTOR, T. H. Cytosine methyklation and the ecology of intragenomic parasites. Trends in Genetics, v. 13, p. 335-340, 1997.

ZHANG, J. Evolution by gene duplication: an update. Trends in Ecology and Evolution v. 18, p. 292-298, 2003.

ZHANG, X.; FIRESTEIN, S. The olfactory receptor gene superfamily of the mouse. Nature Neuroscience, v. 5, p. 124-133, 2002.

ZHANGA, C. -T.; REN ZHANG, R. An isochore map of the human genome based on the Z curve method. Gene, v. 317, p. 127-135, 2003.

ZINGLER, N.; WILLHOEFT, U.; BROSE, H. P.; SCHODER, V.; JAHNS, T.; HANSCHMANN, K. M.; MORRISH, T. A.; LOWER, J.; SCHUMANN, G. G. Analysis of 50 junctions of human Line-1 and Alu retrotransposons suggests an alternative model for 50-end attachment requiring microhomology-mediated end-joining. Genome Research, v. 15, p. 780-789, 2005.

ZUCMAN-ROSSI, J.; BATZER, M. A.; STONEKING, M.; DELATTRE, O.; THOMAS, G. Interethinic polimorphism of EWS intron 6: genome plasticity mediated by Alu retroposition and recombination. Human Genetis, v. 99, p. 357-363, 1997.

Impressão e acabamento Embrapa Informação Tecnológica

O papel utilizado nesta publicação foi produzido conforme a certificação do Bureau Veritas Quality International (BVQI) de Manejo Florestal. Print and finishing
Embrapa Technological Information

The paper used in this publication was produced according to the Bureau Veritas Quality International's (BVQI) Forest Management Certification.



Sugar Phosphate Backbone

Base pai

Nitrogeous base

> Ministério da Agricultura, Pecuária e Abastecimento



