



# Biologia molecular e evolução



# Biologia molecular e evolução

Larissa Barbosa de Paula

© 2017 por Editora e Distribuidora Educacional S.A.  
Todos os direitos reservados. Nenhuma parte desta publicação poderá ser reproduzida ou transmitida de qualquer modo ou por qualquer outro meio, eletrônico ou mecânico, incluindo fotocópia, gravação ou qualquer outro tipo de sistema de armazenamento e transmissão de informação, sem prévia autorização, por escrito, da Editora e Distribuidora Educacional S.A.

**Presidente**

Rodrigo Galindo

**Vice-Presidente Acadêmico de Graduação**

Mário Ghio Júnior

**Conselho Acadêmico**

Alberto S. Santana

Ana Lucia Jankovic Barduchi

Camila Cardoso Rotella

Cristiane Lisandra Danna

Danielly Nunes Andrade Noé

Emanuel Santana

Grasiele Aparecida Lourenço

Lidiane Cristina Vivaldini Olo

Paulo Heraldo Costa do Valle

Thatiane Cristina dos Santos de Carvalho Ribeiro

**Revisão Técnica**

Priscila Perez Domingos

**Editorial**

Adilson Braga Fontes

André Augusto de Andrade Ramos

Cristiane Lisandra Danna

Diogo Ribeiro Garcia

Emanuel Santana

Erick Silva Griep

Lidiane Cristina Vivaldini Olo

---

**Dados Internacionais de Catalogação na Publicação (CIP)**

Paula, Larissa Barbosa de  
P324b Biologia molecular e evolução / Larissa Barbosa de  
Paula. – Londrina : Editora e Distribuidora Educacional S.A.,  
2017.  
176 p.

ISBN 978-85-522-0248-6

1. Biologia molecular. 2. Evolução molecular. I. Título.

CDD 572.8

---

2017

Editora e Distribuidora Educacional S.A.  
Avenida Paris, 675 – Parque Residencial João Piza  
CEP: 86041-100 – Londrina – PR  
e-mail: editora.educacional@kroton.com.br  
Homepage: <http://www.kroton.com.br/>

# Sumário

<b>Unidade 1   Biologia evolutiva</b>	<b>7</b>
Seção 1.1 - Origem e evolução	9
Seção 1.2 - Ciclo celular e diferenciação	21
Seção 1.3 - Genômica comparativa	37
<b>Unidade 2   Material genético e evolução</b>	<b>53</b>
Seção 2.1 - Polimorfismos genéticos	55
Seção 2.2 - Reconstrução filogenética I	65
Seção 2.3 - Reconstrução filogenética II	84
<b>Unidade 3   Mutação, reparo e regulação</b>	<b>99</b>
Seção 3.1 - Erros de replicação	101
Seção 3.2 - Mecanismos de reparo do DNA	112
Seção 3.3 - Regulação gênica	124
<b>Unidade 4   Técnicas de biologia molecular</b>	<b>135</b>
Seção 4.1 - Clonagem e eletroforese	137
Seção 4.2 - Polimerização e sequenciamento	146
Seção 4.3 - Hibridização	157



# Palavras do autor

Prezado aluno,

A partir de agora você dará início ao estudo da Biologia Evolutiva e Molecular, que lhe fornecerá os conhecimentos necessários para compreender a história da evolução, suas causas e seus mecanismos, bem como as técnicas utilizadas para a análise e o entendimento desse processo. Você já reparou que cada espécie apresenta características próprias e parou para pensar na origem e no por quê dessas diferenças? O entendimento dos processos evolutivos e das técnicas de Biologia Molecular permitirá que você responda a questões sobre as hipóteses que podem explicar determinada característica dos organismos e os métodos pelos quais pesquisadores as testaram, aplicando os conhecimentos adquiridos em vivências relacionadas à sua formação.

Para que você assimile todo o conteúdo que será apresentado e adquira a competência de compreender a evolução do genoma, bem como os mecanismos de reparo do DNA e as técnicas de Biologia Molecular envolvidas na investigação de alterações genéticas e moleculares, é necessário que você se dedique ao máximo, por meio da leitura do livro didático, do acesso aos *links* de pesquisa sugeridos e da busca por mais aprendizado em outras fontes de conhecimento.

No decorrer da primeira unidade, conheceremos a origem da vida e as teorias da evolução, bem como o ciclo celular e os processos de diferenciação das células, além de fazer uma comparação entre os diferentes genomas. Na segunda unidade, estudaremos os polimorfismos genéticos e os métodos empregados na reconstrução filogenética. Já na terceira unidade, abordaremos os erros na replicação do DNA e os mecanismos de reparo, bem como os processos de regulação gênica. Por fim, na quarta unidade, entraremos no mundo das técnicas de Biologia Molecular, iniciando com clonagem e eletroforese, passando por polimerização e sequenciamento, e terminando com hibridização. Ao final deste livro didático, você será capaz de melhor entender a evolução molecular da vida e seus complexos processos biológicos. Vamos lá?



# Biologia evolutiva

## Convite ao estudo

Prezado aluno,

A partir de agora, iniciaremos o estudo da Biologia Molecular e da evolução para melhor entender como os organismos foram se modificando, genética e molecularmente, ao longo dos anos, para se adaptarem ao ambiente em que vivem. Ao final desta disciplina, você deverá ter compreendido a evolução do genoma, bem como os mecanismos de reparo do DNA e as técnicas de Biologia Molecular envolvidas na investigação de alterações genéticas e moleculares. Ao final desta unidade, são seus objetivos: conhecer a origem da vida na Terra e as diferenças entre as teorias evolutivas propostas por Lamarck e Darwin, compreender o ciclo celular e os processos de diferenciação das células e entender a importância da genômica comparativa na diferenciação das espécies.

Para compreender o assunto e atingir as competências e os objetivos da disciplina, analisaremos uma situação hipotética que se aproxima dos conteúdos teóricos que serão vistos por você nesta unidade: Letícia é uma garotinha de 10 anos que adora Ciências e é apaixonada por animais. Como presente de aniversário, ela ganhou de seu pai uma visita ao Zoológico de Itatiba, no interior de São Paulo; o maior zoológico particular do Brasil, contendo 500.000m<sup>2</sup> de área verde, dos quais grande parte está inserida em um fragmento de Mata Atlântica. O parque abriga cerca de 1.000 animais, inseridos em mais de 180 espécies; incluindo dezenas ameaçadas de extinção. Na noite anterior ao passeio, a garotinha mal conseguiu dormir, tamanha era a ansiedade de

poder alimentar e tocar alguns animais do parque, além de ver filhotes de várias espécies e acompanhar sua amamentação. Pela manhã, Letícia foi a primeira a acordar e logo pulou da cama para acordar seus pais. A família tomou um café da manhã reforçado, arrumou a mochila da garotinha com água, bolachas, protetor solar e máquina fotográfica, e partiu rumo à Itatiba.

No decorrer desta unidade de ensino, você compreenderá a origem da vida e as teorias evolutivas mais aceitas, os mecanismos envolvidos no ciclo celular e como são realizadas as comparações genômicas entre as diferentes espécies. Na Seção 1.1, abordaremos a origem da vida, as teorias propostas por Lamarck e Darwin para a evolução das espécies e os mecanismos de evolução e seleção natural. Já na Seção 1.2, compreenderemos o ciclo celular e os mecanismos de morte e diferenciação celular. Por fim, na Seção 1.3, trataremos dos métodos empregados na genômica comparativa para a diferenciação das espécies.

# Seção 1.1

## Origem e evolução

### Diálogo aberto

Para darmos continuidade ao seu aprendizado, retomaremos a visita de Leticia e seus pais ao Zoológico de Itatiba: pela manhã, Leticia foi a primeira a acordar e logo pulou da cama para despertar seus pais. A família tomou um café da manhã reforçado, arrumou a mochila da garotinha com água, bolachas, protetor solar e máquina fotográfica, e partiu rumo à Itatiba. Na chegada ao Zoológico, Leticia e seus pais foram recebidos com muita atenção pelos funcionários do local, que lhes passaram algumas instruções e, em seguida, os liberaram para a visita. Durante o passeio pelo ambiente de savana africana, Leticia ficou encantada com as listras das zebras e com o chifre dos rinocerontes, achou graça das pernas longas dos avestruzes e se divertiu com a esguichada de água da tromba dos elefantes. Mas o que mais intrigou a garotinha foi o tamanho do pescoço das girafas: “como pode ser tão comprido?”, indagou. O pai, então, contou à filha que, há milhões de anos, as girafas possuíam o pescoço curto que, com o passar dos anos, foi sendo esticado devido à necessidade de alcançar as folhas nas árvores mais altas. A explicação dada pelo pai de Leticia está correta? Essa é a única explicação existente? Para validar a explicação do pai de Leticia e discorrer sobre outras possíveis teorias, você deve ler o livro didático, dando ênfase ao conteúdo sobre origem da vida e as teorias de Lamarck e Darwin. Vamos lá?

### Diálogo aberto

O universo é formado por matéria viva (seres vivos) e matéria não viva (matéria bruta). A diferença entre ambas consiste em algumas propriedades que os seres vivos apresentam e que não são observadas na matéria bruta:

- a) Composição química mais complexa.
- b) Organização molecular mais estruturada.

c) Reação a estímulos ambientais.

d) Capacidade de absorver matéria e energia do ambiente, a fim de se desenvolver e manter suas funções vitais.

e) Capacidade de manter seu meio interno em condições adequadas de sobrevivência, independentemente das condições do meio externo.

f) Capacidade de crescimento e reprodução.

g) Capacidade de modificação ao longo do tempo (**evolução**), através de adaptações que garantam sua sobrevivência.

Com o avanço da Ciência, muito sobre a origem da vida na Terra foi esclarecido ao longo dos anos, mas muito ainda falta ser esclarecido. Dentre as diversas hipóteses levantadas, as mais consideradas consistem na **criação divina** (vida criada por divindade ou força superior), na **origem extraterrestre** (vida originada fora da Terra e trazida por meteoritos através do espaço) e na **origem por evolução química**, a mais aceita atualmente (a vida surgiu de forma espontânea no planeta, a partir da combinação dos átomos existentes no universo, que resultou em moléculas simples, as quais, por sua vez, se associaram formando moléculas complexas).

A teoria de Aleksandr Oparin vai ao encontro da hipótese de origem por evolução química, propondo uma atmosfera primitiva composta por amônia, metano, hidrogênio e vapor de água. O bioquímico afirmava que conforme as temperaturas do planeta foram diminuindo, o conseqüente vapor de água condensado caiu sobre a crosta, evaporando-se novamente e ocasionando um ciclo de chuvas bastante intenso, acompanhado por fortes descargas elétricas. Essas descargas, juntamente com os raios ultravioleta, modificaram as ligações químicas das moléculas da atmosfera, formando novas substâncias orgânicas, como aminoácidos.

A hipótese clássica e mais aceita sobre os primeiros organismos vivos é a heterotrófica: havia um mar repleto de coacervados (agregados de proteínas e outras moléculas orgânicas) que serviram de alimento a esses seres primitivos. Entretanto, a presença de organismos exclusivamente heterótrofos, em dado momento, esgotaria o estoque de matéria orgânica – o que fez que, antes que isso ocorresse, surgissem seres autótrofos fotossintetizantes por evolução.

Por outro lado, têm-se conhecimento das bactérias autótrofas extremamente simples, que não utilizam a luz solar para produção de matéria orgânica (utilizando a energia proveniente de reações ocorridas entre as substâncias minerais da crosta terrestre), contribuindo para uma hipótese autotrófica. Além disso, o oxigênio liberado através da fotossíntese passou a se acumular na atmosfera. Dessa forma, seres que utilizavam oxigênio para degradar o alimento, surgidos por evolução, passaram a ter maiores condições de sobrevivência.

Mudanças evolutivas quase sempre tornam as espécies mais aptas a sobreviver em seu ambiente. Três evidências comprovam a existência da evolução ao longo dos anos: a anatomia comparada, o estudo dos fósseis e a comparação entre as moléculas que compõem os organismos (bioquímica comparada). À medida que o estudo da anatomia animal foi progredindo, passou-se a perceber que animais muito diferentes quanto ao aspecto externo apresentavam internamente um mesmo plano básico. Comparações mais detalhadas entre os membros de diversos vertebrados mostraram ainda notável semelhança quanto à estrutura básica, evidenciando uma mesma origem, mas funções diferentes, adaptadas ao ambiente no qual vivem. Outra importante evidência evolutiva é a embriologia comparada: no início do desenvolvimento, as semelhanças são impressionantes, mas as diferenças se acentuam consideravelmente à medida que o desenvolvimento ocorre.

O estudo das diversas camadas de rochas que formam a crosta terrestre, das mais antigas às mais recentes, revelou fósseis bastante diferentes, apresentando um aumento gradual de complexidade e diversidade. Rochas muito antigas não apresentam fósseis de organismos atuais, propondo um surgimento mais tardio desses organismos. Do mesmo modo, fósseis muito antigos apresentam espécies que hoje não existem mais, evidenciando sua extinção.

Além da comparação do DNA, a semelhança entre espécies distintas pode ser afirmada através da comparação da composição de uma mesma proteína encontrada nas espécies em questão. A alfa-globina humana, por exemplo, difere em quatro aminoácidos da globina do macaco *Rhesus*. Essa diferença sobe para 17 aminoácidos em relação à globina do boi e para 71 aminoácidos

em relação à globina da carpa – refletindo as diferenças entre os genes que regulam a produção dessa proteína nas quatro espécies. Ainda, em uma proteína comum, ocorre, em média, a substituição de **um** aminoácido a cada **milhão** de anos, permitindo a avaliação, com precisão, da época em que novas espécies apareceram, a partir de um ancestral comum. Aplicando essa informação ao exemplo dado anteriormente, concluímos que a espécie humana divergiu da carpa há muito mais tempo quando comparado ao macaco *Rhesus*.

A adaptação dos seres vivos ao ambiente em que vivem é um fato, porém, a origem da adaptação sempre foi discutida. Na Antiguidade, era defendida a ideia de que as espécies eram fixas e imutáveis (**fixismo**), isto é, já existiam desde a origem do planeta, e a extinção de muitas delas aconteceu devido a eventos especiais, como catástrofes. Lentamente, essa ideia foi dando lugar à outra, baseada em uma substituição gradual de espécies por outras, através de adaptações a ambientes continuamente em processo de mudança (**transformismo**); à medida que o meio muda, muda-se a espécie. Essa ideia tornou-se a base para os pensamentos evolucionistas que surgiriam a partir de então.

Em 1809, opondo-se ao fixismo, Jean Baptiste Lamarck foi um dos primeiros pensadores a tentar explicar cientificamente o mecanismo pelo qual a evolução ocorre. Ele afirmava que as girafas, ao fazerem força para alcançar alimentos em árvores altas, esticaram seu pescoço e passaram essa característica às gerações seguintes, que nasceram gradativamente com pescoços mais compridos. Lamarck estava certo quando afirmou que os fatores ambientais podem modificar os indivíduos (como a exposição prolongada ao sol, que torna a pele mais morena, ou a utilização constante de um músculo, que o faz crescer). Mas enganou-se pensando que tais características adquiridas eram transmitidas à prole: do contrário, camundongos filhos de reprodutores que tiveram seus rabos cortados nasceriam sem rabo; ou filhos de judeus circuncidados não precisariam passar pelo ritual da circuncisão, o que não acontece.

Em 1844, o inglês Charles Darwin entrou em conflito com as ideias da época ao começar a escrever um longo ensaio sobre a seleção natural e sua relação com o surgimento de novas espécies.

De acordo com Darwin, os organismos vivos apresentavam grande capacidade de se reproduzirem, mas muitos morriam antes de chegarem à idade reprodutiva em virtude da limitada quantidade de alimento existente no ambiente em que viviam. Dessa forma, organismos com as mesmas necessidades alimentares competiam entre si, lutando constantemente pela sobrevivência.

Por outro lado, em um mesmo grupo, havia variações hereditárias que podiam ser transmitidas aos descendentes; algumas mais favoráveis do que outras em certo ambiente. Assim, organismos que apresentavam as variações mais favoráveis aquele ambiente apresentavam maiores probabilidades de reprodução e sobrevivência que os demais – transmitindo essas características aos seus descendentes e tornando as gerações futuras sucessivamente mais adaptadas às condições ambientais (**seleção natural**).



### Assimile

**Lamarckismo** → fatores ambientais podem modificar os organismos, e as características adquiridas são transmitidas às gerações futuras.

**Darwinismo** → organismos que apresentam variações mais favoráveis ao ambiente apresentam maiores probabilidades de reprodução e sobrevivência, transmitindo essas variações aos seus descendentes.

Classicamente, “espécies” são os tipos de organismos existentes (cachorro, gato, samambaia, etc.), caracterizadas e diferenciadas umas das outras por sua morfologia. Após a Teoria da Seleção Natural de Darwin, “espécie” passou a ser definida como uma população ou um grupo de populações cujos componentes são capazes (ou apresentam potencial) de se cruzar livremente na natureza e produzir descendentes férteis – não sendo, portanto, capazes de se cruzar com componentes de outra espécie.



### Exemplificando

O cruzamento entre um jumento (*Equus asinus*) e uma égua (*Equus caballus*) é possível. Entretanto, gera um animal (mula) incapaz de gerar descendentes.

O surgimento de novas espécies a partir de uma espécie ancestral é denominado **especiação**. Para que esse processo ocorra, é indispensável que haja um **isolamento geográfico**, isto é, que grupos da espécie original se separem e deixem de se cruzar – podendo ser via migração de grupos para locais diferentes e distantes, ou via aparecimento repentino de barreiras naturais intransponíveis (como rios, vales, montanhas, etc.) (Figura 1.1).

Sendo essa barreira intransponível para ambos os grupos, ficam os indivíduos isolados e impedidos de cruzar entre si. Durante muito tempo, submetidos a pressões de seleção natural diferentes (uma vez que vivem em ambientes diferentes), os genes selecionados em uma das populações não o serão na outra, e vice-versa. Com o passar dos anos, o genoma desses grupos torna-se cada vez mais distinto, e os indivíduos divergem, morfológicamente, cada vez mais. Em determinado nível de diferenciação, essas duas populações formam **raças** (subespécies). Se colocados novamente em contato, o cruzamento entre os indivíduos ainda é possível e gera descendentes férteis.

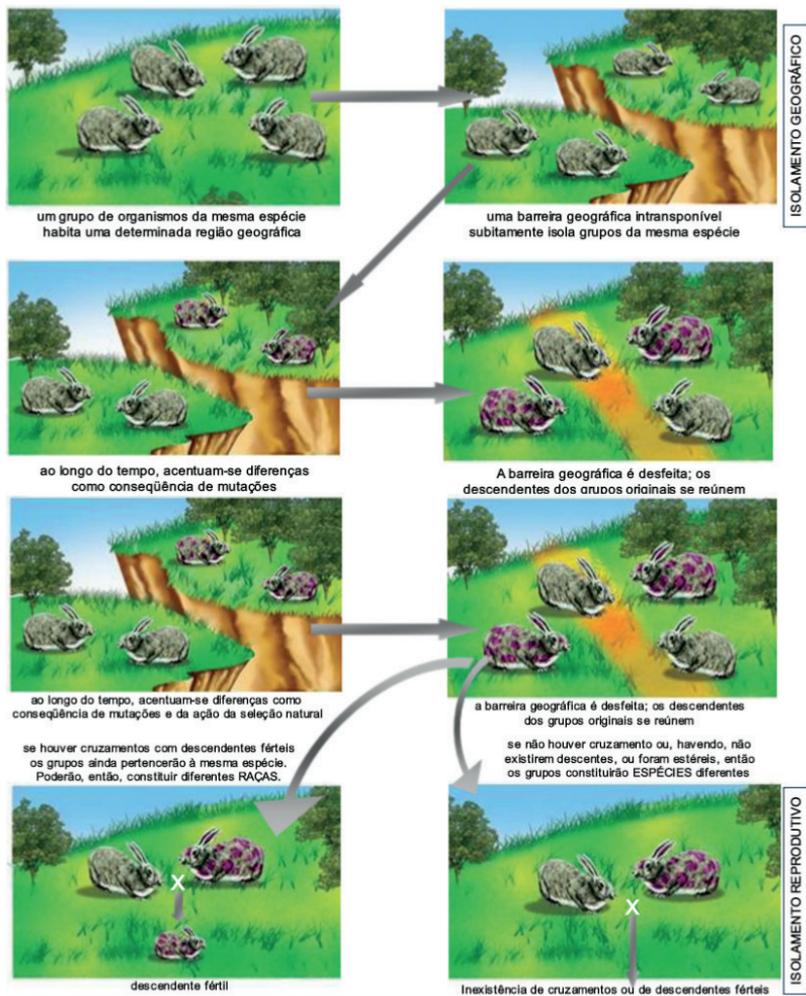


Refleta

O que aconteceria se as espécies ficassem isoladas geograficamente por um período tão longo, que a diferenciação entre elas se tornasse tão grande a ponto de impedir o cruzamento e/ou a geração de descendentes férteis?

Nesse caso, estabelece-se uma condição denominada **isolamento reprodutivo**, a qual gera duas espécies tão diferentes que, a partir de então, evoluirão separadamente (Figura 1.1). Cada uma das recém-formadas pode, por sua vez, sofrer um ciclo semelhante, fragmentando-se em raças (subespécies) e, finalmente, originando novas espécies.

Figura 1.1 | Especiação do tipo alopátrica



Fonte: <<http://slideplayer.com.br/slide/3006288/>>. Acesso em: 22 jun. 2017.

Evidências sugerem que a evolução dos grandes grupos de seres vivos foi possível a partir de um grupo ancestral cujos indivíduos, através do mecanismo de especiação, possibilitaram o surgimento de espécies relacionadas, ou seja, a partir de uma espécie inicial, pequenos grupos conquistaram novos ambientes, sofrendo adaptações que lhes permitiram sobreviver nesses meios.

Desse modo, surgiram novas espécies com muitas características semelhantes à ancestral – fenômeno conhecido como **irradiação adaptativa**. Nessas condições, sempre há um parentesco próximo entre as espécies consideradas.

Outro fenômeno evolutivo é o de **convergência adaptativa**. Um tubarão e um golfinho, morfologicamente, são muito semelhantes. No entanto, pertencem a grupos diferentes (peixe cartilaginoso e mamífero, respectivamente); portanto, essa semelhança morfológica não revela parentesco evolutivo próximo. De que maneira esses animais adquiriram, então, essa semelhança externa? Estando em um mesmo meio (aquático), a atuação desse meio selecionou, em ambas as espécies, uma forma corporal idealmente ajustada à água.

A irradiação adaptativa dos mamíferos originou várias ordens; entre elas, a dos primatas. Um fóssil descoberto nos Estados Unidos com aproximadamente 65 milhões de anos, denominado *Purgatorius unio*, é tido como o originador dos primatas. Do grupo primata ancestral, originaram-se outros dois: prossímios e antropóides. Os iniciantes da linhagem antropóide pertencem ao gênero *Aegyptopithecus*. Drásticas alterações climáticas, ocorridas por volta de 20 milhões de anos atrás, provocaram nesses animais a busca por alimentos nas savanas existentes. Um desses grupos primitivos que habitavam a África possivelmente é o originador da linhagem da qual surgiram os chimpanzés e os homens. Através da genômica comparativa, sugere-se que essas duas espécies devam ter divergido há cerca de 5 milhões de anos, a partir de seu ancestral.

A descoberta de diversos fósseis, a partir de 1924, serviu de base para a definição dos australopitecos, hominídeos de andar ereto e de mãos e dedos semelhantes aos dos homens; porém volume cerebral com cerca de 1/3 daquele do homem moderno. O aumento do volume craniano começou a ser detectado em fósseis de aproximadamente 2 milhões de anos, juntamente com a descoberta de ferramentas simples de pedra. Esse fato foi o pontapé para o enquadramento desta espécie, denominada *Homo habilis*, no gênero *Homo*. Os australopitecos acabaram por desaparecer, enquanto o *Homo habilis* possivelmente foi o precursor do *Homo erectus*, a partir do qual surgiu o *Homo sapiens*. Acredita-se que

o *Homo erectus* tenha se originado na África e se irradiado para a Ásia e a Europa.

Os seres humanos, como são hoje, possivelmente evoluíram paralelamente em várias partes do planeta; com a dispersão dos *Homo erectus* da África para os demais continentes, cerca de 1 milhão de anos atrás. A partir de então, teria ocorrido a diversificação que originou o *Homo sapiens* com suas várias raças. Outra hipótese supõe que os *Homo sapiens* surgiram apenas na África, a partir do *Homo erectus*, deslocando-se em grupos para outros lugares da Terra e originando as diversas etnias atualmente conhecidas.



### Pesquise mais

Ficou curioso(a) sobre outras possíveis teorias da origem da vida na Terra? Leia a Seção 1.2, denominada "Como está relacionada toda a vida na Terra?", e acabe com sua curiosidade!

SADAVA, D. et al. **Vida**: a ciência da biologia. 8. ed. Porto Alegre: Artmed, 2011. v.1.

## Sem medo de errar

A explicação dada pelo pai de Letícia está de acordo com a teoria evolucionista de Lamarck, que afirmava que os fatores ambientais podem modificar os indivíduos e que, no caso das girafas, o pescoço foi se esticando ao longo do tempo porque elas faziam força para alcançar alimentos em árvores altas. Como essa característica era transmitida aos descendentes, as gerações seguintes passaram a nascer gradativamente com pescoços maiores. Entretanto, essa não é a única explicação existente para o longo pescoço desses animais.

Anos mais tarde, outro evolucionista rebateu as ideias de Lamarck ao afirmar que, em um passado muito distante, existiam diferenças entre as girafas: algumas tinham um pescoço mais curto, e outras, um pescoço mais longo. Nos períodos de privação de alimentos, os indivíduos com pescoço mais longo e, portanto,

capazes de pastar a uma ou duas polegadas acima dos demais, foram preservados (**selecionados** pelo ambiente). Seu cruzamento gerou maior número de descendentes, que herdaram as mesmas particularidades corporais, enquanto que os indivíduos com pescoço mais curto e, portanto, menos favorecidos, foram fadados à extinção.

## Avançando na prática

### As mariposas de Manchester

#### Descrição da situação-problema

Por volta de 1850, em Manchester, na Inglaterra, predominava na cidade uma população de mariposas (*Biston betularia*) brancas, com algumas manchas negras. Após a Revolução Industrial, com o aumento da industrialização e consequente aumento da emissão de poluentes na atmosfera, uma população de mariposas escuras passou a ser encontrada em número cada vez maior, tornando-se a forma dominante. Como seria uma explicação plausível para essa alteração de dominância observada após o desenvolvimento industrial?

#### Resolução da situação-problema

Antes da Revolução Industrial e, portanto, na ausência de poluição, os troncos das árvores eram cobertos por líquens brancos. Uma vez que as mariposas brancas conseguiam se camuflar nesses líquens, essa população era menos visível e, conseqüentemente, menos atacada por pássaros predadores, que acabavam atacando principalmente as mariposas escuras, mais expostas. Com a industrialização, a impregnação da fuligem expelida pelas chaminés ocasionou a morte dos líquens, deixando os troncos expostos e mais escuros. Dessa forma, a situação se inverteu: as mariposas escuras passaram a ficar mais camufladas e protegidas dos predadores que as brancas, obtendo maior probabilidade de sobrevivência e reprodução, gerando descendentes escuras. As mariposas brancas, por outro lado, tornaram-se mais visíveis, passando a ser caçadas com maior frequência e, conseqüentemente, vivendo menos, se reproduzindo pouco e gerando menor número de descendentes.

Sendo a cor das mariposas um fator hereditário dependente de um par de genes codificador de dois fenótipos (claro e escuro), o ambiente poluído selecionou naturalmente a população de mariposas mais bem adaptadas a ele. Nas últimas décadas, com a instalação de filtros nas indústrias, promovida por legislações mais severas, os níveis de poluição gasosa foram reduzidos, favorecendo novamente o crescimento de líquens nos troncos das árvores. Como era de se esperar, o retorno da coloração clara dos troncos modificou os critérios de seleção, favorecendo, agora, a população de mariposas brancas em detrimento das escuras.

## Faça valer a pena

**1.** Uma das teorias sobre a origem da vida em nosso planeta apoia a ideia de que os seres vivos surgiram a partir da combinação dos átomos existentes no universo, que resultou em moléculas simples (água, metano, dióxido de carbono, amônia, nitrogênio e hidrogênio), as quais, por sua vez, se associaram formando moléculas complexas.

A respeito da origem da vida na Terra, assinale a alternativa correta.

- a) A formação de coacervados no mar primitivo foi resultado do aparecimento dos primeiros seres heterótrofos.
- b) O surgimento de moléculas complexas, como as proteínas, ocorreu anteriormente ao aparecimento dos aminoácidos.
- c) Uma vez que a escassez de alimentos era grande, os primeiros seres que surgiram eram autótrofos (hipótese autotrófica).
- d) Descargas elétricas e raios ultravioleta tinham papel fundamental na formação das moléculas orgânicas simples.
- e) Devido à grande quantidade de oxigênio nas moléculas de água, a forma primitiva de obtenção de energia era aeróbica.

**2.** A origem da vida é explicada por diversas teorias, as quais podem ser embasadas na criação divina, na origem extraterrestre ou na origem por evolução química. Considere os eventos relativos a esse fenômeno, descritos a seguir.

- I. Obtenção de energia através do processo de fermentação.
- II. Formação de coacervados no mar primitivo.
- III. Obtenção de energia através dos processos de fotossíntese e respiração.
- IV. Estabelecimento do equilíbrio entre seres autótrofos e heterótrofos.

Assinale a alternativa que apresenta a ordem cronológica correta desses eventos.

- a) I, II, III, IV.
- b) I, II, IV, III.
- c) II, I, III, IV.
- d) II, III, IV, I.
- e) IV, III, II, I.

**3.** Considere as teorias descritas a seguir.

I. Cada espécie é composta por um grupo de organismos semelhantes a um determinado tipo ideal.

II. O desaparecimento da visão de peixes habitantes de cavernas escuras é determinado pela ausência de luz.

III. Variações favoráveis permitem a sobrevivência e a reprodução das espécies no ambiente em que vivem, sendo essas variações transmitidas aos seus descendentes.

Assinale a alternativa que apresenta corretamente os idealizadores dessas teorias.

- a) Lineu, Lamarck e Darwin.
- b) Lineu, Darwin e Lamarck.
- c) Darwin, Lamarck e Lineu.
- d) Darwin, Mendel e Lamarck.
- e) Darwin, Lineu e Lamarck.

# Seção 1.2

## Ciclo celular e diferenciação

### Diálogo aberto

Para darmos continuidade ao seu aprendizado, retomaremos a visita de Letícia e seus pais ao Zoológico de Itatiba. Continuando o passeio, a família chegou ao ambiente de pantanal, no qual pode observar antas, tamanduás, veados, cobras e jacarés. Letícia pode segurar um filhote de cobra nas mãos e ficou encantada com os desenhos na pele do animal, a qual achou engraçada por ser escorregadia e gelada. O guia do parque explicou que os répteis não conseguem manter a temperatura do corpo estável, como os humanos, variando, dessa forma, de acordo com o ambiente em que se encontram. A curiosidade da garotinha não parou por aí: ela perguntou por que os dedos dos jacarés eram “colados” uns nos outros. Qual não foi a surpresa de Letícia ao ouvir do funcionário que, durante certo período do desenvolvimento embrionário, os humanos também possuem os dedos “colados”! Qual é o mecanismo celular envolvido no “descolamento” dos dedos humanos? O que poderia ocasionar a permanência dessa condição? Para responder a esses questionamentos, leia o livro didático, especialmente o conteúdo sobre morte celular. Vamos lá?

### Não pode faltar

O corpo humano é composto por cerca de um trilhão de células, todas descendentes de uma única (zigoto), as quais, diariamente, se dividem e substituem suas antecessoras danificadas ou envelhecidas. Desde o século XIX, pesquisadores tentam tornar as células humanas imortais, uma vez que muitas doenças humanas se desenvolvem apenas em células humanas. Em 1951, dois pesquisadores finalmente conseguiram manter em laboratório uma linhagem celular viva, as células HeLa (em homenagem à paciente da qual foram retiradas, Henrietta Lacks) se dividiam várias e várias vezes e passaram então a ser utilizadas para investigação

de câncer, crescimento viral, síntese proteica, efeitos da radiação sobre as células, dentre diversas outras pesquisas.

O ciclo de vida de uma célula (**ciclo celular**) compreende uma série de eventos macromoleculares que leva à divisão da célula e formação de duas novas, constituídas por cromossomos idênticos aos da célula que se dividiu. O ciclo é dividido em três estágios: **interfase** (subdividido nas fases G1, S e G2), **divisão celular** ou **mitose** (subdividido em prófase, metáfase, anáfase e telófase) e **divisão citoplasmática** ou **citocinese**. Em todas as células de um mesmo tipo celular, a duração desse ciclo é aproximadamente a mesma, diferindo de um tipo celular para outro. O crescimento do citoplasma acima de um determinado volume desencadeia a divisão celular, provavelmente devido à incapacidade do núcleo de controlar e/ou coordenar um volume citoplasmático acima do limiar de sua capacidade.

Antes de se dividir, uma célula copia todo seu DNA (replicação) durante a interfase, o intervalo mais longo do ciclo celular e no qual a célula aumenta de massa, praticamente duplica o número de organelas citoplasmáticas e replica seu DNA. Na fase G1, as moléculas de DNA originam moléculas de RNA, as quais passam para o citoplasma e promovem a síntese proteica. Nessa fase, a célula trabalha intensamente para produzir mais substâncias e aumentar seu volume.



### Assimile

G → do inglês *gap*, "intervalo".

**Dica!** Memorize o "G" dessa fase não apenas como *gap*, mas também como *growth* (do inglês, "crescimento").

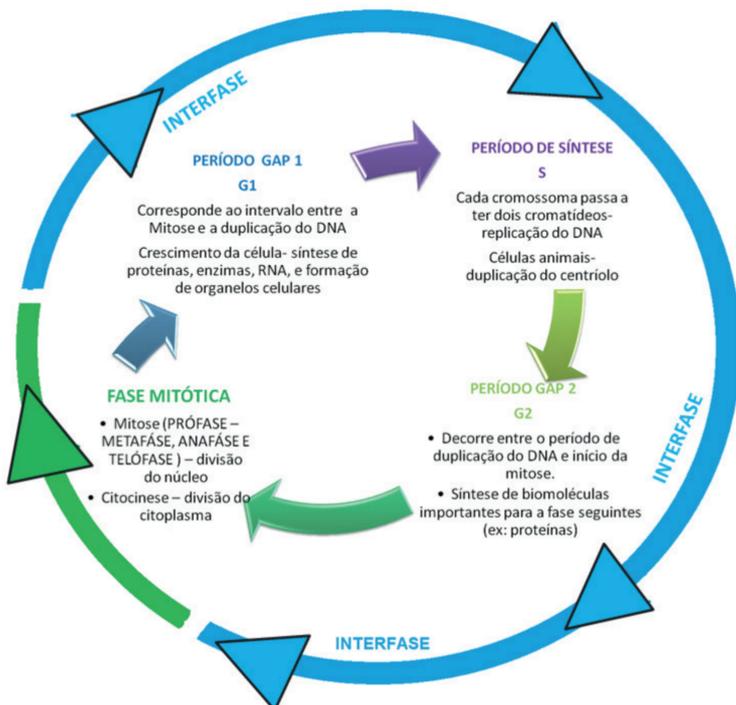
Na fase S (do inglês *synthesis*, "síntese"), as moléculas de DNA param de produzir RNA e se ocupam de sua replicação, isto é, constitui a fase de duplicação do seu material genético, na qual uma molécula de DNA sintetiza outra molécula de DNA, idêntica a ela. Por fim, na fase G2, todos os genes da célula já se encontram duplicados e, então, a célula para a síntese de DNA e volta a produzir mais RNA; visando a maior obtenção de proteínas e retornando

ao seu desenvolvimento – crescendo até atingir a dimensão que desequilibrará a relação **superfície versus volume**, o que, por sua vez, dificultará a absorção ou eliminação de substâncias e induzirá a nova divisão celular (Figura 1.2).



## Exemplificando

Figura 1.2 | Fases da interfase e suas principais características



Fonte: <<https://goo.gl/cbk7V1>>. Acesso em: 20 mar. 2017.

Após a fase G2, a célula entra em mitose já contendo duas cromátides irmãs (unidas pelo centrômero) – uma vez que seus cromossomos foram duplicados durante a interfase (Figura 1.3). A maioria das células animais apresenta, próximo ao núcleo, um centríolo que organiza os microtúbulos enquanto estão sendo formados. Esse centríolo é duplicado minutos antes de

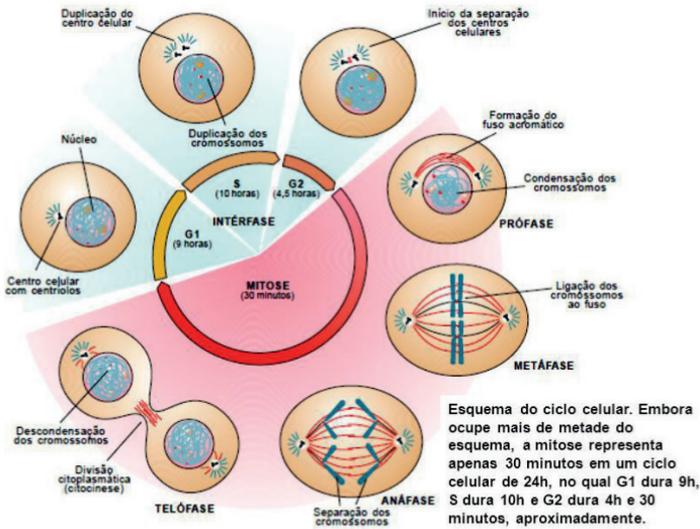
a prófase começar e, durante essa fase, um deles vai para o polo oposto do núcleo. A partir dos centrossomos, os microtúbulos começam a crescer e formar o fuso mitótico. Ao término da prófase, o envoltório nuclear se rompe e os microtúbulos penetram na região nuclear. Uma cromátide de cada cromossomo é presa a microtúbulos que se estendem de um polo do fuso, enquanto sua cromátide irmã é presa a microtúbulos que se estendem do outro polo do fuso.

Na fase seguinte (metáfase), os microtúbulos passam a adicionar e liberar subunidades de tubulina e, à medida que crescem e encolhem, empurram e puxam os cromossomos. Uma vez com o mesmo comprimento, os microtúbulos alinham os cromossomos entre os polos do fuso. Seguindo-se à metáfase, as cromátides irmãs de cada cromossomo se separam e partem opostamente em direção aos polos do fuso – fase denominada “anáfase”, a qual termina quando cada cromossomo e sua duplicata chegam a esses polos.

Por fim, a telófase se inicia quando os dois grupamentos de cromossomos se encontram nos polos do fuso e vesículas derivadas do envelope nuclear antigo se fundem em volta destes grupamentos, enquanto os cromossomos se descondensam. Quando cada conjunto de cromossomos torna-se envolvido por um novo envelope nuclear, dois núcleos se formam.



Figura 1.3 | Fases da mitose e seus respectivos eventos



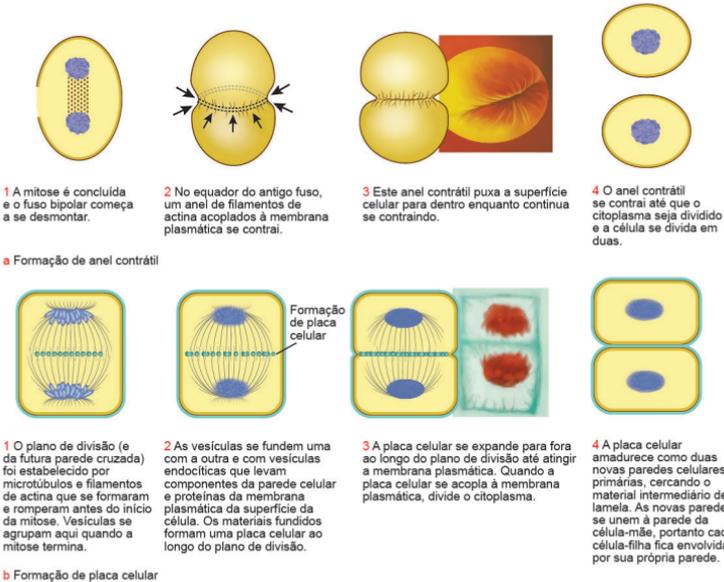
Esquema do ciclo celular. Embora ocupe mais de metade do esquema, a mitose representa apenas 30 minutos em um ciclo celular de 24h, no qual G1 dura 9h, S dura 10h e G2 dura 4h e 30 minutos, aproximadamente.

Fonte: <<https://goo.gl/kGSBfk>>. Acesso em: 20 mar. 2017.

Em células animais, após a mitose ocorre a citocinese (divisão citoplasmática), na qual o citoplasma da célula se divide em dois. No equador do antigo fuso, filamentos de actina e miosina (principalmente) ligados à membrana plasmática se contraem, formando um anel. Esse anel puxa a superfície celular para dentro, enquanto continua se contraindo até que o citoplasma seja dividido. Em células vegetais, uma placa celular se forma entre os polos do fuso e divide o citoplasma quando este atinge e se conecta à parede da célula-mãe (Figura 1.4). Ao final do processo, cada uma das células-filhas encontra-se envolvida por sua própria membrana plasmática e parede celular.



Figura 1.4 | Citocinese em células animais (a) e células vegetais (b)



Fonte: Starr (2012, p. 148).



O que aconteceria se as cromátides irmãs não se separassem como deveriam durante a mitose?

Diversos mecanismos controlam a replicação do DNA, bem como o início e o término da divisão celular. Esses mecanismos de controle estão presentes em determinados pontos do ciclo celular como “freios”, que, quando acionados, param o ciclo naquele ponto e, uma vez liberados, permitem que o ciclo funcione novamente. Algumas proteínas, resultantes dos genes de verificação, são capazes de monitorar se o DNA de uma célula foi correto e completamente copiado, se está danificado e até mesmo se as concentrações dos nutrientes são suficientes para

o crescimento da célula, interagindo para acelerar, retardar ou interromper o ciclo celular. Um exemplo dessas proteínas são as **quinases** que, ao identificarem o DNA rompido ou incompleto, ativam determinadas proteínas em uma cascata de sinalização que interrompe o ciclo celular ou faz a célula morrer.

Se um gene de ponto de checagem sofrer mutação e sua respectiva proteína codificada não funcionar corretamente, a célula poderá pular a fase de interfase e a divisão celular poderá ocorrer repetidamente, sem intervalo. Além disso, um DNA danificado pode ser copiado e esse defeito ser transmitido às células descendentes. Quando todos os mecanismos de checagem falham, uma célula perde o controle sobre seu ciclo celular e seus descendentes podem formar um tumor (massa anormal) no tecido ao redor.

As proteínas que controlam os eventos do ciclo celular são as **ciclins** e as **quinases dependentes de ciclins** que, ao identificarem o DNA rompido ou incompleto, ativam determinadas proteínas em uma cascata de sinalização que interrompe o ciclo celular ou faz a célula morrer. As quinases são moléculas proteicas com atividade enzimática, que adicionam um fosfato a outras proteínas, enquanto que as ciclins são proteínas reguladoras, assim denominadas por serem produzidas e posteriormente destruídas durante algumas fases do ciclo celular.

Há dois grandes grupos de ciclins: as ciclins  $G_1$ , que atuam na fase  $G_1$ , e as ciclins  $M$ , que atuam na mitose. Em determinado momento da fase  $G_1$  (denominado "ponto de partida"), fatores moleculares da matriz extracelular ou outras células estimulam a célula a se dividir. A partir desse ponto, a célula começa a sintetizar ciclina  $G_1$ , que juntamente com a quinase  $Cdk_2$ , forma o complexo *SPF* (sigla das palavras em inglês *S-phase-promoting factor*). Esse complexo é necessário para o início do processo de replicação do DNA, fazendo a célula entrar na fase  $S$ .

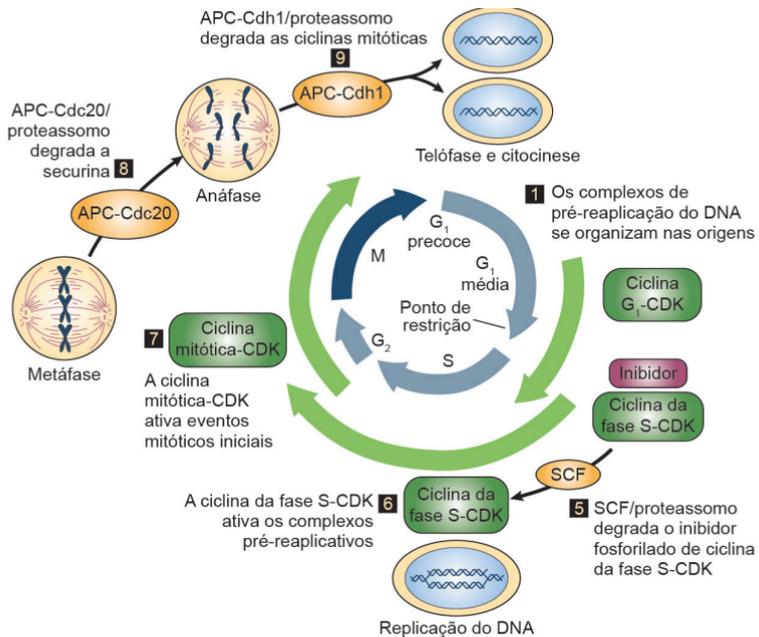
Em determinado momento da fase  $S$ , a ciclina  $G_1$  e outras começam a sofrer proteólise e o complexo *SPF* se desfaz. A partir da fase  $G_2$ , a concentração de ciclina  $M$  começa a aumentar e forma, juntamente com a ciclina  $Cdc_2$ , o complexo *MPF* (*M-phase-promoting factor*), dando início à mitose (Figura 1.5). Quando as proteínas fosforiladas pela ação da  $Cdc_2$  sofrem desfosforilação, a concentração de ciclina  $M$  diminui e determina o fim da mitose

– o complexo *MPF* é, então, desativado na anáfase, dando lugar ao complexo *APC* (*anaphase-promoting complex*), promotor da degradação da ciclina M.



## Exemplificando

Figura 1.5 | Mecanismo de ação das proteínas controladoras dos eventos do ciclo celular



Fonte: Lodish (2005, p. 852).

Danos genéticos podem ocorrer se as células progredirem para a próxima fase do ciclo celular sem que a fase anterior se complete de maneira adequada. Se a anáfase se inicia antes que os dois cinetócoros (estruturas proteicas associadas lateralmente ao centrômero e que dirigem a migração dos cromossomos durante a divisão celular) do cromossomo em replicação estejam ligados aos microtúbulos do polo mitótico oposto, por exemplo, as células-filhas produzidas terão um S-cromossomo a mais ou a menos. Quando isto acontece durante a divisão meiótica que gera

o zigoto humano, pode desencadear a trissomia do cromossomo 21 e resultar na Síndrome de Down. Cada cromátide apresenta uma estrutura proteica associada lateralmente ao centrômero, conhecido como "cinetócoro".

Para minimizar a ocorrência de erros desse tipo, a progressão do ciclo celular é monitorada em diversos **pontos de checagem**, nos quais se garante que os cromossomos estejam intactos e que cada fase do ciclo esteja completa antes que a próxima fase seja iniciada. Um desses pontos, denominado "ponto de checagem do DNA não replicado", identifica a não replicação do DNA e inibe a ativação do complexo MPF; vedando a entrada, em mitose, de células que não tenham completado a síntese do DNA. Outro ponto, denominado "ponto de checagem da ligação ao fuso mitótico", impede a entrada precoce da célula na anáfase, quando somente um cinetócoro de uma cromátide não está associado de maneira correta com os microtúbulos do fuso mitótico.

O **ponto de checagem da segregação dos cromossomos** monitora a localização dos cromossomos-filhos em segregação ao final da anáfase, determinando se a  $Cdc_{14}$  ativa ficará disponível para promover a saída da mitose. Por fim, o **ponto de checagem de dano no DNA** promove a parada do ciclo em resposta ao dano no DNA, até que esse dano seja reparado. A parada em  $G_1$  e em S impede que bases nitrogenadas danificadas sejam copiadas, perpetuando mutações no genoma. A replicação do DNA com dano também causa rearranjos nos cromossomos, que podem contribuir para o desenvolvimento de câncer. Já a parada em  $G_2$  permite que quebras na dupla fita do DNA sejam reparadas antes da mitose.

Em células tumorais, normalmente um ou mais genes de verificação estão ausentes. Genes de verificação capazes de inibir a mitose são chamados de "supressores" porque, quando ausentes, são formados tumores. Ao contrário, quando os genes de verificação estimulam a mitose, recebem o nome de proto-oncogenes, e mutações que alteram seus produtos ou a taxa na qual são formados podem ser responsáveis pela transformação de uma célula normal em tumoral. As proteínas codificadas por genes supressores de tumores, incluindo o *ATM* e o *Chk<sub>2</sub>*, normalmente atuam no ponto de checagem do DNA. Pacientes com mutações

nas duas cópias desses genes desenvolvem cânceres com maior frequência que o normal. Outra proteína supressora (p53) contribui para a parada das células com dano no DNA, e quando esse dano é muito extenso, a proteína também ativa a expressão de genes que levam à morte da célula por apoptose.

A formação de tecidos e órgãos funcionais durante o desenvolvimento dos organismos multicelulares depende de uma série de divisões celulares, semelhante a uma árvore genealógica, denominada "**linhagem celular**"; regulada por fatores intrínsecos e extrínsecos. A linhagem celular se inicia com as **células-tronco**, células não especializadas, que podem indefinidamente originar novas células-tronco e outras mais especializadas. Algumas linhagens celulares contêm células intermediárias (**células progenitoras** ou **precursoras**) cujo potencial para gerar células diferenciadas é mais limitado quando comparado ao das células-tronco.

O zigoto, embora não tenha capacidade de autorrenovação – e, portanto, não possa ser considerado tecnicamente célula-tronco –, é uma célula **totipotente**, capaz de gerar células com propriedades de célula-tronco, as quais, por sua vez, são capazes de gerar todos os tipos celulares do organismo. No estágio embrionário de oito células, cada uma pode originar qualquer tipo celular. Dessa forma, a subdivisão do organismo em partes e tecidos, quando nesse estágio, é reversível; o que não ocorre no estágio de 16 células, quando algumas já se encontram comprometidas com vias de diferenciação específicas.

Em muitos tipos de células, o início da diferenciação definitiva está associado à parada do ciclo celular comumente em G1, sugerindo que a transição para o estado diferenciado pode ser influenciada pelas proteínas do ciclo celular, incluindo as ciclinas e as quinases dependentes de ciclinas. A **diferenciação celular** constitui o processo de especialização das células, as quais passam a exercer funções específicas com grande eficiência.

As modificações celulares que ocorrem na diferenciação são resultado da ativação de determinados genes e da inativação de outros. Os fatores de transcrição essenciais para a diferenciação são específicos para cada tipo celular. Um grupo de fatores de transcrição denominados "fatores determinantes do músculo" (FDM), por exemplo, é fundamental para o início da transcrição

de genes específicos para o músculo. Por outro lado, a inativação seletiva de genes é igualmente importante, uma vez que, conforme a célula se diferencia, ela perde sua potencialidade – e essa perda ocorre através da inativação gênica.

A diferenciação celular é controlada por fatores intra e extracelulares, exigindo, portanto, intensa comunicação célula-célula e célula-ambiente. A compartimentalização precoce de componentes citoplasmáticos no zigoto, por exemplo, ocasiona uma desigualdade que persiste nas clivagens, sendo responsável pelo primeiro passo de diferenciação celular. Fatores extrínsecos resultam de sinais emanados de outras células, da matriz extracelular do organismo em diferenciação ou de agentes provenientes do meio ambiente. Além disso, vários hormônios e fatores de crescimento produzidos em células distantes também interferem na diferenciação e no metabolismo celular.

A formação de novas células acontece pela **morte** de outras preexistentes, para que a quantidade de células adultas se mantenha em equilíbrio. Quando uma célula é destruída por uma injúria qualquer, ocorre o extravasamento de seu citoplasma e a geração de uma resposta tecidual (processo inflamatório), causada por um mecanismo comum denominado “necrose”. Por outro lado, quando a própria célula induz os processos que levam à sua morte, o mecanismo é denominado “apoptose” – sendo que, nesse tipo de morte celular, não há processo inflamatório derivado de necrose. Em processos nocivos ao organismo, a apoptose tem um papel muito importante: células cancerosas, linfócitos com disfunções que geram doenças autoimunes e células com defeitos fisiológicos, por exemplo, são conduzidos à apoptose.

Para não extravasar seu conteúdo citoplasmático e, conseqüentemente, gerar um processo inflamatório, células que sofrem apoptose passam por alguns eventos morfológicos:

- Desorganização do citoesqueleto.
- Condensação e fragmentação do DNA na forma de cromatina.
- Clivagem do núcleo em vários pequenos núcleos, devido à compactação do DNA.
- Formação de várias vesículas, envoltas por unidade de

membrana contendo os pequenos núcleos e citoplasma (corpos apoptóticos), facilitando sua fagocitose pelos macrófagos.

Existem dois principais mecanismos desencadeadores de apoptose. O primeiro envolve substâncias que parecem ser essenciais para que as células se mantenham vivas e, uma vez que essas substâncias estejam em falta, os eventos apoptóticos são desencadeados: uma proteína G é ativada e estimula a fosfolipase C que, ao agir sobre outras substâncias, induz as membranas do retículo liso a liberarem íons cálcio para o meio interno da célula. Dessa forma, o cálcio liberado ativa promotores da apoptose, como endonucleases e proteases, além de impedir que produtos dessa célula afetem as demais durante a apoptose.

O segundo mecanismo ocorre pela ativação de uma família de proteínas catalíticas denominadas "caspases". Substâncias estimuladoras de apoptose atuam aumentando a expressão do gene da caspase-9, indutora de apoptose, enquanto que substâncias inibidoras de apoptose atuam aumentando a expressão do gene *bcl-2*. O gene supressor de tumor *p53* atua na estabilização do DNA na fase G1 do ciclo celular. Se erros no DNA não puderem ser consertados, a proteína *p53*, como vimos anteriormente, induzirá a célula à apoptose, provavelmente para impedir o acúmulo de múltiplas mutações capazes de converter a célula normal em cancerosa. Praticamente todas as células cancerosas apresentam mutações nos dois alelos do gene *p53* ou na via que estabiliza a proteína *p53* em resposta ao dano no DNA. Mutações em *p53*, *ATM* e *Chk2* constituem exemplos dramáticos da importância dos pontos de checagem do ciclo celular na saúde de organismos multicelulares.



### Pesquise mais

Aprofunde seus conhecimentos a respeito da morte celular por apoptose através da leitura do artigo a seguir.

ANAZETTI, M. C.; MELO, P. S. Morte celular por apoptose: uma visão bioquímica e molecular. **Metrocamp Pesquisa**, Campinas, v. 1, n. 1, p. 37-58, 2007. Disponível em: <[http://www.colegiogregormendel.com.br/gm\\_colegio/pdf/2012/textos/3ano/biologia/65.pdf](http://www.colegiogregormendel.com.br/gm_colegio/pdf/2012/textos/3ano/biologia/65.pdf)>. Acesso em: 20 mar. 2017.

## Sem medo de errar

Durante sua visita ao parque, Leticia perguntou ao guia por que os dedos dos jacarés eram “colados” uns nos outros e ficou surpresa ao saber que, durante certo período do desenvolvimento embrionário, os humanos também possuem os dedos “colados”. O mecanismo envolvido no “descolamento” dos dedos humanos é a morte celular por apoptose (morte programada). A apoptose frequentemente é acionada em células que necessitam ser renovadas, sendo crucial durante o desenvolvimento embrionário. Quando em desenvolvimento, embriões humanos apresentam membranas (membranas interdigitais) entre os dedos das mãos e dos pés; mas estas membranas estão ausentes ao nascimento. A apoptose ocorre no interior da célula, através de uma série de reações em cascata, controladas por enzimas denominadas *caspases*, que digerem materiais celulares. Para que este mecanismo seja desencadeado, é necessária a presença de um estímulo fornecido por uma substância produzida por outras células. Havendo a permanência desta condição, isto é, ausência de apoptose nas membranas interdigitais, ocorre sindactilia cutânea – condição caracterizada pela união de dois ou mais dedos, podendo ser resultado de alterações genéticas ou anormalidades herdadas pelos pais.

## Avançando na prática

### Síndrome de Down

#### Descrição da situação-problema

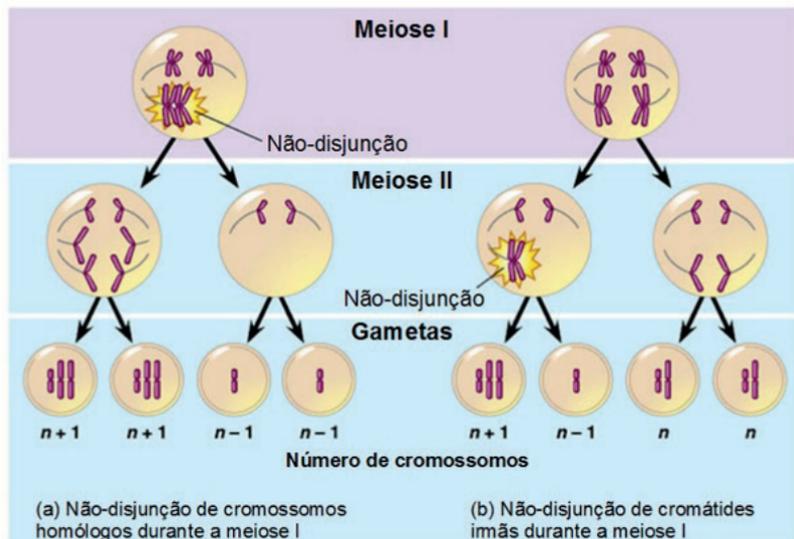
Um casal, cujo filho primogênito é portador da Síndrome de Down, planeja mais um filho. No entanto, eles se preocupam com a possibilidade de o segundo também ser portador da doença e, dessa forma, decidem procurar uma ginecologista especialista em Medicina Reprodutiva. A médica explica ao casal que a Síndrome de Down é uma doença genética causada pela trissomia do cromossomo 21 e que, quando produzido em laboratório através de fertilização in vitro, é possível detectar se o embrião é portador de problemas genéticos antes de ser implantado no útero materno. Nesta técnica, denominada “Diagnóstico Genético

Pré-Implantacional” (PGD), uma ou duas células são retiradas no terceiro dia após a fecundação, quando o embrião encontra-se no estágio de seis a oito células. Qual seria uma das hipóteses para a formação de um cromossomo 21 extra? Por que, durante o estágio de seis a oito células, é possível a retirada de uma ou duas sem que haja prejuízo para o desenvolvimento do embrião?

### Resolução da situação-problema

A progressão do ciclo celular é constantemente monitorada, para garantir que os cromossomos estejam intactos e que cada fase do ciclo esteja completa antes que a próxima fase seja iniciada, evitando, assim, danos genéticos. Se a anáfase tiver início antes que os dois cinetócoros do cromossomo em replicação estejam ligados aos microtúbulos do polo mitótico oposto, esses cromossomos permanecem unidos (**não disjunção dos cromossomos**) durante a divisão meiótica I e são transportados juntos para um dos polos que darão origem aos gametas. Dessa forma, as células-filhas produzidas terão um cromossomo a mais ( $n + 1$ ) ou a menos ( $n - 1$ ), podendo desencadear a trissomia do cromossomo 21 e resultar na Síndrome de Down (Figura 1.6). O Diagnóstico Genético Pré-Implantacional (PDG) é possível, uma vez que, no estágio de seis a oito células, cada uma pode originar qualquer tipo celular, não implicando prejuízos para o desenvolvimento embrionário. Por outro lado, não seria possível no estágio de 16 células, quando algumas já se encontram comprometidas com vias de diferenciação específicas.

Figura 1.6 | Possível hipótese para a formação de um cromossomo 21 extra



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Fonte: <<https://goo.gl/eStwGc>>. Acesso em: 20 mar. 2017.

## Faça valer a pena

1. A seguir, estão apresentadas as fases da mitose:

- ( ) Prófase
- ( ) Metáfase
- ( ) Anáfase
- ( ) Telófase

Dados os eventos listados a seguir:

- (1) Condensação dos cromossomos.
- (2) Formação de novos núcleos.
- (3) Separação das cromátides irmãs.
- (4) Cromossomos duplicados alinhados no equador do fuso.

De acordo com a ordem das fases da mitose, assinale a alternativa que contém a sequência numérica correta dos eventos listados, de cima para baixo.

- a) 1 – 2 – 3 – 4.
- b) 1 – 4 – 3 – 2.
- c) 2 – 1 – 4 – 3.
- d) 3 – 2 – 1 – 4.
- e) 4 – 3 – 2 – 1.

**2.** Em 2001, Leland Hartwell, Richard Timothy Hunt e Paul Nurse receberam o Prêmio Nobel de Medicina e Fisiologia como reconhecimento por seus trabalhos pioneiros envolvendo o ciclo celular, mais precisamente a descoberta de genes e moléculas envolvidos na regulação da divisão celular. Analise as afirmativas a seguir a respeito do ciclo celular e assinale a alternativa correta.

I. A célula perde suas atividades metabólicas quando se encontra em  $G_1$ .

II. Durante a fase  $G_2$ , a síntese de DNA é menos intensa.

III. Durante a fase S, há intensa atividade nucleolar.

IV. O ciclo celular é suspenso na fase S, em células totalmente diferenciadas.

V. Na fase  $G_1$ , a célula tem metade da quantidade de DNA quando comparada à fase  $G_2$ .

a) Apenas I.

b) Apenas III.

c) II e V.

d) I, IV e V.

e) I, II, III, IV e V.

**3.** O ciclo celular constitui um conjunto de eventos pelos quais uma célula passa com o intuito de se duplicar, dando origem a duas novas idênticas a ela. Esses eventos são de extrema importância para o correto funcionamento da célula – erros nessas etapas podem ocasionar morte celular e desenvolvimento de tumores.

Assinale a alternativa que apresenta a ordem correta das fases do ciclo celular.

a) Prófase – Anáfase – Metáfase – Telófase – Interfase – Citocinese.

b) Prófase – Metáfase – Anáfase – Telófase – Interfase – Citocinese.

c) Interfase – Prófase – Metáfase – Anáfase – Citocinese – Telófase.

d) Interfase – Prófase – Anáfase – Metáfase – Telófase – Citocinese.

e) Interfase – Prófase – Metáfase – Anáfase – Telófase – Citocinese.

# Seção 1.3

## Genômica comparativa

### Diálogo aberto

Para darmos continuidade ao seu aprendizado, retomaremos a visita de Leticia e seus pais ao Zoológico de Itatiba. Ao final do dia, restava apenas a visita ao cativeiro dos macacos, onde a família se divertiu com as brincadeiras dos micos-leões-dourados e dos macacos-aranha-do-peito-amarelo. As crianças puderam alimentá-los com bananas e outras frutas e observaram a semelhança entre a amamentação dos primatas e a dos humanos. O guia do parque, ao contar sobre as características e os hábitos dos macacos, relatou que eles são os animais mais parecidos com os humanos, podendo a semelhança chegar a 98,7%. Qual é o componente utilizado do organismo dos seres vivos e como são determinadas essas semelhanças? Para responder a esses questionamentos, leia o livro didático dando ênfase ao conteúdo sobre genômica comparativa. Vamos lá?

### Não pode faltar

Os organismos vivos apresentam muita variação na quantidade de DNA que contêm. Vírus e bactérias, por exemplo, têm muito menos DNA que a maioria dos eucariotos – dentre os quais também existe uma variação considerável. Entretanto, a quantidade de DNA nuclear não corresponde à complexidade do organismo. Algumas salamandras, por exemplo, apresentam 100 vezes mais DNA que o homem. Nos últimos anos, uma variedade de sequências genômicas se tornou disponível, podendo ser utilizada na busca por sinais de mecanismos básicos da evolução, como seleção natural e deriva genética (mudança das frequências alélicas em pequenas populações, de forma totalmente aleatória).

A obtenção de uma sequência completa de nucleotídeos de um segmento de DNA é importante para o entendimento da organização de um gene e sua regulação, sua relação com

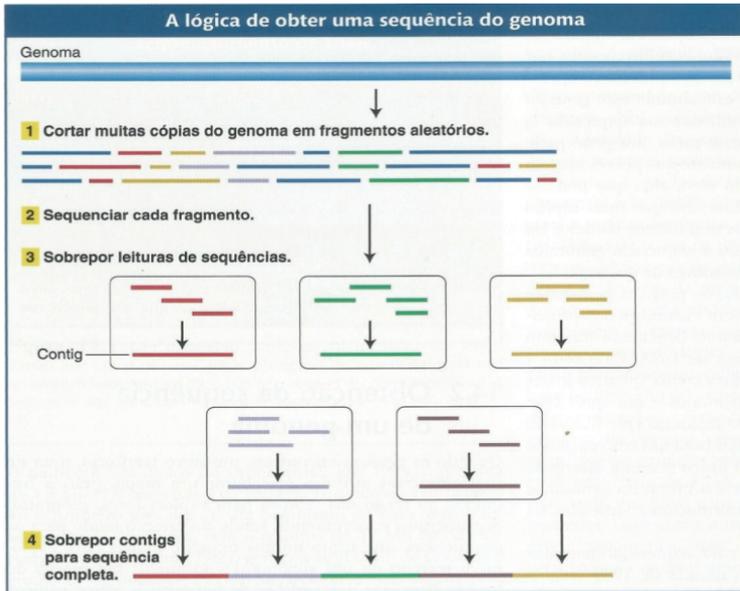
outros genes ou a função do RNA ou da proteína que ele codifica. Assim como na técnica do DNA recombinante e na PCR, o **sequenciamento genômico** explora a complementaridade de pares de bases, juntamente com o entendimento da bioquímica básica da replicação do DNA.

A análise de genomas inteiros contribui para todos os aspectos das pesquisas biológicas, como a possibilidade de localizar genes que contribuem para muitas doenças determinadas por combinações complexas de fatores genéticos. Dessa forma, pesquisas que utilizam um gene de uma espécie cujo genoma já esteja sequenciado demandam apenas a descoberta do local no qual está situado esse gene no mapa genômico, para então determinar sua sequência e, potencialmente, sua função.

Basicamente, a obtenção de uma sequência genômica acontece em quatro etapas: 1) fragmentação das moléculas de DNA do genoma em milhares a milhões de pequenos segmentos aleatórios; 2) leitura da sequência de cada um desses segmentos; 3) encontro, com o auxílio de computador, da sobreposição entre os segmentos nos quais sejam idênticas suas sequências; e 4) continuação da sobreposição de pedaços maiores, até que todos os pequenos segmentos estejam ligados (Figura 1.7). Sendo assim, o maior desafio do projeto genoma é a **montagem de sequência**, isto é, unir todas as leituras individuais em uma sequência consenso, para a qual há concordância (representação verdadeira e precisado genoma em questão).



Figura 1.7 | Etapas para a obtenção de uma sequência genômica



Fonte: Griffiths (2015, p. 424).

A estratégia atual para obtenção e montagem da sequência de um genoma é denominada “sequenciamento *shotgun* de genoma inteiro” (WGS, do inglês *whole-genome shotgun*) e baseia-se na determinação da sequência de muitos segmentos de DNA gerados pela quebra de DNA de longos cromossomos em diversos pequenos segmentos (insertos), seguida da construção de “bibliotecas genômicas”. Posteriormente, os insertos são inseridos em cromossomos acessórios denominados “vetores”, por exemplo, plasmídios, vírus bacterianos modificados e cromossomos artificiais. Uma vez introduzido em células bacterianas, o DNA recombinante (inserto + DNA do vetor) é replicado durante o crescimento e a divisão celular de seu hospedeiro, produzindo muitas cópias idênticas do inserto inserido.

A biblioteca *shotgun* resultante é constituída, portanto, de sequências obtidas a partir de clones selecionados aleatoriamente

da biblioteca genômica inteira, sem qualquer informação sobre onde esses clones estão mapeados no genoma. Em seguida, os fragmentos genômicos nos clones da biblioteca *shotgun* são sequenciados parcialmente. Como a sequência do inserto clonado é desconhecida (objetivo da técnica), os *primers* utilizados são baseados na sequência do DNA vetor adjacente, para orientar a reação de sequenciamento do inserto. Assim, uma grande coleção de curtas sequências aleatórias é formada, algumas, sobrepostas. Essas sequências são montadas em uma sequência consenso que cobre todo o genoma, combinando-se às sequências homólogas compartilhadas por leituras de clones sobrepostos, as quais, por sua vez, são montadas em sequências contínuas maiores denominadas “*contigs*”.



### Exemplificando

A sobreposição de fragmentos da frase “TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE”, de acordo com o espaço a que as sequências correspondem, fornece uma analogia para a montagem da sequência de um genoma.

Figura 1.8 | Sobreposição de fragmentos da frase “TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE”

TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE  
TOTO, I HAVE A FEELING WE’RE NOT IN KANSAS ANYMORE

Fonte: Lewis (2004, p. 447).

Nos genomas animais, contendo estruturas complexas de íntrons e éxons, o desafio da montagem de sequências é muito maior. Assim, diversas “ferramentas” de **Bioinformática** foram criadas para a identificação de genes e a determinação da

composição genética de genomas complexos. Diversos softwares foram desenvolvidos para identificação de genes através de vários critérios de sequência. Entretanto, uma limitação dos softwares atuais é a falha na identificação de promotores – os elementos centrais do promotor são altamente degenerados e, apesar de o complexo de transcrição ser capaz de identificá-lo dentro da célula, os programas computacionais não o são, falhando na identificação desses elementos e quando outras limitações de sequência estão envolvidas (éxons associados, por exemplo).

O método mais relevante empregado pelos programas atuais é a utilização de dados de sequências de cDNA, gerados pela transcrição reversa a partir de RNAm maduros e, portanto, sequências legítimas de éxons. Os cDNAs são utilizados para gerar pequenas leituras de sequência (EST, do inglês *expressed sequence tag*), realizadas a partir de um cDNA maior e obtidas da extremidade 5' ou extremidade 3', principalmente, do cDNA. Sequências de cDNA aleatórias são determinadas através do sequenciamento por *shotgun* e, então, alinhadas sobre as estruturas de *contigs* genômicas – regiões de alinhamento correspondem aos éxons, enquanto a sequência genômica, situada entre as regiões alinhadas, corresponde aos íntrons.



### Refleta

Diz-se que o código genético é degenerado, uma vez que vários códons codificam o mesmo aminoácido. A leucina, por exemplo, pode ser codificada por seis trinças diferentes. Essa é uma característica vantajosa no sentido de tornar possível a ocorrência de mutações que não provocam mudanças no aminoácido codificado, caso a mudança seja para outra trinca que codifica o mesmo aminoácido. Entretanto, constitui certa desvantagem para a identificação e montagem das sequências de DNA de um genoma. Qual seria uma alternativa para esse inconveniente?

Devido à degeneração do código genético, as proteínas relacionadas exibem maior similaridade nas sequências do que os genes que as codificam e, portanto, é possível comparar as sequências das proteínas ao invés das sequências de DNA que as

codificam. O programa de computador utilizado para este fim é conhecido como "BLAST" (do inglês, *Basic Local Alignment Search Tool*), o qual divide uma sequência de proteína em pequenos segmentos e procura, no banco de dados, qualquer sequência similar armazenada – determinando escores altos para aminoácidos idênticos e escores baixos para aminoácidos relacionados (por exemplo, positiva ou negativamente carregados, hidrofóbicos, polares, etc.).

Após a conclusão da pesquisa, o programa classifica as homologias entre a proteína consultada e as várias conhecidas utilizando valores *p*. Quanto mais baixo o valor de *p*, maior a similaridade entre as duas sequências – valores de *p* ao redor de  $10^{-3}$  são considerados evidência significativa de que as proteínas compartilham um ancestral comum. Quando uma proteína não apresenta similaridade significativa com outras proteínas através do BLAST, pode haver ainda o compartilhamento de pequenas sequências com outras proteínas funcionalmente importantes. Esses pequenos segmentos recorrentes em muitas proteínas diferentes são denominados "motivos" e geralmente apresentam funções similares.

Estima-se que o **genoma humano** seja constituído por cerca de 20.500 genes codificadores de proteínas, possuindo um tamanho total de 2,8 bilhões de pares de bases (2,8 gigabases ou Gb). O maior cromossomo, dentre os autossomos, é o cromossomo 2, contendo 237,5 Mb (8,35%), e o menor é o cromossomo 21, contendo 34,2 Mb (1,2%). O cromossomo 1 contém o maior número de genes (3.141), enquanto que o cromossomo 21 contém o menor número (apenas 225 genes).

O evento evolucionário mais significativo do genoma humano foi a formação do cromossomo 2 a partir da fusão entre as porções proximais de dois cromossomos acrocêntricos menores do ancestral comum de humanos e chimpanzês – por muito tempo, acreditou-se que o número cromossômico do *Homo sapiens* era  $2n = 48$ , como os demais antropóides. Além dessa fusão, vários outros rearranjos ocorreram na história evolutiva de humanos, chimpanzês, gorilas e orangotangos.

O genoma humano compartilha 21% dos genes com o genoma de todas as formas de vida celular, sendo "genes de manutenção"

das células, que regulam a maquinaria celular básica. A maioria evolui lentamente e vem sendo copiada ao longo de bilhões de anos, com poucas modificações. Outros 32% são homólogos aos genes de todos os eucariotos, mas não aos genes de bactérias, refletindo a complexidade do metabolismo celular dos eucariotos. 24% são compartilhados com outros animais, mas não com procariotos e eucariotos unicelulares, constituindo genes que controlam o desenvolvimento. Por fim, 22% são compartilhados somente com vertebrados, e apenas 1% ou menos é "exclusivo" dos humanos, não havendo homologia com os genes de outros vertebrados.



### Pesquise mais

Ficou curioso a respeito do genoma humano? Leia o artigo a seguir e acabe com sua curiosidade.

CORRÊA, M. V. O Admirável Projeto Genoma Humano. **PHYSIS: Revista Saúde Coletiva**, Rio de Janeiro, v. 12, n. 2, p. 277-299, 2002. Disponível em: <<http://www.scielo.br/pdf/physis/v12n2/a06v12n2>>. Acesso em: 26 abr. 2017.

Os genomas não são constituídos apenas por sequências que codificam proteínas e RNAs (RNAr e RNAt). Nos humanos, apenas 1,2% do genoma codifica proteínas, predominando sequências parasitas que não codificam produtos que auxiliam a eficiência do funcionamento celular.

Muito do DNA "extra" é composto por elementos genéticos transponíveis ou "genes saltadores", que se deslocam para novas posições no genoma, podendo deixar uma cópia para trás. Cerca de 45% do DNA do genoma humano provém de elementos transponíveis.

Um elemento transponível é um alelo capaz de aumentar sua frequência na população através da transposição para um novo local em um genoma individual e, assim, ser transmitido para a prole. Entretanto, ao se inserir em uma nova posição no genoma, desfaz uma sequência codificadora importante, resultando em uma mutação deletéria "nocaute". Em humanos, transposições já resultaram em formação de tumores e casos de hemofilia. Apesar

de a maioria dos elementos transponíveis atuar como parasitas genômicos e de a maioria das transposições resultar em mutações deletérias, alguns eventos de transposição originam genes novos importantes ou outras mudanças com impacto positivo na aptidão dos organismos.

Assim, podemos concluir que os genomas são **dinâmicos**. Um fenômeno clássico desse dinamismo é a “transferência horizontal de genes”, em que um ou vários genes se mudam de uma espécie para outra, ou seja, a transmissão de alelos ocorre entre espécies e na mesma geração, ao contrário da transmissão “comum” (vertical), na qual os alelos são transmitidos dentro da espécie e de uma geração para a outra.

A **genômica comparativa** busca comparar genomas completos de diferentes organismos, visando a interpretações relacionadas ao seu passado evolucionário, incluindo relações estrutura-função. Dois genes quaisquer com evidente similaridade de sequência, independentemente de serem ou não estreitamente relacionados por função, são denominados “homólogos”. A similaridade de sequência (homologia) implica uma relação evolutiva. Quando dois genes em espécies diferentes compartilham uma relação funcional e de sequências clara, são denominados “ortólogos” (genes derivados de um gene ancestral no último ancestral comum dessas duas espécies). Ainda, quando dois genes são similarmente relacionados uns aos outros, dentro de uma única espécie, são ditos “parálogos”. A maior parte desses genes é originada a partir de eventos de duplicação gênica em um único genoma, seguidos de especialização de uma ou de ambas as cópias do gene ao longo da evolução. Caso a função de um desses genes seja caracterizada em uma espécie, as respectivas informações poderão ser utilizadas para atribuir função gênica a um gene relacionado de uma segunda espécie, ao menos experimentalmente.

Em relação à genômica comparativa em primatas, os dois primatas não humanos estudados (chimpanzé e macaco *Rhesus*) apresentam tamanhos genômicos semelhantes. Através de dados relativos a essas duas espécies e a outras oito espécies de vertebrados, foi encontrada uma assinatura de evolução adaptativa específica para a linhagem humana de 1.240 genes, os quais se expressavam diferentemente nos estágios pré e pós-natal. Além

disso, a comparação de cinco espécies de primatas demonstrou evidências de seleção estabilizadora em 19% a 26% dos genes nos autossomos e 12% a 40% no cromossomo X.

As análises evolutivas das sequências de um genoma são muito mais bem trabalhadas quando se conhece a função dos genes envolvidos – trabalho desenvolvido pela **genômica funcional**, que procura, além de caracterizar a função dos genes, documentar quando são expressos e em que quantidades. O sequenciamento genômico e a genômica funcional permitem que pesquisadores façam comparações entre os conjuntos gênicos inteiros de espécies diferentes e entre os modos como se expressam, contribuindo para o entendimento das modificações genéticas responsáveis por inovações evolutivas.

Por fim, a **genética reversa** envolve uma sequência gênica (provavelmente conhecida a partir de uma sequência genômica) cuja função é desconhecida, com o objetivo de se tentar descobrir qual é sua função. Para tanto, procedimentos de mutagênese direcionada inativam completamente a função de um gene, eliminando-o e observando os efeitos sobre o funcionamento do organismo. Alterações na função do gene mutante revelam aspectos da bioquímica do gene quando este cumpre seu papel normal. Assim, a genética reversa atua analisando mutações e seus efeitos, deduzindo a função normal de um gene ao mostrar como ele não opera corretamente.

## Sem medo de errar

O guia do parque, ao contar as características e os hábitos dos macacos, relatou que eles são os animais mais parecidos com os humanos, podendo a semelhança chegar a 98,7%. Essa e outras semelhanças entre espécies diferentes são determinadas através da **genômica comparativa**, área da Biologia Evolutiva que busca comparar genomas completos de diferentes organismos, **utilizando DNA, RNA ou proteínas**, visando a interpretações relacionadas ao passado evolucionário das espécies em questão, incluindo relações estrutura-função.

## Avançando na prática

### Descobrimo um organismo através de sua sequência gênica

#### Descrição da situação-problema

Em um laboratório de Biologia Molecular de Porto Alegre, foi realizado o sequenciamento de um produto de PCR amplificado de uma amostra de lavado bronqueoalveolar, obtendo-se a seguinte sequência de DNA:

```
ATGGATGTCAATCCGACCTTACTTTTCTTAAAAGTGCCAGCAC
AAAATTGCTATAAGCACAACTTTCCCTTATACCGGAGACCCTCCT
TACAGCCATGGGACAGGAACAGGATACCCATGGATACTGTCAA
CAGGACACATCAGTACTCAGAAAAGGCAAGATGGACAACAAACA
CCGAAACTGGAGCACCGCAACTCAACCCGATTGATGGGCCACT
GCCAGAAGACAATGAACCAAGTGGTTATGCCCAAACAGATTGTG
TATTGGAAGCAATGGCTTTCTTGAGGAATCCCATCCTGGTATT
TTTGAAAACTCGTGTATTGCCCCGATGGAGGTTGTTTCAGCAAAC
ACGAGTAGACAAGCTGACACAAGGCCGACAGACCTATGACTGG
ACTTTAAATAGAAACCAGCCTGCTGCAACAGCATTGGCCAACAC
AATAGAAGTGTTTCAGATCAAATGGCCTCACGGCCAATGAGTCTG
GAAGGCTCATAGACTTCTTAAGGATGTAATGGAGTCAATGAAAA
AAGAAGAAATGGGGATCACAACCTCATTTTCAGAGAAAGAGACGG
GTGAGAGACAATATGACTAAGAAAATGATAACACAGAGAACAATA
GGTAAAAGGAAACAGAGATTGAACAAAAGGAGTTATCTAATTAGA
GCATTGACCCTGAACACAATGACCAAAGATGCTGAGAGAGGGAA
GCTAAAACGGAGAGCAATTGCAACCCCAGGGATGCAAATAAGG
GGGTTTGTATACTTTGTTGAGACACTGGCAAGGAGTATATGTGAG
AAACTTGAACAATCAGGGTTGCCAGTTGGAGGCAATGAGAAGAA
AGCAAAGTTGGCAAATGTTGTAAGGAAGATGATGACCAATTCTCA
GGACACCGAACTTTCTTTGACCATCACTGGAGATAACACCAAAT
GGAACGAAAATCAGAATCCTCGGATGTTTTTGGCCATGATCACAT
ATATGACCAGAAATCAGCCCGAATGGTTCAGAAATGTTCTAAGTA
TTGCTCCAATAATGTTCTCAAACAAAATGGCGAGACTGGGAAAAG
GGTATATGTTTGGAGAGCAAGAGTATGAAACTTAGAACTCAAATAC
CTGCAGAAATGCTAGCAAGCATTGATTTGAAATATTTCAATGATTC
AACAAGAAAGAAGATTGAAAAAATCCGACCGCTCTTAATAGAGGG
GACTGCATCATTGAGCCCTGGAATGATGATGGGCATGTTCAATA
TGTTAAGCACTGTATTAGGCGTCTCCATCCTGAATCTTGACAAA
AGAGATACACCAAGACTACTTACTGGTGGGATGGTCTTCAATCCT
```

Qual é o agente causador da infecção?

### Resolução da situação-problema

1. Acesse a página do Programa BLAST: <<http://blast.ncbi.nlm.nih.gov/Blast.cgi>>.
2. Clique em "Nucleotide Blast".
3. Digite a sequência de DNA em "Enter accession number(s), gi(s), or FASTA sequence(s)".
4. Em "Database", selecione "Nucleotide collection (nr/nt)".
5. Clique em BLAST.
6. Observe os resultados: a sequência de DNA indicada provavelmente pertence ao organismo que aparecerá em primeiro lugar da lista. Este apresentará o escore máximo com 100% de cobertura da sequência e máxima identidade.
7. Defina o nome do organismo identificado e a qual gene a sequência obtida pertence: **VÍRUS INFLUENZA A, GENE POLYMERASE PB1 (PB1)**.

### Faça valer a pena

- 1.** O sequenciamento genômico baseia-se na determinação da sequência de muitos segmentos de DNA, gerados pela quebra de DNA em diversos pequenos segmentos (insertos), os quais, posteriormente, são inseridos em cromossomos acessórios denominados "vetores". Uma vez introduzido em células bacterianas, o DNA recombinante (inserto + DNA do vetor) é replicado durante o crescimento e a divisão celular do hospedeiro, produzindo muitas cópias idênticas do inserto inserido. Analise as opções a seguir e assinale a alternativa que corresponda aquelas que podem ser utilizadas como vetores, para carregar DNA estranho para células hospedeiras.
- I. RNA.
  - II. Vírus.
  - III. PCR.
  - IV. Plasmídeos.
  - V. Agrupamentos de lipídeos.
  - VI. Cosmídeos.
  - VII. Xenotransplante.
  - VIII. Microarranjos de DNA.

- a) I, II e IV.
- b) I, III, V e VII.
- c) II, IV, V e VI.
- d) III, IV, V, VII e VIII.
- e) I, II, III, IV, V, VI, VII e VIII.

**2.** Na Bioinformática, diversos softwares foram desenvolvidos para identificação de genes e determinação da composição genética de genomas complexos. O método mais relevante empregado pelos programas atuais baseia-se na transcrição reversa, através da qual sequências de \_\_\_\_\_ são montadas em um modelo de \_\_\_\_\_. Assinale a alternativa que preencha corretamente as lacunas do texto.

- a) mRNA; DNA.
- b) cDNA; mRNA.
- c) DNA; ribossomo.
- d) Proteína; mRNA.
- e) cDNA; DNA.

**3.** Devido à degeneração do código genético, as proteínas relacionadas exibem maior similaridade nas sequências do que os genes que as codificam e, portanto, é possível comparar as sequências das proteínas ao invés das sequências de DNA que as codificam.

A descrição anterior se refere à qual ferramenta empregada pela Biologia Molecular?

- a) Sequenciamento genômico.
- b) DNA recombinante.
- c) PCR.
- d) BLAST.
- e) Clonagem.

# Referências

ANAZETTI, M. C.; MELO, P. S. Morte celular por apoptose: uma visão bioquímica e molecular. **Metrocamp Pesquisa**, Campinas, v. 1, n. 1, p. 37-58, 2007. Disponível em: <[http://www.colegiogregormendel.com.br/gm\\_colegio/pdf/2012/textos/3ano/biologia/65.pdf](http://www.colegiogregormendel.com.br/gm_colegio/pdf/2012/textos/3ano/biologia/65.pdf)>. Acesso em: 2 abr. 2017.

BIOLOGIA TOTAL. **Divisão Celular**. 2017. Disponível em: <[https://www.biologiatotal.com.br/fmanager/biologiatotal/idade-materna\\_meiose\\_pearson-education-inc.jpg](https://www.biologiatotal.com.br/fmanager/biologiatotal/idade-materna_meiose_pearson-education-inc.jpg)>. Acesso em: 2 jun. 2017.

COLÉGIO VASCO DA GAMA. **Divisão Celular**. 2017. Disponível em: <<http://www.colegiovascodagama.pt/ciencias3c/onze/images/DIVISAOCELULAR.png>>. Acesso em: 2 jun. 2017.

CORRÊA, M. V. O Admirável Projeto Genoma Humano. **PHYSIS: Revista Saúde Coletiva**, Rio de Janeiro, v. 12, n. 2, p. 277-299, 2002. Disponível em: <<http://www.scielo.br/pdf/physis/v12n2/a06v12n2>>. Acesso em: 26 abr. 2017.

COX, M. M.; DOUDNA, J. A.; O'DONNELL, M. Genomas, transcriptomas e proteomas. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas**. Porto Alegre: Artmed, 2012. cap. 8, p. 259-277.

CUNHA, C. Evolução. In: \_\_\_\_\_. **Genética e Evolução Humana**. Campinas, SP: Editora Átomo, 2011. cap. 7, p. 185-220.

\_\_\_\_\_. Evolução Humana. In: \_\_\_\_\_. **Genética e Evolução Humana**. Campinas, SP: Editora Átomo, 2011. cap. 8, p. 221-227.

ESQUEMA DO CICLO CELULAR. 2017. Disponível em: <<https://pt-static.z-dn.net/files/dd0/ac0dac3282a7a0c3f5569c709b033e8e.jpg>>. Acesso em: 2 jun. 2017.

FREEMAN, S.; HERRON, J. C. O padrão da evolução. In: \_\_\_\_\_. **Análise Evolutiva**. 4. ed. Porto Alegre: Artmed, 2009. cap. 2, p. 37-65.

\_\_\_\_\_. Seleção natural darwiniana. In: \_\_\_\_\_. **Análise Evolutiva**. 4. ed. Porto Alegre: Artmed, 2009. cap. 3, p. 90-97.

\_\_\_\_\_. Evolução em locos múltiplos: ligação e sexo. In: \_\_\_\_\_. **Análise Evolutiva**. 4. ed. Porto Alegre: Artmed, 2009. cap. 8, p. 281-317.

\_\_\_\_\_. Evolução em locos múltiplos: ligação e sexo. In: \_\_\_\_\_. **Análise Evolutiva**. 4. ed. Porto Alegre: Artmed, 2009. cap. 9, p. 319-359.

\_\_\_\_\_. A filogenômica e a base molecular da adaptação. In: \_\_\_\_\_. **Análise Evolutiva**. 4. ed. Porto Alegre: Artmed, 2009. cap. 15, p. 575-602.

FUTUYMA, D. J. Evolução molecular. In: \_\_\_\_\_. **Biologia Evolutiva**. 3. ed. Ribeirão Preto: FUNPEC Editora, 2009. cap. 22, p. 625-646.

GRIFFITHS, A. J. F. et al. Genomas e genômica. In: \_\_\_\_\_. **Introdução à Genética**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 14, p. 421-451.

- \_\_\_\_\_. Genomas e genômica. In: \_\_\_\_\_. **Introdução à Genética**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 15, p. 453-478.
- JUNQUEIRA, L. C.; CARNEIRO, J. Divisão de trabalho entre as células: diferenciação. In: \_\_\_\_\_. **Biologia Celular e Molecular**. 9. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 11, p. 233-244.
- LEWIS, R. O Projeto Genoma Humano e a Genômica. In: \_\_\_\_\_. **Genética Humana: Conceitos e aplicações**. 5. ed. Rio de Janeiro: Guanabara Koogan, 2004, cap. 22, p. 447.
- LODISH, H. et al. Regulação do ciclo celular dos eucariotos. In: \_\_\_\_\_. **Biologia Celular e Molecular**. 5. ed. Porto Alegre: Artmed, 2005. cap. 21, p. 849-894.
- \_\_\_\_\_. Origem, linhagem e morte celular. In: \_\_\_\_\_. **Biologia Celular e Molecular**. 5. ed. Porto Alegre: Artmed, 2005. cap. 22, p. 895-932.
- RIDLEY, M. O surgimento da biologia evolutiva. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 1, p. 27-44.
- \_\_\_\_\_. As evidências da evolução. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 3, p. 66-100.
- \_\_\_\_\_. Seleção natural e variação. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 4, p. 101-121.
- \_\_\_\_\_. Conceitos de espécie e variação intra-específica. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 13, p. 375-406.
- \_\_\_\_\_. Especiação. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 14, p. 407-446.
- \_\_\_\_\_. A história da vida. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 18, p. 545-576.
- \_\_\_\_\_. Genômica Evolutiva. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 19, p. 577-592.
- SADAVA, D. et al. Vida: a ciência da biologia. 8. ed. Porto Alegre: Artmed, 2011. 461 p. (v.1: Célula e hereditariedade).
- SALZANO, F. M. Origens. In: \_\_\_\_\_. **Genômica e evolução: moléculas, organismos e sociedades**. São Paulo: Oficina de Textos, 2012. cap. 2, p. 39-56.
- \_\_\_\_\_. Evolução humana. In: \_\_\_\_\_. **Genômica e evolução: moléculas, organismos e sociedades**. São Paulo: Oficina de Textos, 2012. cap. 7, p. 185-195.
- \_\_\_\_\_. O Genoma Humano. In: \_\_\_\_\_. **Genômica e evolução: moléculas, organismos e sociedades**. São Paulo: Oficina de Textos, 2012. cap. 5, p. 143-167.
- STARR, C. et al. Como as células se reproduzem. In: \_\_\_\_\_. **Biologia: unidade e diversidade da vida**. São Paulo: Cengage Learning, 2012. v. 01, cap. 9, p. 140-153.
- UZUNIAN, A.; BIRNER, E. **Biologia**. 3. ed. São Paulo: HARBRA, 2005, p. 324.
- WATSON, J. D. et al. **Genômica comparada e evolução da diversidade animal**. In:

\_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 19, p. 613-641.

\_\_\_\_\_. Técnicas de Biologia Molecular. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 20, p. 660-672.



# Material genético e evolução

## Convite ao estudo

Prezado aluno,

Daremos continuidade à compreensão da Biologia Molecular e Evolução, estudando mais a fundo o material genético e a importância das variações na sequência gênica. Ao final desta unidade, seus objetivos são: entender o que são polimorfismos e qual é a importância deles para a compreensão de doenças humanas e também conhecer os métodos utilizados na reconstrução filogenética e a relação entre os polimorfismos e a reconstrução filogenética.

Para compreender o assunto e atingir as competências e os objetivos propostos pela disciplina, analisaremos uma situação hipotética que se aproxima dos conteúdos teóricos que serão vistos por você nesta unidade: durante sua visita ao Zoológico de Itatiba, Letícia sentiu uma coceira no braço esquerdo, mas a garotinha estava tão empolgada com o passeio, que não deu importância a isso e logo se distraiu com as atrações do parque. No caminho de volta para casa, quando novamente sentiu o braço coçar, queixou-se com sua mãe, que disse “não deve ser nada de mais; apenas uma picadinha de pernilongo”. Entretanto, alguns dias depois, Letícia começou a apresentar febre, erupções cutâneas, dores nos músculos e nas articulações, mal-estar e dores de cabeça. Supondo ser algo mais grave que uma simples gripe, os pais de Letícia decidiram levá-la ao hospital da cidade. Na anamnese, a mãe de Letícia relatou os sintomas da filha à atendente e, ao ser questionada sobre os históricos de doenças/cirurgias e possíveis alergias a medicamentos da

menina, ela relatou reação alérgica ao ácido acetilsalicílico (AAS) e comentou que, quando sofre algum tipo de corte, Letícia sangra muito e apresenta demora no estancamento do sangramento. Diante dessas informações, o médico suspeitou que Letícia poderia ser portadora de hemofilia clássica leve e, por isso, solicitou alguns exames para análise dos fatores de coagulação. Os sintomas apresentados por Letícia e a picada de inseto que ela levou durante sua visita ao parque sugeriram que a garotinha havia contraído dengue, zika ou chikungunya.

No decorrer desta unidade de ensino, você conhecerá a importância dos polimorfismos na reconstrução filogenética. Na Seção 2.1, abordaremos os polimorfismos do DNA para análise de marcadores para mapeamento e para clonagem, bem como a relação entre os polimorfismos e os componentes genéticos de doenças humanas complexas. Já na Seção 2.2, compreenderemos os métodos utilizados na reconstrução filogenética e como acontece a escolha de genes para problemas filogenéticos específicos. Por fim, na Seção 2.3, trataremos de alguns polimorfismos específicos (de isozimas e de fragmentos de restrição), bem como das enzimas de restrição utilizadas para a detecção de polimorfismos no DNA.

# Seção 2.1

## Polimorfismos genéticos

### Diálogo aberto

Uma vez que os testes de coagulação apresentaram valores alterados, o médico solicitou exames mais específicos, explicando à família de Leticia que a hemofilia é uma doença genética que ocorre devido a “pequenas alterações no DNA”, mais especificamente no gene *F8*, localizado na extremidade distal do braço longo do cromossomo X (*padrão de herança recessiva ligada ao cromossomo X*). O médico explicou, ainda, que essas variações na sequência gênica estão presentes em pelo menos 1% da população e que nem sempre causam alterações no fenótipo ou na saúde de seus portadores. O que seriam as “pequenas alterações no DNA” citadas pelo médico? Por que, no caso de Leticia, essas alterações trouxeram prejuízo à sua saúde, interferindo em sua coagulação sanguínea?

### Não pode faltar

Os genes apresentam algumas variações que diferem umas das outras devido a pequenas mudanças na sequência de bases do DNA. Entretanto, muitas dessas variações não se manifestam visivelmente, porque não alteram a proteína codificada de maneira que afete sua função. Partes da sequência de DNA podem variar entre os indivíduos sem necessariamente alterarem seu fenótipo ou seu estado de saúde. Embora um gene “doente” possa se originar de uma nova mutação que surgiu nas gerações anteriores, a maioria das doenças hereditárias ocorre devido a alelos mutados já existentes que foram passados de uma geração para outra, durante muitas gerações.

Uma variação na sequência gênica, presente em pelo menos 1% da população, é chamada de **polimorfismo**. Polimorfismos podem ocorrer em regiões do DNA que codifiquem proteína ou em regiões que não a codifiquem. Existem mais de três milhões

de *polimorfismos de nucleotídeo único* (SNPs, pronunciado "snips"); os quais constituem sítios de uma só base que diferem entre os indivíduos. O genoma humano pode incluir até vinte milhões de SNPs; e é a partir dessas variações, frequentemente sutis, que provém a variedade humana (diferenças na cor dos cabelos, estatura, alergias a medicamentos, tamanho dos pés, etc.). A maioria dos SNPs não produz fenótipos diferentes porque não estão localizados dentro de um gene ou, quando localizados, os dois pares de nucleotídeos (selvagem e mutante) produzem o mesmo produto proteico.

Os SNPs localizados em regiões codificantes de proteínas são classificados em: *sinônimos* (se os alelos distintos codificarem o mesmo aminoácido), *não sinônimos* (se os dois alelos codificarem aminoácidos diferentes) e *sem sentido* ou *nonsense* (se um alelo codificar um códon de parada, enquanto o outro codificar um aminoácido). Fora de regiões codificantes, os SNPs são denominados **silenciosos** e podem ser utilizados como marcadores na genética de populações.

A combinação de alelos de múltiplos loci no mesmo cromossomo homólogo é denominada *haplótipo*. Dois cromossomos homólogos que compartilham o mesmo alelo em cada um dos loci apresentam o mesmo haplótipo. Por outro lado, se dois cromossomos têm genótipos diferentes em até mesmo um dos loci, exibem haplótipos diferentes.



### Assimile

Se o locus *A*, com alelos *A* e *a*, estiver ligado ao locus *B*, com alelos *B* e *b*, existem quatro haplótipos possíveis para o segmento cromossômico no qual esses dois loci estão localizados.

Tabela 2.1 | Haplótipos possíveis para os loci *A* e *B*

<i>A</i>	<i>B</i>
<i>A</i>	<i>b</i>
<i>a</i>	<i>B</i>
<i>a</i>	<i>b</i>

Fonte: elaborada pelo autor.

Um haplótipo denominado “aglomerado estelar”, localizado no cromossomo Y, é encontrado em 8% dos homens asiáticos. Utilizando a taxa de mutação conhecida, pesquisadores estimam que esse haplótipo surgiu entre 700 e 1300 anos atrás, sendo mais comum na Mongólia, provavelmente tendo surgido nesse país e remontando a um homem que lá viveu há cerca de 1000 anos. A distribuição atual desse haplótipo segue os limites geográficos do Império Mongol, estabelecido há, aproximadamente, 1200 anos, sugerindo que todos os homens contemporâneos portadores desse haplótipo sejam descendentes do imperador Genghis Khanou, familiares da linhagem masculina.

Nem sempre alelos mutantes conferem prejuízo ao indivíduo, muitas vezes, a condição recessiva permanece prevalente porque traz alguma vantagem de saúde não correlata (por exemplo, resistência a uma doença infecciosa ou capacidade de sobrevivência frente a uma ameaça ambiental). Essa “vantagem”, que mantém na população um alelo recessivo causador de uma doença, é denominada polimorfismo balanceado, já que o efeito protetor da condição não herdada contrabalança o efeito negativo do alelo deletério. A anemia falciforme e a deficiência da enzima G6PD, por exemplo, são doenças herdadas que conferem proteção contra a malária por tornarem as hemácias do portador inóspitas ao parasita causador dessa doença infecciosa.

A maioria dos genomas contém uma grande quantidade de DNA repetitivo (repetições múltiplas adjacentes de curtas sequências simples de DNA) cuja origem ainda não está totalmente esclarecida, mas é sabido que, em indivíduos diferentes, geralmente estão presentes números diferentes de cópias. Essas repetições são denominadas *polimorfismos de comprimento de sequência simples* (SSLP) ou *repetições in tandem de número variável* (VNTR).

Dois tipos de SSLP são utilizados como marcadores no mapeamento: *minissatélites* e *microsatélites*. Quando o DNA genômico é isolado e fracionado, as sequências repetitivas formam uma fração que é fisicamente separada do restante (fração satélite, separada da parte maior). Marcadores minissatélites baseiam-se na variação do número de repetições *in tandem* de uma unidade repetida de 15 a 100 nucleotídeos de comprimento, enquanto que marcadores microsatélites baseiam-se em números variáveis de

repetições *in tandem* de uma sequência mais simples, geralmente um dinucleotídeo.



### Exemplificando

Tipo mais comum de marcador microssatélite:

5' C-A-C-A-C-A-C-A-C-A-C-A-C-A 3'

3' G-T-G-T-G-T-G-T-G-T-G-T-G-T 5'

A repetição CA/GT ocorre por volta de 100 mil vezes ao longo do genoma, em comprimentos de 1 a 40 unidades. Além das repetições CA, outras dinucleotídicas, bem como tri- e tetranucleotídicas, são encontradas dispersas em todo o genoma humano. Os microssatélites ocorrem com frequências muito altas, e as diferenças alélicas podem ser facilmente identificadas após amplificação por reação em cadeia da polimerase (PCR).

Marcadores moleculares são a chave para orientar pesquisadores na busca por um gene de interesse. A localização do gene para a fibrose cística, por exemplo, foi descoberta através de sua ligação a marcadores moleculares, já conhecidos, localizados no cromossomo 7, levando ao isolamento e ao sequenciamento do gene – resultando, posteriormente, na descoberta de que ele codifica a *proteína reguladora da condutância transmembrana da fibrose cística* (CFTR). Além disso, marcadores moleculares distintos podem ser mapeados relacionando-se a outros e, dessa forma, cria-se um mapa para determinação de um gene utilizando-se fenótipos de interesse.

O primeiro passo no mapeamento de alelos que conferem uma característica dominante particular é a obtenção de amostras de DNA de todos os membros de uma família que contém indivíduos com a doença. O DNA de indivíduos afetados e não afetados é analisado para se determinar um grande número de polimorfismos de DNA conhecidos. O padrão de segregação de cada polimorfismo é, dessa forma, comparado com a segregação das doenças em estudo, a fim de determinar polimorfismos que tendem a segregar com a doença. Os dados de segregação são analisados por

computador e utilizados para o cálculo da probabilidade da ligação entre cada polimorfismo de DNA e o alelo causador da doença.

Qualquer variação na sequência gênica, em uma população, está sujeita à análise genética de populações, incluindo inversões, translocações, deleções ou duplicações, além de presença ou ausência de um elemento de transposição em um locus particular no genoma. Outro polimorfismo bastante comum é o *polimorfismo por inserção-deleção* (indel); que envolve a presença ou ausência de um ou mais nucleotídeos de um locus em um alelo com relação a outro. Ao contrário dos microssatélites, as indels não contêm sequências repetitivas como AGAGAGAG. O genoma da mitocôndria (DNAm) e do cloroplasto (DNAc) de eucariotos também podem apresentar variação genética, especialmente SNPs e microssatélites e podem ser utilizados para acompanhar a história das linhagens femininas, uma vez que, em geral, são de herança materna.

Comparações entre genomas auxiliam na localização de genes envolvidos em doenças. Através da *análise de ligação*, o gene envolvido em uma doença é mapeado com referência a polimorfismos bem caracterizados presentes ao longo do genoma humano. Primeiramente, é localizada a região relacionada com a doença a um cromossomo específico. Em seguida, utilizando um painel que inclui vários loci de SNP caracterizados e mapeados em cada cromossomo, são comparados os genótipos dos indivíduos com e sem a doença, concentrando-se, principalmente, nos membros familiares mais próximos.

Após identificar SNPs particulares mais frequentemente herdados com o gene causador de doença, o gene responsável pode, dessa maneira, ser localizado em um único cromossomo, no qual se encontra a mutação que dá origem à doença. Por fim, a região local que contém o gene é analisada e os genes contidos, identificados. Moléculas de DNA contendo essa região são sequenciadas a partir de muitos indivíduos (portadores da doença ou não), levando à identificação de variantes gênicas consistentemente presentes em indivíduos afetados e ausentes em não afetados.

Outra estratégia para estudos de mapeamento baseia-se no *desequilíbrio de ligação*, que envolve uma doença genética

comumente presente em uma população, decorrente de uma única mutação ocorrida muitas gerações atrás. O cromossomo ancestral em questão conterá polimorfismos ligados muito próximos, conservados por muitas gerações. Polimorfismos mais afastados no cromossomo tenderão a se afastar mais dos genes da doença pela recombinação, enquanto que os mais próximos permanecerão associados. Avaliando a distribuição de marcadores específicos em todos os indivíduos afetados, é possível identificar os marcadores de DNA diretamente associados à doença e, dessa forma, localizar o gene da doença em uma região relativamente pequena.

A expressão de um gene é bastante variável, podendo as variações ocorrer dentro de um mesmo padrão, semelhante para uma população ou individualmente. Assim, a atividade de um par de genes não acontece igualmente em todos os portadores desses genes. O material genético pode sofrer interferências de genes reguladores, de fatores nutricionais, do meio no qual o indivíduo vive e até mesmo de fatores emocionais, através da ação de hormônios e neurotransmissores.



Refleta

O gene DRD4, codificante do receptor do neurotransmissor *dopamina*, apresenta variações que predisõem seus portadores a buscar novos desafios. São os estímulos do meio que aumentam o desejo de aprender, ou o desejo de aprender é que impulsiona os indivíduos a novos estímulos que conduzem ao aprendizado?

Muito se pesquisa sobre a influência mútua de genes e meio ambiente. A *epigenética* estuda os fatores e os gatilhos ambientais que influenciam a atividade dos genes, sendo apontada, de certa forma, como um “neolamarckismo”. Gêmeos monozigóticos, por exemplo, apresentam genes absolutamente idênticos ao nascerem, entretanto, no decorrer de suas vidas, seus hábitos alimentares e comportamentais, a exposição a estímulos diferentes, estresses ambientais e outros fatores levam à metilação ou ao “desligamento” de seus genes, de maneira individual. Após várias décadas, gêmeos idênticos podem apresentar doenças e processos de envelhecimento distintos.

Estudos já demonstraram a influência de pesticidas sobre os genes, ativando ou desativando-os, durante a formação dos gametas, sendo transmitidos para as gerações futuras. Outros estudos buscam diferentes fatores epigenéticos com a mesma capacidade de influenciar as próximas gerações, como tipo de alimentação, períodos de fome e fartura, oscilação entre guerra e tranquilidade, etc. De qualquer forma, pesquisas sugerem que, talvez, nosso estilo de vida atual influencie no desenvolvimento e na saúde de nossos filhos e netos, que ainda nem nasceram.



### Pesquise mais

Saiba mais a respeito da epigenética através da leitura do artigo a seguir.

MULLER, H. R.; PRADO, K. B. Epigenética: um novo campo da genética. **RUBS**, Curitiba, v. 1, n. 3, p. 61-69, set./dez. 2008. Disponível em: <[http://www.colegiogregormendel.com.br/gm\\_colegio/pdf/2012/textos/3ano/biologia/8.pdf](http://www.colegiogregormendel.com.br/gm_colegio/pdf/2012/textos/3ano/biologia/8.pdf)>. Acesso em: 27 jun. 2017.

## Sem medo de errar

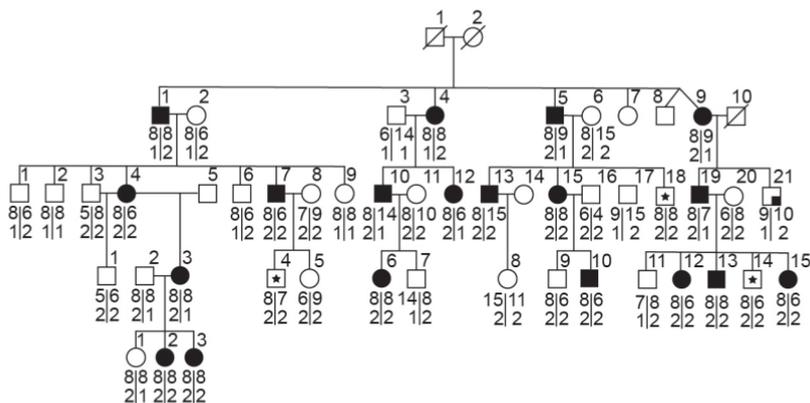
As “pequenas alterações no DNA” citadas pelo médico, ao explicar à família de Leticia a origem da hemofilia, referem-se aos **polimorfismos** genéticos, variações na sequência de DNA, que nem sempre causam alterações no fenótipo ou na saúde de seus portadores. Os eventos mais comuns que dão origem a polimorfismos são deleções, mutações, substituições de base única e variações no número de sequências repetidas. No caso de Leticia, essas variações trouxeram prejuízo à sua saúde porque ocorreram em regiões gênicas que codificam a proteína Fator VIII, a qual, como resultado da alteração em sua produção, apresenta atividade diminuída em relação às demais e interrompe a cascata de coagulação antes da produção do coágulo, tornando o estancamento dos sangramentos mais demorado.

### Mapeando um locus

#### Descrição da situação-problema

O heredograma a seguir, pertencente a uma família com múltiplas gerações de indivíduos com convulsão neonatal familiar benigna (BFNC), foi analisado visando a determinação da localização cromossômica da doença, caracterizada por episódios convulsivos violentos e descontrolados de face, corpo, braços e pernas, durante o primeiro semestre de vida. As convulsões aparentemente não causam efeito nas funções neurológicas e intelectuais e, em cerca de 90% dos casos, os sintomas desaparecem por volta do primeiro ano de vida. Qual seria uma possível metodologia para determinar a localização do cromossomo da doença?

Figura 2.1 | Mapeamento do genoma humano



Fonte: Pasternak (2002, p. 167).

#### Resolução da situação-problema

Apenas utilizando heredogramas, não é possível determinar a localização cromossômica do gene de uma doença – exceto para casos de herança ligada ao sexo. Entretanto, sondas de *polimorfismos de comprimento de fragmentos de restrição* (RFLP) e *polimorfismos de comprimento de sequência simples* (microssatélites), que identificam os loci polimórficos em todo o genoma, permitem estudos de **análise de ligação** entre um locus

polimórfico e um locus de doença. O DNA de cada indivíduo é genotipado para inúmeros marcadores polimórficos diferentes. No caso estudado, de todos os marcadores polimórficos testados, duas sondas (D20S19 e D20S20) mostraram ligação com o locus BFNC. Os alelos dos loci D20S19 e D20S20 de cada membro estão indicados um acima do outro (superiores indicando o locus D20S19 e inferiores indicando o locus D20S20), com a linha vertical separando os haplótipos dos dois cromossomos do indivíduo. Uma vez que os símbolos sólidos (■ e ●) indicam membros afetados, podemos concluir que o haplótipo (8,2) segrega com a doença. Através do cálculo do logaritmo da probabilidade (LOD score), os loci D20S19 e D20S20 estão distantes entre 1 e 4 milhões de pares de bases do locus BFNC e, uma vez que esses loci mapeiam em **20q13.2-13.3**, o locus BFNC deve estar próximo ou situado nessa região cromossômica.

### Faça valer a pena

**1.** Uma variação na sequência gênica presente em pelo menos 1% da população é chamada de polimorfismo. Polimorfismos podem ocorrer em uma parte do DNA que codifique proteína ou em uma parte que não codifique. Os polimorfismos de nucleotídeo único (SNPs) constituem sítios de uma só base que diferem as pessoas.

Estima-se que em humanos existam SNPs, distribuídos mais ou menos aleatoriamente, na casa dos:

- a) 3 milhões.
- b) 5 milhões.
- c) 7 milhões.
- d) 10 milhões.
- e) 15 milhões.

**2.** Os SNPs localizados em regiões codificantes de proteínas são classificados em: \_\_\_\_\_ (se os alelos distintos codificarem o mesmo aminoácido), \_\_\_\_\_ (se os dois alelos codificarem aminoácidos diferentes) e \_\_\_\_\_ (se um alelo codificar um códon de parada, enquanto o outro codificar um aminoácido). Fora de regiões codificantes, os SNPs são denominados \_\_\_\_\_ e podem ser utilizados como marcadores na genética de populações.

Assinale a alternativa que preencha corretamente a lacuna do texto.

- a) Silenciosos – sem sentido – não sinônimos – sinônimos.
- b) Sem sentido – sinônimos – não sinônimos – silenciosos.
- c) Sinônimos – não sinônimos – silenciosos – sem sentido.
- d) Sinônimos – não sinônimos – sem sentido – silenciosos.
- e) Não sinônimos – sinônimos – silenciosos – sem sentido.

**3.** Polimorfismos são variações frequentemente sutis na sequência do DNA presentes em pelo menos 1% da população, que promovem a variedade humana (diferenças na cor dos cabelos, estatura, alergias a medicamentos, tamanho dos pés, etc.).

Assinale a alternativa que contém a sequência numérica correta, de cima para baixo, da definição dos polimorfismos a seguir.

( ) Polimorfismo de nucleotídeo único.

( ) Polimorfismo balanceado.

( ) Polimorfismo de comprimento de sequência simples.

( ) Polimorfismo por inserção-deleção.

(1) Envolve a presença ou ausência de um ou mais nucleotídeos de um locus em um alelo com relação a outro.

(2) O efeito protetor da condição não herdada contrabalança o efeito negativo do alelo deletério.

(3) Sítios de uma só base que diferem entre os indivíduos.

(4) Repetições múltiplas adjacentes de curtas sequências simples de DNA.

a) 2 – 4 – 1 – 3.

b) 3 – 2 – 4 – 1.

c) 1 – 3 – 2 – 4.

d) 1 – 2 – 3 – 4.

e) 4 – 3 – 2 – 1.

## Seção 2.2

### Reconstrução filogenética I

#### Diálogo aberto

Para ter certeza de qual das três doenças Letícia contraiu, o médico solicitou um diagnóstico diferencial, baseado na análise do RNA viral contraído pela garotinha, através de ensaio Triplex RT-PCR (*reação em cadeia da polimerase via transcriptase reversa*), cujo resultado foi positivo para zika e negativo para dengue e chikungunya. Diante da epidemia de zika que começou a se alastrar no Brasil, pesquisadores sequenciaram o genoma completo do zika vírus e chegaram à conclusão de que o vírus causador da epidemia no país não pertence à linhagem africana (continente no qual o vírus foi descoberto, em meados do século passado), mas sim à linhagem asiática, surgida mais recentemente. A análise do genoma viral americano mostrou uma homologia de 99,7% com a cepa asiática responsável pelo surto na Polinésia Francesa alguns anos atrás. Como foram demonstradas as relações de parentesco entre as duas cepas?

#### Não pode faltar

Teorias evolutivas sugerem que todas as formas de vida evoluíram de um único ancestral comum. Sendo assim, a Biologia recentemente vem se concentrando nas conexões evolutivas, isto é, cada espécie é vista como parte de um panorama maior de evolução e não mais como um membro ou representante de uma classificação em uma hierarquia. A questão central dessas relações evolutivas é *quem está relacionado a quem*.

As relações de descendência entre as espécies podem ser representadas através de **árvores filogenéticas**, constituídas de pontos (*nós*) ligados por linhas (*ramos*) e cujas terminações, denominadas *unidades taxonômicas operacionais* (OTUs, do inglês *operational taxonomic units*), podem ser táxons (grupos de organismos) agrupados em famílias, gêneros, espécies ou

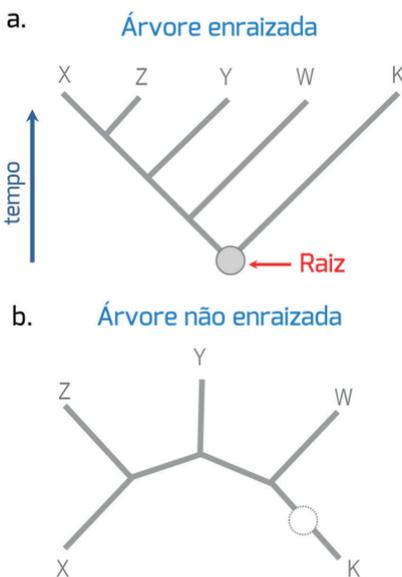
populações. As OTUs são unidas pelos ramos, os quais são conectados por *nós* que representam o ancestral comum mais recente dentre os táxons conectados. Já o nó na base da árvore (*raiz*) corresponde ao ancestral comum de todos os táxons nela representados, e os tamanhos dos ramos geralmente representam uma medida de tempo evolutivo – número de mutações em um ou vários genes, número de alterações gênicas em uma região do genoma, etc.

As árvores filogenéticas podem ser *enraizadas* ou *não enraizadas* (Figura 2.2). Em árvores enraizadas, a presença de uma raiz identifica o sentido da evolução, ou seja, fornece o sentido temporal à árvore. Já árvores não enraizadas não apontam onde está a espécie ancestral de todo o grupo (a “raiz” da filogenia), não indicando precisamente as relações de parentesco e mostrando apenas as relações de proximidade relativa entre as OTUs, a partir de semelhanças compartilhadas.



## Exemplificando

Figura 2.2 | Tipos de árvores filogenéticas



Fonte: <<https://goo.gl/63ZeZW>>. Acesso em: 27 jun. 2017.

À medida que o número de OTUs consideradas aumenta, o número de topologias (padrão de ramificação) possíveis também aumenta consideravelmente, dependendo se a árvore é enraizada ( $N$ ) ou não ( $N^*$ ). Por exemplo, quatro unidades perfazem 15 árvores enraizadas possíveis, enquanto que 10 unidades alcançam a casa dos 34,5 milhões (Tabela 2.2).



## Exemplificando

Tabela 2.2 | Número de possíveis árvores filogenéticas, enraizadas e não enraizadas, para  $n$  OTUs

Nº de OTUs	Árvores enraizadas	Árvores não enraizadas
2	1	1
3	3	1
4	15	3
5	105	15
6	945	105
7	10.395	945
8	135.135	10.395
9	2.027.025	135.135
10	34.459.425	2.027.025

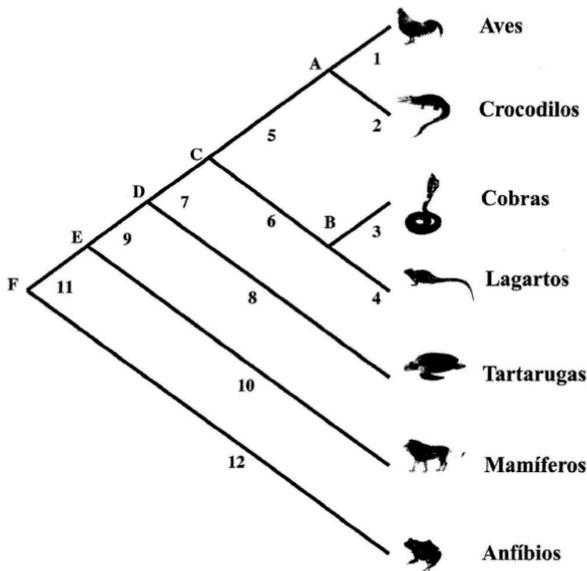
Fonte: Matiola e Fernandes (2012, p. 115).

Entretanto, dentre essas possíveis árvores, apenas uma é verdadeira – sendo necessários métodos matemático-estatísticos para qualquer tipo de análise que se deseja realizar. Quanto maior o número de árvores possíveis, maior é o tempo computacional para estimar as relações entre elas. Quando há muitas rotas evolutivas possíveis, a mais provavelmente correta é a mais curta, ou seja, aquela que assume o menor número de *passos evolutivos* (mudanças de estado dos caracteres).

Grupos que apresentam um ancestral comum mais próximo entre si que com qualquer outro grupo terminal são denominados *grupos-irmãos*. Na Figura 2.3, (Aves + Crocodilos) são grupos-irmãos; assim como (Cobras + Lagartos). Ainda, (Aves + Crocodilos)

é grupo-irmão de (Cobras + Lagartos) e assim por diante. Outro ponto importante de uma árvore filogenética é a *monofilia*. Grupos monofiléticos são grupos que incluem um ancestral comum exclusivo e todos seus descendentes, ou seja, esse ancestral não é ancestral de nenhum outro membro externo ao grupo – por exemplo, o grupo formado por (Mamíferos + Tartarugas + Lagartos + Cobras + Crocodilos + Aves), cujo ancestral comum mais recente é representado pelo nó interno E.

Figura 2.3 | Árvore filogenética enraizada



Fonte: Mاتيoli e Fernandes (2012, p. 114).

Por outro lado, se considerarmos um grupo "Reptilia", formado por (Tartarugas + Lagartos + Cobras + Crocodilos), não obteremos um grupo monofilético, já que um dos descendentes (Aves) da espécie ancestral desse grupo (nó interno D) constitui um grupo excluído dos répteis.

Grupos não monofiléticos são denominados *parafiléticos* ou *polifiléticos*. Um grupo parafilético é constituído por um ancestral comum e apenas parte dos seus descendentes, como o grupo "Reptilia", do qual o grupo (Aves) foi excluído. Já um grupo polifilético é constituído por dois ou mais grupos monofiléticos,

sem que sua espécie ancestral faça parte do grupo, como seria o caso dos “vertebrados homeotérmicos”, incluindo somente os grupos (Aves) e (Mamíferos).

Árvores filogenéticas podem ser construídas com base em características específicas, como a sequência de nucleotídeos nos genomas, um conjunto de detalhes morfológicos ou outras características mensuráveis que diferenciam as espécies em questão. Quaisquer características podem ser utilizadas, mas quanto maior o conjunto de dados, mais sólidos os resultados.

O papel da sistemática filogenética é de, além de descrever a diversidade biológica (função tradicional da taxonomia), organizar o conhecimento sobre essa diversidade, baseando-se no conhecimento das relações de parentesco entre grupos e no conhecimento da evolução de características morfológicas, fisiológicas, citogenéticas, moleculares, comportamentais e ecológicas dos grupos. Atualmente, a técnica é utilizada não apenas para relações históricas de parentesco entre os organismos, mas também como ferramenta preditiva em estudos epidemiológicos.

O ideal de árvores filogenéticas moleculares é que sejam baseadas em genes ou regiões genéticas ortólogas (derivados de um único gene existente no último ancestral comum das espécies comparadas), o que nem sempre é possível, especialmente quando as estruturas consideradas estão mais distantemente relacionadas. Ainda, as regiões em comparação podem ser similares por terem *funções* semelhantes e não origem comum (evolução convergente ou paralela).

Os métodos mais utilizados para a construção de árvores filogenéticas podem ser subdivididos em dois grupos: *caráter-estado* e *matrizes de distâncias*. A partir desses grupos principais, distinguem-se diferentes metodologias baseadas em soluções mais econômicas, como a *máxima parcimônia*, ou que utilizam cálculos probabilísticos, como a *máxima verossimilhança* e a *inferência bayesiana*. Em todo o caso, é necessária uma análise preliminar do conjunto de dados e suas respectivas singularidades para a decisão da metodologia a ser empregada.

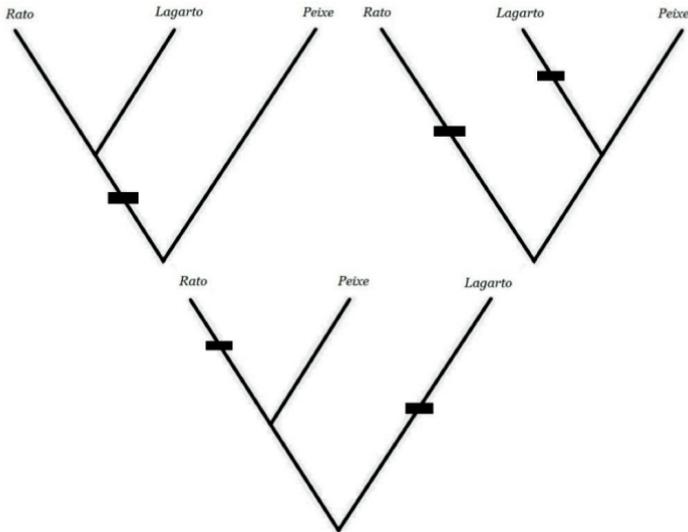


**Cárater-estado** → examina estados alternativos de uma série de características (por exemplo, a posição em uma sequência e o nucleotídeo que nela ocorre).

**Matrizes de distância** → primeiramente convertem as matrizes de características em matrizes de distâncias entre as unidades consideradas.

O método da **máxima parcimônia** se baseia em um modelo de evolução no qual uma mudança é mais provável que duas. Para melhor entendimento desse método, consideraremos três organismos: rato, lagarto e peixe. Em cada uma das topologias possíveis a seguir (Figura 2.4), o traço presente nos ramos indica o desenvolvimento de pulmões.

Figura 2.4 | Topologias possíveis relacionando três organismos



Fonte: Matiolli e Fernandes (2012, p. 119).

De acordo com os princípios da máxima parcimônia, a topologia escolhida seria a primeira, pois assumiria apenas uma mudança, enquanto as outras duas assumiriam duas

mudanças, ou seja, a primeira topologia representa a árvore mais parcimoniosa, requerendo o menor número de mudanças para explicar a característica "pulmão". Entretanto, mais de uma árvore pode apresentar o mesmo comprimento, dessa forma, todas as árvores mais parcimoniosas, com comprimento igual, deverão ser consideradas para a escolha do resultado final. Evidentemente, em algumas situações pode haver árvores mais longas – a questão das árvores serem mais curtas é probabilística, não uma oposição à natureza.

A principal vantagem desse método é sua maneira de interpretar os eventos evolutivos, a qual pode ser facilmente compreendida, já que escolhe o caminho "menos difícil" (aquele que exige menor número de suposições sobre eventos independentes). A desvantagem encontra-se quando o número de filogenias possíveis é muito grande, exigindo grande demanda computacional.

Quando duas espécies apresentam diferenças em relação a sequências homólogas (isto é, trechos de DNA herdados de um ancestral comum), tais diferenças originaram-se a partir de mutações por substituições ocorridas em linhagens que se diversificaram a partir de seu ancestral comum. A quantificação dessas diferenças pode ser expressa em valores ou em distâncias e utilizada para reconstrução filogenética.

Nos **métodos geométricos** (*métodos de distância*), primeiro é necessário o cálculo da distância propriamente dita e, posteriormente, a construção da topologia. O cálculo da distância é realizado estimando-se par a par o número de substituições ocorridas em cada uma das sequências a partir da divergência da sequência ancestral que originou as sequências comparadas, na qual cada 0,10 de distância significa 10% de resíduos que se alteraram em uma ou em outra sequência comparada. Uma das distâncias mais simples é a *distância p*, que constitui uma mera proporção de posições em que as duas sequências diferem.

Para os casos em que a distância  $p$  não pode ser utilizada, devido a taxas de substituição um pouco maiores que a ideal para o problema filogenético em questão, utiliza-se da *distância de Jukes-Cantor*, baseada na distribuição de Poisson e que leva em consideração substituições múltiplas. Entretanto, a distância de Jukes-Cantor é infinita e, portanto, incalculável, quando  $p$

é maior que 0,75. Nessa situação, o mais apropriado é procurar regiões genômicas mais conservadas para o estudo das relações filogenéticas do grupo em questão.

A distância de Jukes-Cantor é baseada em modelos de substituição, nos quais cada base ou aminoácido tem a mesma probabilidade de substituir qualquer outra base ou aminoácido. Uma vez que na prática essa condição muitas vezes não é satisfeita, utiliza-se da distância de *Kimura 2-parâmetros*, que leva em consideração um fator  $a$  mais que as substituições múltiplas, ou seja, a razão de transições e transversões ( $P/Q$ ). Transições são substituições de uma base purínica por outra purínica, assim como de uma base pirimídica por outra pirimídica, enquanto que transversões são substituições de uma base purínica por uma pirimídica e vice-versa.

A distância *Tamura 3-parâmetros*, por sua vez, agrupa todos os parâmetros previamente citados, ou seja, considera substituições múltiplas, taxa  $P/Q$  e conteúdo GC – parecendo, a princípio, ser a distância mais apropriada de todas. Entretanto, quanto maior o número de parâmetros, maior a variância associada; sendo esta distância aconselhada apenas para casos em que  $P/Q$  e conteúdo GC diferem bastante do esperado.

Uma vez construída a matriz de distâncias par a par, parte-se para a escolha do algoritmo de reconstrução da árvore propriamente dita. Da mesma forma que, para o cálculo de distâncias, existem diversos algoritmos para reconstrução de topologias. Vale ressaltar que a utilização de um determinado método para o cálculo das distâncias não influencia no algoritmo a ser empregado, mas pode influenciar nos tamanhos dos ramos dentro da topologia ou na própria topologia. Os tamanhos dos ramos podem ser calculados com base nas distâncias entre as sequências.

A máxima verossimilhança é um método que demanda um **modelo probabilístico** de evolução de caracteres e que pode levar em consideração: frequência de bases, taxa de substituição entre pares de nucleotídeos em uma sequência de DNA, proporção de sítios invariáveis e heterogeneidade de taxas de substituição entre sítios. Tais parâmetros são estimados diretamente a partir dos dados e utilizados no cálculo da probabilidade de um estado de caráter ser substituído por outro, de acordo com o modelo

evolutivo de escolha. A árvore final deverá ser aquela com a maior probabilidade de que os resultados tenham sido originados em conformidade com o modelo evolutivo escolhido (ou seja, aquela com maior verossimilhança).

Entretanto, os parâmetros considerados não são as topologias em si, mas os tamanhos dos ramos para cada topologia. Assim, a verossimilhança é maximizada variando os tamanhos dos ramos em cada topologia, mas, para a escolha da árvore mais verossimilhante, é preciso calcular a probabilidade relacionada a diferentes topologias e cada uma delas com as variações nos tamanhos dos ramos e demais parâmetros do modelo evolutivo. A árvore filogenética (topologia + comprimento dos ramos) que apresentar a maior verossimilhança será considerada a melhor estimativa histórica da evolução das sequências.

Os modelos que utilizam a máxima verossimilhança mais frequentemente empregados em reconstruções filogenéticas envolvem a utilização de parâmetros que regem a evolução das sequências. O *modelo de um tipo de substituição*, por exemplo, baseia-se em uma sequência gênica na qual os nucleotídeos A, C, G e T ocorrem em frequências iguais, e a probabilidade de substituição de um nucleotídeo para outro, em um intervalo de tempo, depende apenas da taxa de substituição. O *modelo de dois parâmetros*, por outro lado, baseia-se na premissa de que transições ocorrem mais frequentemente que transversões e, portanto, considera as taxas de transição  $\alpha$  separadamente das taxas de transversão  $\beta$ . Uma vez que a frequência dos quatro nucleotídeos não é similar (a cadeia leve do genoma mitocondrial de vertebrados, por exemplo, apresenta uma proporção de nucleotídeos G reduzida em relação aos demais nucleotídeos), o *modelo proporcional* foi desenvolvido para o cálculo da probabilidade de substituição, levando-se em conta a frequência desigual de bases. O *modelo HKY85* e o *modelo F84* foram desenvolvidos a fim de se distinguir diferenças nas taxas de transição e transversão, sendo que o modelo HKY85 também considera desigualdades na frequência de bases.

A variação na composição de bases não reflete diferenças apenas na taxa de transições e transversões, mas também na taxa de transições entre pirimidinas e transições entre purinas – sendo o *modelo TN93* empregado para adequar essa diferença a modelos

de evolução de sequências. Por fim, o *modelo geral de reversão ao longo do tempo* (GTR, do inglês *General Time Reversible*) considera a frequência desigual entre os nucleotídeos e as taxas de substituição diferentes entre os seis possíveis tipos de substituição reversíveis entre A, C, G e T.

Encontrar a árvore mais verossímil implica a averiguação de todas as topologias possíveis e das variações de comprimento dos ramos e valores dos parâmetros do modelo evolutivo para cada topologia, o que é inviável para um grande número de táxons, já que o número de possibilidades é exorbitante e o processo demandaria um tempo extremamente longo. Uma estratégia para essa questão foi desenvolvida com o intuito de reduzir o tempo gasto no cálculo de probabilidades das árvores, a denominada *Quebra-Cabeça de Quartetos* (em inglês, *Quartet Puzzling*).

Dessa forma, é possível examinar as relações entre todos os possíveis conjuntos de táxons. Para uma amostra com quatro táxons (A, B, C e D), por exemplo, três topologias são possíveis. A probabilidade de cada uma é examinada, e aquela que apresentar a maior probabilidade será escolhida. Feito isso para todos os possíveis quartetos da amostra, as melhores árvores obtidas para cada um dos quartetos são utilizadas em um "quebra-cabeça" com a intenção de encaixar as topologias encontradas em uma topologia global.

A inconveniência dessa estratégia é que ela pode não encontrar a árvore com a maior verossimilhança, mas, por outro lado, encontra uma árvore em um curto período de tempo e que pode ser utilizada como árvore inicial em métodos de rearranjos na busca pela mas também mais provável. Esse processo pode ser realizado por computador através de algoritmos que produzirão rearranjos em uma árvore inicial (no caso, a obtida através do "quebra-cabeça" de quartetos).

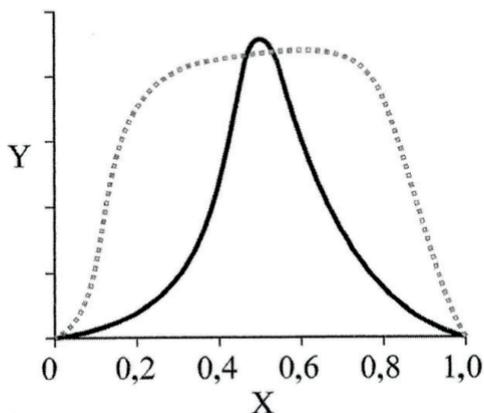
Novos algoritmos têm sido desenvolvidos para acelerar as buscas de verossimilhança, quando se trata de conjuntos de dados compostos por grande número de sequências. Implantados em programas computacionais, os algoritmos genéticos criam, em dado intervalo de tempo, uma amostra de soluções, na qual cada uma é composta de diversas variáveis (topologia, comprimento de ramos e parâmetros do modelo evolutivo).

Outros programas utilizam algoritmos que buscam o ponto mais alto nas curvas de distribuição das probabilidades, maximizando a verossimilhança dos dados a serem explicados por uma topologia, dado um modelo evolutivo. Dessa forma, é possível obter uma árvore inicial e estimar as variáveis de interesse, posteriormente “podando-se” um sub-ramo dessa árvore e o reinserindo em uma posição alternativa – assim, as variáveis da nova árvore são estimadas e comparadas com a inicial. Caso haja melhora considerável da verossimilhança, a nova topologia passará a ser considerada a melhor, e o processo será repetido determinadas vezes.

As vantagens da máxima verossimilhança em relação à máxima parcimônia e aos métodos geométricos são a menor variância (o método é menos afetado por erros de amostragem, mesmo com sequências curtas) e a tendência a uma maior robustez frente às violações do modelo evolutivo (uma vez que os processos evolutivos atuantes em determinado sítio ocorrem de maneira semelhante em muitos outros). A principal desvantagem encontra-se no grande tempo despendido para o cálculo das probabilidades, especialmente quando as análises envolvem um grande número de OTUs ou genomas completos. Nesses casos, a máxima parcimônia e os métodos geométricos tornam-se os únicos métodos possíveis de reconstrução filogenética para muitas sequências, devido à maior velocidade computacional que apresentam.

Uma alternativa aos métodos tradicionais vistos até agora é a **inferência bayesiana**, na qual informações a priori assumem diversas formas de como atribuir maior probabilidade a determinada topologia, assumir uma distribuição pré-delimitada para os valores possivelmente assumidos pelos parâmetros do modelo evolutivo ou restringir quais OTUs deverão ser mais relacionadas entre si, excluindo-se outras. Outra diferença entre os dois métodos é que a inferência bayesiana não estima a altura da probabilidade posterior, mas sim o volume (densidade) da superfície da probabilidade posterior. Tomemos como exemplo as probabilidades marginais de duas árvores obtidas através de inferência bayesiana, como mostrado na Figura a seguir (Figura 2.5).

Figura 2.5 | Árvores filogenéticas obtidas por inferência bayesiana



Fonte: Matioli e Fernandes (2012, p. 148).

Se levássemos em consideração a altura do pico, concluiríamos que a árvore 1 (linha reta) tem a maior probabilidade posterior. Entretanto, sua área sob a superfície da probabilidade posterior é muito menor em relação à árvore 2 (linha tracejada), a qual é favorecida, dessa maneira, pelo método bayesiano. Além disso, a árvore 2 é apoiada por um conjunto maior de valores para o parâmetro  $x$ , demonstrando que a estimativa de parâmetros contém incertezas e a árvore “verdadeira” pode ser próxima à árvore mais verossímil ou a com maior probabilidade posterior.

Ambas as estatísticas, clássica e bayesiana, utilizam probabilidades para a avaliação da confiança estatística da ocorrência de determinado evento. No entanto, a estatística bayesiana utiliza probabilidade como mensuração direta da incerteza associada com um parâmetro – somente dados observados são relevantes no cálculo da probabilidade, sendo irrelevantes eventos tão ou mais extremos que os observados. Ainda, a probabilidade posterior do método bayesiano reproduz a distribuição de valores do parâmetro de interesse e não a distribuição dos dados, como na estatística clássica. A inferência bayesiana tem sido bastante aplicada em diversas áreas da Biologia Evolutiva, especialmente em estudos envolvendo relógio molecular, seleção positiva e reconstrução de estados ancestrais.



Ao longo desta seção, conhecemos diversos métodos possíveis para a reconstrução de árvores filogenéticas e vimos que árvores baseadas em caracteres moleculares demandam um algoritmo de reconstrução aplicado a um determinado conjunto de sequências. Mas como **escolher o gene** mais apropriado para um dado problema filogenético?

Com a alta disponibilidade de sequências gênicas nas últimas décadas, dados moleculares tornaram-se uma importante fonte de informação para diversas áreas da Biologia e da Medicina. Mais especificamente na sistemática, esses dados podem ser utilizados para resolução de problemas filogenéticos envolvendo grandes grupos, filogenia dos grandes primatas e ordens de mamíferos, relações genealógicas das populações humanas, dentre outros.

Os dados utilizados em uma reconstrução filogenética molecular podem ser um conjunto de sequências de nucleotídeos ou de aminoácidos, uma vez que carregam a informação gênica entre gerações e, portanto, revelam a história evolutiva das linhagens. A decisão por um ou por outro deve ser tomada com base na taxa de evolução do gene e no tempo de divergência das espécies que serão estudadas (sequências de nucleotídeos apresentam taxa de evolução mais rápida que sequências de aminoácidos). Em todo o caso, a escolha do gene é muito mais importante que a escolha do método porque se a escolha daquele for cautelosa, qualquer método de reconstrução filogenética terá boas chances de encontrar a árvore verdadeira.

Um dos principais fatores a ser considerado é a *homologia*, ou seja, o ideal é que comparações entre diferentes espécies sejam a partir de caracteres homólogos (que descendem de um mesmo ancestral). Os processos que podem gerar genes homólogos são a duplicação gênica e a especiação. Genes originados por duplicação são denominados *parálogos*, enquanto que genes que passam a ter histórias evolutivas independentes, por especiação, são denominados *ortólogos*. Caso o problema em questão seja de eventos de duplicação em famílias ou superfamílias de genes, por exemplo, devem ser analisados genes parálogos de uma única

espécie. Em contrapartida, para reconstrução filogenética de grupos taxonômicos, devem ser escolhidos genes ortólogos.

Sequências consideradas homólogas devem ser *alinhas* a fim de que a posição de cada base comparada também seja homóloga. Devido a perdas ou ganhos de trechos ao longo dos genes, faz-se necessária a inserção de intervalos (hifens) nas sequências das espécies que perderam regiões ou das espécies que não ganharam tais intervalos.

Programas computacionais auxiliam nesse processo, alinhando, primeiramente, as sequências mais semelhantes, em seguida, as que se conectam às anteriormente alinhadas e, assim por diante, até que todas as sequências sejam alinhadas. O alinhamento final constitui um ótimo indicador da adequação do gene escolhido para o problema filogenético em questão.



## Assimile

No alinhamento mostrado na figura a seguir (A), observamos diversos indels (inserções/deleções), indicando uma evolução muito rápida do gene para escolhê-lo na reconstrução de uma filogenia. A solução, nesse caso, é escolher um gene mais conservado. Por outro lado, sequências muito conservadas (B) são pouco informativas para reconstruções filogenéticas.

Figura 2.6 | Reconstruções

```

                                     A
seq1  A-CCTCACCGACCG-----TATTCTCTAC----AAACCACAAAGATATTG
seq2  ATAATCAC-G-C-GC-GACTATT-TCT-CCAGACAGACAGACCACAA--A
seq3  ATGATCAC--ACCGCTG-CTAT-CT-TACGGACCACAAAGATATTGGAAC
seq4  AT-TTCACCGACTGCTGACT-TT-TC-A---ACCAC-----AAAG
seq5  A--TTCACTTCCGCTGAC-AT-CT--ACAAACCACA--GATATTGGAAC

                                     B
seq1  ATGTTCACCGACCGCTGACTATTCTCTACAAACCACAAAGATATTGGAAC
seq2  ATGTTCACCGACCGCTGACTATTCTCTACAGACCACAAAGATATTGGAAC
seq3  ATGTTCACCGACCGCTGACTATTCTCTACAAACCACAAAGATATTGGAAC
seq4  ATGTTCACCGACTGCTGACTATTCTCTACAAACCACAAAGATATTGGAAC
seq5  ATGTTCACCGACCGCTGACTATTCTCTACAAACCACAAAGATATTGGAAC
```

Fonte: Mاتيoli e Fernandes (2012, p. 161).

Além do alinhamento, outra questão importante é que a proporção de diferenças máximas entre os pares de sequências alinhadas seja de 5% a 40% entre os pares de uma sequência de nucleotídeos. O limite mínimo (5%) é necessário para que haja variação suficiente para resolver a filogenia, enquanto o limite máximo (40%) corresponde ao limite de variabilidade para um alinhamento confiável.

Em todos os casos, o ideal é que, antes do início do sequenciamento, uma pesquisa sobre o gene em questão seja realizada em bibliografias ou bancos de sequências, para uma avaliação inicial da variabilidade genética. Se a variabilidade for adequada para o nível da filogenia a ser estudada, continua-se o processo com o gene selecionado; caso contrário, se o gene for muito variável ou muito conservado, procura-se outro mais adequado.



### Pesquise mais

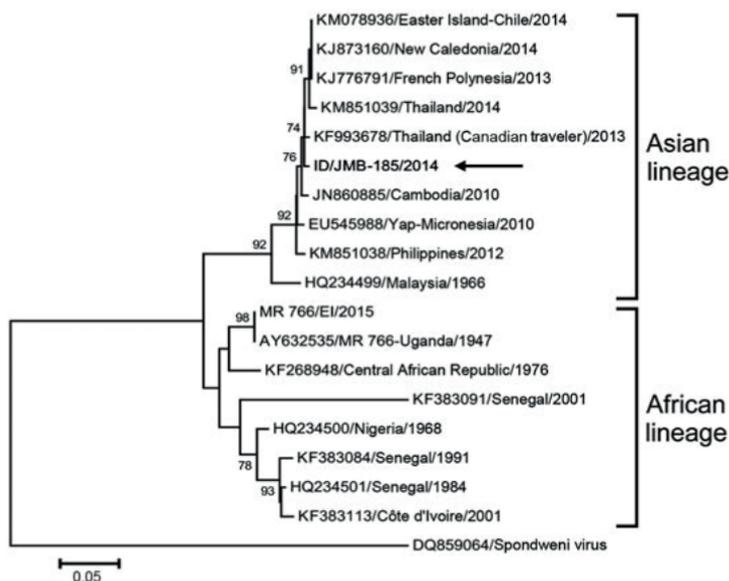
Quer saber mais a respeito dos métodos para construção de árvores filogenéticas e dos marcadores moleculares utilizados na filogenia? Faça a leitura do documento a seguir.

BUSO, G. S. C. **Marcadores moleculares e análise filogenética**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2005. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/bitstream/doc/187120/1/doc137.pdf>>. Acesso em: 27 jun. 2017.

## Sem medo de errar

Para chegar à conclusão de que o vírus causador da epidemia de zika no Brasil não pertencia à linhagem africana, mas sim à linhagem asiática, surgida mais recentemente, as relações de parentesco foram demonstradas através da **construção de uma árvore filogenética**, como a mostrada na figura a seguir.

Figura 2.7 | Construção de uma árvore filogenética



Fonte: <<https://wwwnc.cdc.gov/eid/images/15-1915-F1.jpg>>. Acesso em: 27 jun. 2017.

Sequências de DNA do vírus encontrado no país (ID/JMB-185/2014, destacado pela seta) foram comparadas a sequências depositadas no GenBank e, dessa forma, foi estabelecida a relação do vírus em questão com linhagens asiáticas.

## Avançando na prática

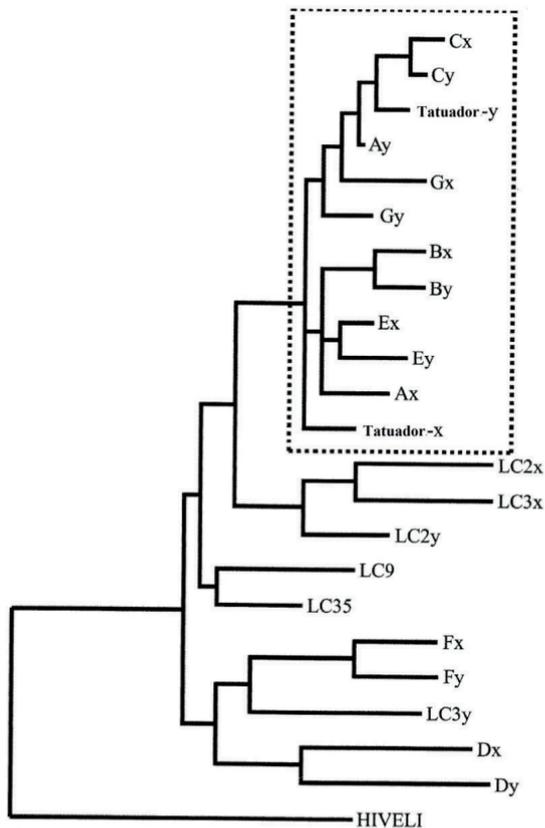
### Tatuagem viral

#### Descrição da situação-problema

Um tatuador de Campinas – São Paulo, portador do vírus HIV, sempre tomou todas as medidas de segurança cabíveis em seus atendimentos, mas há cerca de um mês, suspeitou ter contaminado alguns de seus clientes. Pesquisadores do Laboratório de Pesquisa em AIDS, pertencente ao Departamento de Clínica Médica da Faculdade de Ciências Médicas da Universidade Estadual de Campinas (UNICAMP), isolaram o vírus do tatuador, dos clientes suspeitos e de outras pessoas infectadas, não clientes do tatuador,

e sequenciaram alguns genes desses vírus. Utilizando o método da máxima verossimilhança, os pesquisadores construíram a árvore filogenética mostrada na figura a seguir.

Figura 2.8 | Árvore filogenética



Fonte: adaptada de Matioli e Fernandes (2012, p. 142).

Sabendo-se que as letras de A até G representam os clientes do tatuador; x e y representam duas linhagens diferentes de vírus sequenciadas; LC se refere às outras pessoas infectadas (não clientes do tatuador) e HIVELI constitui a sequência de uma linhagem africana utilizada como grupo externo, à qual conclusão os pesquisadores chegaram em relação à contaminação dos clientes do tatuador?

## Resolução da situação-problema

De acordo com a árvore filogenética obtida, é possível concluir que alguns dos clientes (A, B, C e E, destacados pelo retângulo na Figura 2.8) foram contaminados pelo tatuador. Entretanto, as sequências obtidas do vírus dos demais clientes (D e F) mostraram-se mais relacionadas às sequências do vírus dos não clientes do tatuador, indicando que a contaminação destes dois não aconteceu no estúdio de tatuagem.

### Faça valer a pena

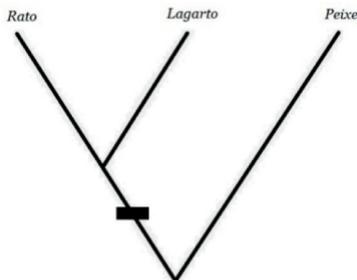
**1.** As relações de descendência entre as espécies podem ser representadas através de *árvores filogenéticas*, as quais podem ser enraizadas ou não. À medida que o número de unidades consideradas aumenta, o número de topologias possíveis (padrão de ramificação) também aumenta consideravelmente.

Qual é o número possível de árvores filogenéticas, levando-se em consideração quatro unidades/características?

- a) 5.
- b) 10.
- c) 15.
- d) 20.
- e) 25.

**2.** Um biólogo, em seu estudo sobre a evolução de grupos de animais, analisou as sequências gênicas de três organismos (rato, lagarto e peixe) e obteve, ao final de seu trabalho, a construção da árvore filogenética representada na figura a seguir.

Figura 2.9 | Árvore filogenética para três organismos



Fonte: Matiolli e Fernandes (2012, p. 119).

O traço horizontal, presente em um dos ramos da árvore filogenética obtida pelo biólogo, corresponde a uma:

- a) Ancestralidade.
- b) Extinção.
- c) Monofilia.
- d) OTU.
- e) Característica derivada.

**3.** Os métodos mais utilizados para a construção de árvores filogenéticas podem ser subdivididos em dois grupos: um que analisa diretamente os estados alternativos de uma série de características (*caráter-estado*) e outro que, primeiramente, converte esses padrões em *matrizes de distâncias*.

Assinale a alternativa que corresponda a um método para reconstrução filogenética pertencente ao grupo dos métodos geométricos.

- a) Máxima parcimônia.
- b) Máxima verossimilhança.
- c) Inferência bayesiana.
- d) Distância  $p$ .
- e) Método de dois parâmetros.

## Seção 2.3

### Reconstrução filogenética II

#### Diálogo aberto

Após o diagnóstico de hemofilia de Leticia, os pais da garotinha foram aconselhados a realizar exames genéticos para confirmação da condição de portadores do gene da doença, já que se trata de um distúrbio genético-hereditário ligado ao cromossomo X e, uma vez Leticia sendo hemofílica ( $XhXh$ ), ambos os pais devem conter o gene recessivo. O resultado do mapeamento genético deu positivo para a mãe de Leticia ( $XHXh$ ) e negativo para o pai ( $XHY$ ), gerando uma discussão familiar a respeito da paternidade da menina. Qual seria uma possível técnica de Biologia Molecular a ser utilizada na investigação da paternidade de Leticia? Em que se baseia essa técnica?

#### Não pode faltar

Diversas metodologias para a descrição da variabilidade genética foram desenvolvidas nos últimos anos, visto ser a caracterização e o entendimento da variabilidade genética intra e interpopulacional uma das principais questões ligadas à Biologia Evolutiva. Algumas variações fenotípicas com herança mendeliana clássica (polimorfismo dos grupos sanguíneos do sistema ABO, por exemplo) eram passíveis de serem estudadas, mas limitavam os organismos possíveis, os quais, além de possuírem os marcadores adequados, deveriam ser adaptados às condições laboratoriais para acompanhamento durante gerações.

Em 1966, a *eletroforese de isozimas* revolucionou a genética de populações, possibilitando acesso a um vasto número de locus (posição que um gene ocupa no cromossomo) em qualquer organismo dos mais variados ambientes e utilizando apenas uma amostra de tecido do organismo em questão. Atualmente, a técnica constitui uma importante ferramenta aplicável a questões evolutivas, por exemplo, demonstrando que diferentes genótipos

alozímicos (obtidos por alozimas, isto é, formas variantes de uma enzima codificada por diferentes alelos de um locus) podem apresentar diferentes valores adaptativos, sendo possível correlacioná-los com fatores ambientais.

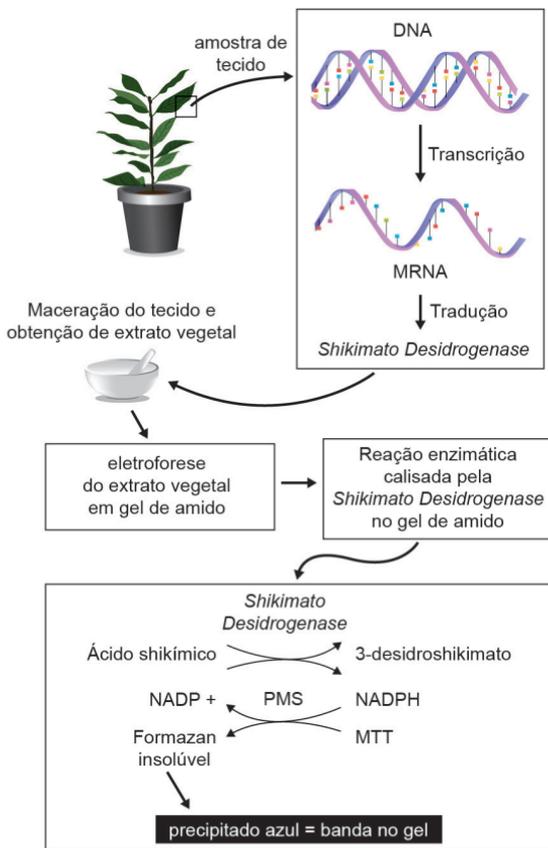
**Isozimas** são enzimas que diferem em relação à sequência de aminoácidos, mas que catalisam a mesma reação química. Essas enzimas vêm sendo utilizadas na Biologia Evolutiva através de duas formas: estudo minucioso de um locus e suas variantes alélicas (*alozimas*) e utilização de muitos locus para estudos populacionais. Na comparação entre espécies ou populações, utilizando-se polimorfismos enzimáticos, diversos locus são amostrados em cada indivíduo analisado, sendo que os métodos de análise tratam todos os locus como adaptativamente neutros – embora alguns polimorfismos sejam mantidos por forças seletivas (a maior parte da variabilidade observada é neutra), novas mutações podem ser fixadas por processos estocásticos e os polimorfismos observados são transitórios.

O princípio básico da eletroforese de isozimas consiste no fato de que pequenas mudanças na sequência do DNA causam mudanças na sequência de aminoácidos que, por sua vez, podem alterar a estrutura da proteína e, conseqüentemente, sua mobilidade eletroforética. Uma vantagem dessa técnica é a possibilidade de amostragem de vários indivíduos ao mesmo tempo e, assim, a obtenção de diversos locus analisados para cada um.

A extração das enzimas é feita a partir de tecidos, órgãos ou organismos inteiros como pequenos insetos, por exemplo, macerados em solução tampão e aplicados diretamente nos géis (amido ou poliacrilamida). A distância de migração das enzimas é observada por coloração histoquímica, que inclui a presença do substrato específico da enzima, coenzimas, cofatores e um corante capaz de precipitar, oxidar ou fluorescer diante da reação principal (Figura 2.10).



Figura 2.10 | Extração de enzimas para eletroforese de isoformas



Fonte: <<https://goo.gl/9V4GZq>>. Acesso em: 27 jun. 2017.

Os resultados da eletroforese são obtidos sob a forma de um zimograma (padrão de bandas) e, para sua interpretação, aplica-se o conhecimento da estrutura quaternária de enzimas (Figura 2.11).

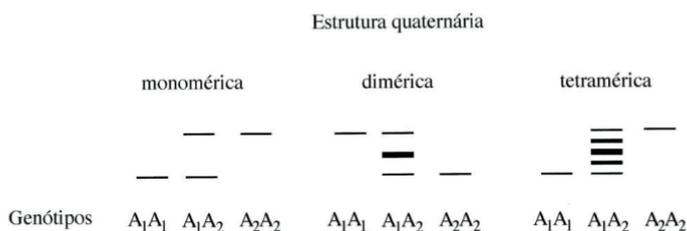
- Enzima monomérica → duas bandas no heterozigoto.
- Enzima dimérica → três bandas no heterozigoto.
- Enzima tetramérica → cinco bandas no heterozigoto.

Os géis, corados para diferentes sistemas enzimáticos, podem ser analisados através da interpretação genética das bandas encontradas e genótipos individuais são anotados. Assim, os dados podem ser aproveitados para a obtenção de todos os parâmetros que consideram frequências gênicas e genóticas, além de quase a totalidade das análises desenvolvidas para variação quantitativa.



## Exemplificando

Figura 2.11 | Estrutura quaternária de enzimas para interpretação de zimogramas



Fonte: Matioli e Fernandes (2012, p. 167).

O estudo de isozimas permite quantificar níveis de variabilidade genética, estimar fluxo gênico, elucidar limites interespecíficos, estabelecer relações evolutivas entre diferentes táxons, dentre outros, apresentando custo relativamente baixo, sendo adaptável para qualquer grupo de organismos e permitindo a análise de um amplo número de indivíduos e muitos locus de cada amostra. Em contrapartida, os níveis de variabilidade obtidos podem ser subestimados, já que mutações no material genético nem sempre alteram a estrutura da proteína e nem todas as mudanças na sequência de aminoácidos provocam alterações na mobilidade eletroforética.

Outro ponto importante é que a variabilidade difere entre os táxons superiores e os índices de variabilidade podem ser alterados dependendo dos locus considerados nos estudos comparativos, sendo alguns mais polimórficos que outros. Além disso, a variabilidade gênica parece estar associada à heterogeneidade ambiental: espécies generalistas, com vasta distribuição geográfica, ampla mobilidade e exploradoras de diversos nichos ecológicos, apresentam maior variabilidade que espécies especialistas, com distribuição geográfica restrita e baixa mobilidade.

Ainda, dados de variabilidade podem fornecer informações sobre eventos históricos que influenciaram a estrutura genética das populações. Baixos índices de heterozigiosidade nos locus enzimáticos, por exemplo, podem indicar eventos recentes de afunilamento populacional, principalmente quando espécies relacionadas apresentam níveis muito mais altos de variabilidade. Estudos alozímicos também podem fornecer informações importantes à inferência de padrões de fluxo gênico e estruturação de acasalamentos, através da análise da distribuição geográfica de frequências alozímicas, além de dados sobre variação geográfica surgida por mecanismos evolutivos locais, que envolvem interações de mecanismos genéticos e processos ambientais.

Distâncias genéticas são medidas estatísticas relativas à diferenciação entre populações, utilizando-se dados de frequências genotípicas. Frequências gênicas e genotípicas de indivíduos tomados ao acaso em populações naturais, obtidas através da eletroforese de enzimas, podem estimar a diferenciação intra e interespecífica das populações em questão. Essas frequências podem ser transformadas em índices que permitem a estimativa do grau de distância genética ou de similaridade entre espécies e populações distintas. Além do mais, variantes alélicas encontradas isoladas em uma população ou táxon podem ser utilizadas como um meio de identificação em casos de espécies de difícil identificação morfológica.

O emprego da eletroforese de isozimas em pesquisas de caracterização genética de populações apresentou um declínio nos últimos anos devido à popularização e diminuição dos custos de técnicas de análise de DNA (como polimorfismos de comprimento de fragmento de restrição (RFLP) e sequenciamento direto, aplicados conjuntamente com a técnica de reação em cadeia da polimerase (PCR)); mas a técnica ainda ganha destaque por revelar padrões únicos, quando comparados a métodos baseados na tecnologia de DNA.



**Refleta**

Qual é o marcador molecular mais adequado para estudos de padrões de variabilidade genética em populações naturais?

Evidências da seleção natural, por exemplo, parecem ser mais relevantes em marcadores alozímicos que em marcadores de DNA nuclear e mitocondrial, embora a variabilidade do DNA nuclear seja mais homogênea. A escolha do marcador molecular mais adequado para análises de variabilidade genética em populações naturais é uma das grandes dificuldades enfrentadas pelos pesquisadores, visto que essa escolha, muitas vezes, é influenciada pela disponibilidade de recursos e experiências pessoais prévias.

Outra questão é a maior potencialidade do DNA nuclear para estudos de polimorfismos quando comparado à eletroforese de alozimas – entretanto, a maioria dos marcadores empregados nos métodos baseados na tecnologia do DNA é anônima e não tem funções bioquímicas e fisiológicas conhecidas, o que não acontece com as isozimas, que continuam sendo uma das melhores opções para estudos de espécies próximas e populações conspecíficas (pertencentes à mesma espécie).



### Pesquise mais

Aprofunde seus conhecimentos sobre marcadores moleculares na análise de variabilidade genética de populações através da leitura do artigo a seguir.

CARNEIRO, P. L. S. et al. Oscilação genética em populações submetidas a métodos de seleção tradicionais e associados a marcadores moleculares. **Revista Brasileira de Zootecnia**, Viçosa, v. 35, n. 1, p. 84-91, jan./fev. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-35982006000100010](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-35982006000100010)>. Acesso em: 27 jun. 2017.

Marcadores alozímicos, indiscutivelmente, representaram um grande passo na compreensão de processos micro e macroevolutivos, uma vez que não são profundamente afetados pela interação com o ambiente. No entanto, para estudos de estrutura populacional e outras aplicações para análises intraespecíficas são necessários níveis suficientes de variabilidade, e as análises de isozimas não se apresentam suficientemente variáveis em alguns organismos.

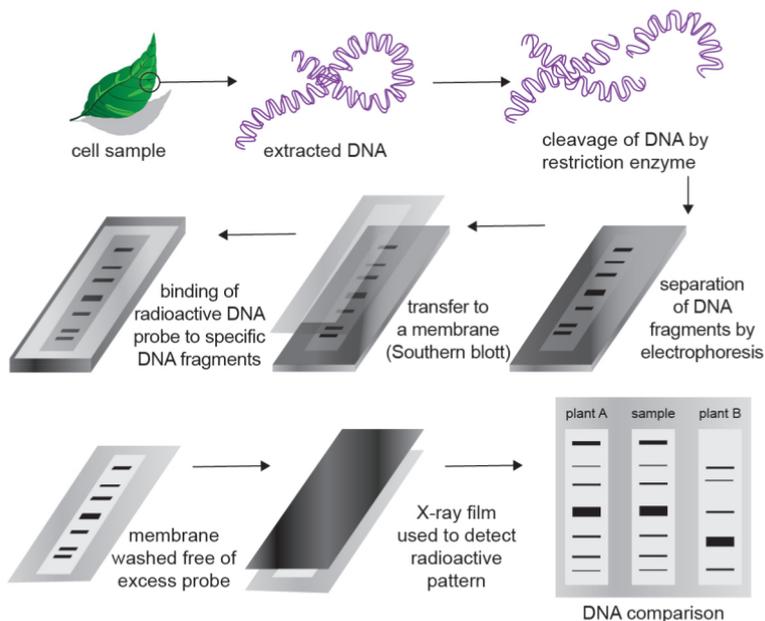
No final da década de 1960, grupos de pesquisadores isolaram e purificaram enzimas de diversas linhagens de bactérias capazes de digerir sequências específicas de DNA dupla fita contendo quatro, cinco ou seis pares de bases de comprimento. Essas enzimas, denominadas **enzimas de restrição** (*endonucleases*), protegem as células bacterianas contra a invasão de DNA exógeno e, através de um sistema específico de metilação, o DNA da própria bactéria é protegido.

Enzimas de restrição podem ser utilizadas como marcadores para a detecção de polimorfismos no DNA. Esses marcadores, denominados RFLP (do inglês, *Restriction Fragment Length Polymorphism*), consistem no padrão de fragmentos produzidos pelas endonucleases ao digerirem um determinado material genético. Os RFLP têm sido amplamente utilizados na medicina forense, na investigação de vínculo genético (determinação de paternidade) e no diagnóstico de doenças hereditárias, apresentando diversas vantagens em comparação a marcadores morfológicos, citológicos ou alozímicos: a) são herdados como marcadores mendelianos livres de efeitos pleiotrópicos (um só gene determina mais de uma característica fenotípica), b) não são afetados pelo ambiente, c) podem ser obtidos em elevado número, e d) possuem distribuição aleatória no genoma.

A técnica de RFLP consiste na clivagem (corte) de moléculas de DNA pela ação de enzimas de restrição, separação dos fragmentos gerados por eletroforese e visualização dos fragmentos em forma de bandas, na qual cada uma corresponde a um grupo de moléculas com mesmo tamanho (Figura 2.12). Diferentes indivíduos apresentam padrões diferentes de tamanho de fragmentos, sendo essa variabilidade resultado de inserções, deleções, rearranjos ou substituições de base na sequência de nucleotídeos. Essas alterações, na sequência de nucleotídeos, são reconhecidas por determinadas endonucleases, alterando o número de sítios de clivagem e, conseqüentemente, o tamanho dos fragmentos.



Figura 2.12 | Técnica de RFLP



Fonte: <<https://goo.gl/Gq7UGd>>. Acesso em: 27 jun. 2017.

Pelo fato de o “agente revelador” (mais comumente, brometo de etídeo) se intercalar em moléculas de DNA, fragmentos maiores apresentam bandas mais intensas, tornando possível uma correlação direta entre a intensidade da banda e o tamanho do fragmento com a quantidade de DNA. Outra metodologia utiliza uma substância radioativa (fósforo ou enxofre) para marcar as moléculas de DNA antes de serem fragmentadas – mas, nesse caso, a intensidade das bandas não apresenta relação direta com o tamanho dos fragmentos, os quais são detectados através da exposição do gel a um filme de raios X.

Nos casos em que a quantidade de material genético é baixa, por exemplo, devido à má preservação do material ou à pouca quantidade de tecido disponível para a extração do DNA, emprega-se a técnica de *Southern blot*, cujo diferencial está na transferência

dos fragmentos para uma membrana de náilon ou nitrocelulose, após a separação em gel de agarose. A membrana é hibridizada com uma sonda de DNA fria ou radioativa, e as condições de hibridação são definidas com base no grau de similaridade entre o DNA da amostra e o DNA da sonda. Se a sonda utilizada for fria, a revelação acontece por imunoensaio e há a precipitação de corante na própria membrana, enquanto se for radioativa, ocorre por exposição da membrana a um filme de raios X.

O DNA de cada organismo apresenta um padrão de fragmentos específicos, denominado *perfil de digestão*, sendo que, entre várias amostras, podem ser detectados perfis invariáveis (padrões de fragmentos idênticos para uma mesma enzima de restrição) ou perfis variáveis (padrões de bandas diferentes entre organismos). Uma mesma endonuclease pode gerar diferenças nos perfis de digestão entre indivíduos, causadas por alterações na quantidade e na distribuição dos sítios de restrição. Cada enzima de restrição tem um sítio específico de reconhecimento, no qual é realizada a clivagem, sendo, geralmente, sequências de nucleotídeos simétricas, contendo de quatro a seis pares de bases.



### Exemplificando

A enzima *Eco* RI reconhece uma sequência de seis pares de base e, ao cortá-la, gera extremidades coesivas:

Figura 2.13 | Clivagem da enzima *Eco* RI



Fonte: adaptada de <<https://goo.gl/qTCkvC>>. Acesso em: 27 jun. 2017.

A enzima *Rsa* I, por sua vez, reconhece uma sequência de quatro pares de base e a corta ao meio (setas), gerando pontas retas.

Figura 2.14 | Clivagem da enzima Rsa I



Fonte: adaptada de <<https://goo.gl/2rVYVw>>. Acesso em: 27 jun. 2017.

O número de fragmentos gerados e seus respectivos tamanhos podem ser alterados devido à substituição de bases dentro dos sítios de reconhecimento ou a alterações no DNA causadas por adições ou deleções de sequências (indels), criando ou eliminando sítios de restrição em uma determinada sequência para uma enzima em particular. Esses rearranjos de sequência podem alterar o padrão de fragmentos para várias enzimas simultaneamente. O conhecimento de sítios de restrição é bastante útil na caracterização do DNA, pois fornece pontos específicos, cujas distâncias entre eles são conhecidas. Esse mapeamento auxilia na localização de genes, na clonagem de regiões de interesse em vetores moleculares e funciona como marcador em estudos evolutivos.

Diversos genes são utilizados para inferências filogenéticas e para diagnóstico e prevenção de algumas doenças. Uma das sequências mais estudadas através de RFLP são as *sequências de cópia simples* (DNAsc<sub>n</sub> – *single copy nuclear DNA*), além de sequências repetidas. Regiões não codificantes (íntrons) também vêm sendo analisadas, por serem mais variáveis que as codificantes.

### Sem medo de errar

O resultado do mapeamento genético gerou uma discussão familiar a respeito da paternidade de Letícia porque sua mãe é portadora do gene da hemofilia ( $X^H X^h$ ) e seu pai não o é ( $X^H Y$ ). Se Letícia fosse filha dele, seria normal ( $X^H X^H$ ) ou apenas portadora ( $X^H X^h$ ) e não hemofílica ( $X^h X^h$ ). A paternidade de Letícia poderia ser confirmada através da **técnica de RFLP**, que se baseia na análise de

polimorfismos e consiste na clivagem de moléculas de DNA pela ação de enzimas de restrição, seguida da separação dos fragmentos gerados por eletroforese e visualização dos fragmentos em forma de bandas. Cada pessoa apresenta pequenas mutações na sequência de DNA que a diferem das outras pessoas e que alteram o tamanho dos fragmentos clivados pelas enzimas de restrição, dessa maneira, indivíduos podem ser diferenciados pelo comprimento desses fragmentos. Da mesma forma, cada indivíduo herda algumas dessas sequências (polimorfismos) do pai e algumas da mãe, possibilitando, dessa forma, a investigação de vínculo genético.

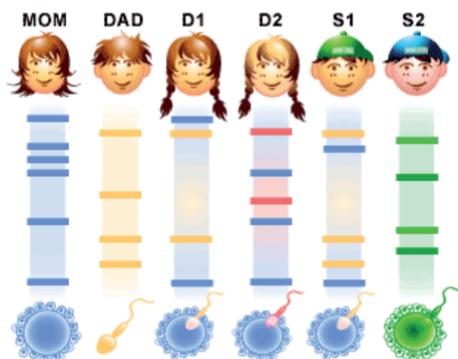
## Avançando na prática

### Quem é quem?

#### Descrição da situação-problema

Um estagiário da Clínica Genética Médica, em Campinas – SP, pediu a seu supervisor que o deixasse tentar solucionar um caso de investigação de vínculo genético por si próprio, a fim de testar seus conhecimentos. O supervisor concordou e repassou ao estagiário o caso de uma família composta pelos pais e quatro filhos (dois meninos e duas meninas), dentre os quais dois são filhos do casal, um é filho somente da mãe (fruto de seu casamento anterior) e um é adotado. Analisando o resultado da técnica de RFLP empregada no caso (Figura 2.15), a que conclusões o estagiário chegou a respeito de quem é cada filho do casal?

Figura 2.15 | Resultado da técnica de RFLP



Fonte: <<https://goo.gl/4NHhsq>>. Acesso em: 27 jun. 2017.

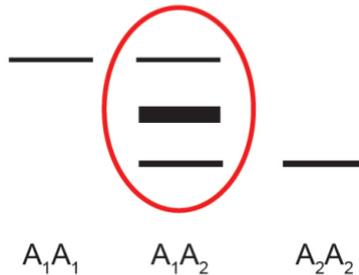
## Resolução da situação-problema

A partir do resultado da técnica de RFLP, é possível observar que a filha D1 compartilha polimorfismos tanto com a mãe (bandas em azul) quanto com o pai (bandas em amarelo), assim como o filho S1 – indicando que ambos são os filhos do casal. A filha D2 compartilha polimorfismos somente com a mãe (bandas em azul), indicando ser a filha de seu casamento anterior. Por fim, o filho S2 não compartilha polimorfismos com nenhum dos pais (bandas em verde), indicando ser o filho adotado pelo casal.

### Faça valer a pena

**1.** Os resultados da eletroforese de isozimas são obtidos sob a forma de um *zimograma* (padrão de bandas). Os géis, corados para diferentes sistemas enzimáticos, podem ser analisados através da interpretação genética das bandas encontradas, e genótipos individuais são anotados. A isozima identificada na figura a seguir é classificada como uma enzima:

Figura 2.16 | Zimograma da isozima identificada



Fonte: Mاتيoli e Fernandes (2012, p. 167).

- a) Monomérica.
- b) Dimérica.
- c) Trimérica.
- d) Tetramérica.
- e) Pentamérica.

**2.** Em 1968, pesquisadores descobriram uma classe especial de enzimas, denominadas *enzimas de restrição* (endonucleases), que protegem as células bacterianas contra a invasão de DNA exógeno, bem como o DNA da própria bactéria. Além dessa proteção, enzimas de restrição são capazes de clivar (cortar) sequências de DNA, as quais podem ser utilizadas como marcadores para a detecção de polimorfismos no DNA. Assinale a alternativa que corresponda à denominação desses marcadores.

- a) Minissatélites.
- b) Microssatélites.
- c) VNTR.
- d) RFLP.
- e) DNAscñ.

**3.** O DNA de cada organismo apresenta um padrão de fragmentos específicos, sendo que, entre várias amostras, podem ser detectados padrões de fragmentos idênticos para uma mesma enzima de restrição ou padrões de bandas diferentes entre organismos.

O padrão de fragmentos específicos abordado no texto anterior é denominado:

- a) Padrão de herança.
- b) Padrão de sequências.
- c) Perfil de digestão.
- d) Heredograma.
- e) Padrão ouro.

# Referências

BIOTECH. **Técnica de RFLP**. 2017. Disponível em: <<http://www.biotech.ubc.ca/MolecularBiology/DNAfingerprintherbs/rflp.gif>>. Acesso em: 27 jun. 2017.

BUSO, G. S. C. **Marcadores moleculares e análise filogenética**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2005. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/bitstream/doc/187120/1/doc137.pdf>>. Acesso em: 27 jun. 2017.

CARNEIRO, P. L. S. et al. Oscilação genética em populações submetidas a métodos de seleção tradicionais e associados a marcadores moleculares. **Revista Brasileira de Zootecnia**, Viçosa, v. 35, n. 1, p. 84-91, jan./fev. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-35982006000100010](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-35982006000100010)>. Acesso em: 27 jun. 2017.

CENTERS FOR DISEASE CONTROL AND PREVENTION. **Construção de uma árvore filogenética**. 2017. Disponível em: <<https://wwwnc.cdc.gov/eid/images/15-1915-F1.jpg>>. Acesso em: 27 jun. 2017.

COX, M. M.; DOUDNA, J. A.; O'DONNELL, M. Genomas, transcriptomas e proteomas. In: **Biologia Molecular: Princípios e técnicas**. Porto Alegre: Artmed, 2012. p. 259-277.

\_\_\_\_\_. Genomas, transcriptomas e proteomas. In: \_\_\_\_\_. Porto Alegre: Artmed, 2012. p. 282-291.

CUNHA, C. Interação entre genes e meio ambiente. In: \_\_\_\_\_. **Genética e Evolução Humana**. Campinas: Átomo, 2011. p. 171-183.

FREEMAN, S.; HERRON, J. C. Estimando árvores evolutivas. In: \_\_\_\_\_. **Análise Evolutiva**. 4. ed. Porto Alegre: Artmed, 2009. p. 111-137.

FUTUYMA, D. J. A árvore da vida: classificação e filogenia. In: \_\_\_\_\_. **Biologia Evolutiva**. 3. ed. Ribeirão Preto: FUNPEC Editora, 2009. p. 87-125.

\_\_\_\_\_. Variação. In: \_\_\_\_\_. 3. ed. Ribeirão Preto: FUNPEC Editora, 2009. p. 239-247.

GENÉTICA VIRTUAL. **Extração de enzimas para eletroforese de isozimas**. 2017. Disponível em: <<http://files.geneticavirtual.webnode.com.br/200000076-c266dc45a7/Digitalizar0005%201.jpg>>. Acesso em: 27 jun. 2017.

GRIFFITHS, A. J. F. et al. Mapeamento de cromossomos eucarióticos por recombinação. In: **Introdução à Genética**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. p. 116-120.

\_\_\_\_\_. Genética de populações. In: \_\_\_\_\_. 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. p. 552-555.

LEWIS, R. Mudanças de frequências alélicas. In: \_\_\_\_\_. **Genética Humana: Conceitos e aplicações**. 5. ed. Rio de Janeiro: Guanabara Koogan, 2004. p. 282-283.

LODISH, H. et al. Técnicas de genética molecular e genômica. In: \_\_\_\_\_. **Biologia Celular e Molecular**. 5. ed. Porto Alegre: Artmed, 2005. p. 390-395.

MATIOLI, S. R.; FERNANDES, F. M. C. **Biologia molecular e evolução**. 2. ed. Ribeirão Preto: Holos, 2012.

MULLER, H. R.; PRADO, K. B. Epigenética: um novo campo da genética. **RUBS**, Curitiba, v. 1, n. 3, p. 61-69, set./dez. 2008. Disponível em: <[http://www.colegiogregormendel.com.br/gm\\_colegio/pdf/2012/textos/3ano/biologia/8.pdf](http://www.colegiogregormendel.com.br/gm_colegio/pdf/2012/textos/3ano/biologia/8.pdf)>. Acesso em: 27 jun. 2017.

PASTERNAK, J. J. Mapeamento físico e genético do genoma humano. In: \_\_\_\_\_. **Genética Molecular Humana: mecanismos das doenças hereditárias**. São Paulo: Manole, 2002. p. 159-180.

RIDLEY, M. A reconstrução da filogenia. In: \_\_\_\_\_. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. p. 447-493.

SALZANO, F. M. Evolução molecular. In: \_\_\_\_\_. **Genômica e evolução: moléculas, organismos e sociedades**. São Paulo: Oficina de Textos, 2012. p. 72-74.

SCIENCE CREATIVE QUARTERLY. **Resultado da técnica de RFLP**. 2017. Disponível em: <<https://i0.wp.com/www.scq.ubc.ca/wp-content/DNAfingerprintfamily.gif?w=900>>. Acesso em: 27 jun. 2017.

SIGMA-ALDRICH. **Enzima Eco RI**. 2017. Disponível em: <[http://www.sigmaaldrich.com/content/dam/sigma-aldrich/product5/192/ecor-i.eps/\\_jcr\\_content/renditions/ecor-i-medium.jpg](http://www.sigmaaldrich.com/content/dam/sigma-aldrich/product5/192/ecor-i.eps/_jcr_content/renditions/ecor-i-medium.jpg)>. Acesso em: 27 jun. 2017.

\_\_\_\_\_. **Enzima Rsa I**. 2017. Disponível em: <[https://tools.thermofisher.com/content/sfs/gallery/high/RsaI\\_FD\\_recognition.jpg](https://tools.thermofisher.com/content/sfs/gallery/high/RsaI_FD_recognition.jpg)>. Acesso em: 27 jun. 2017.

UFRGS. **Tipos de árvores filogenéticas**. 2017. Disponível em: <<https://www.ufrgs.br/bioinfo/wp-content/uploads/2014/03/figura4-5.png>>. Acesso em: 27 jun. 2017.

# Mutação, reparo e regulação

## Convite ao estudo

Prezado aluno,

Daremos continuidade à compreensão da Biologia Molecular e Evolução, estudando as mutações e as maneiras pelas quais o DNA é reparado. Seus objetivos, ao final da leitura desta unidade, são: compreender os principais erros de replicação e os mecanismos de reparo do DNA; bem como a regulação gênica e a regulação pós-transcricional. Para compreender o assunto e atingir as competências e os objetivos da disciplina, analisaremos uma situação hipotética que se aproxima dos conteúdos teóricos que serão vistos por você nesta unidade: após a revelação da paternidade de Letícia, o casal ficou afastado por 10 meses. Letícia conheceu seu pai biológico e ficou bastante empolgada com sua irmã mais nova por parte de pai, mas continuou convivendo frequentemente com seu pai de criação. O casal acabou se reconciliando; mas superada a crise no casamento, a família se viu diante de outro problema: Marcos, pai adotivo de Letícia, começou a apresentar fortes dores abdominais, formigamento nas mãos e nos pés, dores articulares, fraqueza muscular, irritabilidade, agressividade, declínio cognitivo (memória, atenção e aprendizagem), perda de peso, vômito e alucinações, que evoluíram para delírios, convulsões, paralisias e coma, deixando-o hospitalizado por dois meses.

No decorrer desta unidade de ensino, conheceremos mais a respeito de mutação, reparo de DNA e regulação gênica. Na Seção 3.1, conheceremos alguns agentes mutagênicos e os erros de replicação por eles ocasionados. Já na Seção 3.2,

compreenderemos os mecanismos pelos quais o material genético consegue minimizar ou eliminar estes erros. Por fim, na Seção 3.3, trataremos das estratégias de regulação da transcrição e regulação pós-transcricional.

# Seção 3.1

## Erros de replicação

### Diálogo aberto

Durante sua anamnese no hospital, Marcos relatou ao médico de plantão que trabalhava há 15 anos como soldador em uma fábrica de tubos de ferro. Questionado pelo médico se utilizava os devidos equipamentos de proteção (roupas de manga comprida, avental, máscara de solda, touca, luvas, óculos de proteção e protetores auriculares), Marcos revelou que nem sempre utilizava a máscara e os protetores auriculares, porque dificultavam sua comunicação com os demais colegas durante o trabalho. O que este descuido de Marcos pode ter provocado? Qual é a relação entre a profissão de Marcos e os sintomas apresentados por ele?

### Não pode faltar

Por ser muito extenso (cerca de três bilhões de pares de bases), o genoma humano não é muito estável, sofrendo alterações durante sua síntese ou sendo alvo da ação de agentes físico-químicos provenientes do meio externo ou de produtos do próprio metabolismo. Os genes são constituídos pelas bases nitrogenadas adenina, timina, citosina e guanina, cujas alterações nas sequências dessas bases originam variações do mesmo gene, os *alelos*.

As alterações sofridas pelo DNA, denominadas mutações, podem ser induzidas ou espontâneas. Mutações induzidas são resultado da ação de agentes físico-químicos denominados **agentes mutagênicos**, enquanto mutações espontâneas resultam de agentes mutagênicos presentes no ambiente celular (incluindo aqueles provenientes do próprio metabolismo da célula) ou de erros ocorridos durante a replicação do DNA (independentes da ação de agentes mutagênicos).



Os principais agentes mutagênicos são compostos presentes na dieta e em determinadas drogas e medicamentos, cigarros e bebidas alcoólicas, além de fatores ambientais, como radiação solar, agrotóxicos e efluentes industriais.

Para saber mais sobre os efeitos de cada um desses fatores, leia o artigo a seguir.

DÜSMAN, E. et al. Principais agentes mutagênicos e carcinogênicos de exposição humana. **SaBios: Rev. Saúde e Biol.**, v. 7, n. 2, p. 66-81, 2012. Disponível em: <<http://revista.grupointegrado.br/revista/index.php/sabios2/article/view/943/438>>. Acesso em: 5 ago. 2017.

As mutações são classificadas em:

- Somáticas → ocorrem em qualquer célula do organismo, exceto células germinativas.
- Germinativas → ocorrem nos gametas e são transmitidas aos descendentes.
- Autossômicas → ocorrem em genes localizados nos cromossomos autossômicos.
- Ligadas ao X → afetam genes localizados no cromossomo X.
- Com sentido trocado (*missense mutation*) → troca de um nucleotídeo na região codificadora de um gene, resultando na criação de um novo códon que, por sua vez, corresponderá a um aminoácido diferente.
- Sem sentido (*nonsense mutation*) → substituição do códon original por um códon de parada, resultando na terminação precoce da proteína traduzida.
- Silenciosas → alteram o códon, mas não resultam na troca do aminoácido correspondente na proteína.
- Transições → substituição de purina por purina ou de pirimidina por pirimidina.
- Transversões → substituição de purina por pirimidina ou vice-versa.
- Modificadoras da fase de leitura (*frameshift mutation*) →

adição ou deleção de um ou alguns pares de bases não múltiplos de três, alterando a fase de leitura de todas as trincas posteriores ao sítio mutado.

- Perda de função → reduz ou elimina a função do produto gênico.
- Nulas → resultam na perda total da função do produto gênico.
- Ganho de função → resulta em um produto gênico com função aumentada ou completamente nova.
- Neutras → ocorrem em regiões desprovidas de genes e não afetam produtos gênicos ou a expressão dos genes.
- Visíveis → afetam uma característica morfológica (alteram fenótipo).
- Nutricionais → resultam na perda da capacidade de sintetizar um aminoácido ou uma vitamina.
- Bioquímicas → resultam em quebras nas vias bioquímicas de transformação enzimática.
- Comportamentais → afetam os padrões de comportamento de um organismo.
- Reguladoras → afetam a regulação da expressão de um gene, ativando ou suprimindo sua expressão de maneira inadequada.
- Letais → capazes de interromper um processo essencial à sobrevivência do organismo, podendo ser condicionais restritivas (letais em um determinado ambiente) ou condicionais permissivas (não letais em outro ambiente).

Para que ocorra uma mutação, é necessário que haja erros na incorporação dos nucleotídeos durante a replicação do DNA e que a lesão em um nucleotídeo, após um ciclo replicativo, não seja reparada. As mutações mais simples, denominadas mutações pontuais, envolvem a alteração de um único nucleotídeo (substituição de bases) ou a inserção ou deleção de um nucleotídeo ou de um número reduzido de nucleotídeos (mutações indel), sendo resultado do mau funcionamento do sistema de replicação e reparo do DNA.

Outro processo que pode provocar mutações envolve a formação de lesões no DNA, que fazem que uma base nitrogenada se emparelhe com uma base não correspondente (por exemplo, uma lesão na guanina faz que esta se emparelhe com uma timina

e não com uma citosina). Entretanto, uma lesão, por si só, não configura uma mutação; a lesão somente provocará uma mutação se, após um ciclo replicativo, ela não for reparada. Ainda, nem todas as lesões são mutagênicas, uma vez que podem gerar um produto que ainda tende a se emparelhar com a base nitrogenada correspondente.



### Refleta

A alteração da sequência de nucleotídeos de um gene sempre altera a sequência de aminoácidos de uma proteína?

Transições e transversões podem alterar a sequência de nucleotídeos de um gene sem alterar a sequência de aminoácidos de uma proteína, uma vez que o código genético é degenerado – ou seja, mais de um códon pode codificar para um mesmo aminoácido (Figura 3.1). Uma mutação em um gene com o códon AGG, envolvendo a substituição da adenina por citosina, por exemplo, não provocará a alteração na sequência de aminoácidos da proteína codificada por este gene (*mutação silenciosa*).



### Exemplificando

Figura 3.1 | Aminoácidos e seus respectivos códons codificantes

AMINOÁCIDO	CÓDONS
alanina	GCU, GCC, GCA, GCG
arginina	CGU, CGC, CGA, CGG, AGA, AGG
aspártico	GAU, GAC
fenilalanina	UUU, UUC
leucina	UUA, UUG, CUU, CUC, CUA, CUG
lisina	AAA, AAG
metionina e códon de iniciação	AUG
serina	UCU, UCC, UCA, UCG, AGU, AGC
tirosina	UAU, UAC
triptofano	UGG

Fonte: <<https://goo.gl/8mykyJ>>. Acesso em: 5 jun. 2017.

Por outro lado, algumas substituições podem ocasionar o surgimento precoce de um códon de terminação (*stop códon*); sendo estes: UAA, UAG e UGA. Esta mutação, como por exemplo, a substituição da citosina do códon CAG por uma timina – gerando o stop códon UAG e codificando para o término da síntese proteica – é denominada “mutação sem sentido” (*nonsense mutation*, em inglês). Desta maneira, a proteína gerada será menor que a original, já que todos os aminoácidos codificados por códons posteriormente à mutação não serão incorporados à proteína mutante – gerando proteínas com baixa ou nenhuma atividade. Quanto mais próxima a mutação *nonsense* da extremidade 3' da matriz de leitura aberta, mais possivelmente a proteína resultante terá alguma atividade biológica.

Em regiões codificadoras para polipeptídeos, deleções e inserções de nucleotídeos em números não múltiplos de três alteram toda a proteína após a mutação, uma vez que alteram o quadro aberto de leitura de um gene e podem ocasionar o surgimento precoce de um códon de parada – gerando proteínas com baixa ou nenhuma atividade, assim como as mutações *nonsense*. Estas mutações, denominadas *mudanças de matriz de leitura* (*frameshift mutations*), fazem com que toda a sequência de aminoácidos posteriores ao sítio mutado não tenha relação com a sequência original de aminoácidos, resultando em perda completa da estrutura e da função da proteína.

Além dos erros de replicação, lesões espontâneas (danos ocorridos naturalmente no DNA), como despurinação e desaminação, também são capazes de gerar mutações. A despurinação consiste na perda de uma purina, interrupção da ligação glicosídica entre a base e a desoxirribose, e perda subsequente de uma adenina ou guanina. Células de mamífero perdem cerca de 10 mil purinas espontaneamente em um ciclo celular de 20 horas; entretanto, contam com um sistema de reparo eficiente que remove os sítios apurínicos resultantes. Já a desaminação consiste na produção de uma uracila a partir de uma citosina, e os resíduos de uracila não reparados tenderão a se parear com uma adenina durante a replicação, resultando na conversão de um par GC em AT (transição GC → AT). Ainda, outro tipo de lesão espontânea capaz de gerar mutações é o dano oxidativo ao DNA e aos precursores do DNA (GTP, por exemplo), causado por

espécies reativas de oxigênio, tais como radicais superóxido ( $O_2^-$ ), peróxido de hidrogênio ( $H_2O_2$ ) e radicais hidroxila ( $OH$ ).

A luz ultravioleta geralmente é causa de danos a bases nucleotídicas na maioria dos organismos, gerando vários tipos diferentes de alterações no DNA (*fotoprodutos*), enquanto que a radiação ionizante ocasiona a formação de moléculas ionizadas e excitadas que danificam o DNA. As moléculas produzidas pelos efeitos da radiação ionizante na água produzem os maiores danos devido à natureza aquosa dos sistemas biológicos. Ainda, a radiação ionizante pode danificar o DNA diretamente, através da quebra de ligações *N*-glicosídicas, ocasionando a formação de sítios apurínicos ou apirimidínicos, além de quebras de filamentos (responsáveis pela maioria dos efeitos letais deste tipo de radiação).

Quando um nucleotídeo é incorretamente incorporado à sequência gênica durante a replicação, ocorre uma pequena distorção na estrutura do DNA, reconhecida pela *exonuclease de revisão de leitura* – tipo especial de nuclease que se move no sentido contrário ao da polimerização e cliva o nucleotídeo incorreto, dando “uma nova chance” à DNA polimerase. Esta mudança na geometria entre o 3'-OH e o nucleotídeo a ser incorporado diminui a velocidade de adição de novos nucleotídeos e aumenta a velocidade da atividade exonucleásica. Existe um período de tempo determinado, após cada ciclo replicativo, para que cada nucleotídeo incorretamente incorporado seja removido e substituído pelo correto. Caso esta **revisão de leitura proteica** não funcione neste curto período de tempo, a molécula de DNA recém-sintetizada será metilada resultando em mutação.

Da mesma maneira que a síntese de DNA, a revisão de leitura pode acontecer sem a dissociação do DNA da polimerase. As exonucleases atuam como o botão “delete” de um teclado, removendo os erros mais recentes e diminuindo o surgimento de uma base incorreta de  $10^5$  para a cada  $10^7$  nucleotídeos adicionados. O nível de precisão é ainda melhorado através da atuação do sistema de reparo de *malpareamentos pós-replicação*, que veremos mais adiante. Entretanto, alguns nucleotídeos incorretamente adicionados conseguem escapar da detecção pelas exonucleases e geram um malpareamento entre a fita molde e a recém-sintetizada.

Os **erros nas trincas de aminoácidos**, um exemplo bastante comum de fonte de mutações, constituem repetições de trincas de sequências de nucleotídeos CGG e CAG em determinados genes, resultando em doenças progressivamente mais severas nos filhos e netos dos indivíduos afetados, como distrofia muscular, síndrome do X frágil e doença de Huntington – esta última causada pela repetição de resíduos de glutamina (códon CAG) na proteína mutante, a qual dificulta a transcrição nos neurônios cerebrais, inclusive a transcrição do gene para o receptor de um neurotransmissor. Nestas doenças, as repetições de trinucleotídeos ocorrem nas matrizes de leitura abertas dos transcritos do gene com mutação, ocasionando expansões ou contrações do número de repetições de um único aminoácido – as repetições CAG, por exemplo, codificam uma repetição de poliglutamina.

Alguns agentes mutagênicos não são incorporados ao DNA, mas alteram uma base de maneira a formar um malpareamento específico, como é o caso, por exemplo, do etilmetanossulfato (EMS) e da nitrosoguanidina (NG). O **reparo no pareamento genético** é realizado pelo *sistema de reparo de malpareamentos*, o qual deve verificar o genoma por inteiro. Uma vez que os malpareamentos são temporários, já que quando geram mutações são eliminados no ciclo de replicação seguinte, o sistema de reparo deve encontrá-los e repará-los rapidamente e de forma precisa, substituindo o nucleotídeo incorreto da fita recém-sintetizada e não o nucleotídeo correto da fita molde.

Na *E. coli*, por exemplo, o discernimento de qual dos dois nucleotídeos deve ser substituído é feito através da marcação da fita molde por uma hemimetilação temporária: a enzima Dam metilase metila os resíduos A em ambas as fitas da sequência 5'-GATC-3', amplamente distribuída por todo o genoma. Quando a forquilha de replicação passa pelo DNA metilado nos sítios GATC de ambas as fitas, a dupla de fitas-filha de DNA está hemimetilada (metiladas somente na fita molde), até que a Dam metilase alcance e metile a fita recém-sintetizada, e esta seja marcada por não ter o grupo metil – sendo, portanto, reconhecida como a fita a ser reparada.

As células eucarióticas apresentam diversas proteínas para corrigir malpareamentos, sendo algumas específicas para

malpareamentos simples e outras específicas para reconhecer pequenas inserções ou deleções resultantes de equívocos durante a replicação. Porém, eucariotos não apresentam o mecanismo elaborado de hemimetilação para marcação da fita molde, como a *E. coli*. Como o sistema de reparo reconhece, então, qual das duas fitas deve ser reparada?



### Assimile

A síntese da fita-filha ocorre de maneira descontínua, através da formação de fragmentos de Okazaki unidos ao DNA previamente sintetizado pela DNA ligase. Anteriormente à etapa de ligação, estes fragmentos estão separados do DNA já sintetizado por uma quebra parecida com a gerada pela MutH (proteína que se liga a sítios hemimetilados) na fita recém-sintetizada de *E. coli*. O sistema de reparo eucariótico repara os malpareamentos em moldes artificiais que contêm uma quebra de forma seletiva na fita que contém a clivagem. Assim, as proteínas MSH (homólogas humanas da MutS de *E. coli*) interagem com um componente do grampo deslizante no replissomo e são recrutadas para o sítio de síntese descontínua de DNA na fita-filha – interação esta que também recruta proteínas de reparo de malpareamentos para a extremidade 3' da fita molde.

### Sem medo de errar

O descuido de Marcos em não utilizar rotineiramente a máscara de solda durante o trabalho pode ter provocado uma **intoxicação por chumbo**. A profissão de Marcos (soldador em uma fábrica de ferros) o deixa em contato direto com este metal pesado. A inalação dos vapores tóxicos emanados pelo chumbo quando este é aquecido a altas temperaturas e entra em contato com o ar atmosférico permite que este agente mutagênico atinja os tecidos moles, podendo se depositar nos ossos, dentes e cabelos, além de ser distribuído entre as estruturas celulares e se ligar fortemente às mitocôndrias, em um importante fenômeno de distribuição e toxicidade celular. O chumbo pode provocar um grupo de doenças conhecidas por *porfirias*, por ser capaz de inibir a produção de hemoglobina, afetando diversas reações enzimáticas essenciais para a síntese da heme. Três enzimas apresentam

atividade inibida pelo chumbo: *5-aminolevulinato desidratase (ALA)*, *coproporfirinogênio (COPRO) III oxidase* e *ferroquelatase*. Em consequência, há uma maior produção e excreção dos precursores 5-aminolevulinato sintetase e coproporfirina, com aumento de protoporfirina circulante. Nas hemácias, a síntese diminuída de mono-oxigenases (citocromos P450) compromete a oxidação de drogas e o chumbo se liga à hemoglobina. A ferroquelatase, responsável por catalisar a inserção de ferro na protoporfirina IX, é extremamente sensível ao chumbo. Entretanto, a inibição desta enzima é um fator limitante da taxa para a síntese de hemoglobina, uma vez que a protoporfirina IX se acumula nos eritrócitos, constituindo cerca de 95% das porfirinas não ligadas ao ferro nas hemácias. Desta forma, uma diminuição na atividade da ferroquelatase provoca aumento do substrato (protoporfirina eritrocitária) nas células vermelhas.

## Avançando na prática

### Fragilidade cromossômica

#### Descrição da situação-problema

Um jovem de 25 anos apresenta, desde seu nascimento, retardo mental, distúrbios de aprendizagem, fala repetitiva, hiperatividade, timidez, ansiedade social, atenção de curta duração, retardos de linguagem e temperamento explosivo. Ao final da adolescência e início da fase adulta, passou também a apresentar face longa e estreita, orelhas proeminentes, mandíbula longa e testículos aumentados. O jovem é portador de uma doença genética hereditária transmitida por sua mãe. De que doença se trata e quais são suas causas?

#### Resolução da situação-problema

A doença apresentada pelo jovem é a **síndrome do X frágil**, segunda causa genética mais frequente de retardo mental (atrás apenas da síndrome de Down), provocada por um **distúrbio de repetição de trincas**. Em indivíduos afetados pela doença, a área do X frágil contém de 200 a 2000 repetições da sequência CGG do DNA, como parte de um gene chamado *gene do retardo mental do X frágil* (FMR1). O gene FMR1 codifica a *proteína do retardo*

*mental do X frágil* (FMRP), a qual, quando anormal, se liga a diversas moléculas diferentes de RNAm, cujas proteínas codificadas são essenciais para o funcionamento neuronal, desabilitando-as. Em indivíduos não afetados, a área do X frágil apresenta de 6 a 50 repetições CGG, enquanto homens transmissores (indivíduos que herdam a anomalia cromossômica, porém assintomáticos) e mulheres com sintomas brandos apresentam uma pré-mutação contendo de 50 a 200 repetições CGG.

## Faça valer a pena

**1.** Associe corretamente as mutações, a seguir, às definições dadas:

- ( ) Silenciosas
- ( ) Ligadas ao X
- ( ) Com sentido trocado
- ( ) Nulas
- ( ) Neutras
- ( ) Sem sentido

(1) Ocorrem em regiões desprovidas de genes e não afetam produtos gênicos ou a expressão dos genes.

(2) Troca de um nucleotídeo na região codificadora de um gene, resultando na criação de um novo códon que, por sua vez, corresponderá a um aminoácido diferente.

(3) Alteram o códon, mas não resultam na troca do aminoácido correspondente na proteína.

(4) Afetam genes localizados no cromossomo X.

(5) Resultam na perda total da função do produto gênico.

(6) Substituição do códon original por um códon de parada, resultando na terminação precoce da proteína traduzida.

Assinale a alternativa que contém a sequência numérica correta, de cima para baixo.

- a) 1 – 2 – 3 – 4 – 5 – 6.
- b) 3 – 4 – 2 – 5 – 1 – 6.
- c) 3 – 1 – 6 – 5 – 4 – 2.
- d) 4 – 5 – 2 – 6 – 1 – 3.
- e) 5 – 6 – 4 – 3 – 2 – 1.

**2.** Regiões codificadoras para polipeptídeos, deleções e inserções de nucleotídeos em números não múltiplos de três alteram toda a proteína após a mutação e podem ocasionar o surgimento precoce de um códon de parada, gerando proteínas com baixa ou nenhuma atividade.

O texto anterior se refere a que tipo de mutação ocorrida no material genético?

- a) Com sentido trocado.
- b) Sem sentido.
- c) Silenciosa.
- d) Pontual.
- e) Mudanças de matriz de leitura.

**3.** Quando um nucleotídeo é incorretamente incorporado à sequência gênica durante a replicação, ocorre uma pequena distorção na estrutura do DNA. Esta mudança na geometria entre o 3'-OH e o nucleotídeo a ser incorporado diminui a velocidade de adição de novos nucleotídeos. Existe um período de tempo determinado, após cada ciclo replicativo, para que cada nucleotídeo incorretamente incorporado seja removido e substituído pelo correto.

Que tipo especial de proteína é responsável pelo reconhecimento da distorção na estrutura do DNA citada anteriormente?

- a) Enzima de restrição.
- b) Desoxirribonuclease.
- c) Ribonuclease.
- d) Exonuclease de revisão de leitura.
- e) Nucleotidase.

## Seção 3.2

### Mecanismos de reparo do DNA

#### Diálogo aberto

A exposição a agentes mutagênicos no ambiente de trabalho pode alterar o estado de saúde dos indivíduos que nele trabalham. A maioria desses agentes exibe uma variedade de mutações características, dependentes de diversos fatores, incluindo a natureza das alterações primárias no DNA (modificações de base, quebras de filamentos, incorporações de bases modificadas, etc.) e os efeitos secundários subsequentes, causados pela resposta do organismo a essas alterações. A determinação do potencial carcinogênico de produtos químicos utilizando animais é lenta e trabalhosa. Uma vez que a maioria dos agentes causadores de tumores é mutagênica, o efeito carcinogênico de produtos químicos pode ser avaliado através da capacidade de estes produtos causarem mutações. Cite um teste simples para avaliação dos efeitos carcinogênicos e potencial mutagênico de produtos químicos, como o chumbo. Em que consiste esse teste?

#### Não pode faltar

Todas as células estão expostas a uma imensa variedade de agentes mutagênicos, com potencial de danificar o DNA e provocar mutações. Dentre as **lesões por ação química**, destacam-se os análogos de bases, substâncias mutagênicas com estrutura suficientemente similar às bases normais (de modo a serem metabolizadas e incorporadas ao DNA durante a replicação), porém suficientemente diferentes para aumentar a frequência de pareamentos incorretos e, conseqüentemente, de mutações.

Outros agentes mutagênicos atuam através da modificação química das bases do DNA, como o ácido nitroso, que reage com as bases constituídas de grupamentos amínicos e causa transições AT↔GC. Moléculas de corantes aromáticos, como as *acridinas*, induzem a um tipo diferente de mutação, intercalando-se no DNA

– isto é, introduzindo-se entre pares de bases adjacentes da dupla hélice e, desta forma, causando inserção ou deleção de um ou mais pares de bases. O resultado destas mutações é a alteração da fase de leitura na tradução, a não ser que a inserção ou a deleção ocorra em números múltiplos de três.

A *alquilação* gera como dano a transferência de grupamentos metílico ou etílico para os sítios reativos das bases e dos fosfatos da cadeia de DNA, sendo a guanina um dos sítios de alquilação mais vulneráveis, sujeita à metilação do oxigênio ligado ao seu carbono 6 e gerando O<sup>6</sup>-metil-guanina – a qual parece incorretamente com a timina, trocando GC por AT quando o DNA é replicado.



### Refleta

Diante das inúmeras maneiras pelas quais o DNA pode ser danificado (erros de replicação, luz ultravioleta, radiação ionizante, espécies reativas de oxigênio, etc.), como a vida conseguiu superar todos esses contratempos e evoluir por bilhões de anos?

Os organismos vivos têm inúmeros mecanismos eficientes de reparo do DNA, para identificar e corrigir as lesões antes que provoquem o bloqueio da replicação ou mutação. Do contrário, as células não sobreviveriam tanto tempo. O mais importante dentre eles constitui a revisão de leitura proteica realizada pelas DNA polimerases durante a replicação do material genético, removendo nucleotídeos incorretamente incorporados na sequência gênica. Caso deseje relembrar o mecanismo de ação deste sistema de reparo, volte à Seção 3.1 desta unidade.

Outra via de reparo é a **reversão direta de lesões do DNA**, que regenera a base normal. A luz ultravioleta (UV), por exemplo, pode produzir o fotodímero mutagênico pirimidina ciclobutano (CPD), o qual pode ser reparado pela enzima *CPD fotoliase*, a qual se liga ao fotodímero e o divide a fim de regenerar as bases originais, em um mecanismo de reparo denominado *fotorreativação*. As enzimas alquiltransferases também são capazes de reverter diretamente as lesões do DNA, através da remoção de alguns grupos alquila adicionados, à posição O-6 da guanina, por mutágenos como a nitrosoguanidina (NG) e o etilmetanossulfonato (EMS).



Quer saber mais sobre os efeitos carcinogênicos da radiação UV? Leia o artigo a seguir.

SGARDI, F. C.; CARMO, E. D.; ROSA, L. E. B. Radiação ultravioleta e carcinogênese. **Revista de Ciências Médicas**, v. 16, n. 4-6, p. 245-250, 2007. Disponível em: <<https://seer.sis.puc-campinas.edu.br/seer/index.php/cienciasmedicas/article/view/1050/1026>>. Acesso em: 15 jun. 2017.

Alguns sistemas de reparo utilizam-se das propriedades de complementaridade antiparalela para reparar segmentos danificados ao seu estado não danificado e, portanto, são chamados de *sistemas de reparo dependentes de homologia*. Esses sistemas removem uma base ou um segmento maior da cadeia de DNA e os substitui por um segmento de nucleotídeo recém-sintetizado complementar ao filamento oposto que serve de molde.

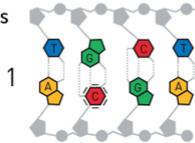
Um dos sistemas dependentes de homologia é o sistema de **reparo por excisão de base**, o mecanismo mais importante para remoção de bases incorretas ou danificadas após o sistema de revisão de leitura proteica. Nesse sistema, as enzimas *DNA glicosilases* clivam ligações base-açúcar, liberando as bases incorretas e gerando sítios apurínicos ou apirimidínicos. Em seguida, uma endonuclease cliva o filamento danificado antes do sítio e uma terceira enzima (*desoxirribofosfodiesterase*) remove o trecho do açúcar-fosfato vizinho, de maneira que a DNA polimerase preencha o espaço com nucleotídeos complementares ao outro filamento. Por fim, a DNA ligase une o novo nucleotídeo (Figura 3.2).



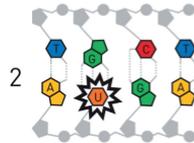
Figura 3.2 | Reparo por excisão de bases

### Reparo por excisão de bases

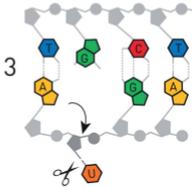
A excisão de bases repara o DNA quando a base do nucleotídeo está danificada. Por exemplo, a citosina.



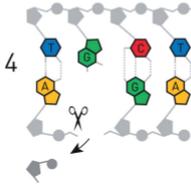
1 A citosina pode ser danificada e formar uma uracila



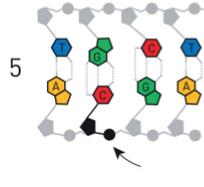
2 A citosina não se emparelha com a guanina.



3 A enzima, glicosilase, encontra e remove a base de uracila.



4 Outras enzimas removem o resto do nucleotídeo da fita de DNA



5 A DNA polimerase preenche e a DNA liase sela a fita de DNA.

Fonte: <<https://goo.gl/Mr9fo9>>. Acesso em: 15 jun. 2017.

Quando o dano é mais complexo que um simples dano de base passível de ser reparado por excisão, como é o caso da adição de CPD por luz UV, o sistema de reparo por excisão de bases não é capaz de corrigi-lo. Uma vez que a DNA polimerase não continua a síntese de DNA devido a estes danos, há um bloqueio na replicação – e forquilhas de replicação bloqueadas podem gerar morte celular. Para reverter esta situação, uma via alternativa é o *reparo por excisão de nucleotídeo* (NER), capaz de desbloquear a replicação e a transcrição, reparando o dano.

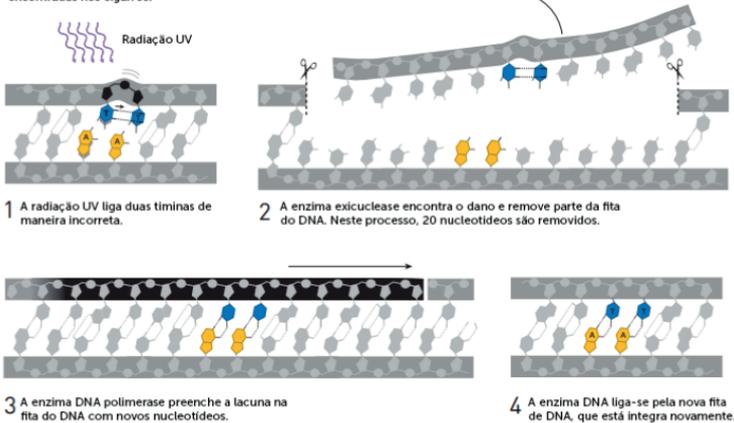
O NER, apesar de complexo, pode ser dividido em quatro etapas: 1) reconhecimento da(s) base(s) danificada(s); 2) montagem de complexo multiproteico no local; 3) clivagem do filamento danificado alguns nucleotídeos antes e depois do sítio danificado, com posterior remoção dos nucleotídeos entre os cortes; e 4) utilização do filamento não danificado como molde para a DNA polimerase, seguida de ligação do filamento (Figura 3.3).



Figura 3.3 | Reparo por excisão de nucleotídeo (NER)

### Reparo por excisão de nucleotídeos

A excisão de nucleotídeos repara danos causados ao DNA por radiação UV ou substâncias carcinogênicas como aquelas encontradas nos cigarros.



Fonte: <<https://goo.gl/VgmuR9>>. Acesso em: 15 jun. 2017.

O *reparo por recombinação* é ainda mais sofisticado que os reparos por excisão, sendo acionado quando ambas as fitas estão danificadas. Nesta situação, uma fita não pode servir de molde para o reparo da outra e, portanto, a informação deve ser obtida a partir da sequência do cromossomo homólogo. Entretanto, o reparo por recombinação opera apenas quando o homólogo do cromossomo danificado está presente na célula. E se um cromossomo é danificado no início do ciclo celular, antes que a replicação gênica tenha produzido seu homólogo? Neste caso, um sistema livre de falhas, denominada *união de extremidades não homólogas* (NHEJ, do inglês *non homologous end joining*), une diretamente as duas extremidades do DNA danificado através de desalinhamentos entre as fitas simples presentes nas extremidades danificadas. Assim, as regiões de fita simples são removidas por nucleases e as lacunas, preenchidas pela DNA polimerase.

Quando a progressão de uma DNA polimerase, durante a replicação, é bloqueada devido às bases danificadas, uma polimerase translesão copia sobre o sítio da lesão, de maneira independente

do pareamento de bases entre as fitas molde e recém-sintetizada – sendo este mecanismo um sistema de reparo de último recurso, já que a *síntese translesão* é mutagênica. Uma vez que este reparo por recombinação está finalizado, o sistema de reparo por excisão de nucleotídeos tem outra chance de reparar o dímero de timinas.



## Assimile

Quadro 3.1 | Sistemas de reparo do DNA

Tipo	Lesão	Enzima
Reparo de malpareamentos	Erros de replicação	MSH, MLH e PMS
Fotorreativação	Dímeros de pirimidinas	DNA fotoliase
Reparo por excisão de base	Base danificada	DNA glicosilase
Reparo por excisão de nucleotídeos	Dímero de pirimidinas e adutos volumosos nas bases	XPC, XPA, XPD, ERCCI-XPF e XPG
Reparo de quebras de fita dupla	Quebras de fita dupla	RecA e RecBDC, em <i>E.coli</i>
Síntese de DNA translesão	Dímero de pirimidinas ou sítio apurínico	DNA polimerases da família Y, em <i>E.coli</i>

Fonte: adaptado de Watson (2006).

Durante a meiose, quando os cromossomos homólogos são pareados antes da primeira divisão celular, ocorre uma permuta física de sequências de DNA entre os cromossomos, denominada *crossing over*; a qual constitui um dos resultados da **recombinação homóloga**. A frequência de *crossing over* entre dois genes no mesmo cromossomo depende da distância física entre estes genes: quanto maior a distância entre eles, maior a frequência de permuta. Além de proporcionar variações genéticas, a recombinação homóloga permite que as células recuperem sequências perdidas por lesões no DNA, substituindo a região danificada por uma fita de DNA não danificada de um cromossomo homólogo. Além disso, fornece um mecanismo para o reinício das forquilhas de replicação interrompidas ou danificadas. Por último, tipos especiais de recombinação regulam a expressão de alguns genes.

A recombinação deve estar completa antes da primeira divisão celular, permitindo que os cromossomos homólogos se alinhem de maneira correta e sejam, assim, separados – do contrário, não conseguem se alinhar corretamente (*não disjunção*), gerando uma elevada incidência de perda cromossômica. Durante este processo, as cromátides-irmãs permanecem pareadas, sendo separadas na segunda divisão nuclear. O resultado dessa divisão são quatro gametas, cada qual com sua cópia de cromossomos. Gametas com número maior ou menor de cromossomos não se desenvolvem após a fertilização, portanto, uma falha na recombinação homóloga reflete diretamente em baixa fertilidade.

Outra consequência da recombinação é a conversão gênica, na qual uma cópia do gene *A* foi convertida em *a* ou vice-versa, produzindo gametas com os genótipos *A a a a* ou *A A A a*. O correto (ausência de conversão gênica) seria a formação de dois gametas contendo o alelo *A* e dois contendo o alelo *a*, produzindo gametas com o genótipo *A A a a*.

Existem alguns processos genéticos que rearranjam as sequências de DNA, originando uma estrutura genômica mais dinâmica, uma vez que a molécula de DNA é muito estável. A **recombinação sítio-específica** consiste em muitas reações nas quais um segmento de DNA definido é rearranjado. Esse segmento que será deslocado contém elementos de sequência curtos, denominados *sítios de recombinação*, nos quais ocorre a troca de DNA. Um exemplo clássico é a integração do genoma do bacteriófago lambda ( $\lambda$ ) no DNA cromossômico bacteriano. A recombinação sítio-específica pode resultar em três tipos distintos de rearranjos de DNA: 1) inserção de um segmento de DNA em um sítio específico; 2) remoção de um segmento de DNA; ou 3) inversão de um segmento de DNA. O resultado da recombinação (inserção, remoção ou inversão) dependerá da organização dos sítios de reconhecimento na(s) molécula(s) de DNA que participa (m) do processo.

Células e vírus utilizam a recombinação sítio-específica para diversas funções biológicas: inserção de DNA no cromossomo hospedeiro durante a infecção (fagos); alteração da expressão gênica, possibilitando a expressão de dois genes alternativos através da inversão de um segmento de DNA; manutenção da

integridade estrutural de moléculas de DNA circulares durante os ciclos de replicação, recombinação homóloga e divisão celular, dentre outras.

Muitos compostos são agentes carcinogênicos em potencial, sendo de extrema importância a existência de sistemas-modelo válidos, para uma avaliação eficiente e efetiva da carcinogênese dos compostos. Na década de 1970, Bruce Ames percebeu uma forte correlação entre a capacidade de os compostos causarem câncer e a de causarem mutações, observando que a medida das taxas de mutação nos sistemas bacterianos seria um modelo efetivo para avaliação da mutagenicidade de compostos, como primeiro passo para a detecção de potenciais carcinógenos.

O **teste de Ames** utiliza uma linhagem de *Salmonella typhimurium*, contendo uma mutação *nonsense* ou *frameshift*, por exemplo, no operon (grupos de um ou mais genes estruturais expressos a partir de um promotor), responsável pela biossíntese de histidina. As células desta linhagem tornam-se incapazes de se multiplicar e formar colônias em meios de cultura livres de histidina. No entanto, se forem tratadas com um composto mutagênico e, portanto, possivelmente carcinogênico, este composto ocasionará a reversão da mutação *nonsense* ou *frameshift* (dependendo de sua natureza) em um número pequeno de células mutantes.

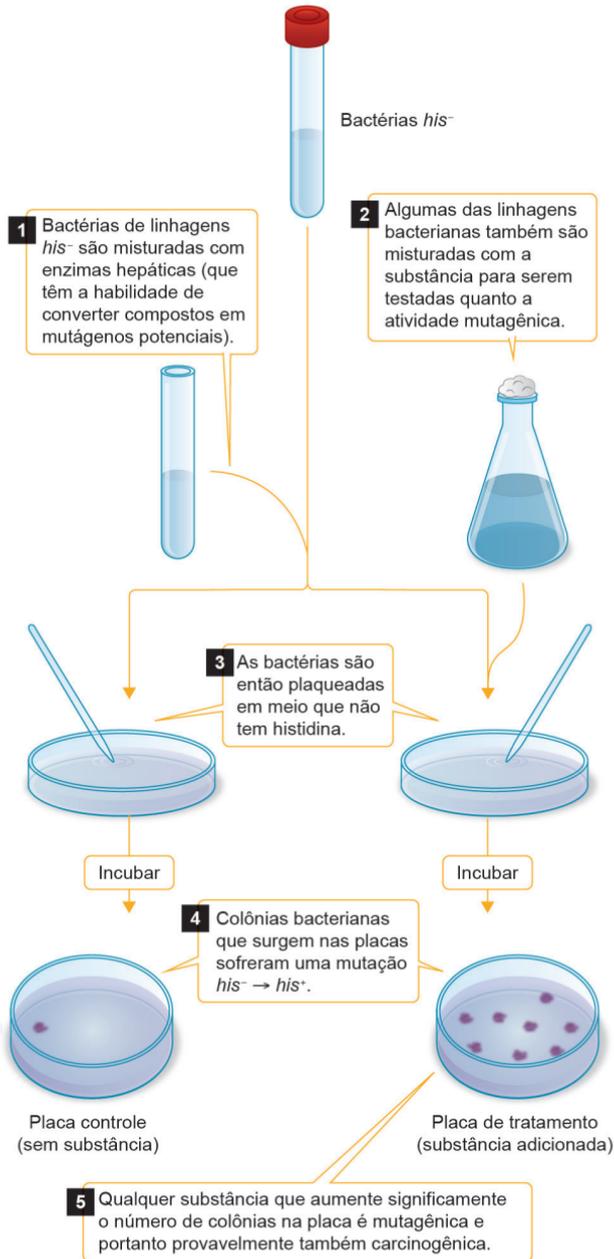
Essa reversão retoma a capacidade celular de se multiplicar, formando colônias em meio de cultura livre de histidina. Quanto mais potente o mutagênico, maior o número de colônias. Alguns compostos carcinogênicos não são, por si só, mutagênicos, mas são convertidos em agentes mutagênicos no fígado, órgão que metaboliza substâncias exógenas. Na identificação de compostos convertidos em mutagênicos no fígado, o teste de Ames trata os potenciais mutagênicos com uma mistura que contém enzimas hepáticas.

## Sem medo de errar

Um teste simples para avaliação dos efeitos carcinogênicos e potencial mutagênico de produtos químicos como o chumbo é o **teste de Ames**. Esse teste consiste na medida das taxas de mutação em sistemas bacterianos para avaliar a mutagenicidade de compostos, como um primeiro nível de detecção de potenciais carcinógenos. Alguns metabólitos carcinogênicos produzidos

no corpo são agentes mutagênicos. Esses metabólitos são produzidos no fígado e as reações enzimáticas que convertem os carcinógenos em metabólitos bioativos não ocorrem em bactérias. Porém, tratando-se linhagens especiais de *Salmonella typhimurium* com extratos de fígado de rato que contêm enzimas metabólicas, essas linhagens adquirem um dos vários alelos mutantes de um gene responsável pela síntese de histidina, capazes de "reverter" (retornar ao fenótipo selvagem) apenas através de alguns eventos mutacionais adicionais. As bactérias tratadas de cada uma destas linhagens são, então, expostas ao composto-teste e cultivadas em meio de cultura sem histidina. A falta deste nutriente assegura que somente cresçam os indivíduos revertentes que contêm a substituição de base apropriada ou mudança de matriz de leitura. O número de colônias em cada placa e o número total de bactérias testadas são determinados, permitindo o cálculo da frequência de reversão. Compostos produtores de metabólitos que induzem níveis elevados de reversão relativos a extratos de fígado de controles não tratados são claramente mutagênicos e, possivelmente, carcinogênicos.

Figura 3.4 | Teste de Ames



Fonte: <<https://mutacaogenica.files.wordpress.com/2010/05/22.jpg>>. Acesso em: 15 jun. 2017.

## Xeroderma pigmentoso

### Descrição da situação-problema

Uma jovem de 17 anos começou a desenvolver muitas pequenas lesões semelhantes às sardas pré-cancerosas, sendo extremamente sensível à luz do sol. Ela procura uma dermatologista, que faz um levantamento de seu histórico familiar e a aconselha a fazer um mapeamento genético, no qual é detectado *xeroderma pigmentoso*, uma doença recessiva autossômica que torna seus portadores propensos a desenvolver cânceres pigmentados de pele durante a vida. Diversos genes podem ser mutados, gerando este fenótipo. Qual é a causa da doença apresentada pela jovem?

### Resolução da situação-problema

O xeroderma pigmentoso é causado por **defeitos no reparo de excisão de nucleotídeo**. Quando um dano ao DNA, por exemplo, a adição de pirimidina ciclobutano (CPD), causada pela radiação ultravioleta, é mais complexo e não pode ser corrigido pelo sistema de reparo por excisão de bases, a DNA polimerase não continua a síntese de DNA e ocorre um bloqueio na replicação. Forquilhas de replicação bloqueadas podem gerar morte celular se não revertidas pelo sistema de reparo por *excisão de nucleotídeo* (NER), capaz de desbloquear a replicação e a transcrição, reparando o dano. No xeroderma pigmentoso, os defeitos nesse sistema de reparo o tornam incapaz de desbloquear a replicação e a transcrição e, conseqüentemente, de reparar o dano no DNA dos portadores da doença.

## Faça valer a pena

**1.** Um dos mais consideráveis sistemas de reparo do DNA dependentes de homologia é o sistema de *reparo por excisão de base*, o mecanismo mais importante para remoção de bases incorretas ou danificadas, após o sistema de revisão de leitura proteica.

Qual é a enzima envolvida nesse sistema, responsável por liberar as bases incorretas?

- a) DNA fotoliase.
- b) DNA glicosilase.
- c) DNA polimerase.
- d) DNA helicase.
- e) MSH.

**2.** Gera como dano a transferência de grupamentos metílico ou etílico para os sítios reativos das bases e dos fosfatos da cadeia de DNA, sendo a guanina um dos sítios mais vulneráveis, sujeita à metilação do oxigênio ligado ao seu carbono 6 e gerando O<sup>6</sup>-metil-guanina.

A descrição apresentada no texto faz referência a que tipo de lesão do DNA?

- a) Despurinação.
- b) Desaminação.
- c) Dano oxidativo.
- d) Alquilação.
- e) Despirimidinação.

**3.** Quando uma fita não pode servir de molde para o reparo da outra, a informação deve ser obtida a partir da sequência do cromossomo homólogo. Entretanto, este sistema de reparo opera apenas quando o homólogo do cromossomo danificado está presente na célula.

O sistema de reparo em questão, acionado quando ambas as fitas estão danificadas, trata-se de:

- a) Reparo por excisão de base.
- b) Reparo por excisão de nucleotídeo.
- c) Reparo por recombinação.
- d) Reparo de quebras de fita dupla.
- e) Reparo por malpareamentos.

## Seção 3.3

### Regulação gênica

#### Diálogo aberto

Como vimos na Seção 3.1, o chumbo pode provocar um grupo de doenças conhecidas por porfirias, por ser capaz de afetar diversas reações enzimáticas essenciais para a síntese da heme, inibindo a produção de hemoglobina. Neste caso, trata-se de uma forma adquirida da doença, a qual apresenta duas outras formas adquiridas (forma adquirida da porfiria cutânea tardia e forma crônica da tirosinemia tipo I). As formas hereditárias da doença constituem: porfiria ala desidratase, porfiria aguda intermitente, coproporfiria hereditária e porfiria variegata (classificadas como porfirias agudas), além de porfiria eritropoética congênita, porfiria cutânea tardia, porfiria hepatoeritropoética e protoporfiria eritropoética (classificadas como porfirias cutâneas). As formas hereditárias são resultado de defeitos em um processo genético essencial para a definição de quais genes serão ativados ou não, em determinados momentos e tecidos. Que processo genético é este?

#### Não pode faltar

A estrutura e o funcionamento de uma célula dependem das proteínas que ela contém, sendo o controle da expressão gênica um aspecto fundamental da biologia molecular da célula. Controlando o início da transcrição, a célula pode regular a proteína que será produzida, bem como a velocidade desta produção. A **regulação da transcrição** é essencial para que a maquinaria enzimática e os componentes celulares se adaptem a alterações ambientais e nutricionais. A regulação gênica frequentemente é mediada por proteínas que reagem a sinais ambientais, aumentando ou diminuindo a transcrição de genes específicos.

Quando a transcrição de um gene é *ativada*, ambos RNA e proteína(s) por ele codificados são produzidos em taxas maiores, enquanto a transcrição é reprimida, o RNA mensageiro (RNAm,

responsável pela transferência de informações do DNA até o local de síntese das proteínas) e a(s) proteína(s) por ele codificados são sintetizados em taxas menores. Desta maneira, organismos unicelulares normalmente sintetizam apenas as proteínas necessárias para sua sobrevivência sob determinadas condições. Já organismos multicelulares controlam a expressão de seus genes visando assegurar que o gene correto seja expresso na célula correta e no momento correto, durante o desenvolvimento embrionário e na diferenciação tecidual.

Cerca de metade dos genes procarióticos encontram-se agrupados em operons; cada um codificando as enzimas envolvidas em uma via metabólica particular. Uma vez que um operon bacteriano é transcrito a partir de um ponto inicial sob a forma de um único RNAm, a **regulação gênica em procariotos** procede com todos os genes contidos neste operon regulados de maneira coordenada, ou seja, ativados ou reprimidos com a mesma intensidade. Para iniciar a transcrição, a RNA polimerase deve estar associada a um fator sigma ( $\sigma$ ), o qual funciona como fator de iniciação.

O processo de utilização da lactose consiste em dois componentes: genes estruturais que codificam os produtos necessários para o transporte e o metabolismo da lactose, e elementos regulatórios (gene *lacI*, operador e promotor). Juntos, estes dois componentes formam o *operon lac*.

Quando a *E. coli* se encontra em um ambiente deficiente em lactose, por exemplo, a síntese do RNAm *lac* é reprimida, fazendo que a energia celular não seja desperdiçada na síntese de enzimas que as células não poderão utilizar. Em um ambiente contendo lactose e glicose, a bactéria metaboliza glicose, preferencialmente, por se tratar da molécula central do metabolismo de carboidratos. A lactose somente é metabolizada em taxas elevadas se, na presença de lactose, houver depleção significativa de glicose no meio.

Na ausência de glicose, as células de *E. coli* respondem sintetizando AMP cíclico (AMPC). Conforme a concentração de AMPC aumenta, este se liga a um sítio sobre cada unidade da proteína CAP, provocando uma alteração conformacional que possibilita a ligação da proteína ao sítio de CAP na região controladora de transcrição *lac*. O complexo CAP-AMPC ligado interage, dessa forma, com a polimerase ligada ao promotor, estimulando a taxa de iniciação da

transcrição e levando à síntese de altos níveis de RNAm *lac*, bem como das enzimas codificadas pelo *operon lac*.



### Assimile

**Controle regulador negativo** → proteína repressora bloqueia a transcrição, ligando-se ao DNA no sítio operador (sistema *lac*). A regulação negativa constitui uma maneira direta de o sistema *lac* desligar genes na ausência de açúcares no ambiente.

**Controle regulador positivo** → necessita de fatores proteicos para ativar a transcrição, por exemplo, a repressão catabólica.

Diversas proteínas reguladoras pertencem a famílias proteicas que apresentam *motifs* de ligação ao DNA muito parecidos. Outras partes das proteínas, como seus domínios de interação proteína-proteína, tendem a ser menos semelhantes. A especificidade da regulação gênica depende das interações químicas entre as cadeias laterais de aminoácidos e os grupos químicos nas bases do DNA.



### Refleta

Procaríotos e eucariotos compartilham muitos aspectos da regulação gênica: ambos utilizam proteínas de ligação ao DNA com sequências específicas para modular a taxa de transcrição. Entretanto, em vista do maior genoma e da maior gama de propriedades, a regulação nos eucariotos é mais complexa e requer mais tipos de proteínas reguladoras e de interações com regiões reguladoras adjacentes no DNA. Qual é a principal diferença na regulação destes dois grupos?

A diferença mais significativa na **regulação gênica em eucariotos** é que o DNA é compactado em nucleossomos, formando *cromatina*, um elemento essencial na regulação gênica. Em bactérias, o início da transcrição é evitado ou reduzido caso a ligação da RNA polimerase esteja bloqueada, normalmente devido à ligação de uma proteína reguladora repressora. Proteínas reguladoras ativadoras aumentam a ligação da RNA polimerase a promotores, onde é necessária pouca ajuda. Em contrapartida, em eucariotos a maquinaria transcrricional não se liga ao promotor na ausência de outras proteínas reguladoras.

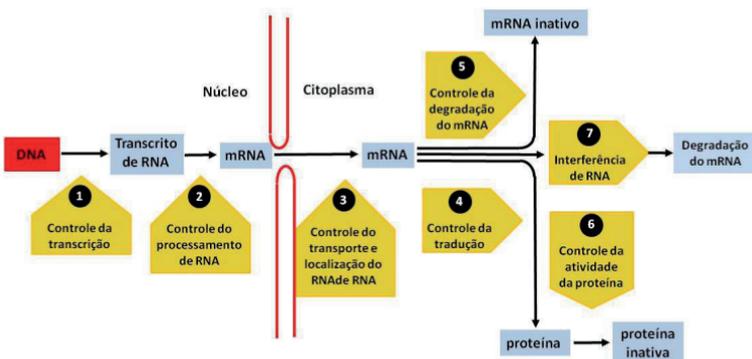
Desta forma, a estrutura da cromatina frequentemente deve ser alterada para ativar a transcrição.

Há uma notável variação entre os genes quando um está ligado (transcrito) ou desligado (não transcrito) e em relação à proporção em que o transcrito precisa ser feito. Por exemplo, certos genes são transcritos somente durante um estágio do desenvolvimento, enquanto outros somente na presença de uma infecção viral. A maioria dos genes em uma célula eucariótica está desligada em algum momento. Portanto, a regulação gênica eucariótica deve garantir que a expressão da maioria dos genes no genoma esteja desligada em algum momento e gerar milhares de padrões da expressão gênica, podendo ocorrer através de diversas etapas (Figura 3.5).



## Exemplificando

Figura 3.5 | Etapas da regulação gênica eucariótica



Fonte: <<https://goo.gl/Nh9J8e>>. Acesso em: 25 jun. 2017.

Três RNA polimerases diferentes efetuam a transcrição gênica nos eucariontes. A *RNA polimerase I* transcreve apenas os genes repetidos *in tandem* que codificam os RNA 18S, 5,8S e 28S no nucléolo. A *RNA polimerase II*, além de transcrever a maioria dos genes de RNA pequeno nuclear (RNAsn), transcreve também todos os genes codificantes de proteína, decididamente envolvidos nas respostas celulares a fatores ambientais e desenvolvimentistas. A RNA polimerase III, por sua vez, transcreve alguns pequenos genes de RNA, tais como os que codificam RNAt, RNA 5S ribossomal e o

RNAsn U6, cujos transcritos são produzidos em todos os tipos de células e, portanto, atuam como genes de manutenção.

Células embrionárias seguem vias de desenvolvimento diferentes devido à expressão de diferentes conjuntos de genes. Basicamente, a regulação gênica embrionária é realizada no início da transcrição, através de três mecanismos principais: localização do RNAm, contato célula-célula e difusão de moléculas sinalizadoras secretadas. A localização do RNAm é realizada pela ligação de sequências 3'-UTR específicas às extremidades crescentes dos microtúbulos. No contato célula-célula, uma molécula sinalizadora se liga à membrana e altera a expressão gênica das células vizinhas, através da ativação de uma via de sinalização celular. Os gradientes extracelulares das moléculas de sinalização celular secretadas podem estabelecer múltiplos tipos celulares durante o desenvolvimento de um tecido ou órgãos complexos. Esses gradientes produzem gradientes intracelulares de fatores de transcrição ativados que, por sua vez, controlam a expressão gênica de maneira dependente da concentração.

A regulação gênica também pode ocorrer após o início da transcrição (**regulação gênica pós-transcricional**), incluindo o alongamento da transcrição e da tradução. Em eucariotos, a maioria dos RNAs precisa sofrer processamento e, em alguns casos, padrões alternativos de processamento originam diferentes produtos proteicos. Outra forma de regulação gênica envolve pequenas moléculas de RNA que inibem a expressão de genes homólogos. Estes RNAs incluem RNAs reguladores, utilizados no desenvolvimento animal, e outros gerados em plantas quando acometidas por infecção viral. Os mecanismos pelos quais esses RNAs inibem a expressão de genes podem envolver a destruição do RNAm, a inibição da tradução e a modificação dos nucleossomos nos promotores destes genes, dirigida pelo RNA. Essa estratégia de repressão é a base da técnica de interferência por RNA (RNAi), mecanismo celular responsável pelo silenciamento gênico pós-transcricional que atua sobre o RNAm.

Um tipo importante de RNA não codificante presente em plantas e animais é denominado microRNA (miRNA), podendo, no caso dos seres humanos – que produzem mais de 400 miRNAs diferentes –, regular ao menos um terço de todos os genes que codificam

proteínas, por pareamento de bases com RNAm específicos. Um único miRNA é capaz de regular um conjunto de diversos RNAm, com a vantagem de que um miRNA ocupa relativamente pouco espaço no genoma, quando comparado a um gene codificante de uma proteína reguladora transcricional. Embora descobertos recentemente, os miRNAs já representam papel importante na regulação da expressão gênica.



### Pesquise mais

Quer saber mais sobre MicroRNAs e seu papel na regulação da expressão gênica? Leia o artigo indicado a seguir.

RICARTE FILHO, J. C. M.; KIMURA, E. T. MicroRNAs: nova classe de reguladores gênicos envolvidos na função endócrina e câncer. **Arquivos Brasileiros Endocrinologia & Metabologia**, v. 50, n. 6, p. 1102-1107, 2006. Disponível em: <<http://dx.doi.org/10.1590/S0004-27302006000600018>>. Acesso em: 25 jun. 2017.

## Sem medo de errar

As formas hereditárias das porfirias porfiria ala desidratase, porfiria aguda intermitente, coproporfiria hereditária, porfiria variegata, porfiria eritropoética congênita, porfiria cutânea tardia, porfiria hepatoeritropoética e protoporfiria eritropoética constituem defeitos na **regulação da expressão gênica**, processo essencial para que a maquinaria enzimática e os componentes celulares se adaptem a alterações ambientais e nutricionais. A regulação gênica frequentemente é mediada por proteínas que reagem a sinais ambientais, aumentando ou diminuindo a transcrição de genes específicos. Quando a transcrição de um gene é *ativada*, ambos, RNA e proteína(s), por ele codificados são produzidos em taxas maiores, enquanto a transcrição é reprimida, o RNAm e a(s) proteína(s) por ele codificados são sintetizados em taxas menores. Desta maneira, organismos unicelulares normalmente sintetizam apenas as proteínas necessárias para sua sobrevivência sob determinadas condições. Já organismos multicelulares controlam a expressão de seus genes visando assegurar que o gene correto

seja expresso na célula correta e no momento correto, durante o desenvolvimento embrionário e na diferenciação tecidual.

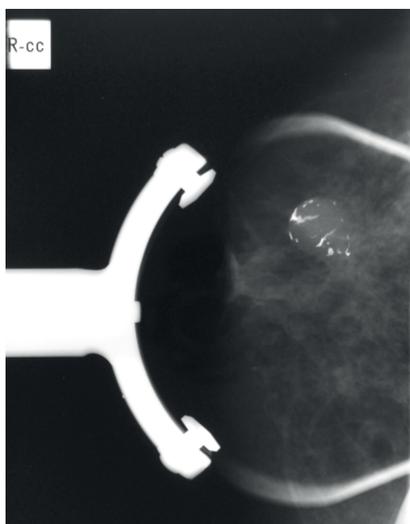
## Avançando na prática

### Fibroadenoma

#### Descrição da situação-problema

Paciente de 35 anos, sexo feminino, sem filhos, há cerca de um ano notou a presença de um nódulo na mama esquerda, o qual aumentava progressivamente de volume com o passar do tempo, atingindo o diâmetro atual de 6cm. Nega dor e secreção papilar e faz uso de anticoncepcional oral há 5 anos. Ao exame físico, apresentou nódulo de 6 x 5cm de diâmetro no quadrante superior esquerdo, firme, indolor e aderido à pele, porém não aderido à fáscia e aos músculos, com o restante do parênquima multinodular, sem nódulos dominantes. Ainda, apresentou linfonodo de 2cm no maior diâmetro, firme, indolor e não aderido. O exame radiográfico demonstrou nódulo hiperdenso e ovalado, com calcificações interiores (Figura 3.6). A paciente realizou mamografia e ultrassonografia complementares, além de biópsia percutânea de fragmento. O resultado obtido revelou ser o nódulo um fibroadenoma.

Figura 3.6 | Exame radiográfico



Fonte: <<https://goo.gl/9WNVih>>. Acesso em: 15 jul. 2017.

Sabendo-se que as células epiteliais e estromais dos fibroadenomas apresentam origem policlonal, tratando-se de processos hiperplásicos (aumento do número de células em um órgão ou tecido), qual seria o provável mecanismo desregulado desencadeador desta patologia?

### Resolução da situação-problema

Fibroadenomas são formados por células epiteliais e estromais de origem policlonal, de um processo hiperplásico cujos mecanismos envolvidos em seu desenvolvimento ainda não são completamente conhecidos. Estudos recentes revelam o possível papel de fatores de crescimento e seus receptores no desencadeamento de doenças benignas de mama, incluindo os fibroadenomas. Hormônios esteroides participam do desenvolvimento da glândula mamária e podem estar envolvidos na regulação do ciclo celular e na alteração da transcrição gênica responsável pela **regulação gênica** do ciclo celular mamário. A análise da expressão gênica envolvida na regulação do ciclo celular pode fornecer informações sobre as interações dos hormônios esteroides com os demais fatores reguladores do ciclo e da diferenciação celular.

### Faça valer a pena

**1.** Em um ambiente que contém lactose e glicose, a bactéria metaboliza glicose, preferencialmente, por se tratar da molécula central do metabolismo de carboidratos. A lactose somente é metabolizada em taxas elevadas se, na presença de lactose, houver depleção significativa de glicose no meio.

Na ausência de glicose, as células de *E. coli* respondem sintetizando:

- a) Amido.
- b) Galactose.
- c) Frutose.
- d) ATP.
- e) AMP cíclico.

**2.** Proteínas reguladoras ativadoras aumentam a ligação da RNA polimerase a promotores em que é necessária pouca ajuda. Em contrapartida, em eucariotos a maquinaria transcricional não se liga ao promotor na ausência de outras proteínas reguladoras. Desta forma, a estrutura da \_\_\_\_\_ frequentemente deve ser alterada para ativar a transcrição.

Assinale a alternativa que preenche corretamente a lacuna do texto.

- a) Base nitrogenada.
- b) Cromatina.
- c) Histona.
- d) Proteína.
- e) Enzima.

**3.** São moléculas de fita simples que contêm de 19 a 25 nucleotídeos não codificadores de proteínas, que agem como potentes reguladores pós-transcricionais da expressão gênica em animais e plantas. Estudos recentes enfatizam a importância dessas moléculas ao relatar alterações na expressão em diferentes patologias humanas.

De que molécula, importante na regulação da expressão gênica de eucariotos, o texto trata?

- a) DNA.
- b) RNAm.
- c) RNAt.
- d) RNAr.
- e) MicroRNA.

# Referências

COX, M. M.; DOUDNA, J. A.; O'DONNELL, M. Mutação e reparo do DNA. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas.** Porto Alegre: Artmed, 2012. cap. 12, p. 409-424.

\_\_\_\_\_. Mutação e reparo do DNA. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas.** Porto Alegre: Artmed, 2012. cap. 12, p. 424-444.

\_\_\_\_\_. Regulação do fluxo de informação. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas.** Porto Alegre: Artmed, 2012. cap. 19, p. 667-696.

\_\_\_\_\_. A regulação da expressão gênica em bactérias. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas.** Porto Alegre: Artmed, 2012. cap. 20, p. 697-732.

\_\_\_\_\_. A regulação transcricional da expressão gênica em eucariotos. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas.** Porto Alegre: Artmed, 2012. cap. 21, p. 733-766.

\_\_\_\_\_. A regulação pós-transcricional da expressão gênica em eucariotos. In: \_\_\_\_\_. **Biologia Molecular: Princípios e técnicas.** Porto Alegre: Artmed, 2012. cap. 22, p. 767-804.

FCM – UNICAMP. **Exame radiográfico.** 2017. Disponível em: <[http://www.fcm.unicamp.br/drpixel/sites/fcm.unicamp.br/drpixel/files/Fibroadenoma\\_Calc.jpg](http://www.fcm.unicamp.br/drpixel/sites/fcm.unicamp.br/drpixel/files/Fibroadenoma_Calc.jpg)>. Acesso em: 5 ago. 2017.

GRIFFITHS, A. J. F. et al. Regulação da expressão gênica em bactérias e seus vírus. In: \_\_\_\_\_. **Introdução à Genética.** 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 11, p. 331-357.

\_\_\_\_\_. Regulação da expressão gênica em eucariotos. In: \_\_\_\_\_. **Introdução à Genética.** 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 12, p. 359-387.

\_\_\_\_\_. Mutação, reparo e recombinação. In: \_\_\_\_\_. **Introdução à Genética.** 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 16, p. 479-491.

\_\_\_\_\_. Mutação, reparo e recombinação. In: \_\_\_\_\_. **Introdução à Genética.** 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 16, p. 491-503.

LEWIS, R. Mutação gênica. In: \_\_\_\_\_. **Genética Humana: Conceitos e aplicações.** 5. ed. Rio de Janeiro: Guanabara Koogan, 2004. cap. 11, p. 207-223.

\_\_\_\_\_. Mutação gênica. In: \_\_\_\_\_. **Genética Humana: Conceitos e aplicações.** 5. ed. Rio de Janeiro: Guanabara Koogan, 2004. cap. 11, p. 223-229.

LODISH, H. et al. Controle transcricional da expressão gênica. In: \_\_\_\_\_. **Biologia Celular e Molecular.** 5. ed. Porto Alegre: Artmed, 2005. cap. 11, p. 441-486.

\_\_\_\_\_. O controle gênico pós-transcricional e o transporte nuclear. In: \_\_\_\_\_. **Biologia Celular e Molecular.** 5. ed. Porto Alegre: Artmed, 2005. cap. 12, p. 487-527.

MATIOLI, S. R.; FERNANDES, F. M. C. (Eds.). Estabilidade do material genético: mutagênese e reparo. In: \_\_\_\_\_. **Biologia Molecular e Evolução**. Ribeirão Preto: Holos, Editora / Sociedade Brasileira de Genética, 2012. cap. 05, p. 45-54.

MUTAÇÃO GÊNICA. **Teste de Ames**. 2017. Disponível em: <<https://mutacaogenica.files.wordpress.com/2010/05/22.jpg>>. Acesso em: 5 ago. 2018.

NANOCELL. **Reparo por excisão de bases**. 2017. Disponível em: <<http://www.nanocell.org.br/wp-content/uploads/2015/11/Screen-Shot-2015-11-03-at-1.38.26-PM.png>>. Acesso em: 15 ago. 2017.

\_\_\_\_\_. **Reparo por excisão de nucleotídeo (NER)**. 2017. Disponível em: <<http://www.nanocell.org.br/wp-content/uploads/2015/11/Screen-Shot-2015-11-03-at-1.38.39-PM.png>>. Acesso em: 5 ago. 2017.

RICARTE FILHO, J. C. M.; KIMURA, E. T. MicroRNAs: nova classe de reguladores gênicos envolvidos na função endócrina e câncer. **Arquivos Brasileiros Endocrinologia & Metabologia**, v. 50, n. 6, p. 1102-1107, 2006. Disponível em: <<http://dx.doi.org/10.1590/S0004-27302006000600018>>. Acesso em: 25 jun. 2017.

SGARDI, F. C.; CARMO, E. D.; ROSA, L. E. B. Radiação ultravioleta e carcinogênese. **Revista de Ciências Médicas**, v. 16, n. 4-6, p. 245-250, 2007. Disponível em: <<https://seer.sis.puc-campinas.edu.br/seer/index.php/cienciasmedicas/article/view/1050/1026>>. Acesso em: 15 jun. 2017.

WATSON, J. D. et al. Mutabilidade e reparo do DNA. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 9, p. 235-242.

\_\_\_\_\_. Mutabilidade e reparo do DNA. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 9, p. 246-258.

\_\_\_\_\_. Recombinação homóloga em nível molecular. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 10, p. 259-292.

\_\_\_\_\_. Recombinação sítio-específica e transposição do DNA. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 11, p. 293-310.

ZAHA, A.; FERREIRA, H. B.; PASSAGLIA, L. M. P. (Orgs.). Mutação e reparação do DNA. In: \_\_\_\_\_. **Biologia Molecular Básica**. Porto Alegre: Artmed, 2012. cap. 7, p. 133-137.

\_\_\_\_\_. Mutação e reparação do DNA. In: \_\_\_\_\_. **Biologia Molecular Básica**. Porto Alegre: Artmed, 2012. cap. 7, p. 146-162.

\_\_\_\_\_. Recombinação genética. In: \_\_\_\_\_. **Biologia Molecular Básica**. Porto Alegre: Artmed, 2012. cap. 8, p. 163-184.

\_\_\_\_\_. Controle da expressão gênica em procariotos. In: \_\_\_\_\_. **Biologia Molecular Básica**. Porto Alegre: Artmed, 2012. cap. 13, p. 277-300.

\_\_\_\_\_. Controle da expressão gênica em eucariotos. In: \_\_\_\_\_. **Biologia Molecular Básica**. Porto Alegre: Artmed, 2012. cap. 14, p. 301-318.

# Técnicas de biologia molecular

## Convite ao estudo

Prezado aluno,

Daremos continuidade à compreensão da biologia molecular e evolução, estudando as técnicas empregadas em biologia molecular. Seus objetivos, ao final da leitura desta Unidade 4, são compreender os principais procedimentos envolvidos na análise de genes e genomas para diagnóstico de doenças, estudar a origem e a evolução das espécies, dentre outros. Para compreender o assunto e atingir as competências e os objetivos da disciplina, analisaremos uma situação hipotética que se aproxima dos conteúdos teóricos que serão vistos por você nesta unidade.

Hemofilias são doenças causadas pela deficiência dos fatores VIII (hemofilia A) ou IX (hemofilia B) da coagulação, podendo ser hereditárias ou adquiridas – estas últimas, mais raras, resultam do desenvolvimento de autoanticorpos associados a doenças autoimunes, câncer ou causas de origem idiopática. A hemofilia A, tipo de hemofilia da qual Letícia é portadora, é caracterizada pela recorrência de hemorragias principalmente articulares e musculares. Nos casos mais graves, ocorrem hemorragias internas no sistema nervoso central (SNC). A frequência e a gravidade dos episódios dependem do nível residual do fator VIII plasmático e de sua atividade funcional, sendo estes parâmetros determinantes para a classificação da hemofilia em grave, moderada e leve.

No decorrer desta unidade de ensino, conheceremos mais a respeito da clonagem e do sequenciamento de DNA, bem como das técnicas de hibridização. Na Seção

4.1, compreenderemos como ocorre a extração de DNA e a utilização de vetores para clonagem de DNA, bem como a técnica de eletroforese. Já na Seção 4.2, conheceremos as técnicas empregadas na polimerização e no sequenciamento gênico. Por fim, na Seção 4.3, trataremos das técnicas de hibridização.

# Seção 4.1

## Clonagem e eletroforese

### Diálogo aberto

Embora os aspectos clínicos e bioquímicos da hemofilia A sejam conhecidos há décadas, as bases moleculares da doença foram compreendidas somente a partir do início da década de 1980, através da caracterização do gene que codifica o fator VIII, localizado na extremidade do braço longo do cromossomo X e cujo índice de polimorfismo é baixo. Os polimorfismos observados nesse gene são polimorfismos de nucleotídeo único (SNPs) e têm repetições *in tandem* de número variável (VNTR), os quais apresentam relevância clínica, uma vez que podem ser utilizados para rastreamento de um gene mutado em uma família afetada. Qual foi a provável técnica utilizada na descoberta do gene que codifica o fator VIII?

### Não pode faltar

Como vimos na Seção 2.3, da Unidade 2 deste livro, a descoberta de uma classe de enzimas bacterianas denominadas *enzimas de restrição* ou *endonucleases*, no final da década de 1960, revolucionou a biologia molecular, proporcionando avanços na medicina forense, na investigação de vínculo genético (determinação de paternidade) e no diagnóstico de doenças hereditárias, apresentando diversas vantagens em comparação com marcadores morfológicos, citológicos ou alozímicos. A dupla fita de DNA sofre **clivagem por enzimas de restrição** em determinadas sequências nucleotídicas, sendo o próprio DNA bacteriano protegido da clivagem pela modificação química dessas sequências. Cada enzima de restrição tem um sítio específico de reconhecimento, no qual é realizada a clivagem, sendo, geralmente, sequências de nucleotídeos simétricas, contendo de quatro a seis pares de bases (Figura 4.1).



A enzima *Eco RI* reconhece uma sequência de seis pares de base e, ao cortá-la, gera extremidades coesivas.

Figura 4.1 | Clivagem da enzima de restrição *Eco RI*



Fonte: <<https://goo.gl/GnPRia>>. Acesso em: 1 set. 2017.

A enzima *Rsa I*, por sua vez, reconhece uma sequência de quatro pares de base e a corta ao meio (setas), gerando pontas retas.

Figura 4.2 | Clivagem da enzima de restrição *Rsa I*



Fonte: adaptada de <<https://goo.gl/tmUK2P>>. Acesso em: 1 set. 2017.

Após a clivagem, os pequenos fragmentos de DNA obtidos podem ser separados uns dos outros através de eletroforese, técnica bioquímica utilizada para separação de moléculas com base em sua carga elétrica e em seu peso molecular, em função da aplicação de um campo elétrico. Moléculas com carga negativa migram para o polo positivo (ânodo), enquanto que moléculas com carga positiva migram para o polo negativo (cátodo).

Na **eletroforese convencional**, as substâncias a serem separadas estão em solução ou suspensão, não sendo empregado algum tipo de suporte. Esta técnica é bastante limitada, devido às soluções estarem sujeitas a uma série de influências físicas do ambiente que lhes causam perturbações, tais como ondas mecânicas e movimentos de convecção do líquido, pelo aquecimento da solução causado pela aplicação da diferença de potencial. Já na **eletroforese em gel**, a mistura de fragmentos de DNA é colocada em uma das extremidades de um bloco de gel de agarose ou

poliacrilamida, contendo uma rede microscópica de poros, na qual é aplicada uma voltagem. Uma vez que o DNA tem carga negativa, os fragmentos migram em direção ao eletrodo positivo, sendo que os fragmentos maiores migram mais lentamente, impedidos pela matriz de agarose.

As amostras de DNA são misturadas a um corante composto por azul de bromofenol e glicerol, utilizado para aumentar a densidade das amostras, impedindo que saiam dos poços, colorindo-as e servindo como indicador do processo eletroforético. Uma mistura de fragmentos de DNA, de tamanhos variáveis e equidistantes entre si, é colocada no primeiro poço para servir de comparação aos fragmentos da amostra, uma vez que contém fragmentos de tamanhos já conhecidos.

Géis de poliacrilamida têm poros menores que os de agarose, sendo utilizados para a separação de proteínas e pequenos fragmentos de DNA e RNA (cerca de 5 a 500 pb). As principais vantagens da utilização de géis de poliacrilamida em relação aos de agarose são, além dos poros menores e mais regulares, a possibilidade de aplicação de maiores quantidades de amostra, a utilização para recuperação e purificação de DNA e RNA, devido à relativa pureza de seus ingredientes, e a maior resistência em relação à agarose.

Em algumas horas, os fragmentos ficam espalhados ao longo do gel, de acordo com o tamanho, formando uma “escada” de bandas. Por fim, uma parte do gel que contém a banda de interesse pode ser cortada usando uma lâmina de bisturi ou gilete, e o DNA pode ser extraído. A **extração de DNA** é o procedimento que serve de base para o fornecimento do material que será posteriormente analisado. A extração consiste basicamente em quatro etapas: 1) lise da célula, para exposição do DNA; 2) remoção dos lipídeos da membrana celular, com detergente; 3) remoção de proteínas da membrana celular, através da adição de protease; e 4) precipitação do DNA.



**Refleta**

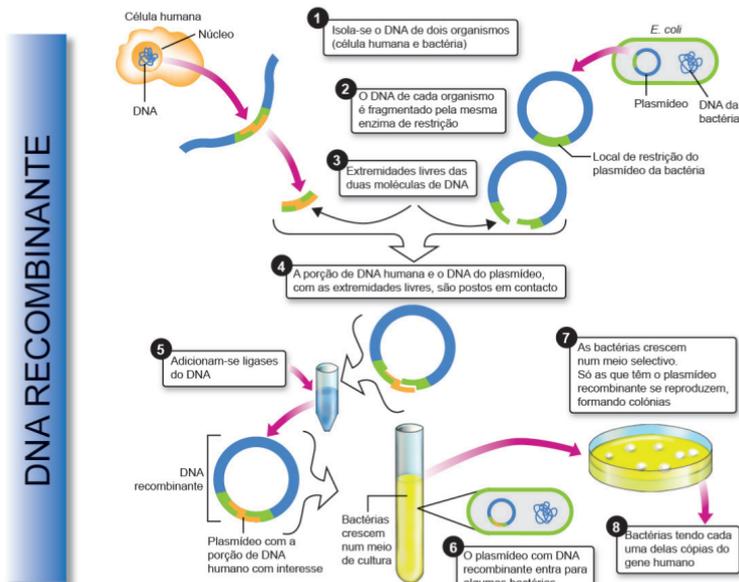
Para estudos que envolvam análises de expressão gênica e caracterização de transcritos, a extração de DNA seria a melhor metodologia?

A extração e purificação do RNA é o primeiro passo para estudos que envolvam análises de expressão gênica e caracterização de transcritos, consistindo em algumas etapas: separação dos ácidos nucleicos dos demais componentes celulares; separação do RNA do DNA contaminante, resultando em um grau de pureza aceitável; e manutenção da integridade molecular do RNA. O processo envolve um forte desnaturante para lisar as células, solubilizar seus componentes e desnaturar ribonucleases (RNases) endógenas.

Um fragmento de DNA pode ser clonado, isto é, gerar várias cópias idênticas a ele, tornando possível a separação física, de uma determinada extensão de DNA (em geral, um gene de interesse), do restante do DNA celular – no qual este fragmento estaria presente em quantidades muito pequenas, quando comparado à massa final da sequência amplificada. A **clonagem de DNA** constitui um dos principais avanços da biologia molecular, tornando possível a tecnologia do DNA recombinante.

DNA recombinante é qualquer molécula de DNA construída em laboratório, contendo fragmentos diversos unidos em novas combinações através da ação da DNA ligase. Uma vez que o DNA tem a mesma estrutura química em todos os organismos, esta tecnologia permite a união de fragmentos de DNA de qualquer fonte, produzindo moléculas de DNA não encontradas na natureza. Ainda, uma vez que a união dos fragmentos é feita pela DNA ligase, a célula não é capaz de detectar que os fragmentos originalmente eram separados, tratando o DNA recombinante como uma molécula única. O DNA recombinante, por sua vez, pode ser introduzido no DNA de uma célula hospedeira (mais comumente uma bactéria, por se dividir rapidamente) e se replicar, transcrevendo-se como uma parte normal do próprio DNA celular.

Resumidamente, o processo de gerar um DNA recombinante pode ser dividido em quatro etapas: 1) produção dos fragmentos desejados de DNA, através de clivagem com enzimas de restrição, 2) inserção desses fragmentos em um vetor, 3) introdução desse vetor em um hospedeiro e 4) identificação, seleção e caracterização dos clones recombinantes (Figura 4.3).

**Figura 4.3 | Tecnologia do DNA recombinante**


Fonte: <[http://images.slideplayer.com.br/1/288898/slides/slide\\_6.jpg](http://images.slideplayer.com.br/1/288898/slides/slide_6.jpg)>. Acesso em: 1 set. 2017.

Entretanto, é mais fácil manipular, copiar e purificar um DNA recombinante quando ele é mantido como uma molécula independente, separada do cromossomo bacteriano. Para tanto, são utilizados **vetores** para manter o DNA estranho em uma célula bacteriana. Os dois tipos de vetores mais utilizados na clonagem são os *fagos* e os *plasmídeos*.

“Fagos” é um tipo derivado de vírus bacteriano que injeta o DNA no hospedeiro, no qual se replica, sendo o fago lambda ( $\lambda$ ) amplamente utilizado, por ser capaz de aceitar grandes fragmentos. Já os plasmídeos mais utilizados são moléculas de DNA circular relativamente pequenas e capazes de se replicar dentro de uma bactéria, independentemente do cromossomo bacteriano. Para ser útil, um vetor deve conter algumas propriedades:

1. Tem capacidade de infectar uma célula hospedeira.
2. Contém uma origem de replicação de DNA, que lhe permite se replicar independentemente do cromossomo hospedeiro.

3. Contém uma marca de seleção, que permite que as células contendo o vetor (e o DNA nele inserido) sejam identificadas facilmente.

4. Apresenta sítios únicos para uma ou mais enzimas de restrição, permitindo que os fragmentos de DNA sejam inseridos em um local definido dentro de um vetor.

5. Assim que transformadas, as células devem ser selecionáveis preferencialmente por crescimento das células hospedeiras em meio sólido contendo agente seletivo.

Após a reprodução de centenas de milhões de cópias do plasmídeo, obtidas a partir da duplicação bacteriana, o fragmento de interesse pode ser recuperado separando-se o DNA plasmidial do cromossomo bacteriano, utilizando uma endonuclease apropriada, seguido da separação entre DNA plasmidial e fragmento, através de eletroforese em gel.

O DNA eucariótico contém íntrons, cuja descoberta de sua localização no gene não é tão simples, sendo, portanto, utilizado RNA mensageiro para estudos de genes eucarióticos e sua expressão, uma vez que os íntrons já foram excluídos. Entretanto, o RNAm não pode ser clonado diretamente, pois as enzimas de restrição e a DNA ligase cortam e copiam apenas DNA fita dupla. Por outro lado, a enzima *transcriptase reversa*, uma enzima de replicação de determinados tipos de vírus, é capaz de transformar RNAm em DNA, utilizando o RNAm como molde para montar uma fita de DNA complementar (DNAc). Como qualquer outro DNA, o DNAc com fita dupla pode ser cortado com enzimas de restrição, e os fragmentos, dessa forma, podem ser inseridos em um vetor de clonagem utilizando DNA ligase.



**Pesquise mais**

Saiba mais a respeito da clonagem humana e suas implicações através da leitura do artigo a seguir.

ZATZ, M. Clonagem humana: conhecer para opinar. **Pesquisa FAPESP**, São Paulo, n. 73, p. 8-14, mar. 2002. Disponível em: <[http://www.revistapesquisa.fapesp.br/wp-content/uploads/2002/03/08\\_PERGUNTASF.pdf](http://www.revistapesquisa.fapesp.br/wp-content/uploads/2002/03/08_PERGUNTASF.pdf)>. Acesso em: 1 set. 2017.

## Sem medo de errar

As bases moleculares da hemofilia A foram compreendidas somente a partir do início da década de 1980, após a **clonagem** e caracterização do gene que codifica o fator VIII (F8). O gene compreende 186.000 pares de bases, distribuídos entre 26 éxons e 25 íntrons, os quais variam de 69 a 3.106 pb. Toda a sequência do gene compreende por volta de 9 kb de éxons e 177 kb de íntrons, correspondendo, juntos, a cerca de 0,1% de todo o cromossomo X. Os polimorfismos observados nesse gene são polimorfismos de nucleotídeo único (SNPs) e repetições *in tandem* de número variável (VNTR), os quais têm relevância clínica, uma vez que podem ser utilizados para rastreamento de um gene mutado em uma família afetada. Bancos de dados acessíveis disponibilizam uma relação completa de mutações publicadas, dados fenotípicos e métodos para rastreamento de mutações, além de uma relação de polimorfismos e uma revisão sobre a patologia molecular da doença. Atualmente, existem mais de mil mutações associadas à hemofilia A, envolvendo substituições, inserções e deleções espalhadas por todo o gene. As mutações mais comumente encontradas são a inversão dos íntrons 1 e 22, encontradas, respectivamente, em cerca de 5% e 50% dos casos de hemofilia A grave.

## Avançando na prática

### Insulina sintética

#### Descrição da situação-problema

Diabetes mellitus é uma doença metabólica crônica, caracterizada pela produção insuficiente de insulina pelo pâncreas ou sua utilização de forma incorreta pelo organismo. O número de portadores gira em torno de 200 milhões de pessoas, representando 9% da mortalidade mundial. Quase todos os pacientes requerem tratamento farmacológico, uma vez que as células  $\beta$  (beta) das ilhotas de Langerhans do pâncreas sofrem falência com o passar do tempo. Esse tratamento consiste na aplicação de injeções contendo insulina, hormônio cuja função é reduzir os níveis de glicose no sangue. Anos atrás, a insulina utilizada para o tratamento da diabetes era extraída de pâncreas suíno, entretanto, com o

avanço de técnicas de biologia molecular, tornou-se possível a produção desse hormônio em laboratório. Qual foi e em que se baseia a técnica pioneira para produção de insulina sintética?

### Resolução da situação-problema

A primeira técnica desenvolvida para a produção de insulina sintética foi a **tecnologia do DNA recombinante**. Esta tecnologia permite a união de fragmentos de DNA de qualquer fonte, produzindo moléculas de DNA não encontradas na natureza. Além disso, uma vez que a união dos fragmentos é feita pela DNA ligase, a célula não é capaz de detectar que os fragmentos originalmente eram separados, tratando o DNA recombinante como uma molécula única. O DNA recombinante, por sua vez, pode ser introduzido no DNA de uma célula hospedeira (mais comumente uma bactéria, por se dividir rapidamente) e se replicar, transcrevendo-se como uma parte normal do próprio DNA celular. Na produção de insulina, o gene humano é isolado, clivado por enzimas de restrição e inserido em um plasmídeo de *E. coli*. O plasmídeo é, dessa forma, inserido na bactéria e, conforme esta célula se replica, o gene humano da insulina é codificado e o hormônio, produzido. Por ser mais parecida com a insulina produzida pelo organismo humano, a insulina sintética obtida através dessa técnica oferece baixo índice de rejeição quando comparada às insulinas de origem animal e redução dos efeitos colaterais, contribuindo para um aumento na qualidade de vida dos portadores da doença.

### Faça valer a pena

**1.** Enzimas de restrição clivam o DNA em determinadas seqüências nucleotídicas, sendo que cada enzima tem um sítio específico de reconhecimento, no qual é realizada a clivagem (geralmente seqüências de nucleotídeos simétricas, contendo de quatro a seis pares de bases). Assinale a alternativa que contém o corte produzido pela endonuclease *Eco RI*.

As alternativas foram retiradas de: NZYTECH. **Institucional**. 2017. Disponível em: <<https://www.nzytech.com>>. Acesso em: 4 set. 2017.

- a) 5'...AG↓CT...3'  
3'...TC↓GA...5'

- b) 5'...GC↓GGCCGC...3'  
3'...CGCCGG↓CG...5'
- c) 5'...A↓AGCTT...3'  
3'...TTCGA↓A...5'
- d) 5'...GGGCC↓C...3'  
3'...C↓CCGGG...5'
- e) 5'...GGGCC↓C...3'  
3'...C↓CCGGG...5'

**2.** DNA recombinante é qualquer molécula de DNA construída em laboratório, contendo fragmentos diversos unidos em novas combinações através da ação da DNA ligase. Uma vez que o DNA tem a mesma estrutura química em todos os organismos, esta tecnologia permite a união de fragmentos de DNA de qualquer fonte, produzindo moléculas de DNA não encontradas na natureza. Resumidamente, o processo de gerar um DNA recombinante pode ser dividido em quatro etapas:

1. Identificação, seleção e caracterização dos clones recombinantes.
2. Introdução deste vetor em um hospedeiro.
3. Produção dos fragmentos desejados de DNA, através de clivagem com enzimas de restrição.
4. Inserção destes fragmentos em um vetor.

Assinale a alternativa que corresponda à ordem correta das etapas da tecnologia do DNA recombinante.

- a) 1 – 2 – 3 – 4.
- b) 2 – 4 – 3 – 1.
- c) 2 – 1 – 4 – 3.
- d) 3 – 4 – 2 – 1.
- e) 4 – 3 – 2 – 1.

**3.** Para estudos de genes eucarióticos e sua expressão, utiliza-se RNA mensageiro ao invés de DNA, uma vez que, no RNAm, os íntrons já foram excluídos. Entretanto, o RNAm não pode ser clonado diretamente, pois as enzimas de restrição e a DNA ligase cortam e copiam apenas DNA fita dupla. Por outro lado, uma enzima de replicação de determinados tipos de vírus é capaz de transformar RNAm em DNA.

Qual é a enzima de replicação capaz de transformar RNAm em DNA?

- a) DNA polimerase.
- b) *Eco RI*.
- c) Transcriptase reversa.
- d) *Rsa I*.
- e) RNase.

## Seção 4.2

### Polimerização e sequenciamento

#### Diálogo aberto

O histórico familiar e a ocorrência de hemorragias constituem os principais sinais clínicos para o diagnóstico da hemofilia. No entanto, entre 20% a 30% dos pacientes não apresentam história familiar. Os avanços nas técnicas moleculares possibilitaram a caracterização da doença no nível molecular, através da utilização de DNA ou RNA, auxiliando o diagnóstico laboratorial, uma vez que o gene F8A apresenta tamanho extenso, complexidade genômica e grande diversidade de mutações. Quais seriam as possíveis técnicas de biologia molecular utilizadas para a detecção de mutações envolvendo pequenas alterações na sequência de DNA (no geral, mutações pontuais) e detecção da inversão do intron 22 de um portador da hemofilia A? Explique sucintamente os princípios básicos destas técnicas.

#### Não pode faltar

Vimos, na Seção 2.1 da Unidade 2 deste livro, que uma variação na sequência gênica, presente em pelo menos 1% da população, é chamada de polimorfismo, podendo ocorrer em regiões do DNA que codifiquem proteína ou em regiões que não a codifiquem. A detecção de polimorfismos foi amplamente aperfeiçoada com o desenvolvimento da técnica de **reação em cadeia da polimerase (PCR, do inglês *polymerase chain reaction*)**, reduzindo o tempo de execução dos experimentos, bem como seus custos e sua complexidade. A técnica consiste em uma reação de polimerização em cadeia, a partir da qual se obtém o enriquecimento de um fragmento específico de DNA através de sua duplicação em modo exponencial.

A principal vantagem da PCR é a possibilidade de obtenção de quantidades muito grande de fragmentos específicos de DNA. A quantidade de material genético a ser utilizada depende

essencialmente da quantidade aproximada de DNA-alvo existente na amostra, embora a técnica seja tão sensível que, teoricamente, uma única molécula de DNA pode servir como molde para a amplificação.

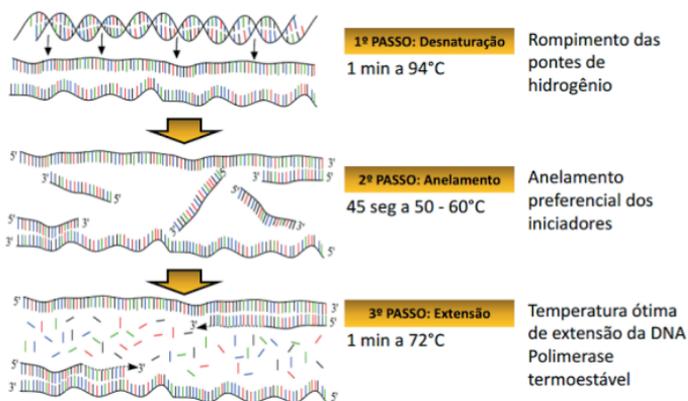
Outro ponto importante é o tamanho dos fragmentos amplificados, uma vez que fragmentos maiores que 3 a 4 mil pares de bases podem induzir a erros – sendo aconselhável, nestes casos, que se utilize uma mistura de polimerases da qual uma pequena fração tem a propriedade de “rever” o processo de alongação, removendo nucleotídeos mal incorporados, responsáveis pela interrupção da síntese pela *Taq* polimerase do DNA. Ainda, a proporção dos trifosfatos de desoxirribonucleotídeos (dNTP) deve ser igual à proporção presente na molécula a ser sintetizada.

O princípio da técnica consiste em três etapas básicas, repetidas em ciclos por diversas vezes: 1) desnaturação térmica (separação das fitas) do DNA molde; 2) anelamento de oligonucleotídeos sintéticos (*primers*) a cada uma das fitas do DNA molde, funcionando como iniciadores da reação de polimerização; e 3) polimerização das novas fitas de DNA a partir de cada *primer*, utilizando cada um dos quatro dNTP como substrato. O processo de amplificação envolve repetidos ciclos, resultando, dessa forma, na duplicação do fragmento de DNA a cada ciclo. Uma vez que os fragmentos recém-sintetizados também são complementares e capazes de se ligar aos *primers* após desnaturação, cada ciclo sucessivo dobra a quantidade de DNA sintetizada no ciclo anterior. O resultado é um acúmulo exponencial do fragmento específico aproximadamente a  $2^n$ , onde “n” é igual ao número de ciclos.

Cada fase da reação tem uma temperatura ótima, sendo cerca de 95°C para a fase de desnaturação térmica, cerca de 60°C para a fase de anelamento do *primer* e cerca de 72°C para a fase de polimerização da nova fita, pela enzima DNA polimerase (Figura 4.4).



Figura 4.4 | Fases e respectivas temperaturas da PCR



Fonte: <<https://goo.gl/XnXaJf>>. Acesso em: 1 set. 2017.

A termolabilidade dessa enzima foi um dos principais problemas técnicos encontrados a princípio, uma vez que as DNA polimerases até então conhecidas perdiam rapidamente sua atividade enzimática em elevadas temperaturas. A solução para esse problema foi o isolamento de enzimas termoestáveis, como a *Taq* polimerase (purificada de *Thermus aquaticus*), a partir de organismos termofílicos, como bactérias nativas de fontes termais ou chaminés vulcânicas do fundo do mar.



Quais seriam os inconvenientes da falta de especificidade da PCR?

A falta de especificidade da reação pode resultar em PCRs com arrastos (fragmentos com vários tamanhos) após a eletroforese em gel, sendo a solução, nestes casos, os seguintes procedimentos, em ordem dos mais prováveis de resultar em sucesso:

1. Aumento da temperatura de anelamento, podendo ser mais alta nos primeiros ciclos e mais baixa nos ciclos subsequentes.

2. Diminuição de DNA-alvo na amostra.
3. Diminuição na concentração de cloreto de magnésio.
4. Diminuição na concentração dos oligonucleotídeos.
5. Diminuição na concentração de dNTPs.

A falta de especificidade também pode acarretar em “nenhuma” amplificação (presença de um arrasto, correspondente a uma enormidade de fragmentos diferentes amplificados, porém não visualizados), sendo as soluções, nestes casos, iguais às anteriormente apresentadas. Por outro lado, se houver excesso de especificidade, o anelamento dos *primers* com as cadeias alvo não acontece, e, portanto, não ocorre amplificação. As soluções, nestes casos, são inversas às anteriormente apresentadas, na mesma ordem.

Atualmente, existem diversos métodos que exploram o princípio da PCR, em vista de sua grande versatilidade e potência: varredura e/ou sequenciamento rápido de insertos de DNA; estudos de expressão qualitativa e quantitativa de genes específicos; identificação de RNAm gerado por meio de processamento alternativo do pré-RNAm; amplificação e clonagem de sequências genômicas a partir de pequenas quantidades de DNA genômico complexo; síntese química de genes por meio da amplificação e junção de partes de uma sequência de DNA conhecida; diagnóstico, através da detecção de sequências de DNA específicas de determinado patógeno; teste de paternidade e mutagênese sítio dirigida, utilizando iniciadores com sequências de nucleotídeos modificadas.

Como podemos perceber, a PCR pode ser utilizada em diversas abordagens que permitem seu emprego no diagnóstico pré-natal (particularmente, nas doenças herdadas) e de doenças infecciosas (causadas por vírus, bactérias, fungos e protozoários), tipagem para transplante de órgãos e susceptibilidade para doenças autoimunes específicas (detecção de polimorfismos), diagnóstico e prognóstico de câncer (detecção de oncogenes e genes supressores de câncer), criminalística (*DNA fingerprint*) etc.

Uma variante da PCR é a PCR em tempo real (**PCR Real Time**) ou PCR quantitativa (PCR-Q), cujas maiores aplicações são os estudos de expressão gênica, além de constituir uma importante

ferramenta para diagnóstico molecular clínico, atuando na detecção de vírus ou bactérias e mensuração de suas respectivas cargas virais e bacterianas, bem como na determinação do estágio de um câncer, genotipagem e estudo de variações do número de cópias. A técnica compreende uma amplificação convencional, porém a detecção do resultado é realizada ao longo de cada ciclo. Para tanto, é adicionada na reação uma molécula fluorescente ou um oligonucleotídeo marcado. Conforme a amplificação vai ocorrendo, um sinal fluorescente é emitido ao final de cada ciclo, captado por um sistema óptico e convertido em um gráfico de amplificação.

As vantagens da técnica são a grande capacidade de quantificação de ácidos nucleicos mesmo em pequenas quantidades iniciais, possibilitando a detecção de menos de cinco cópias da sequência-alvo, além da rapidez dos equipamentos de leitura e a não necessidade de manipulação do produto pós-PCR, minimizando as chances de contaminação. As desvantagens vigoram entre as mesmas da PCR convencional; porém, mais evidentes devido à própria sensibilidade da reação.

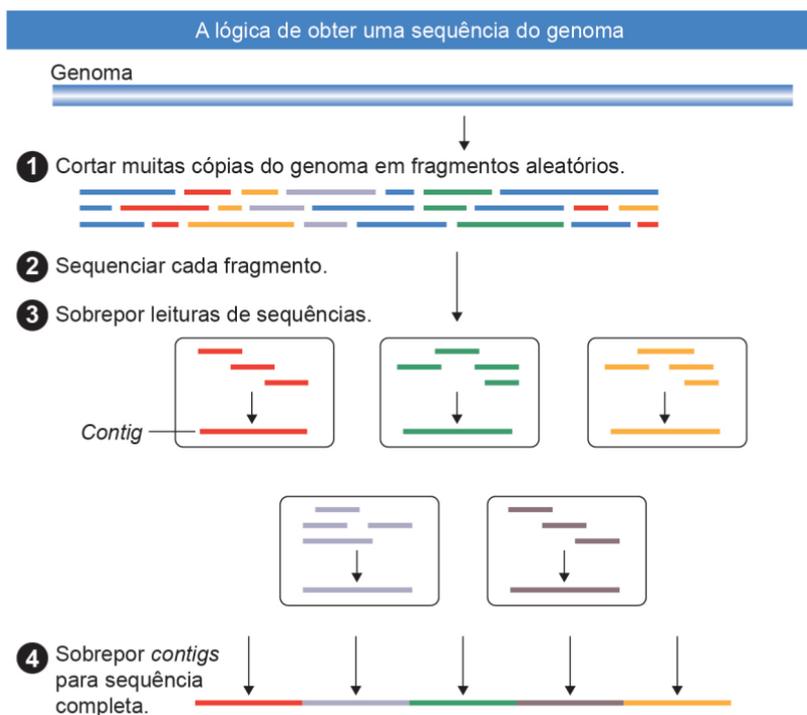
Diversos fatores podem interferir na PCR-Q, dentre eles: precisão de pipetagem (a quantidade de DNA/DNAc colocado inicialmente deve ser a mesma para todas as amostras); qualidade e integridade do DNA/DNAc; desenho dos *primers*, o qual pode influenciar na temperatura de hibridação e na formação de estruturas secundárias; condições de termociclagem; qualidade dos reagentes e tamanho do fragmento amplificado (ideal até 150 pb).

Outras modificações da técnica original consistem na reação da *transcriptase reversa seguida de reação em cadeia da polimerase* (RT-PCR; síntese de uma fita de DNAc utilizando RNAm como molde, catalisada por uma transcriptase reversa); *PCR multiplex* (mais de um segmento genômico amplificado em uma única reação, cada qual com seu par de *primers* específico) e *Nested-PCR* (segmento genômico primeiramente amplificado de forma abrangente, copiando até mesmo sequências localizadas fora dele, e, utilizando-se deste primeiro produto, amplificação da real sequência-alvo; concomitantemente ou em duas reações separadas – *Semi-Nested PCR*).

Como vimos na Seção 1.3 da Unidade 1 deste livro, a obtenção de uma sequência completa de nucleotídeos de um segmento de DNA é importante para o entendimento da organização de um gene e sua regulação, sua relação com outros genes ou a função do RNA ou da proteína que ele codifica. Assim como na técnica do DNA recombinante e na PCR, o **sequenciamento de DNA** explora a complementaridade de pares de bases, juntamente com o entendimento da bioquímica básica da replicação do DNA.

A análise de genomas inteiros contribui para todos os aspectos das pesquisas biológicas, como a possibilidade de localizar genes que contribuam para muitas doenças determinadas por combinações complexas de fatores genéticos. Desta forma, pesquisas utilizando um gene, de uma espécie cujo genoma já esteja sequenciado, demandam apenas a descoberta do local no qual está situado esse gene no mapa genômico para então determinar sua sequência e, potencialmente, sua função.

Basicamente, a obtenção de uma sequência genômica ocorre em quatro etapas: 1) fragmentação das moléculas de DNA do genoma em milhares a milhões de pequenos segmentos aleatórios; 2) leitura da sequência de cada um desses segmentos; 3) encontro, com o auxílio de computador, da sobreposição entre os segmentos nos quais suas sequências sejam idênticas; e 4) continuação da sobreposição de pedaços maiores, até que todos os pequenos segmentos estejam ligados (Figura 4.5). Sendo assim, o maior desafio do projeto genoma é a *montagem de sequência*, isto é, unir todas as leituras individuais em uma sequência consenso, para a qual há concordância (representação verdadeira e precisa do genoma em questão).



Fonte: Griffiths (2015, p. 424).

No final da década de 1970, Frederick Sanger desenvolveu um método com mais rapidez e acurácia, conhecido como **sequenciamento de Sanger**, que se baseia na extensão enzimática da fita de DNA e sua inibição através da inserção de um nucleotídeo análogo deficiente no grupamento 3'OH, conhecido como "didesoxinucleotídeo" (ddNTP). O método inicial desenvolvido por Sanger baseava-se em uma marcação radioativa que poderia ser feita em dNTPs (normalmente, citosinas), *primers* ou ddNTPs terminadores. Com o aperfeiçoamento tecnológico, a detecção dos fragmentos com marcação radioativa foi substituída pela marcação com compostos fluorescentes (fluorocromos). Desta forma, cada ddNTP terminador passou a ter uma cor específica, através da incorporação de um fluorocromo à sua estrutura, substituindo as quatro reações necessárias para o sequenciamento de uma fita de DNA por apenas uma – enquanto que na detecção radioativa são necessárias quatro canaletas para

cada sequenciamento (uma para cada ddNTP terminador), na detecção fluorescente apenas uma canaleta é necessária, já que cada ddNTP tem uma cor diferente.

Sequenciadores automáticos associados a computadores e softwares específicos distinguem cada fluorocromo e armazenam o nucleotídeo correspondente em bancos de dados, deixando o resultado pronto para o pesquisador. Os sequenciamentos passaram a ser de até 700 bases com alta qualidade, comparados aos sequenciamentos de no máximo 100 bases realizados nos primórdios da técnica. Atualmente, os equipamentos mais sofisticados são capazes de realizar 96 sequenciamentos a cada duas horas.

Uma nova tecnologia de sequenciamento, denominada **“sequenciamento de nova geração”** (NGS), foi desenvolvida em meados de 2005 e evoluiu rapidamente, uma vez que promove o sequenciamento de DNA em plataformas capazes de gerar informações sobre milhões de pares de bases em uma única corrida. Duas plataformas de sequenciamento já têm ampla utilização em todo o mundo: 454 FLX e Solexa, com capacidade de gerar informações muitas vezes maior que o sequenciamento de Sanger e com maior economia de tempo e custo, devido ao uso da clonagem *in vitro* e de sistemas de suporte sólido para as unidades de sequenciamento. A clonagem *in vitro* em suporte sólido possibilita que milhares de leituras sejam produzidas de uma só vez com a plataforma 454 FLX ou Solexa.

O sequenciamento na plataforma 454 FLX é realizado a partir de uma combinação de reações enzimáticas iniciada a partir da liberação de um pirofosfato oriundo da adição de um dNTP à cadeia. Em seguida, esse pirofosfato é convertido em ATP pela enzima *ATP sulfúrilase*, sendo utilizado pela enzima *luciferase* na oxidação da *luciferina*, produzindo um sinal de luz que pode ser capturado por uma câmera *charge-coupled device* (CCD) acoplada. Esta plataforma é capaz de sequenciar facilmente genomas pequenos como os de bactérias e alguns eucariotos, produzindo as maiores leituras e, portanto, sendo a mais utilizada.



Conheça uma das possíveis aplicações do sequenciamento de nova geração através da leitura do artigo a seguir.

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, Santa Maria, v. 40, n. 3, p. 735-744, 2010. Disponível em: <<https://repositorio.unesp.br/bitstream/handle/11449/29442/S0103-84782010000300040.pdf?sequence=1&isAllowed=y>>. Acesso em: 1 set. 2017.

## Sem medo de errar

Para a detecção de mutações envolvendo pequenas alterações na sequência de DNA (no geral, mutações pontuais), utiliza-se comumente da amplificação por **PCR** seguida de um ensaio de mobilidade eletroforética e/ou de um sequenciamento. A técnica consiste em uma reação de polimerização em cadeia, a partir da qual se obtém o enriquecimento de um fragmento específico de DNA através de sua duplicação em modo exponencial. Já para a detecção da inversão do íntron 22, vêm sendo utilizadas as técnicas de PCR longa e, mais recentemente, **PCR multiplex**. A **PCR longa** consiste em uma técnica cuja extensão é mais longa que a PCR convencional (acima de 5 kb). Já a PCR multiplex consiste em uma reação de amplificação contendo dois ou mais conjuntos de *primers* desenhados para detecção de múltiplas sequências-alvo em uma mesma amostra.

## Avançando na prática

### HPV: Papiloma Vírus Humano

#### Descrição da situação-problema

Uma vez dentro da faixa etária especificada (meninas de 9 a 14 anos de idade), Letícia foi levada por sua mãe a um posto de saúde para se vacinar contra o papiloma vírus humano, HPV, prevenindo-se contra infecção por HPV e câncer de colo do útero. O HPV pode se instalar em qualquer região do corpo através de uma porta de

entrada, como uma microabrasão na pele ou mucosa. Além da região genital, o vírus pode se instalar nos olhos, na boca, na faringe, nas vias respiratórias, na uretra, no reto, no ânus e líquido amniótico. A quantificação do HPV oferece informações importantes para o acompanhamento da eficácia do tratamento. Qual seria uma possível técnica para detecção e quantificação do HPV?

### Resolução da situação-problema

A detecção e a quantificação do HPV podem ser realizadas através da técnica de **PCR em tempo real** combinada à técnica de RFLP (*Restriction Fragment Length Polymorphism*). A PCR *Real Time* é uma técnica amplamente utilizada para estudos de expressão gênica. Além de constituir uma importante ferramenta para diagnóstico molecular clínico, ela atua na detecção de vírus ou bactérias e mensuração de suas respectivas cargas virais e bacterianas, bem como na determinação do estágio de um câncer, genotipagem e estudo de variações do número de cópias.

### Faça valer a pena

**1.** A falta de especificidade da reação pode resultar em PCRs com arrastos (fragmentos com vários tamanhos) após a eletroforese em gel, sendo a solução, nestes casos, os seguintes procedimentos:

1. Diminuição de DNA-alvo na amostra.
2. Diminuição na concentração dos oligonucleotídeos.
3. Diminuição na concentração de cloreto de magnésio.
4. Aumento da temperatura de anelamento, podendo ser mais alta nos primeiros ciclos e mais baixa nos ciclos subsequentes.
5. Diminuição na concentração de dNTPs.

Assinale a alternativa que corresponda à ordem correta dos procedimentos mais prováveis de resultar em sucesso.

- a) 2 – 4 – 1 – 3 – 5.
- b) 2 – 3 – 4 – 5 – 1.
- c) 3 – 5 – 2 – 4 – 1.
- d) 4 – 1 – 3 – 2 – 5.
- e) 5 – 4 – 3 – 2 – 1.

**2.** A análise de genomas inteiros contribui para todos os aspectos das pesquisas biológicas, como a possibilidade de localizar genes que contribuam para muitas doenças determinadas por combinações complexas de fatores genéticos. Basicamente, a obtenção de uma sequência genômica se dá em quatro etapas:

1. Continuação da sobreposição de pedaços maiores, até que todos os pequenos segmentos estejam ligados.
2. Fragmentação das moléculas de DNA do genoma em milhares a milhões de pequenos segmentos aleatórios.
3. Encontro, com o auxílio de computador, da sobreposição entre os segmentos nos quais sejam idênticas suas sequências.
4. Leitura da sequência de cada um destes segmentos.

Assinale a alternativa que corresponda à ordem correta das etapas de um sequenciamento genômico.

- a) 2 – 4 – 3 – 1.
- b) 2 – 3 – 4 – 1.
- c) 3 – 2 – 4 – 1.
- d) 4 – 1 – 3 – 2.
- e) 4 – 3 – 2 – 1.

**3.** O sequenciamento de nova geração (NGS) foi desenvolvido em meados de 2005 e evoluiu rapidamente, uma vez que promove o sequenciamento de DNA em plataformas capazes de gerar informações sobre milhões de pares de bases em uma única corrida. O sequenciamento na plataforma 454 FLX é realizado a partir de uma combinação de reações enzimáticas iniciada a partir da liberação de um pirofosfato oriundo da adição de um dNTP à cadeia.

Qual é a enzima responsável pela conversão do pirofosfato em ATP?

- a) ATP sintase.
- b) ATP sulfúrilase.
- c) Luciferase.
- d) ATP-glicose-fosfo-transferase.
- e) ATP sintetase.

## Seção 4.3

### Hibridização

#### Diálogo aberto

A probabilidade de uma mulher ser hemofílica, como é o caso de Leticia ( $X^hX^h$ ), é muito pequena, uma vez que o sexo feminino possui dois cromossomos X e, normalmente, apenas um dos cromossomos é afetado – como é o caso da mãe de Leticia ( $X^HX^h$ ), apenas portadora do gene da doença. Se Leticia se casar com um homem não afetado ( $X^HY$ ), suas filhas mulheres serão apenas portadoras ( $X^HX^h$ ), enquanto que seus filhos homens serão hemofílicos ( $X^hY$ ). Por outro lado, se Leticia se casar com um homem também hemofílico ( $X^hY$ ), tanto as filhas mulheres quanto os filhos homens serão hemofílicos ( $X^hX^h$  e  $X^hY$ , respectivamente). Em ambos os casos, é altamente recomendável que Leticia engravide através de fertilização in vitro (FIV), para que possa, antes da transferência do embrião para o útero, realizar o diagnóstico genético pré-implantacional (PGD), exame que detecta uma condição genética específica, como a hemofilia. Qual é a técnica de biologia molecular utilizada neste procedimento? Em que se baseia esta técnica?

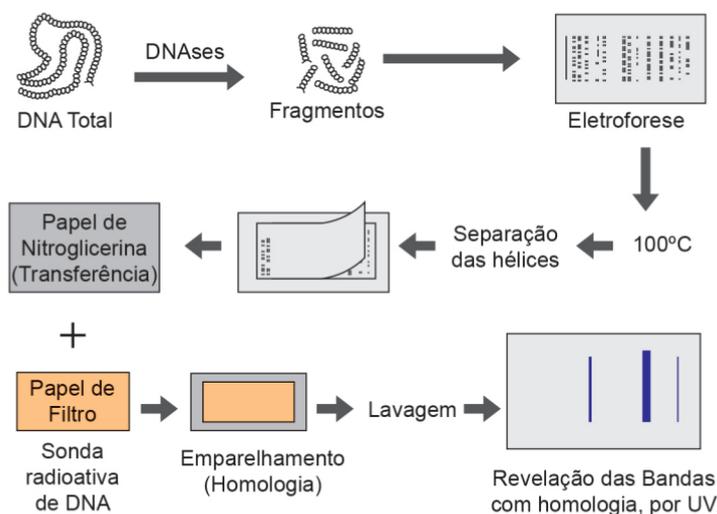
#### Não pode faltar

O grau de semelhança entre sequências de DNA ou RNA pode ser determinado através de técnicas de *hibridização*, normalmente empregadas para determinar a distância gênica entre duas espécies, mas também para analisar genes expressos e não expressos. A análise dos genes de interesse acontece pelo emparelhamento de sequências de DNA complementares a uma sonda (probe), constituída por uma sequência de nucleotídeos complementares, desenvolvida a partir de segmentos conhecidos do DNA ou RNA que se deseja identificar e marcada por radioatividade, quimioluminescência ou anticorpos. A utilização dessas sondas possibilita a caracterização e identificação do DNA baseado em

suas propriedades, destacando-se o emparelhamento de bases através de pontes de hidrogênio e a desnaturação/renaturação das cadeias complementares. As técnicas da biologia molecular e biotecnologia que utilizam a hibridização como princípio do método são diversas: *southern blotting*, *western blotting*, *northern blotting*, hibridização fluorescente in situ (FISH), dentre outras.

O ***southern blotting*** se baseia na hibridação entre moléculas de DNA (fixas em um suporte) com sondas marcadas, visando a localização de sequências específicas de DNA. A técnica é possível devido à estrutura em duplas fitas das moléculas de DNA, as quais, quando aquecidas ou submetidas a um tratamento com soluções desnaturantes, tendem a se separar. As moléculas desnaturadas, uma vez submetidas a um resfriamento em condições controladas, tendem a se renaturar novamente. Desta forma, a hibridação entre moléculas de DNA de diferentes origens é possível a partir do momento em que tais DNAs apresentam regiões complementares de sequências significantes.

A técnica de *southern blotting* consiste em uma modificação da técnica de RFLP (*Restriction Fragment Length Polymorphism*): após a clivagem da molécula de DNA pelas enzimas de restrição e da separação por eletroforese em gel de agarose, o DNA é imerso em solução alcalina para desnaturar os fragmentos de DNA de fita dupla, os quais, posteriormente, são transferidos para uma membrana de náilon ou nitrocelulose a qual são aderidos, criando uma impressão ou marca. Os fragmentos são aderidos à membrana em posições equivalentes às que eles assumiram no gel após a eletroforese. A membrana é hibridizada com uma sonda de DNA fria ou radioativa, e as condições de hibridação são definidas com base no grau de similaridade entre o DNA da amostra e o DNA da sonda. Se a sonda utilizada for fria, a revelação ocorre por imunoensaio e precipitação de corante na própria membrana, enquanto que se for radioativa, ocorre por exposição da membrana a um filme de raios-X (Figura 4.6).

Figura 4.6 | Técnica *southern blotting*

Fonte: <<http://www.scielo.br/img/revistas/sp/v33n1/01f4.gif>>. Acesso em: 1 set. 2017.

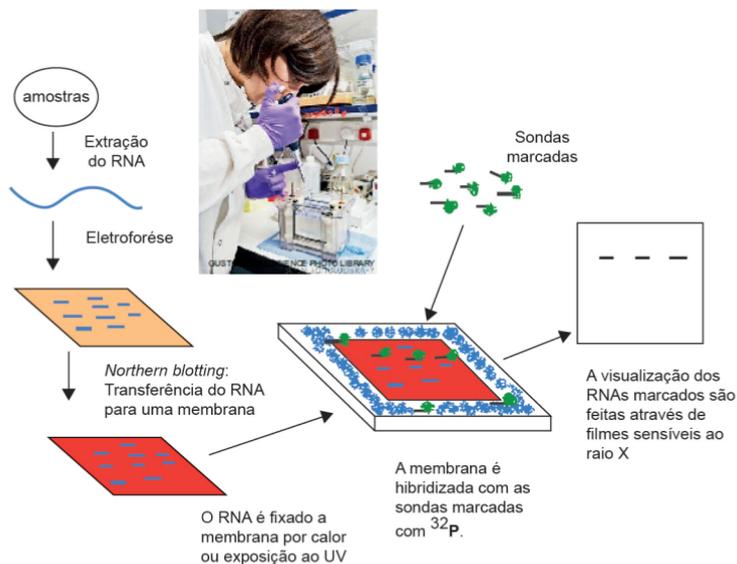
Já o **western blotting** consiste na imunodeteção de proteínas em filtro de nitrocelulose, rotineiramente utilizado para detecção de proteínas e em extratos proteicos, detecção de produtos de degradação proteica, caracterização de soros e anticorpos monoclonais, varredura e caracterização de polipeptídeos recombinantes, determinação de massa molecular, caracterização de epítopos, dentre outros. A técnica consiste na detecção de pequenas quantidades de proteínas, as quais se encontram adsorvidas em uma membrana de nitrocelulose, através da reação com anticorpos específicos desenvolvidos para o reconhecimento do polipeptídeo em análise. O anticorpo é adicionado à membrana e ali reage por um determinado período de tempo, após o qual uma fração fica retida na região da membrana que contém o polipeptídeo específico.

Em uma segunda etapa, o anticorpo adsorvido especificamente sobre o polipeptídeo pode ser detectado com um antissoro, isto é, um antianticorpo específico (segundo anticorpo ou anticorpo conjugado) quimicamente modificado através do acoplamento covalente de uma ou mais moléculas de uma enzima como a fosfatase alcalina, por exemplo. Este reagente atua na detecção do anticorpo específico e permite sua visualização, em razão da atividade enzimática associada. Com o auxílio de substratos sintéticos para a enzima conjugada, torna-se possível a visualização da região do filtro no qual o polipeptídeo do ensaio está localizado, uma vez que o substrato utilizado é capaz de precipitar e formar uma camada colorida no local exato no qual se encontra o segundo anticorpo e, portanto, o anticorpo específico, conseqüentemente, do polipeptídeo que se deseja detectar.

O ***northern blotting***, por sua vez, é utilizado para a detecção de moléculas de RNA separadas por eletroforese e imobilizadas em um suporte, permitindo a análise da expressão de genes em diferentes estágios de desenvolvimento ou de diferentes tipos celulares – uma vez que são imobilizadas moléculas de RNA, incluindo os RNAm. A principal diferença em relação às técnicas anteriormente citadas consiste no fato de as moléculas de RNA serem polímeros de fita simples. Assim, o pareamento intramolecular de bases é frequente e promove um dobramento da molécula ou o pareamento de bases presentes em moléculas distintas, formando regiões de híbridos RNA-RNA. Dessa forma, a migração do RNA em géis de agarose não depende somente de seu tamanho, mas também do estado conformacional da molécula (Figura 4.7). Como esta técnica é realizada com um excesso de sonda de DNA, a quantidade de hibridização se relaciona à quantidade de RNAm presente na amostra original, permitindo que as quantidades relativas dos RNAm sejam determinadas.



Figura 4.7 | Técnica *northern blotting*

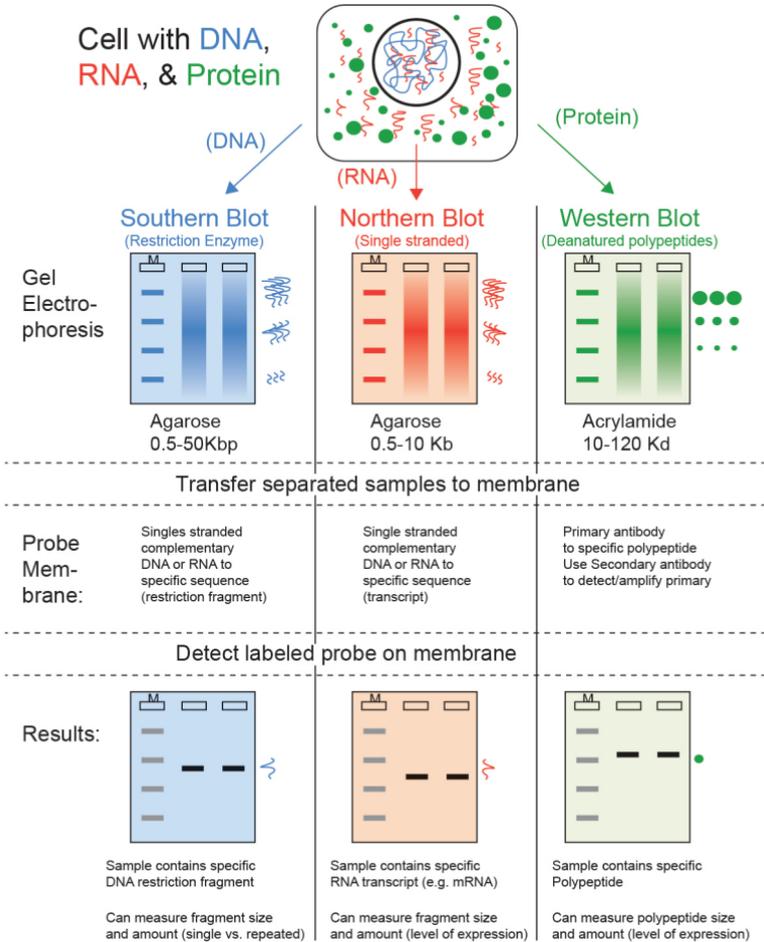


Fonte: <<https://goo.gl/uM9fnD>>. Acesso em: 1 set. 2017.



A figura e a tabela a seguir mostram uma comparação entre as três técnicas.

Figura 4.8 | Comparação entre as técnicas *southern blotting*, *northern blotting* e *western blotting*



Fonte: <<https://bio.libretexts.org/@api/deki/files/5357/Fig8.16.png?revision=1>>. Acesso em: 1 set. 2017.

Quadro 4.1 | Comparação entre as técnicas *southern blotting*, *northern blotting* e *western blotting*

	<i>Southern blot</i>	<i>Northern blot</i>	<i>Western blot</i>
Gel	Agarose	Agarose	Acrilamida
Probe	DNA ou RNA fita simples complementar a uma sequência específica (fragmento de restrição).	DNA ou RNA fita simples complementar a uma sequência específica (transcrito).	Anticorpo primário para polipeptídeo específico, seguido de segundo anticorpo para detecção/ amplificação do primário.
Resultados	Amostra contém fragmentos de restrição de DNA específicos.  Pode mensurar o tamanho do fragmento e a quantidade.	Amostra contém transcritos de RNA específicos.  Pode mensurar o tamanho do fragmento e a quantidade.	Amostra contém polipeptídeos específicos.  Pode mensurar o tamanho do polipeptídeo e a quantidade.

Fonte: elaborado pelo autor.

Para se clonar um gene, é preciso identificar seus fragmentos correspondentes entre todos os clones armazenados em uma biblioteca. Este procedimento é realizado utilizando-se de uma sonda de DNA, cuja sequência é complementar a uma parte do gene de interesse, podendo esta sonda ser utilizada para identificar as colônias de células que tenham os clones contendo aquela região do gene. Este processo é denominado *hibridização em colônia*, no qual uma biblioteca de DNAc (DNA complementar) conterà milhares de insertos diferentes, cada qual contido em um vetor comum. Após a transformação de uma linhagem bacteriana hospedeira apropriada à biblioteca, as células são distribuídas sobre placas de Petri contendo meio de cultura sólido. Desta forma, cada célula se multiplica e origina colônias de células isoladas, e cada célula pertencente a uma determinada colônia contém o mesmo vetor e o mesmo inserto da biblioteca.

Posteriormente, o mesmo tipo de membrana positivamente carregada das técnicas de *southern* e *western blotting* é utilizado

para fixar pequenas quantidades de DNA para incubação com sonda. Neste caso, a membrana é pressionada em cima das placas com as colônias, e uma parte das células de cada colônia, bem como seu DNA, é transferida para a membrana. Assim, a membrana retém uma amostra de cada clone de DNA posicionado na membrana, em um padrão semelhante ao padrão de colônias da placa, garantindo que o clone desejado, uma vez identificado pela incubação da membrana com a sonda, possa ser facilmente identificado através da colônia de células que o contém, e que o plasmídeo contendo o inserto de DNA apropriado possa ser purificado.



Refleta

Como aconteceram os grandes avanços na Biologia Molecular?

Os grandes avanços na Biologia Molecular aconteceram por causa dos diversos experimentos realizados com fagos e bactérias. Os trabalhos pioneiros com a bactéria pneumococo, por exemplo, resultaram na descoberta de que o material genético era o DNA. Como vimos, diversas técnicas empregadas na biologia molecular e na biotecnologia utilizam estes microrganismos para a clivagem do DNA essencial às técnicas de extração e clonagem de DNA, como vetores na técnica de DNA recombinante, dentre diversas outras aplicações.

As bactérias podem ser analisadas através de ferramentas da citologia, como a microscopia por imunofluorescência (para localização de proteínas em células fixadas com anticorpos específicos), microscopia com fluorescência utilizando Proteína Fluorescente Verde (para localização de proteínas em células vivas) e **hibridização in situ por fluorescência (FISH – fluorescence in situ hybridization)**, para localização de regiões cromossômicas e de plasmídeos nas células.

A FISH permite a localização física precisa de genes ou sequências de DNA diretamente nos cromossomos presentes em preparações citológicas. A técnica, aplicada em metáfases humanas e núcleos interfásicos, utilizando sondas de DNA, tornou-se uma ferramenta essencial nos estudos do genoma humano, permitindo uma localização cromossômica regional bastante precisa de genes

de cópia única ou sequências repetidas de DNA. Ainda, métodos de marcação não isotópica de sondas de DNA permitem que a FISH seja realizada em qualquer laboratório de citogenética.

A hibridização *in situ* oferece diversas perspectivas para o mapeamento gênico, permitindo atualmente o mapeamento de qualquer gene (ou sequência de DNA) que já tenha sido clonado. Uma das vantagens é a possibilidade de detectar alterações tanto em metáfases e pró-metáfases quanto em núcleos interfásicos, amnióticos; inclusive em material fixado em parafina.

Desta forma, a técnica se consolidou como ferramenta de fundamental importância para a obtenção de um diagnóstico rápido das alterações cromossômicas numéricas e das doenças genéticas ligadas ao sexo; em líquido amniótico, vilos coriais e restos ovulares a fresco ou fixados em parafina. Dentre as possíveis aplicações da FISH, estão o diagnóstico precoce das aneuploidias cromossômicas numéricas, determinação do sexo fetal, determinação de doenças genéticas em casos específicos e identificação do cromossomo *philadelphia* presente na leucemia mieloide crônica.

Baseadas nos mesmos princípios, outras duas técnicas foram desenvolvidas para ampliar o poder da técnica de FISH na análise dos cromossomos humanos: *hibridação genômica comparativa* (CGH – *comparative genomic hybridization*) e cariotipagem espectral (SKY). A CGH possibilita o aumento da resolução na detecção de alterações cromossômicas e melhora do diagnóstico sem causa definida, como retardo mental, diferenciação secundária feminina, menopausa precoce e alguns tumores. Atualmente, a técnica tem sido bastante utilizada na análise de doenças desenvolvidas em consequência de acidentes nucleares que causam danos cromossômicos não detectados pela análise cariotípica simples.

A CGH permite analisar, em um único experimento, alterações do número de cópias de DNA de todo o genoma, utilizando, para tanto, uma amostra de DNA do paciente e uma amostra de DNA controle (cariotipicamente normal). As alterações são classificadas como perdas e ganhos de segmentos cromossômicos, em um nível de detalhamento 100 vezes maior que a citogenética clássica. Por convenção, a técnica baseia-se na hibridação competitiva *in situ* entre tecido normal e tecido de interesse, nos quais as regiões

alteradas do DNA são quantificadas ao longo dos cromossomos na metáfase.

As imagens são capturadas por câmara CCD e avaliadas em um programa especial de análise de imagens. A intensidade do sinal de diferentes fluorocromos é quantificada em diferentes tons de cinza ao longo de um único cromossomo. Os segmentos gênicos são quantificados em perda ou ganho de informações, os quais podem ser duplicações ou deleções (perdas) e ampliações (ganhos) de segmentos do DNA. Nas deleções, as razões de fluorescência são menores devido à ausência de sequências do DNA genômico encontrado em uma região específica do cromossomo, e são marcadas em verde. Já as ampliações apresentam razão de fluorescência maior, devido ao número de cópias de DNA encontradas em regiões específicas quando comparadas com o DNA controle, sendo marcadas em vermelho. Por fim, as regiões nas quais não houve perdas ou ganhos são destacadas em azul.

A exata identificação da região cromossômica deletada ou duplicada é fundamental para estabelecer a relação genótipo-fenótipo. A cariotipagem espectral (SKY) foi desenvolvida objetivando colorir simultaneamente cada cromossomo de uma cor, utilizando sondas de "pintura cromossômica" de cada um dos cromossomos humanos, marcadas por combinações de fluorocromos distintos. As 24 sondas desenvolvidas são misturadas em um coquetel, pré-hibridizadas com DNA rico em sequências repetitivas (evitando hibridizações cruzadas) e, então, as metáfases do DNA da amostra são hibridizadas. As imagens são capturadas por um microscópio de fluorescência, passando por um interferômetro, e os dados são analisados em um software específico, o qual gera uma imagem em que cada cromossomo é marcado com uma cor diferente.

A limitação desta técnica consiste na incapacidade de detectar rearranjos intracromossômicos, como inversões e deleções. Ainda, a faixa de detecção é limitada pela resolução do cariótipo analisado, e os centrômeros e braços curtos dos cromossomos acrocêntricos não sofrem hibridização devido à supressão de sequências repetitivas. A SKY tem sido bastante utilizada no esclarecimento de alterações citogenéticas previamente detectadas no cariótipo com bandeamento G (técnica para detecção de possíveis alterações numéricas e/ou estruturais em cromossomos obtidos de diferentes

passagens de culturas de células-tronco mesenquimais), incluindo cromossomos marcadores, translocações e rearranjos complexos.



### Pesquise mais

Saiba mais a respeito da importância da citogenética no diagnóstico de doenças causadas por alterações cromossômicas através da leitura do artigo a seguir.

CHAVES, T. F.; NICOLAU, L. S. Citogenética e cariotipagem humana. **Revista Saúde e Desenvolvimento**, [S.l.], v. 4, n. 2, p. 57-66, 2013. Disponível em: <<https://www.uninter.com/revistasaude/index.php/saudeDesenvolvimento/article/view/229/189>>. Acesso em: 1 set. 2017.

### Sem medo de errar

O diagnóstico genético pré-implantacional (PGD) é um exame que consiste na biópsia de embriões pós-fertilização *in vitro*, no qual algumas células são retiradas para análise e diagnóstico de alterações cromossômicas antes da implantação do embrião na cavidade uterina, podendo também ser utilizado para identificação do sexo fetal. No PGD, comumente é empregada a **hibridização *in situ* por fluorescência (FISH)**, técnica utilizada no diagnóstico precoce das aneuploidias cromossômicas numéricas, determinação do sexo fetal, determinação de doenças genéticas em casos específicos, identificação do cromossomo *philadelphia* presente na leucemia mieloide crônica, dentre outros. A FISH permite a localização física precisa de genes ou sequências de DNA diretamente nos cromossomos presentes em preparações citológicas, aplicada em metáfases humanas e núcleos interfásicos, utilizando sondas de DNA. A técnica oferece diversas perspectivas para o mapeamento gênico, permitindo, atualmente, o mapeamento de qualquer gene (ou sequência de DNA) que já tenha sido clonado. Uma das vantagens é a possibilidade de detectar alterações tanto em metáfases e pró-metáfases quanto em núcleos interfásicos, amnióticos, inclusive em material fixado em parafina. A hibridação genômica comparativa (CGH), uma técnica recente baseada nos mesmos princípios da FISH, permite a análise de todos os cromossomos de um embrião, auxiliando em diagnósticos que até então não eram possíveis.

### Desvendando as técnicas

#### Descrição da situação-problema

Um estagiário do Laboratório de Citogenética Humana e Citogenômica, da Faculdade de Ciências Médicas da Universidade Estadual de Campinas (UNICAMP), em sua tarde de organização das geladeiras do laboratório, encontrou alguns frascos contendo sondas de DNA (probes) sem identificação de qual sonda se tratava. No frasco 1, estava escrito "fragmento de restrição"; no frasco 2, "transcrito" e, no frasco 3, "anticorpo primário para polipeptídeo". A que técnicas empregadas em biologia molecular se relacionavam as sondas de DNA contidas nos frascos 1, 2 e 3, respectivamente?

#### Resolução da situação-problema

O frasco 1 continha sondas compostas por DNA ou RNA fita simples complementar a uma sequência específica (fragmento de restrição), utilizadas na técnica de **southern blotting**. O frasco 2 continha sondas compostas por DNA ou RNA fita simples complementar a uma sequência específica (transcrito), utilizadas na técnica de **northern blotting**. Por fim, o frasco 3 continha sondas compostas por anticorpos primários para polipeptídeo específicos, utilizadas na técnica de **western blotting**. O *southern blotting* se baseia na hibridação entre moléculas de DNA (fixas em um suporte) com sondas marcadas, visando a localização de sequências específicas de DNA. Já o *western blotting* consiste na imunodeteção de proteínas em filtro de nitrocelulose; rotineiramente utilizado para detecção de proteínas e em extratos proteicos, detecção de produtos de degradação proteica, caracterização de soros e anticorpos monoclonais, varredura e caracterização de polipeptídeos recombinantes, determinação de massa molecular, caracterização de epítomos, dentre outros. O *northern blotting*, por sua vez, é utilizado para a detecção de moléculas de RNA separadas por eletroforese e imobilizadas em um suporte, permitindo a análise da expressão de genes em diferentes estágios de desenvolvimento ou de diferentes tipos celulares – uma vez que são imobilizadas moléculas de RNA, incluindo os RNAm.

## Faça valer a pena

**1.** Utilizada para a detecção de moléculas de RNA separadas por eletroforese e imobilizadas em um suporte, permitindo a análise da expressão de genes em diferentes estágios de desenvolvimento ou de diferentes tipos celulares – uma vez que são imobilizadas moléculas de RNA, incluindo os RNAm.

O texto anterior faz referência à qual técnica comumente empregada na biologia molecular?

- a) PCR.
- b) Sequenciamento de DNA.
- c) *Southern blotting*.
- d) *Western blotting*.
- e) *Northern blotting*.

**2.** O grau de semelhança entre sequências de DNA ou RNA pode ser determinado através de técnicas de *hibridização* normalmente empregadas para determinar a distância gênica entre duas espécies, mas também para analisar genes expressos e não expressos.

Assinale a alternativa que contém, dentre as propriedades do DNA, aquelas que são essenciais à técnica de hibridização.

- a) Grande estabilidade em meio aquoso e replicação autônoma *in vitro*.
- b) Complementaridade das bases e universalidade das unidades monoméricas.
- c) Capacidade de hibridização com RNA e susceptibilidade a agentes mutagênicos.
- d) Estrutura unifilar e capacidade de replicação.
- e) Sequências lineares idênticas e mesma proporção de nucleotídeos em toda a extensão da molécula.

**3.** As técnicas de biologia molecular e biotecnologia que utilizam a hibridização como princípio do método são diversas: *southern blotting*, *western blotting*, *northern blotting*, hibridização fluorescente *in situ* (FISH), dentre outras. O *southern blotting* se baseia na hibridação entre moléculas de DNA (fixas em um suporte) com sondas marcadas, visando a localização de sequências específicas de DNA.

Assinale a alternativa que contém informações corretas a respeito da técnica de *southern blotting*.

- a) Os fragmentos de DNA obtidos após clivagem com enzimas de restrição são separados por eletroforese em gel de agarose e corados

com brometo de etídio antes de serem hibridizados em membrana de nitrocelulose.

b) Após separação por eletroforese em gel de poliacrilamida, as bandas obtidas dos fragmentos de DNA são cortadas e transferidas para uma membrana de nitrocelulose.

c) O *southern blotting* não apresenta sensibilidade suficiente para detectar polimorfismos relacionados a alterações de clivagem, quando estas são muito pequenas.

d) A principal diferença entre o *southern blotting* e o *western blotting* é que o primeiro utiliza gel de poliacrilamida e o segundo, gel de agarose.

e) O *southern blotting* consiste na imunodeteção de proteínas em filtro de nitrocelulose.

# Referências

ALBERTS, B. et al. Análise de genes e genomas. In: \_\_\_\_\_. **Fundamentos da Biologia Celular**. 3. ed. Porto Alegre: Artmed, 2011. cap. 10, p. 327-336.

CALMARIA TEMPESTADE. **Técnica northern blotting**. 2017. Disponível em: <[https://calmariatempestade.files.wordpress.com/2012/06/northern\\_blot\\_eletroforese.png](https://calmariatempestade.files.wordpress.com/2012/06/northern_blot_eletroforese.png)>. Acesso em: 1 set. 2017.

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, Santa Maria, v. 40, n. 3, p. 735-744, 2010. Disponível em: <<https://repositorio.unesp.br/bitstream/handle/11449/29442/S0103-84782010000300040.pdf?sequence=1&isAllowed=y>>. Acesso em: 1 set. 2017.

CHAVES, T. F.; NICOLAU, L. S. Citogenética e cariotipagem humana. **Revista Saúde e Desenvolvimento**, [S.l.], v. 4, n. 2, p. 57-66, 2013. Disponível em: <<https://www.uninter.com/revistasaude/index.php/saudeDesenvolvimento/article/view/229/189>>. Acesso em: 1 set. 2017.

GRIFFITHS, A. J. F. et al. Genomas e genômica. In: \_\_\_\_\_. **Introdução à Genética**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2015. cap. 14, p. 424.

MALACINSKI, G. M. Transcrição. In: \_\_\_\_\_. **Fundamentos de Biologia Celular**. Rio de Janeiro: Guanabara Koogan, 2011. cap. 8, p. 148-149.

\_\_\_\_\_. DNA recombinante e engenharia genética: moldagem molecular dos genes. In: \_\_\_\_\_. **Fundamentos de Biologia Celular**. Rio de Janeiro: Guanabara Koogan, 2011. cap. 15, p. 312-330.

MATIOLI, S. R.; FERNANDES, F. M. C. (Eds.). Métodos baseados em PCR para análise de polimorfismos de ácidos nucleicos. In: \_\_\_\_\_. **Biologia Molecular e Evolução**. Ribeirão Preto: Holos, Editora / Sociedade Brasileira de Genética, 2012. cap. 19, p. 181-190.

NZYTECH. **Institucional**. 2017. Disponível em: <<https://www.nzytech.com>>. Acesso em: 4 set. 2017.

PASTERNAK, J. J. Tecnologia do DNA recombinante. In: \_\_\_\_\_. **Genética Molecular Humana: mecanismos das doenças hereditárias**. São Paulo: Manole, 2002. cap. 5, p. 133-141.

\_\_\_\_\_. Tecnologia do DNA recombinante. In: \_\_\_\_\_. **Genética Molecular Humana: mecanismos das doenças hereditárias**. São Paulo: Manole, 2002. cap. 05, p. 141-144.

SIDEPLAYER. **Tecnologia do DNA recombinante**. 2017. Disponível em: <[http://images.slideplayer.com.br/1/288898/slides/slide\\_6.jpg](http://images.slideplayer.com.br/1/288898/slides/slide_6.jpg)>. Acesso em: 1 set. 2017.

SIELO. **Técnica southern blotting**. 2017. Disponível em: <<http://www.scielo.br/img/revistas/sp/v33n1/01f4.gif>>. Acesso em: 1 set. 2017.

SIGMA-ALDRICH. **Clivagem das enzimas de restrição Eco RI e Rsa I**. 2017. Disponível

em: <[http://www.sigmaaldrich.com/content/dam/sigma-aldrich/product5/192/ecor-i.eps/\\_jcr\\_content/renditions/ecor-i-medium.jpg](http://www.sigmaaldrich.com/content/dam/sigma-aldrich/product5/192/ecor-i.eps/_jcr_content/renditions/ecor-i-medium.jpg)>. Acesso em: 1 set. 2017.

STARR, C. et al. Estudo e manipulação de genomas. In: \_\_\_\_\_. **Biologia: unidade e diversidade da vida**. São Paulo: Cengage Learning, 2012. v. 1, cap. 7, p. 92-94.

WATSON, J. D. et al. Técnicas de biologia molecular. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 20, p. 647-655.

\_\_\_\_\_. Técnicas de biologia molecular. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 20, p. 651-657.

\_\_\_\_\_. Técnicas de biologia molecular. In: \_\_\_\_\_. **Biologia Molecular do Gene**. 5. ed. Porto Alegre: Artmed, 2006. cap. 20, p. 658-666.

ZATZ, M. Clonagem humana: conhecer para opinar. **Pesquisa FAPESP**, São Paulo, n. 73, p. 8-14, mar. 2002. Disponível em: <[http://www.revistapesquisa.fapesp.br/wp-content/uploads/2002/03/08\\_PERGUNTASF.pdf](http://www.revistapesquisa.fapesp.br/wp-content/uploads/2002/03/08_PERGUNTASF.pdf)>. Acesso em: 1 set. 2017.











ISBN 978-85-522-0248-6



9 788552 202486 >