

# Como nascem os genes?



Imagem gerada pelo DALL-E, uma ferramenta de inteligência artificial desenvolvida pela OpenAI

**Diego Trindade de Souza<sup>1</sup>, Maria Dulcetti Vibranovski<sup>2</sup>, Sergio Russo Matioli<sup>2</sup>**

<sup>1</sup>Instituto SENAI de Inovação – Biotecnologia, São Paulo, SP

<sup>2</sup>Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Campus do Butantã, São Paulo, SP

Autor para correspondência - Maria Dulcetti Vibranovski, mdv@ib.usp.br

**Palavras-chave:** origem de genes, duplicação gênica, embaralhamento de éxons, retrotransposição

Os genes são transmitidos de uma célula para outra e para a próxima geração pelo processo de replicação do material genético, o DNA. Entretanto, neste artigo, abordaremos a questão do surgimento dos genes do ponto de vista evolutivo, ou seja, o aparecimento de um gene novo que não é herdado dos ancestrais. Os mecanismos pelos quais novos genes podem surgir incluem a duplicação gênica, a duplicação genômica, a **retrotransposição**, a transferência gênica horizontal e o “embaralhamento de éxons”. Além disso, genes podem surgir “do zero”, a partir de regiões genômicas que anteriormente não possuíam qualquer função.

Embora a definição de gene tenha sofrido várias modificações conceituais, desde os fatores mendelianos da Genética clássica até o advento da Biologia Molecular e da Genômica, neste artigo, para fins didáticos, restringiremos a definição de gene para: um trecho de uma molécula de DNA que resulta na transcrição de um RNA e que, posteriormente, é traduzido em um polipeptídeo. Apesar dessa restrição, os mecanismos considerados podem estar envolvidos na evolução de outras categorias de elementos genéticos.

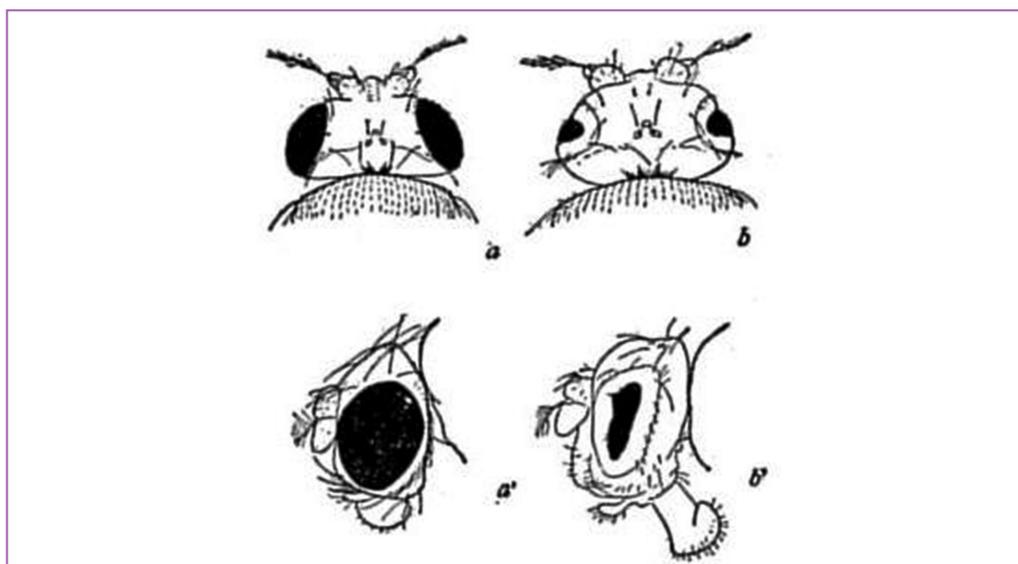
## Evolução dos genes por duplicação gênica

A ideia de que novos genes podem surgir a partir de genes existentes surgiu praticamente junto com a Teoria Cromossômica

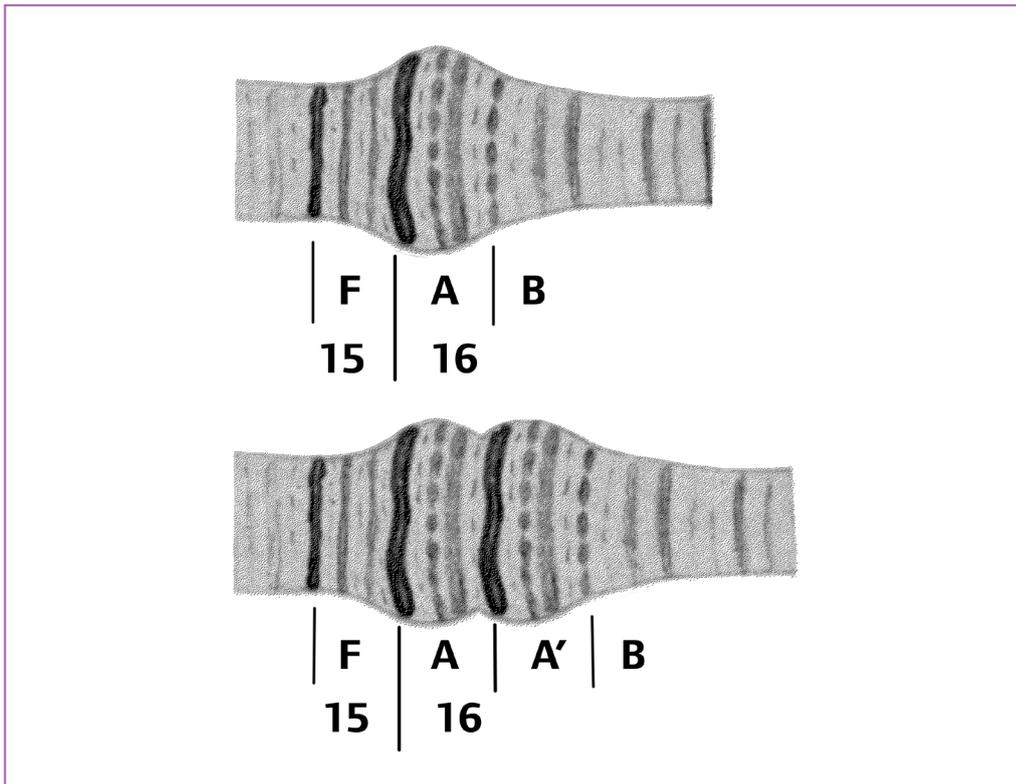
da Herança, no início do século XX. Em 1914, foi descrita uma mutação no cromossomo X de *Drosophila melanogaster* que conferia aos olhos de machos hemizigóticos e fêmeas homozigóticas um aspecto de barra (Figura 1). Por esse motivo, a mutação foi denominada Bar (barra em inglês). Mais tarde, verificou-se que a mutação mapeava em uma região do cromossomo X dessa espécie, que apresentava uma repetição das bandas evidenciadas nos cromossomos politênicos de suas larvas, indicando a duplicação de uma região cromossômica (Figura 2). A partir desse tipo de evidência, o biólogo evolucionista inglês J.B.S. Haldane, em 1932, postulou que genes novos se originavam desta maneira, a partir de genes preexistentes. Atualmente nós sabemos que, nas populações humanas atuais, o número de cópias de alguns genes nos genomas apresenta variações, fenômeno conhecido como variação no número de cópias (*copy number variation*, em inglês).

**Éxon** - trecho da sequência de uma molécula de DNA que é transcrito e traduzido, que é separado de outros trechos que não são traduzidos, os íntrons, no processo de *splicing* e que, portanto, permanece no RNAm maduro que será traduzido em polipeptídeo.

**Retrotransposição** - fenômeno no qual um elemento transponível integrado no genoma, inicialmente transcrito em um RNA, por meio da ação de uma transcriptase reversa (enzima que catalisa a síntese de uma molécula de DNA tendo como molde uma molécula de RNA) acaba tendo uma cópia de sua sequência incorporada em outra posição no genoma de um organismo.



**Figura 1.** Aspecto dos olhos de *Drosophila melanogaster*, à esquerda (a e a') fenótipo selvagem e à direita (b e b') mutante Bar em fêmea homozigótica. Fonte: Morgan, domínio público. ([https://commons.wikimedia.org/wiki/File:Critique\\_of\\_the\\_Theory\\_of\\_Evolution\\_Fig\\_031.jpg](https://commons.wikimedia.org/wiki/File:Critique_of_the_Theory_of_Evolution_Fig_031.jpg)).

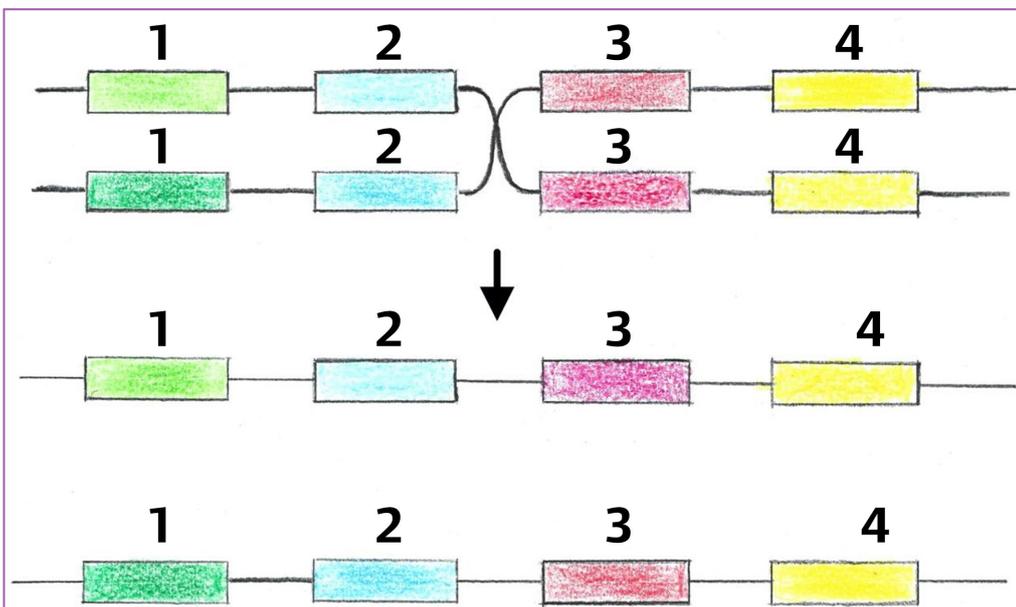


**Figura 2.** Esquema dos cromossomos politênicos de *Drosophila melanogaster*, mostrando o trecho duplicado no cromossomo X que corresponde à mutação Bar. Fonte: modificado a partir de [http://ftp.flybase.org/flybase/associated\\_files/Maps/Dmel\\_Bridges/Dmel\\_X\\_Bridges.png](http://ftp.flybase.org/flybase/associated_files/Maps/Dmel_Bridges/Dmel_X_Bridges.png). Mapa dos cromossomos de *Drosophila melanogaster*. (Journal of Heredity 29(1):12 suplemento, 1938).

## Origem das duplicações gênicas

Durante o processo de recombinação genética, que ocorre durante a meiose dos organismos diploides com reprodução sexuada,

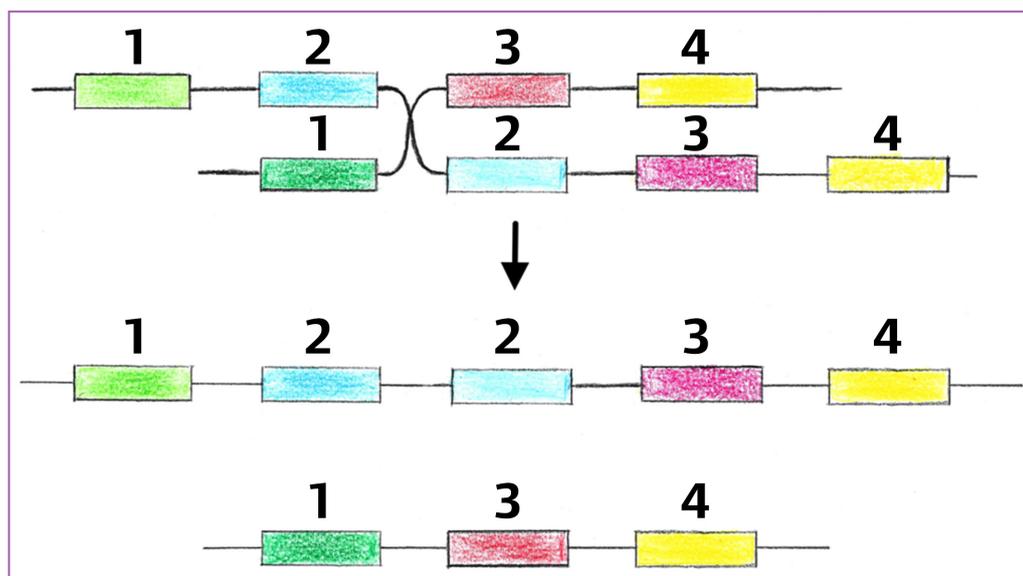
há o emparelhamento dos cromossomos homólogos. Esse emparelhamento é muito preciso e, em princípio, poderia ser mapeado nucleotídeo a nucleotídeo. Durante o emparelhamento dos cromossomos homólogos, pode haver a formação do *crossing-over*, que é o processo pelo qual uma fita dupla de DNA de uma cromátide liga-se à fita dupla de outra cromátide (Figura 3).



**Figura 3.** Emparelhamento correto, durante a meiose, entre cromátides homólogas, mostrando o fenômeno de *crossing-over* com as cromátides recombinantes que resultaram.

Ocasionalmente, o emparelhamento das fitas duplas de DNA das cromátides homólogas pode não ser preciso, havendo um deslocamento local dos genes de uma determinada região de uma cromátide em relação à outra. Se, na região do emparelhamento deslocado,

houver um *crossing-over*, os cromossomos resultantes serão diferentes em termos das cópias dos genes nos cromossomos. Um dos cromossomos resultantes terá um trecho duplicado e outro terá esse trecho ausente (Figura 4).



**Figura 4.** Emparelhamento desigual, não homólogo, durante a meiose, entre cromátides homólogas, com a ocorrência de *crossing-over*. As cromátides resultantes apresentam quantidades diferentes de genes, em que uma delas possui duas cópias do gene 2 e a outra tem a ausência deste gene.

O destino dos cromossomos resultantes desse processo dependerá do efeito da ausência ou da redundância dos genes envolvidos no fenótipo dos indivíduos portadores de tais cromossomos. Caso o trecho “perdido” contenha um gene essencial para a sobrevivência dos indivíduos da espécie em questão, os cromossomos com tal deleção tenderão a ser eliminados da população. Por outro lado, se essa duplicação local desse trecho conferir uma vantagem para os indivíduos portadores, a tendência será de que o trecho duplicado aumente de frequência durante as gerações subsequentes.

Ao longo do tempo, cada uma das cópias do gene duplicado pode acumular modificações, resultado do processo de mutação, a ponto de passarem a ter funções diferentes. Existem muitos exemplos de genes que se acredita terem surgido a partir de duplicações gênicas. Um exemplo clássico é o caso das **globinas** dos vertebrados. Nas **hemácias** dos vertebrados, a hemoglobina é constituída por quatro cadeias proteicas sendo duas cadeias alfa e duas cadeias beta.

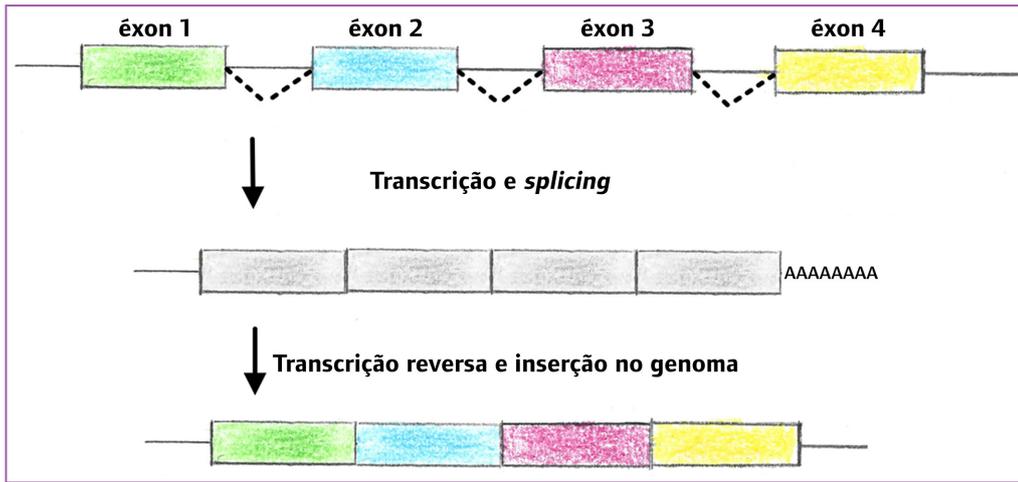
As seqüências de resíduos de aminoácidos das cadeias são semelhantes em 42% delas. A probabilidade de que tais cadeias tenham surgido de forma independente é tão ínfima que isso sequer é levado em consideração!

Na Figura 5, um alinhamento das seqüências de resíduos de aminoácidos com as cadeias alfa e beta das globinas humanas está representado, no qual se pode verificar a grande semelhança entre essas cadeias. No caso das hemoglobinas humanas, durante as fases embrionária e fetal, as cadeias das globinas expressas são diferentes daquelas que estão presentes após o nascimento. Ao analisar as propriedades bioquímicas das hemoglobinas com globinas embrionárias e fetais, percebe-se que existem diferenças nas afinidades com as quais elas se ligam ao oxigênio. Essas diferenças são resultado de adaptações às necessidades respiratórias dessas fases, com o embrião e o feto realizando trocas gasosas com o sangue da mãe, enquanto, após o nascimento, as trocas gasosas são realizadas no nível pulmonar.

**Globina** - macromolécula proteica solúvel cujo nome se refere ao seu formato globular, que faz parte da hemoglobina. Várias pesquisas importantes foram feitas com as globinas devido à facilidade de purificar a hemoglobina, que é vermelha.

**Hemácia** - também conhecida como célula vermelha do sangue, possui grandes quantidades de hemoglobina, proteína com grande afinidade por oxigênio molecular com efeito importante nas trocas gasosas entre os tecidos.

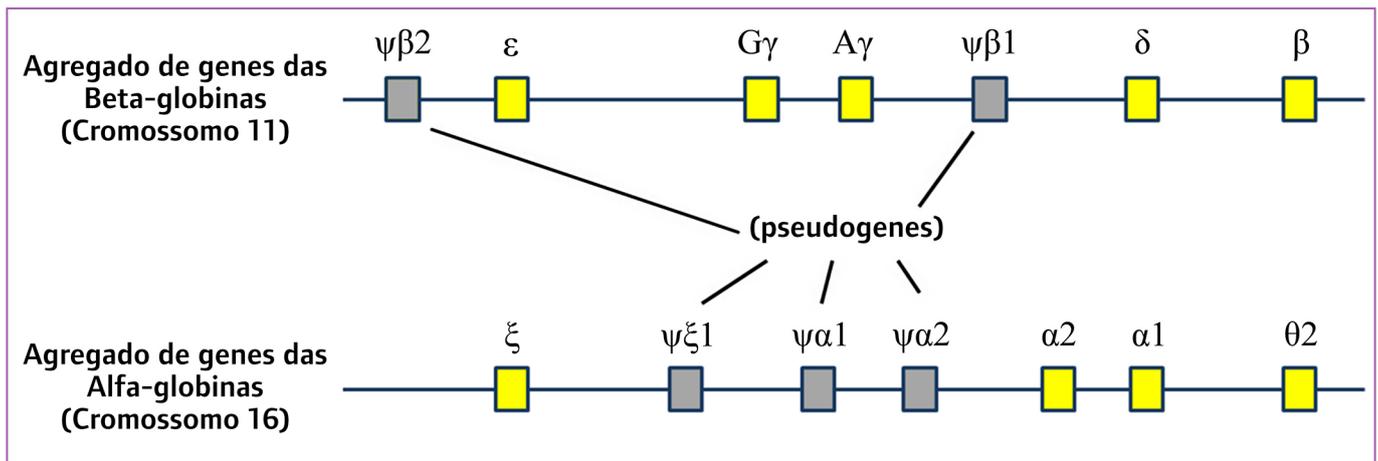




**Figura 6.** Mecanismo de origem de um retrogene. A presença ocasional de uma transcriptase reversa pode fazer com que haja a síntese de uma molécula de DNA que, se incorporada ao genoma, gera um retrogene. Eles são caracterizados pela ausência de regiões intrônicas e com repetições de adeninas, característica de rNAm. Fonte: baseada em figura de Vibranovski e Long (2016), com permissão dos autores.

Os genes que se originam pelo processo de duplicação gênica e permanecem no mesmo genoma são conhecidos como genes parálogos, que é um tipo de homologia em que as

cópias não divergem ao longo da formação e diversificação das espécies (como no caso dos genes ortólogos), mas divergem coexistindo no mesmo genoma.



**Figura 7.** Localização, no genoma humano, de genes que compõem a família das globinas. Na hemoglobina de adultos, as cadeias alfa são codificadas pelos genes  $\alpha 1$  e  $\alpha 2$  e as cadeias beta são codificadas pelo gene  $\beta$ . No embrião, as cadeias épsilon e zeta são expressas (genes  $\epsilon$  e  $\xi$ ). No feto, as cadeias gama (genes  $G\gamma$  e  $A\gamma$ ) são expressas. Fonte: traduzido e adaptado de [https://commons.wikimedia.org/wiki/File:Globin\\_Gene\\_Cluster.png](https://commons.wikimedia.org/wiki/File:Globin_Gene_Cluster.png). (acesso permitido se fonte citada).

## Pseudogenes

Durante o estudo das famílias gênicas, observou-se que havia sequências de nucleotídeos nos genomas muito semelhantes às sequências de genes componentes das famílias gênicas funcionais, mas que não eram transcritas ou traduzidas para moléculas proteicas funcionais. Essas sequências foram denominadas pseudogenes. Sua presença é interpretada como resultado de duplicações gênicas em que, durante ou após a duplicação, ocorreram alterações, resultado de processos mutacionais, que inviabilizaram sua atividade

ou expressão correta. Essas alterações podem incluir a presença de códons de parada, a ausência de códons de iniciação de tradução, a ausência de região promotora, entre outras. Os pseudogenes podem ser considerados "fósseis moleculares" que atestam a natureza dinâmica da evolução no nível molecular.

Normalmente, os pseudogenes têm seus nomes iniciados com a letra grega "psi" ( $\psi$ ). Assim, na família das globinas, a  $\psi$ -globina trata-se de um pseudogene com sequência semelhante ao gene de uma globina. No genoma humano, existem cerca de quatro vezes mais pseudogenes do que genes funcionais!

## Origem de genes por duplicação genômica

Se há genes que se originaram por duplicações gênicas que resultaram em genes que foram se diferenciando durante a evolução, quando ocorre o fenômeno de **poliploidização**, o mesmo pode acontecer, mas em escala muito maior, com genomas que inicialmente são completamente redundantes. Pela análise dos genomas de organismos atuais, sabe-se que no passado houve um ou mais eventos de poliploidização. Durante a evolução dos vertebrados, é postulado que houve dois ciclos de tetraploidização. Isso se deve à observação de que há quatro grupos da família gênica dos genes **Hox**, enquanto nos **grupos irmãos** dos vertebrados há somente um membro dessa família. Durante a história evolutiva da levedura *Saccharomyces cerevisiae* (fermento de pães e cervejas), também há sólidas evidências de que houve um ciclo de tetraploidização.

## Transferência gênica horizontal

Ao longo das gerações, os genes são transferidos de maneira "vertical", de progenitores para descendentes, como tradicionalmente esquematizam os heredogramas. Entretanto, pode haver circunstâncias nas quais genes podem ser transferidos "horizontalmente" ou "lateralmente", para um indivíduo de uma espécie diferente. Durante a evolução dos procariotos, houve muitas ocorrências em que genes ou porções dos genomas se originaram a partir de linhagens oriundas de espécies diferentes. Isso se deve aos diferentes mecanismos moleculares que permitem essa transferência entre os procariotos. Desde os primórdios da biologia molecular, conheceu-se o mecanismo de transformação, quando Frederick Griffith, em 1928, mostrou que havia uma substância química que poderia alterar geneticamente um microrganismo, a bactéria *Streptococcus pneumoniae*. Pos-

teriormente, Avery, MacLeod e McCarty, em 1944, demonstraram que o princípio transformante de Griffith era o DNA. No processo conhecido como transformação, basta haver material genético no meio onde vivem os microrganismos para que estes o incorporem em seu material genético pela simples absorção através da parede celular (Figura 8). A transformação acontece quando há alguma desestruturação momentânea na parede bacteriana. Esse fenômeno é aproveitado para transformações em laboratório, para fins biotecnológicos, com a aplicação de corrente elétrica (eletroporação, formação de poros por eletricidade) ou uso de sais, como os de cálcio.

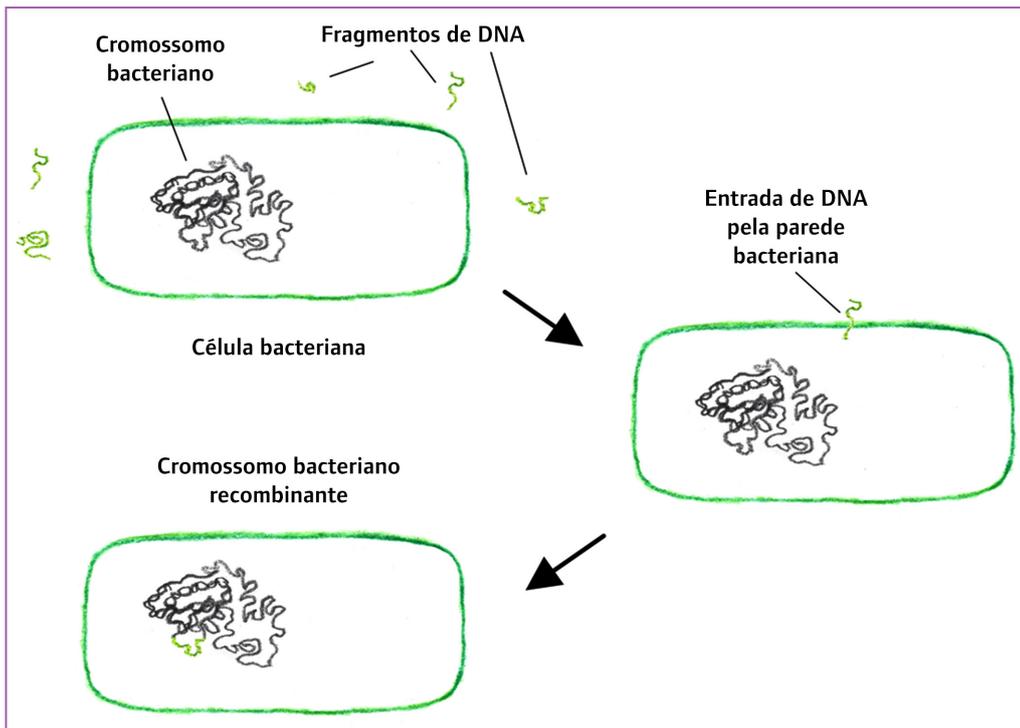
Além da transformação, pode haver transferência de material genético mediada por agentes como plasmídeos ou vírus. Os plasmídeos são sequências de DNA circulares que podem conter genes e têm a propriedade de serem transmitidos de bactéria para bactéria quando ocorre o fenômeno da conjugação. Na conjugação, mostrada na Figura 9, duas bactérias compartilham uma porção de citoplasma através de uma conexão aberta entre as duas paredes celulares. Quando esse processo ocorre entre bactérias de duas espécies diferentes, caracteriza-se a transferência gênica horizontal. Esse tipo de alteração genética tem grande importância na saúde pública, pois plasmídeos podem portar fatores de resistência a antibióticos, fazendo com que linhagens de bactérias se tornem resistentes sem que tenha ocorrido evolução para isso em sua própria espécie, mas sim em outra espécie, que pode ser muito diferente!

Outro tipo de processo que pode estar envolvido na transferência gênica horizontal é conhecido como transdução, que envolve a transferência de genes de um microrganismo para outro através da ação de vírus. Quando os vírus infectam uma célula, seu genoma pode ser recombinado com uma porção do genoma do hospedeiro. Quando vírus recombinantes infectam outro hospedeiro, que pode ser de uma espécie distinta, podem integrar no genoma do novo hospedeiro genes que estavam presentes na espécie anteriormente infectada. (Figura 10).

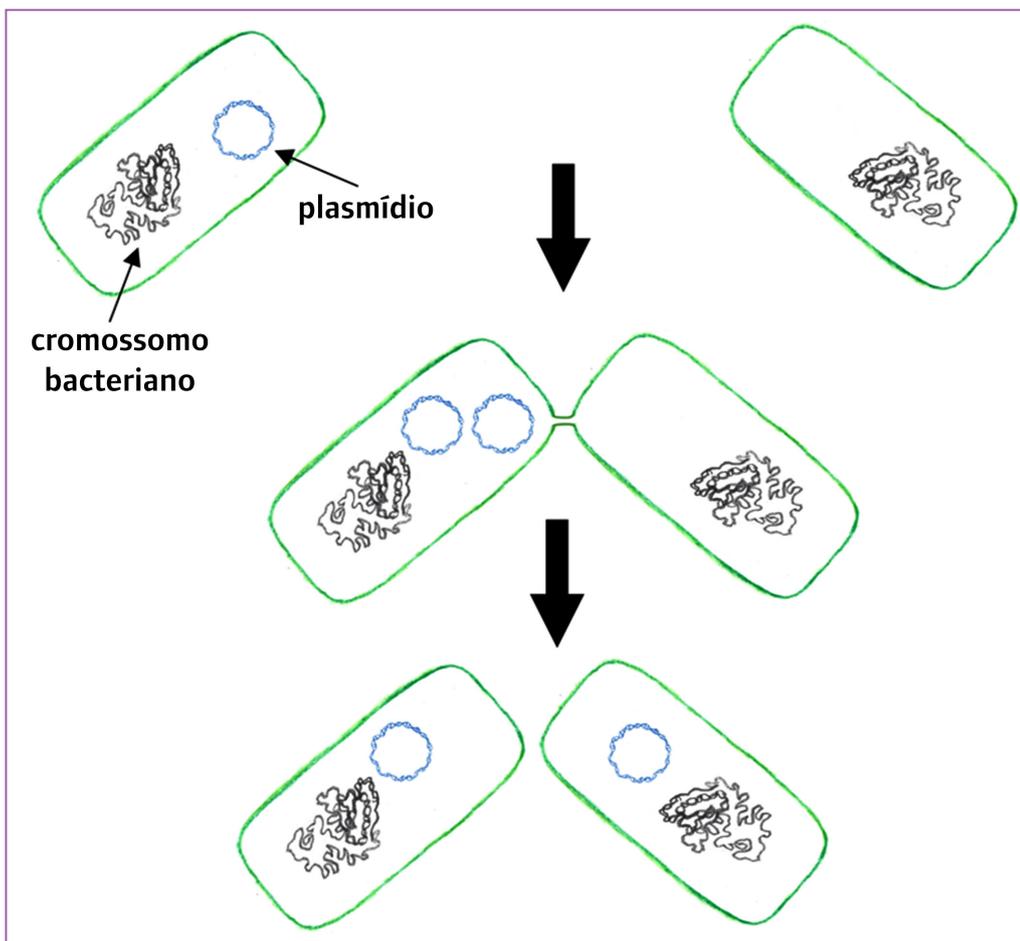
**Poliploidização** - fenômeno no qual há a produção de descendentes com uma ploidia (número de genomas por célula) maior que a ploidia dos ancestrais. Por exemplo, em uma espécie cuja ploidia é  $2n$  (diploide), é produzido indivíduo  $4n$  (tetraploide).

**Genes Hox** - genes que fazem parte da determinação, durante o desenvolvimento embrionário, de membros de grande parte dos animais multicelulares e de como cada um dos segmentos corpóreos desenvolverá suas estruturas.

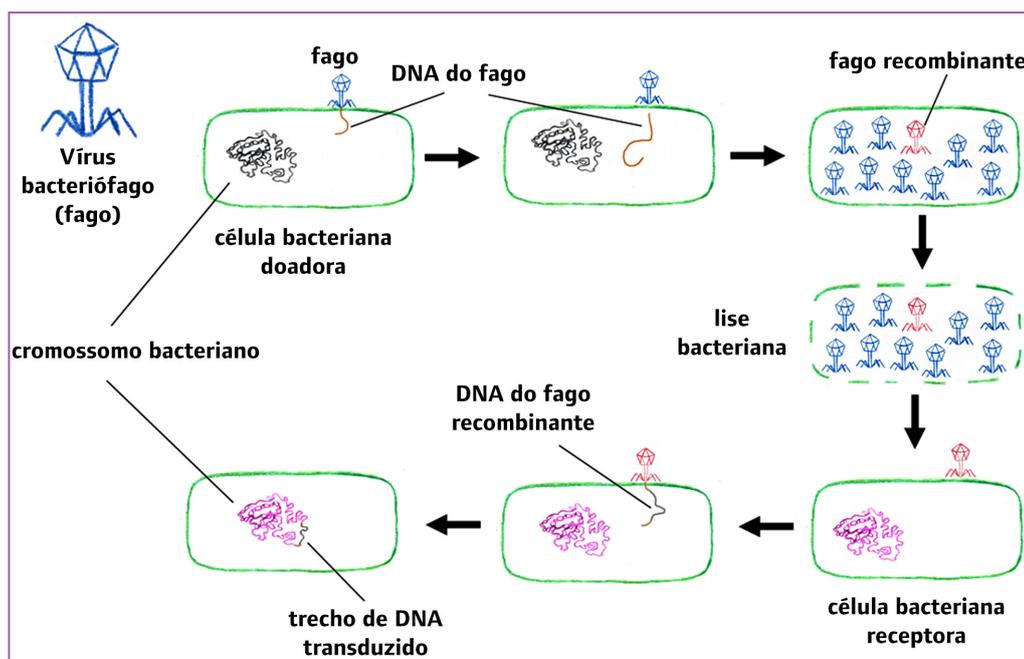
**Grupos irmãos** - grupos taxonômicos que compartilham um ancestral comum recente em uma filogenia.



**Figura 8.** Transformação. Células bacterianas podem, em certas situações, absorver fragmentos de DNA que estão no meio em que vivem.



**Figura 9.** Conjugação entre bactérias em que há a transferência de plasmídeos entre elas.



**Figura 10.** Transdução. Ela ocorre quando um trecho de DNA de origem bacteriana acaba sendo incorporado no DNA de um vírus que, quando infecta outra bactéria, pode transferir uma parte do material genético da bactéria infectada anteriormente. Fonte: Baseada e traduzida de [https://commons.wikimedia.org/wiki/File:Transduction\\_image.pdf](https://commons.wikimedia.org/wiki/File:Transduction_image.pdf).

De modo geral, existem evidências abundantes de eventos de transferência gênica horizontal em organismos procariontes. Nos organismos eucariontes, esse tipo de evento é bem mais raro. No entanto, em mamíferos, há evidências de que um gene tenha sido introduzido em seu genoma a partir de um gene existente em um vírus. Nos vírus, esse gene codifica uma proteína do envelope, estrutura proteica que recobre seu material genético. Nos mamíferos, o mesmo gene codifica uma proteína chamada sincitina, que é expressa na superfície das células da placenta, com a função de adesão entre elas. A comparação entre genomas de diferentes espécies de mamíferos, de diversos grupos, evidenciou que essa transferência do gene da sincitina ocorreu independentemente, em duas ocasiões, durante a evolução dos carnívoros e a evolução dos primatas.

## Evolução de genes por "embaralhamento de éxons"

Muitas proteínas que tiveram sua estrutura determinada revelaram uma grande complexidade. Uma inspeção mais detalhada de sua

estrutura mostrou que essa complexidade se deve à organização em módulos. Assim como uma casa pode estar dividida em compartimentos como sala, quarto, banheiro e cozinha, proteínas complexas têm seus módulos, que podem agir independentemente. Esses módulos proteicos são denominados domínios. Na Figura 11, está representada uma proteína com estrutura multidomínio, a **piruvato quinase**. Além disso, embora não seja universal, existe uma correspondência entre os domínios proteicos e sua codificação no material genético em éxons definidos.

Você já brincou com blocos de montar? Imagine que cada bloco representa uma pequena parte de um gene chamada éxon. Assim como é possível encaixar blocos de maneiras diferentes para criar todos os tipos de estruturas, nosso sistema biológico pode embaralhar os éxons para originar genes que codificam proteínas novas e diversas. Esse processo é chamado de embaralhamento de éxons. Nos próximos parágrafos, vamos nos aprofundar nesse conceito biológico e entender como ele funciona.

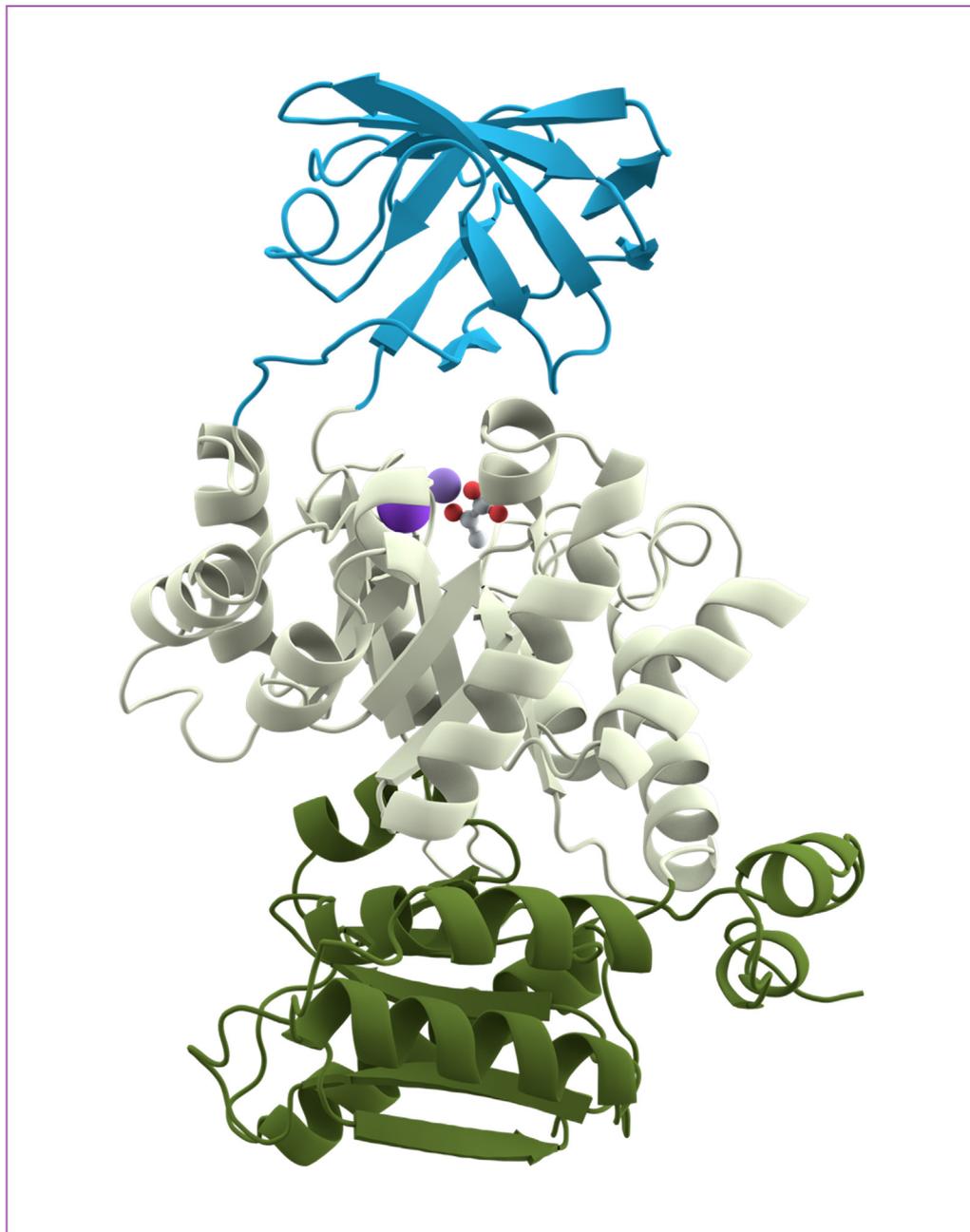
Inicialmente proposto por Walter Gilbert em 1978, o embaralhamento de éxons é um processo natural que permite o rearranjo dos éxons, ou seja, trechos dos genes de organismos eucarióticos. Vale lembrar que, nesses

**Piruvato quinase** - enzima participante da glicólise, via metabólica responsável pela obtenção de energia a partir da glicose e que catalisa a transferência do fosfopiruvato para o ADP, formando ATP.

genes, os éxons são intercalados com íntrons, regiões não codificadoras, e são unidos durante o processo de *splicing* de RNA para formar um RNA mensageiro (RNAm) maduro que pode ser traduzido em uma proteína.

O embaralhamento de éxons pode ocorrer por meio de alguns processos que podem produzir novas combinações que resultam

em proteínas diferentes. Os éxons podem ser duplicados dentro do genoma, levando à criação de cópias adicionais deles, conforme mostrado na figura 4, só que, ao invés de envolver genes, neste caso somente éxons são envolvidos. O éxons duplicados podem então ser incorporados a um novo contexto genômico, contribuindo para a diversidade de proteínas.



**Figura 11.** Representação esquemática de uma molécula multidomínio (Piruvato quinase), mostrando os seus três domínios em cores diferentes. Em azul, o domínio é regulador, ou seja, determina a velocidade da reação catalisada. Os domínios representados em branco e verde são responsáveis pela reação, que ocorre em duas etapas: a remoção do fosfato do fosfopiruvato e a ligação do fosfato com a molécula de ADP. Fonte: [https://commons.wikimedia.org/wiki/File:Pyruvate\\_kinase\\_protein\\_domains.png](https://commons.wikimedia.org/wiki/File:Pyruvate_kinase_protein_domains.png).

Conforme já mencionado, existem nos genomas elementos que têm a propriedade de se mover, os elementos genéticos móveis. Caso esse movimento se dê por meio da transcrição reversa de seu intermediário de RNA, chamamos esses elementos de retrotransposons. Os retrotransposons, como os elementos LINE-1, podem facilitar o embaralhamento de éxons ao mobilizar esses segmentos para novos locais genômicos, carregando-os “de carona” com sua própria sequência. Em uma associação com a edição de texto, é como se eles pudessem copiar e colar éxons de um local para outro, criando estruturas éxon-íntron.

Pode haver ainda a recombinação não homóloga. Esse tipo de recombinação envolve a troca de DNA entre cromossomos não homólogos. Ela pode resultar na união de éxons de diferentes genes, resultando em novas combinações.

Os íntrons também fornecem alvos para a recombinação. A recombinação dentro dos íntrons pode reunir éxons de diferentes genes sem interromper as sequências de codificação, facilitando o embaralhamento de éxons.

## Qual é o impacto do embaralhamento de éxons para os seres vivos?

O embaralhamento de éxons desempenha um papel fundamental na evolução, aumentando a diversidade genética. Essa diversidade ajuda os organismos a se adaptarem a novos ambientes e a desenvolverem novas funções, essenciais para a sobrevivência. Além disso, permite a rápida evolução de novas características, podendo conferir aos organismos uma vantagem em ambientes em constante mudança. Basicamente, a teoria do embaralhamento de éxons sugere que a recombinação intrônica cria combinações de domínios, contribuindo para a evolução de novas funções proteicas.

O fenômeno de embaralhamento de éxons não ocorre apenas durante a evolução. Nosso

sistema imunológico usa o embaralhamento de éxons para criar a grande variedade de anticorpos que possuímos. Os anticorpos são proteínas denominadas imunoglobulinas, que possuem uma região invariável e uma região chamada *hipervariável*. A região hipervariável é composta por combinações de trechos traduzidos de éxons, permitindo que cada indivíduo tenha um enorme repertório de anticorpos. Esses anticorpos reconhecem e neutralizam microrganismos responsáveis por diferentes infecções.

## Embaralhamento de éxons como ferramenta biotecnológica

O embaralhamento de éxons não é apenas um fenômeno natural interessante, mas também pode ser utilizado como uma ferramenta biotecnológica com grande potencial para inovar na engenharia genética e no desenvolvimento de proteínas com potencial terapêutico em diversas áreas. Esse processo pode ser aproveitado para criar proteínas com novas funções ou melhorar as propriedades de proteínas existentes, contribuindo significativamente para avanços na medicina e na biotecnologia.

O processo de embaralhamento de éxons pode ser empregado na evolução dirigida de proteínas. Os cientistas usam o embaralhamento de éxons para criar genes que codificam novas proteínas em laboratório. Ao combinar diferentes éxons, eles podem produzir proteínas com funções aprimoradas ou totalmente novas, que podem ser utilizadas no desenvolvimento de novos medicamentos.

No desenvolvimento de novas variedades de vegetais, o embaralhamento de éxons ajudou a desenvolver novos genes que melhoram a produção de energia e outras funções vitais. Por exemplo, alguns genes em girassóis que ajudam as células vegetais a obter energia com mais eficiência evoluíram naturalmente por meio do embaralhamento de éxons. Esse processo não só aumenta a eficiência energética, mas também contribui para a resistência

a estresses ambientais, como seca e salinidade. Além disso, o embaralhamento de éxons pode levar ao desenvolvimento de novas vias metabólicas que permitem às plantas sintetizarem compostos importantes, como hormônios de crescimento e defensivos naturais contra pragas e doenças. A biotecnologia moderna aproveita esses mecanismos naturais para criar plantas transgênicas com características desejáveis, como maior produtividade, resistência a herbicidas e melhor qualidade nutricional.

## Como identificar o embaralhamento de éxons

A identificação do embaralhamento de éxons pode ser um processo complexo que pode envolver a combinação de ferramentas de bioinformática, técnicas de biologia molecular e genômica comparativa.

O embaralhamento de éxons pode ser evidenciado comparando-se as estruturas de éxon-íntron dos genes em diferentes espécies. A semelhança significativa nas sequências de éxons, juntamente com diferenças na posição de íntrons, pode indicar embaralhamento de éxons.

Ao se construir árvores filogenéticas para éxons e íntrons separadamente pode-se observar grandes incongruências entre elas. Essas incongruências entre as árvores filogenéticas podem sugerir eventos de embaralhamento de éxons. Existem alguns *softwares*, como GENSCAN, Augustus ou GeneMark que são utilizados para prever estruturas de éxons e íntrons. É em uma segunda etapa basta procurar por éxons que apresentem homologia com éxons de outros genes.

Ferramentas como Pfam ou InterPro podem identificar domínios de proteínas conservadas dentro dos éxons. Se esses domínios forem encontrados em contextos de genes diferentes, isso pode sugerir sua origem por embaralhamento de éxons. É possível ainda utilizar os dados de sequenciamento para mapear as junções de éxon-éxon. Junções inesperadas podem indicar embaralhamento de éxons.

## Surgimento de Genes "de Novo"

O surgimento de genes "de novo" é um fenômeno intrigante e significativo na biologia evolutiva. O estudo de genes "de novo" desafia a visão tradicional de que novos genes surgem principalmente através da duplicação e divergência de genes existentes. Esses genes surgem a partir de sequências de DNA não codificadoras que não possuíam funções proteicas prévias, normalmente anteriormente localizadas em regiões intergênicas ou dentro de íntrons de genes existentes. Isto é, esses genes podem emergir "do zero" e desempenhar funções críticas nos organismos e facilitar a adaptação dos organismos a novos ambientes.

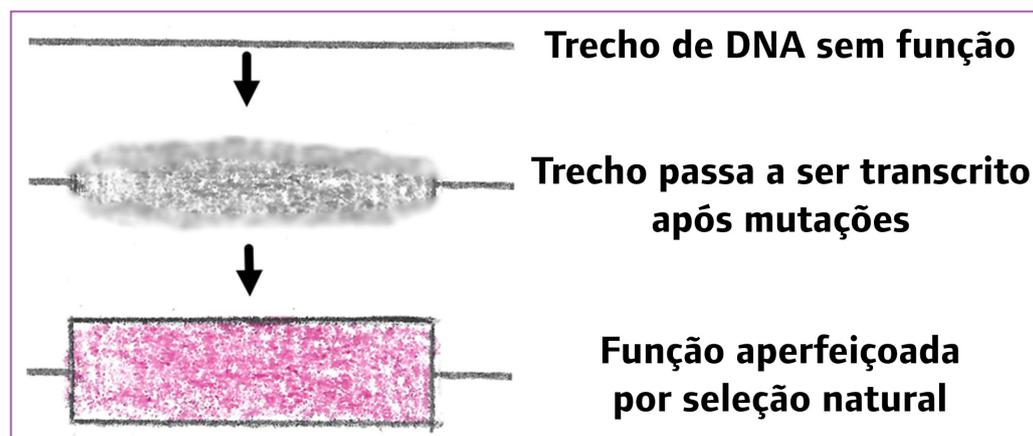
Genes "de novo" podem surgir através de várias etapas, incluindo a transcrição de regiões que antes não eram codificadoras no genoma, a formação de quadros de leitura abertos (**ORFs**) e a subsequente seleção de proteínas funcionais que oferecem vantagens adaptativas. Por exemplo, em estudos com *Drosophila melanogaster*, novos genes frequentemente mostram expressão específica em testículos, sugerindo um papel na reprodução e evolução sexual. Um exemplo de genes "de novo" em humanos bem documentado é o gene *MYEOV*, que está associado a vários tipos de câncer.

A identificação de genes "de novo" começa com o sequenciamento de genomas de diferentes organismos para encontrar todas as sequências de DNA. Pesquisadores buscam quadros de leitura abertos (ORFs) e comparam essas sequências com genomas de outras espécies para verificar a existência de genes semelhantes (homólogos). Se não encontrarem homólogos, a sequência é considerada um candidato a gene "de novo". Para validar esses genes, os cientistas verificam se a sequência de DNA é transcrita em RNA e se esse RNA é traduzido em uma proteína funcional, usando técnicas como a espectrometria de massa. Eles também realizam experimentos para confirmar a função da nova proteína, por exemplo, desligando o gene e observando os efeitos no organismo.

**ORF** - sigla em inglês de *open reading frame*, em português, quadro aberto de leitura. Trata-se de uma sequência de DNA que se inicia por um códon de iniciação, tem vários códons sucessivos que correspondem a aminoácidos e se encerra com um códon de parada da tradução. Logo, é uma sequência de DNA que potencialmente pode codificar um polipeptídeo.

Os desafios para identificar genes “de novo” incluem a dificuldade de detectar homólogos distantes, erros na montagem dos genomas e a possível perda paralela de genes em várias linhagens. Isso levanta a possibilidade de que muitos genes considerados “de novo” possam,

na verdade, ser derivados de fragmentos divergentes de genes mais antigos que foram perdidos na maioria das linhagens ou revividos em apenas uma. Apesar desses desafios, a compreensão dos genes “de novo” é crucial para entender a evolução e adaptação dos organismos.



**Figura 12.**

Esquema que representa o surgimento de um gene “de novo”. Um trecho sem qualquer função, por ação de mutações, passa a ser transcrito e/ou traduzido e, gradualmente, passa a desempenhar alguma função.

## Surgimento do primeiro gene

Podemos considerar essa questão como uma das grandes questões até então não respondidas dentro do âmbito da Ciência. Dado que os seres vivos são entidades que necessariamente possuem a propriedade de reprodução e de transmitir suas características para seus descendentes, evidentemente a origem do primeiro gene aconteceu de forma concomitante com o surgimento do primeiro ser vivo! Embora seja muito difícil sabermos como foi exatamente o primeiro ser vivo, através das propriedades que conhecemos dos seres vivos atuais, podemos especular como poderia ter sido o primeiro organismo. Desde a década de 1960 vem sendo proposta a ideia de um “mundo de RNA”, onde os primeiros seres vivos seriam constituídos por esse tipo de macromolécula. Isso se baseou no fato de que moléculas de RNA podem tanto transmitir informação genética como catalisar reações químicas, que poderiam inclusive permitir algum tipo de autorreplicação. Assim, um RNA autorreplicante seria o candidato natural para ser o primeiro gene e também o primeiro ser vivo. Ainda não há qualquer evidência experimental nesse sentido, ou seja, apesar de diversas tentativas, ainda não se conseguiu sintetizar, em laboratório, uma molécula com tais propriedades.

Caso isso seja algum dia demonstrado, com certeza será considerado um dos maiores avanços científicos de todos os tempos, sem contar o seu enorme potencial biotecnológico.

## Para saber mais

MATIOLI, S.R. e FERNANDES, F.M.C. *Biologia molecular e evolução*. Holos, editora e SBG, 2012. (<https://srmatioli.ib.usp.br/biolmolevol/index.html>)

SOUZA, D.T. *Origem de genes recentes, uma abordagem com PSSMs deterioradas e arquiteturas de domínio proteico*. Tese de doutorado, Universidade de São Paulo, 2016. ([https://www.teses.usp.br/teses/disponiveis/95/95131/tde-16012017-170749/publico/Origem\\_de\\_genes\\_recentes\\_uma\\_abordagem\\_com\\_PSSMs\\_deterioradas\\_e\\_arquiteturas\\_de\\_dominio\\_proteico\\_souza\\_dt.pdf](https://www.teses.usp.br/teses/disponiveis/95/95131/tde-16012017-170749/publico/Origem_de_genes_recentes_uma_abordagem_com_PSSMs_deterioradas_e_arquiteturas_de_dominio_proteico_souza_dt.pdf)).

VIBRANOVSKI, M.D. E LONG, M. *Origination and Evolution of Genes on the Sex Chromosomes*. *Encyclopedia of Evolutionary Biology*. 2016, 117-126. (<https://doi.org/10.1016/B978-0-12-800049-6.00172-4>)

## Agradecimentos

Os autores agradecem às sugestões e às críticas feitas no texto pelas Profas. Dras. Regina Célia Mingroni Netto e Eliana Dessen. M.D.V. agradece à FAPESP pelo apoio recebido conforme processo 2023/14607-6. A redação de alguns trechos foi melhorada com o emprego da plataforma ChatGPT (<https://openai.com/>).