Genética na Escola – ISSN: 1980-3540

Genes sobrepostos

Tiago Campos Pereira[1,2]

[1]Depto. de Biologia, FFCLRP, Universidade de São Paulo. Av. Bandeirantes, 3900. Monte Alegre, Ribeirão Preto - SP. CEP 14040-901

[2]Programa de Pós-Graduação em Genética, FMRP, Universidade de São Paulo. Av. Bandeirantes, 3900. Monte Alegre, Ribeirão Preto - SP. CEP 14040-901

Autor para correspondência - tiagocampospereira@ffclrp.usp.br

Palavras-chave: genes sobrepostos, expressão gênica, silenciamento, RNA não codificador

Genética na Escola – ISSN: 1980-3540

Os genes codificam proteínas por meio dos mecanismos de transcrição e tradução, sendo que a tradução se baseia no código genético, que é um sistema por meio do qual cada três nucleotídeos (equivalentes a um códon no RNA mensageiro) correspondem a um determinado aminoácido incorporado na síntese de um polipeptídeo. Tipicamente, uma determinada região do cromossomo possui apenas um gene codificador de proteína, por isso os genes não se sobrepõem. Neste artigo veremos uma característica notável dos genomas – a possibilidade de haver sobreposição de dois ou mais genes em um mesmo segmento do DNA, com todos os desafios biológicos e de criptografia associados a essa situação.

O código genético

Uma sequência de DNA consegue codificar uma proteína por meio dos mecanismos de transcrição e tradução, sendo que a tradução se baseia no código genético – um sistema de correspondência por meio do qual um conjunto de três bases nitrogenadas sequenciais no gene, e no respectivo RNA mensageiro, representa um aminoácido. Desta forma, o código genético permite que, com trincas formadas por combinações entre quatro tipos de bases nitrogenadas, seja possível haver 64 combinações diferentes, que correspondem a 20 aminoácidos diferentes. Informações detalhadas sobre como esse sistema funciona podem ser encontradas no artigo da revista Genética na Escola intitulado "Código genético - o código dos vinte" (v. 4 n. 1,2009; https://doi.org/10.55838/1980-3540.ge.2009.70).

Fase de leitura ou quadro de leitura - forma como uma sequência de RNA mensageiro pode ser lida em trincas. Cada RNA mensageiro pode possuir até três fases de leitura (+1, +2 e +3). É importante destacar que não existe a fase de leitura +4, pois isso equivaleria à fase de leitura +1 omitindo-se a primeira trinca (e o primeiro aminoácido).

Por convenção, diz-se que o RNA mensageiro é lido na **fase de leitura** +1, ou seja, a adenina (A) do códon de início (AUG) é designada como sendo a primeira base da **região codificadora** (CDS), a uracila é a segunda base e assim por diante (Figura 1).

Região codificadora - algumas partes do gene codificam um polipeptídeo e essas partes, em conjunto, são denominadas como região codificadora; entretanto, outras partes não codificam (por exemplo os íntrons, a extremidade 5′ não traduzida e a extremidade 3′ não traduzida). No RNA mensageiro, a região codificadora, cuja sigla é CDS, começa no códon de iniciação (AUG) e se encerra no códon de parada (que pode ser UAA, UAG ou UGA).

Uma das características gerais do código genético é que ele tipicamente não apresenta sobreposição – cada base do RNA mensageiro faria parte de apenas um códon. Isso, por sua vez, levaria à impossibilidade de que dois ou mais genes codificadores de proteínas pudessem se **sobrepor**.

Sobreposição - situação em que dois ou mais genes compartilham um único trecho do genoma.

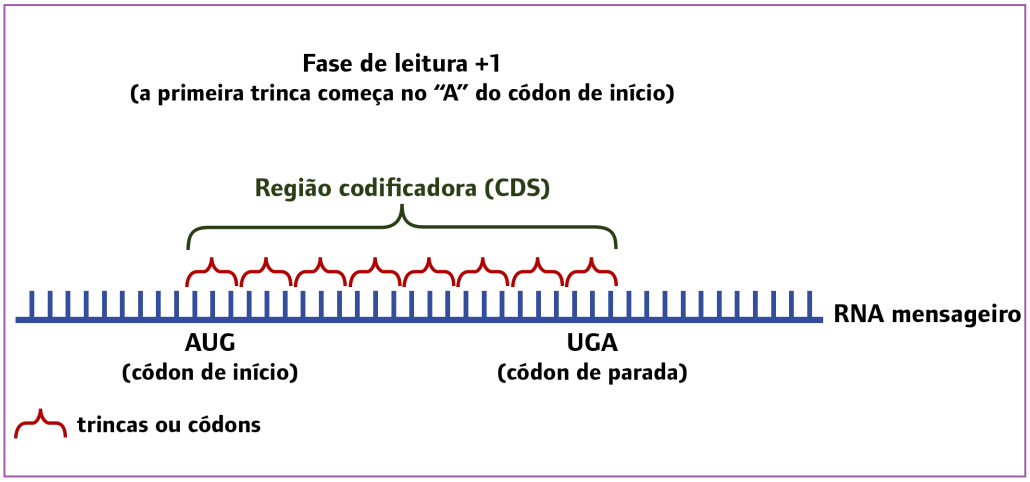

Figura 1.
Fase de leitura. O RNA mensageiro é traduzido no ribossomo por meio do código genético. A primeira base (adenina) do códon de iniciação (AUG) estabelece a fase de leitura ou quadro de leitura: as trincas serão lidas sempre tendo a adenina como nucleotídeo de referência "1". O conjunto de todas as trincas responsáveis por codificar uma proteína é chamado de região codificadora ou CDS.

Genética na Escola – ISSN: 1980-3540

Tipos de sobreposição gênica

A Biologia, contudo, é cheia de belas exceções. Dois ou mais genes podem, sim, se sobrepor de diferentes maneiras. Por exemplo, pode haver sobreposições na mesma orientação ou na orientação inversa, pontuais, parciais ou totais (Figura 2).

Sobreposições pontuais ou parciais são relativamente comuns, por incluírem um número muito pequeno de bases sobrepostas. Entretanto, sobreposições extensas ou totais são bem menos frequentes.

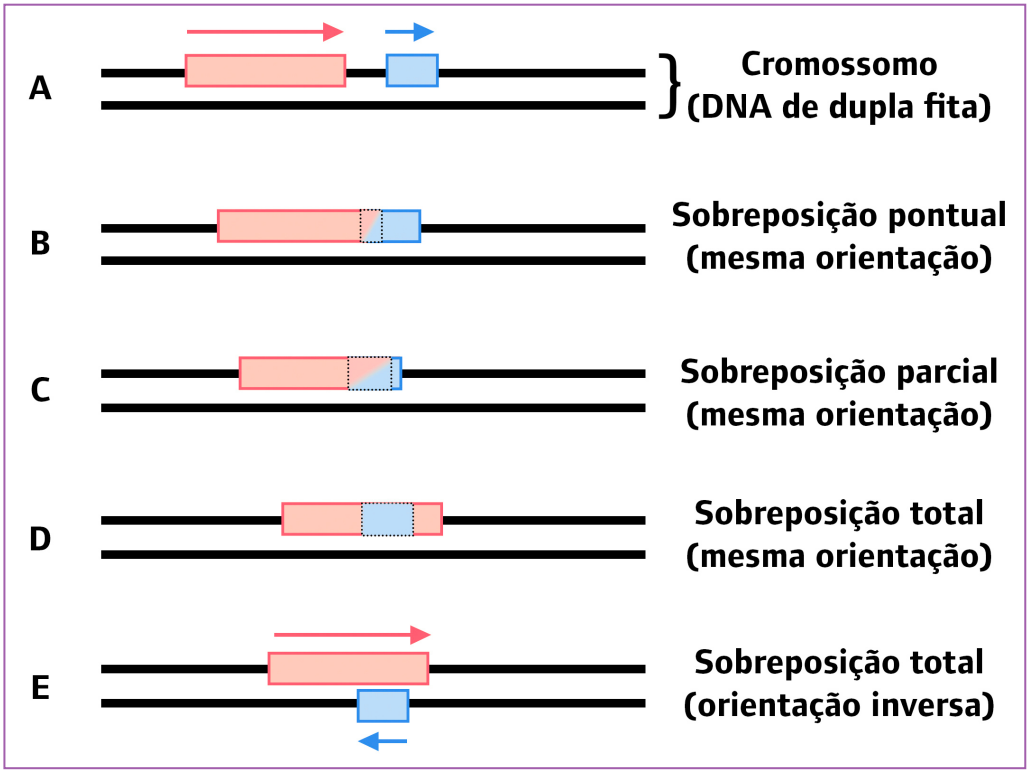

Figura 2.
Tipos de sobreposição gênica. Tipicamente, pensamos em genes como sequências de DNA independentes, que não se sobrepõem **(A)**. Entretanto, eles podem se sobrepor de diferentes maneiras. Por exemplo, pontualmente, compartilhando um ou poucos nucleotídeos **(B)**; parcialmente, compartilhando uma região maior **(C)**; ou até mesmo um estando localizando dentro do outro **(D)**. É importante também lembrarmos que como muitos genomas são de DNA de dupla fita, é possível que a sequência de código de um gene esteja localizada em uma fita de DNA, ao passo que o outro gene sobreposto a ele esteja na outra fita **(E)**. Nesse caso, como as cadeias de DNA são antiparalelas, isto é, uma segue a direção 5′→3′ e a outra a direção 3′→5′, naturalmente cada um desses dois genes é transcrito em uma direção específica (vide setas laranja e azul). As caixas retangulares com bordas pontilhadas indicam as regiões de sobreposição.

A problemática da sobreposição de genes codificadores de proteínas

Por que é tão difícil sobrepor genes codificadores de proteínas? A seguir, veremos que este tipo de situação é um problema biomatemático realmente desafiador.

Por exemplo, quando observamos a sequência nucleotídica de um gene (considerada a fase +1) e tentamos lê-la na fase +2, notamos que surgem ao acaso diversos códons de parada, que impedem que a fase +2 codifique, de fato, uma cadeia proteica de tamanho biologicamente relevante (Figura 3). Isso também ocorre se tentarmos ler na fase +3.

Esse surgimento de códons de parada **prematuros** é quase inevitável e ocorre porque a leitura de dois códons subsequentes na fase +1 pode, espontaneamente, gerar trincas UAG, UGA ou UAA (que são os três códons de parada) quando lidos na fase +2 ou +3 (Figura 3).

Prematuro - que surge antes do esperado. Neste caso, refere-se a um códon de parada que surge precocemente na sequência após o início da tradução, de forma a interromper antecipadamente a síntese proteica.

Genética na Escola – ISSN: 1980-3540

Figura 3.
Códons de parada dificultam a sobreposição gênica. Consideremos uma sequência codificadora de um RNA mensageiro (em preto; separado em trincas de acordo com a fase +1; apenas a parte inicial da sequência do RNAm é mostrada). Quando ela é traduzida na fase de leitura +1, iniciando a contagem dos códons a partir da primeira base (A, de AUG), observamos a proteína resultante (em azul, apenas a parte inicial da sequência é mostrada). Porém, ao tentarmos ler essa mesma sequência na fase de leitura +2, iniciando a contagem dos códons a partir da segunda base (U, de AUG), notamos que surge um códon de parada (retângulo vermelho) já no começo da cadeia polipeptídica (em vermelho). Algo semelhante ocorre quando tenta-se ler na fase de leitura +3 (em verde). Esses códons de parada não estão presentes na sequência do RNA mensageiro quando ele é lido na fase +1; porém, emergem quando ele é lido nas demais fases (destaques em verde e rosa na sequência do RNAm). Cada letra (M, A, P, S, W, P, G, R etc.) localizada à frente de "Fase +1", "Fase +2" e "Fase +3", representa um diferente aminoácido presente nos polipeptídeos correspondentes.

Nós podemos brincar com esse desafio e tentar encontrar uma sequência nucleotídica codificadora de proteína que, mesmo na fase +2, não apresente as trincas de parada. Com esforço, paciência ou com o uso de um computador, pode-se obter uma sequência nucleotídica que codifique polipeptídeos distintos nessas duas fases de leitura. Porém, algo muito importante precisa ser levando em conta: provavelmente as duas sequências proteicas geradas artificialmente, a partir de uma sequência nucleotídica igualmente **sintética**, não apresentarão função biológica alguma (Figura 4).

Sintético - artificial, não natural, produzido artificialmente pelo homem.

Em termos práticos, resolvemos o problema matemático, porém não o problema biológico. Obter duas sequências proteicas funcionais, isto é, polipeptídeos que exerçam papéis na célula, a partir de uma mesma sequência nucleotídica é um desafio notório.

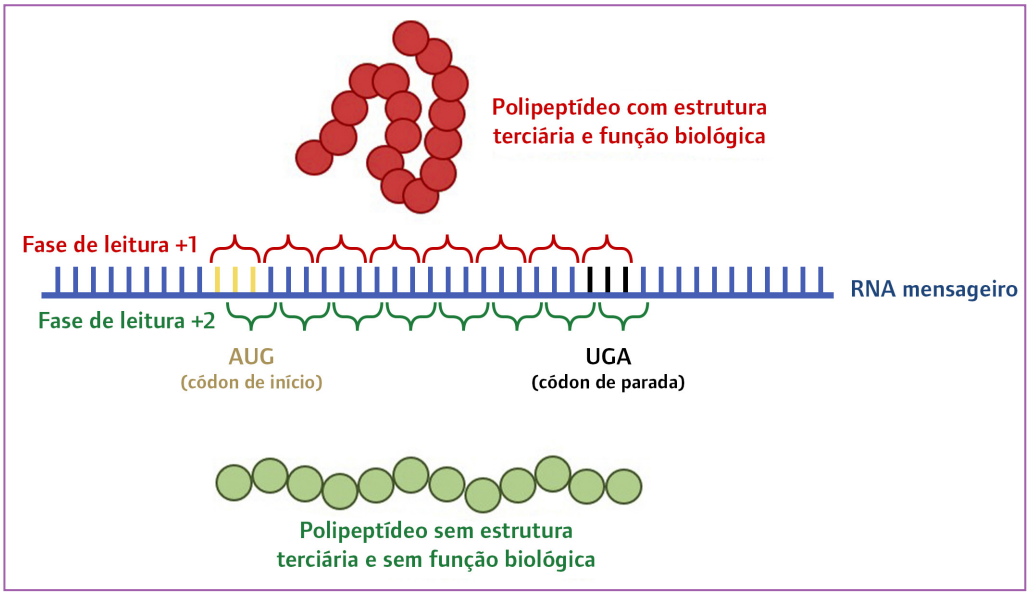

Figura 4.
Os desafios da sobreposição gênica. É possível criar uma sequência nucleotídica (gene) artificial que codifique duas (ou mais) cadeias polipeptídicas longas (em vermelho e em verde), evitando códons de parada prematuros. Entretanto, provavelmente uma (ou ambas) cadeia polipeptídica artificial não será capaz de exercer um papel biológico (atuar como enzima, por exemplo) devido à ausência de uma estrutura terciária adequada (em verde).

Genética na Escola – ISSN: 1980-3540

Exemplos reais de genes sobrepostos

Um dos exemplos mais conhecidos de genes sobrepostos encontra-se no genoma da usina energética da célula humana, a mitocôndria.

O genoma mitocondrial (DNAmt) humano é pequeno, com aproximadamente 17 mil pares de bases, disposto em um DNA de dupla fita circular. Ele é ordens de grandeza menor do que o genoma nuclear humano, que possui aproximadamente 3,3 bilhões de pares de bases.

Mesmo sendo tão pequeno, ele é capaz de codificar 13 proteínas, além de dois RNAs ribossômicos e 22 RNAs transportadores. Todos os 13 polipeptídeos codificados pelo DNAmt estão envolvidos com as funções respiratórias oxidativas das mitocôndrias (ou seja, com a produção de energia – ATP).

Curiosamente, duas dessas proteínas – a subunidade 6 da ATPase (ou ATP6) e a subunidade 8 da ATPase (ou ATP8) – são codificadas por genes parcialmente sobrepostos (Figura 5)[2].

O gene da ATP8 possui 207 pares de bases de extensão, codificando um polipeptídeo de 68 aminoácidos (além de um códon de parada). Por sua vez, o gene da ATP6 possui 681 pares de bases de extensão, codificando um polipeptídeo de 226 aminoácidos (além de códon de parada). Esses dois genes se sobrepõem ao longo de uma região de 46 pares de bases (Figura 5), a qual é lida nas fases +1 e +3. As fases correspondem, respectivamente, à ATP8 (com 15 aminoácidos da cadeia ou 22% do tamanho total do gene) e à ATP6 (com 16 aminoácidos da cadeia ou 7% do tamanho total do gene).

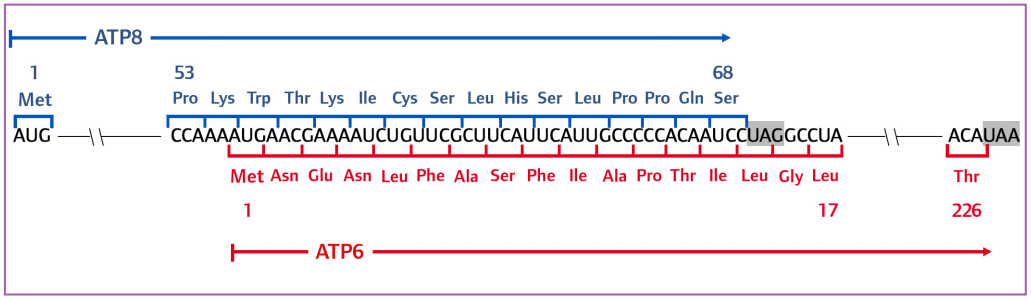

Figura 5.
Exemplo real de sobreposição gênica no DNA mitocondrial. O genoma mitocondrial codifica algumas proteínas, entre elas a ATP6 e ATP8, cujos genes se sobrepõem parcialmente. Em azul, a ATP8, com seu primeiro aminoácido metionina (Met 1) até o último aminoácido serina (Ser 68). A ATP6 (em vermelho) é produzida a partir da mesma sequência, porém traduzida a partir de outra fase de leitura, iniciando-se em outro nucleotídeo. Ela começa com uma metionina (Met 1) e se estende até a treonina na posição 226 (Thr 226). É interessante notar que a cadeia de ATP6 não é interrompida pelo códon de parada (em cinza) da ATP8 porque ele é relido como dois outros códons (Leucina – Leu e Glicina – Gly) gerados com a mudança da fase de leitura. Demarcados em cinza: códons de parada.

Outro belíssimo exemplo de sobreposição gênica é observado no genoma do vírus ϕX174 (ϕ é uma letra grega, pronuncia-se *Phi*). ϕX174 é um **bacteriófago** cujo genoma é de DNA de fita simples de 5386 nucleotídeos de extensão, com 11 genes codificadores de proteínas, alguns dos quais apresentam sobreposição[3]. Em especial, há dois genes sobrepostos a um terceiro, sendo que cada RNA mensageiro é traduzido em uma fase de leitura distinta (Figura 6).

O primeiro deles é o gene da proteína A, de 513 aminoácidos, codificada por uma região de 1543 nucleotídeos traduzidos na fase +1. O segundo é o gene da proteína K, de 56 aminoácidos, codificada por uma região de 171 nucleotídeos traduzidos na fase +2. O terceiro é o gene da proteína B, de 120 aminoácidos, codificada por uma região de 363 nucleotídeos traduzidos na fase +3.

Bacteriófago - vírus que infecta bactérias.

Tanto o gene codificador da proteína B, quanto o gene da proteína K, se sobrepõem ao gene codificador da proteína A. Em especial, o nucleotídeo na posição 51 do genoma de ϕX174 faz parte dos três genes, sendo um caso de sobreposição tripla (Figura 6).

Genética na Escola – ISSN: 1980-3540

É interessante notar que genes sobrepostos tendem a emergir em genomas ultrapequenos, mitocondriais e virais, nos quais o espaço é muito limitante, de tal forma que a sobreposição de genes é uma alternativa interessante e vantajosa. Entretanto, é importante lembrar que a sobreposição gênica ocorre em diversas outras espécies, porém geralmente são muito pequenas ou não envolvem a sobreposição de regiões codificadoras.

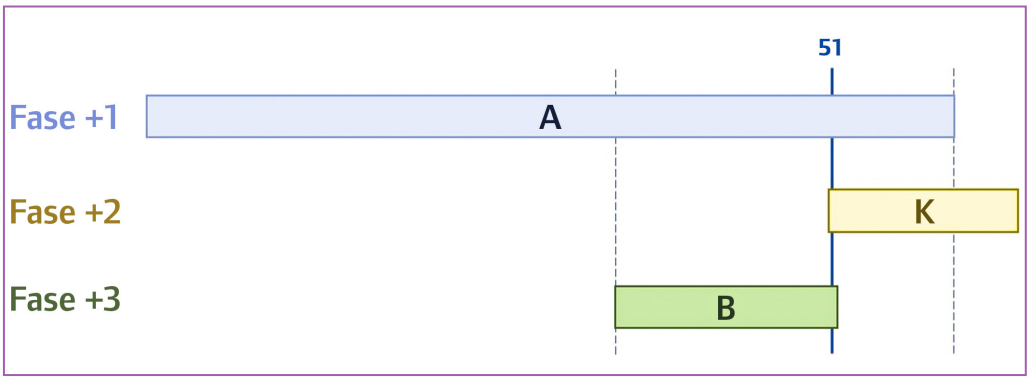

Figura 6.
Sobreposição gênica em vírus. O bacteriófago ϕX174 possui vários genes sobrepostos, entre eles os genes codificadores das proteínas A, B e K. O nucleotídeo de número 51 (indicado) faz parte dos três genes, exemplificando uma sobreposição pontual tripla. As linhas tracejadas evidenciam as regiões de sobreposição.

Questões biológicas referentes a genes sobrepostos

Geralmente, quando uma mutação ocorre em um gene codificador de uma proteína, um aminoácido pode ser alterado. Por exemplo, consideremos a trinca GAC codificadora do aminoácido aspartato. Uma mutação do tipo A→T irá converter a trinca em GUC, codificando um aminoácido (valina) totalmente distinto na cadeia polipeptídica. Entretanto, em genes sobrepostos, uma única mutação pode afetar aminoácidos nos produtos de dois (ou mais genes) simultaneamente e isso pode chegar a ser letal.

Espécies parasitas (vírus e outros microrganismos unicelulares, por exemplo) tendem a apresentar genomas menores devido a um benefício reprodutivo. Quanto menor o genoma, mais rapidamente ele é replicado, produzindo uma quantidade maior de descendentes. Nesse sentido, a sobreposição gênica favorece a redução genômica, sendo, portanto, vantajosa.

Contudo, por outro lado, a sobreposição gênica impõe restrições, pois a vastíssima maioria das mutações que ocorrerem na região sobreposta terão um efeito negativo ou **deletério**. Assim sendo, há forças contrárias que atuam nos genomas de espécies parasitas, levando à sobreposição de alguns genes (visando a compactação genômica), porém de maneira não excessiva (devido à dificuldade biomatemática de se sobrepor genes codificadores de proteínas e ao custo biológico relacionado às mutações nessas regiões).

Deletério - destrutivo, prejudicial, nocivo ao organismo.

Qual é o limite da sobreposição gênica?

Inicialmente, poderíamos supor que o limite de sobreposição seria de três genes codificadores de proteínas em uma mesma sequência de DNA. Entretanto, deve-se lembrar que, nos genomas celulares, o DNA se apresenta como dupla fita, de tal forma que uma fita poderia conter três genes sobrepostos, e a outra fita, convencionalmente chamada de negativa, poderia codificar mais três outros genes, cujos RNAs mensageiros seriam traduzidos nas fases de leitura -1, -2 e -3 (Figura 7).

Genética na Escola – ISSN: 1980-3540

Portanto, em tese, uma mesma região do DNA de dupla fita poderia codificar seis proteínas totalmente diferentes. Contudo, o desafio **criptográfico** é verdadeiramente gigantesco.

Criptográfico - relativo à criptografia, isto é, à codificação de mensagens. O código genético é uma forma de criptografia.

Conclusões

Genômica - ciência que estuda os genomas dos organismos, por meio do sequenciamento genético, identificação de genes e análises comparativas entre genomas de diferentes espécies.

Uma das estratégias vantajosas para a compactação dos genomas é a sobreposição gênica, porém existem ganhos e perdas no processo. O entendimento de que um gene pode estar localizado dentro de outro gene também tem implicações para a **genômica**. Isto é, quando um novo genoma é sequenciado e seus genes são anotados, outros genes ainda podem existir "escondidos" dentro dos genes já **anotados**.

Anotado - identificado. A anotação gênica se refere à identificação de um gene em um genoma, o que consiste em revelar onde começa e onde termina a sequência nucleotídica de um gene dentro de um determinado cromossomo ou genoma.

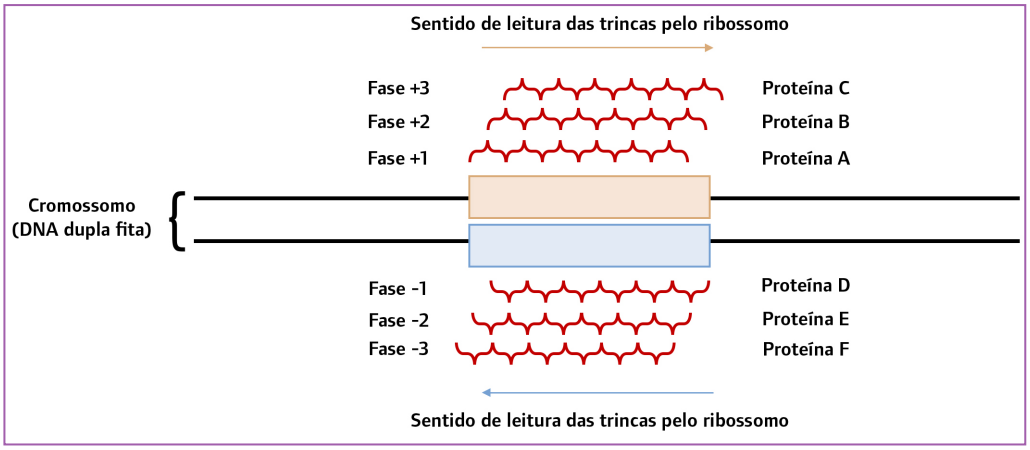

Figura 7.
Seis possíveis fases de leitura. Considerando-se um cromossomo (DNA de dupla fita), cada cadeia de DNA pode ser lida em três fases distintas, porém de polaridades inversas (positiva e negativa), totalizando seis fases: +1, +2, +3, -1, -2 e -3. Em teoria, cada uma dessas fases poderia codificar uma proteína distinta (de A a F).

Por fim, a sobreposição gênica também tem uma aplicação biotecnológica interessante, pois permite posicionar mais de um gene em uma mesma sequência de DNA que será introduzida em um organismo, com técnicas de engenharia genética, otimizando, assim, o processo.

A sobreposição de genes codificadores é algo realmente fantástico na Biologia. Para evidenciar esse fascínio, nada melhor que uma analogia: imagine ter em suas mãos um belo romance, de 300 páginas, lido a partir da primeira letra, da primeira palavra (Fase de leitura +1). Agora, imagine que a primeira letra tenha sido eliminada e que todas as demais letras do livro são reagrupadas em novas palavras (porém mantendo a ordem sequencial). Você percebe, então, que o livro contém uma segunda história, por exemplo, uma aventura espacial. Uma sobreposição de histórias totalmente diferentes, em um mesmo texto. Deslumbrante, não?

Para saber mais

Wright BW, Molloy MP, Jaschke PR. Overlapping genes in natural and engineered genomes. *Nat Rev Genet*. 2022 Mar;23(3):154-168.

He J, Ford HC, Carroll J, Douglas C, Gonzales E, Ding S, Fearnley IM, Walker JE. Assembly of the membrane domain of ATP synthase in human mitochondria. *Proc Natl Acad Sci U S A*. 2018 Mar 20;115(12):2988-2993.

Smith M, Brown NL, Air GM, Barrell BG, Coulson AR, Hutchison CA 3rd, Sanger F. DNA sequence at the C termini of the overlapping genes A and B in bacteriophage phi X174. *Nature*. 1977 Feb 24;265(5596):702-5.