Genética na Escola – ISSN: 1980-3540

Esteganografia molecular – envio de mensagens secretas com o DNA

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: DNA, criptografia, sigilo, codificação

Genética na Escola – ISSN: 1980-3540

O envio de mensagens privadas, a guarda de conteúdos sigilosos e a preservação de **dados pessoais sensíveis** demandam técnicas que ocultem essas informações de criminosos ou de nações adversárias. Tipicamente, duas técnicas não excludentes podem ser utilizadas para este fim – a criptografia e a esteganografia. Neste artigo, a esteganografia molecular com base em DNA será detalhada, em um exemplo real no qual a criptografia molecular também foi utilizada. Essas técnicas emergem como alternativas muito interessantes e promissoras para um mundo em que a proteção de dados é essencial.

Dados pessoais sensíveis - de acordo com a Lei Geral de Proteção de Dados (LGPD), dados pessoais sensíveis são aqueles que, se forem revelados, podem gerar algum tipo de discriminação. Por exemplo: sexo, raça, informações relacionadas à saúde, orientação sexual, religião e posicionamento político.

A esteganografia

A esteganografia (do grego *steganós* – "coberto ou oculto", *graphia* – "escrita") refere-se à ocultação de mensagens em um objeto comum, ou em um texto maior e não secreto, tal como uma carta convencional ou um livro, por exemplo. Isto é, esconder uma informação importante de tal forma que apenas o remetente e o destinatário sabem que ela existe e onde ela se encontra (Figura 1). A esteganografia é importantíssima em tempos de guerra, durante os quais as mensagens enviadas às tropas militares de um país não devem ser identificadas pelo exército inimigo. A esteganografia também pode ser útil em sistemas comerciais, nos quais a proteção de dados pessoais sensíveis do consumidor é muito importante.

Possivelmente, um dos primeiros registros do uso da esteganografia data de 440 a.C, no livro "Histórias", do grego Heródoto. Esse autor cita dois exemplos, sendo que, em um deles, Histieu envia uma mensagem gravada no couro cabeludo (raspado) de um de seus servos de confiança, destinada ao seu vassalo Aristágoras. Após o cabelo ter crescido, Histieu envia o homem à cidade de Mileto, onde deve encontrar-se com Aristágoras e dizer-lhe que raspasse sua cabeça e a olhasse.

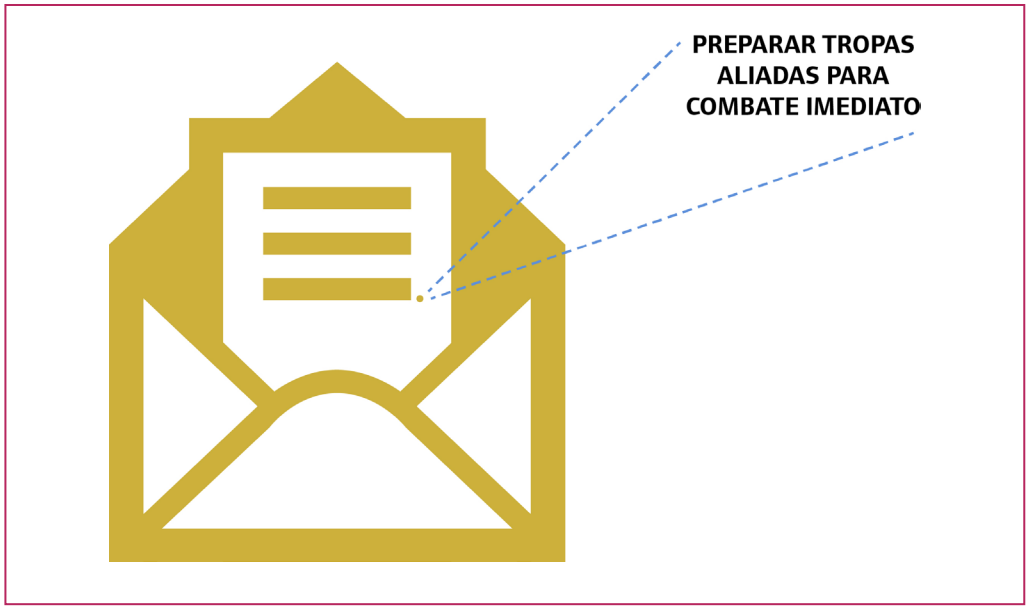

Figura 1.
Objetivo da esteganografia. Esta técnica tem como meta ocultar mensagens sigilosas – neste caso PREPARAR TROPAS ALIADAS PARA COMBATE IMEDIATO – dentro de um contexto maior e aparentemente não secreto (uma carta comum). Neste exemplo, a mensagem foi escondida por meio da sua miniaturização e seu posicionamento no sinal de pontuação que finaliza a carta.

Genética na Escola – ISSN: 1980-3540

Diferença entre esteganografia e criptografia

Ambas a esteganografia (do grego, *escrita encoberta*) e a criptografia (do grego, *escrita secreta*) têm um ponto em comum – esconder uma determinada mensagem. Entretanto, elas correspondem a conceitos tecnicamente distintos.

A criptografia tem como princípio a codificação de uma mensagem em uma linguagem desconhecida, que ninguém pode compreender facilmente (Figura 2). Assim, pode-se, por exemplo, representar as letras do alfabeto usando símbolos gráficos ($ @ * & %) ou uma mistura de letras de diferentes alfabetos (hebraico – א ב ג ד; grego – α φ ψ π ξ). Dessa forma, a criptografia não se preocupa em esconder o fato de existir uma mensagem secreta. Seu objetivo é tornar essa mensagem indecifrável por outros.

Por sua vez, a esteganografia tem como preceito ocultar a existência de uma mensagem de natureza sigilosa (Figura 1). Assim, objetiva-se esconder a mensagem, que pode estar escrita em uma linguagem conhecida, dentro de um suporte (texto, papel, livro etc.), de tal forma que as pessoas enxergam uma estrutura aparentemente não sigilosa (o texto, o papel ou o livro), sem ter ideia de que há ali uma mensagem secreta.

Naturalmente, existe a possibilidade de se usar ambas as técnicas de maneira conjugada, isto é, cifrando uma mensagem (criptografia) e em seguida ocultando-a em um **suporte físico** (estenografia).

Suporte físico - qualquer estrutura na qual uma mensagem pode ser registrada, tal como papel, livro, quadro, carta, fotografia ou outros objetos comuns.

Figura 2.
Objetivo da criptografia. Esta abordagem tem como finalidade codificar uma mensagem – neste caso OURO NO CAMPO LESTE – por meio de símbolos ou letras de outros alfabetos. As respectivas palavras (codificadas e decodificadas) estão apresentadas nas mesmas cores em **B** e **C**.

Tipos convencionais de esteganografia

As técnicas mais comuns de esteganografia envolvem esconder um texto ou uma imagem dentro de um suporte físico maior. Talvez um dos exemplos mais simples de esteganografia seja a de embeber uma mensagem em um texto maior e aparentemente inócuo, por exemplo, ocultando a mensagem a cada dez letras do texto visível, como pode ser visto a seguir: o texto inofensivo – ALEXANDRE TINHA UM TRENZINHO BELO. SEMPRE, COM ALEGRIA, GUARDAVA-O ONDE SÓ ELE SABIA. – contém uma mensagem esteganografada – ATÉ LOGO.

Genética na Escola – ISSN: 1980-3540

Outro princípio esteganográfico bastante conhecido e muito usado por crianças é o de escrever palavras usando suco de limão como tinta, um cotonete como pincel e uma folha branca de papel como suporte físico. Após escrever a mensagem e deixar o líquido sobre o papel secar, a mensagem desaparece, restando apenas a folha em branco. Contudo, ao se passar um ferro quente sobre o papel, a mensagem é revelada (Figura 3).

Figura 3.
Uso lúdico de esteganografia. Um exemplo muito conhecido de técnica esteganográfica é a escrita de mensagens em folha de papel utilizando suco de limão como tinta. Após secar, a mensagem desaparece; porém, com o uso de um ferro de passar quente, a mensagem é revelada. Ferro de passar – foto de cottonbro studio: https://www.pexels.com/pt-br/foto/casa-lar-residencia-apartamento-4107220/.

Porém, há muitas outras técnicas esteganográficas, tal como ocultar uma imagem complexa e detalhada – por exemplo, um mapa com a localização de uma unidade militar inimiga para fabricação de bombas – em uma carta comum. Nessa estratégia, a fotografia aérea de alta resolução é feita, passada por um processo de miniaturização, até que a mesma seja menor que o sinal de pontuação do tipo "ponto final", que encerra esta frase. Após a carta remetida chegar ao destinatário, ele poderá visualizar o mapa com detalhes por meio de lentes e ampliação (Figura 4).

Figura 4.
Uso militar de esteganografia. O envio de um mapa detalhado de uma cidade, dentro da qual se encontra uma unidade de fabricação de armas, pode ser feito de maneira oculta, por miniaturização e posicionamento em um sinal de pontuação. Mapa: foto de Anna Tarazevich: https://www.pexels.com/pt-br/).

Genética na Escola – ISSN: 1980-3540

Outro princípio esteganográfico se baseia no uso de cores para ocultar mensagens. Nesse sistema, todos os pontos que formam cada uma das letras da mensagem sigilosa são de um determinado pigmento de cor muito específico, azul claro, por exemplo. Em seguida, todo o restante do espaço da folha é coberto por diversos pontos aleatórios, em diversas outras cores, exceto o tom específico de azul. O quadro resultante pode parecer uma simples coleção multicolorida de pontos que não formam nenhuma imagem definida ou mensagem aparente. Entretanto, ao se iluminar o quadro com um filtro de luz azul claro na tonalidade específica, a mensagem enrustida emerge nesta cor, enquanto todos os demais pontos permanecem na cor preta, pois absorvem a luz azul (Figura 5).

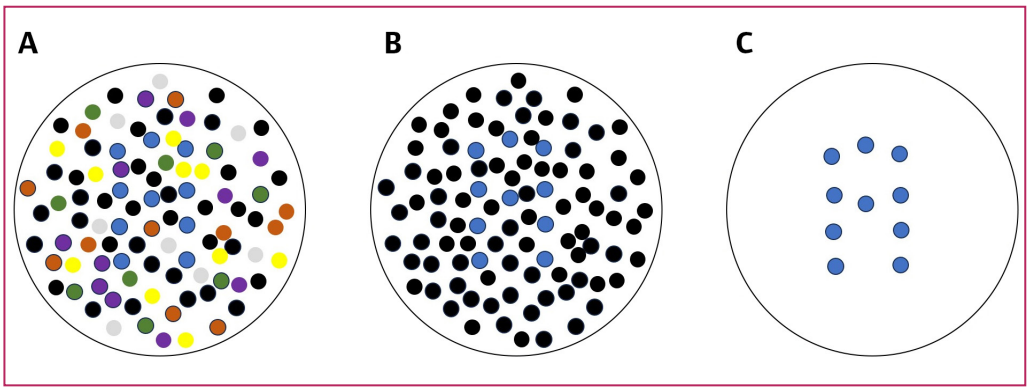

Figura 5.
Esteganografia com uso de cores. É possível escrever uma mensagem utilizando pontos de uma cor específica (neste caso, azul claro), misturados com uma miríade de outros pontos aleatórios em diversas cores. Assim, a mensagem (neste caso, a letra "A") desaparece. Entretanto, ao se iluminar a superfície com uma luz com filtro azul claro, todos os pontos ficam em preto e apenas os pontos em azul claro ficam nesta cor. Assim é possível ler a mensagem oculta.

Curiosamente, mensagens podem ser escondidas em outros suportes além de estruturas sólidas, ou outros meios além da luz/visão. Mensagens de áudio podem ser encobertas dentro de uma gravação de uma conversa aparentemente cotidiana. Neste contexto, uma mensagem de voz pode ser fragmentada e dispersa, de maneira específica, dentro de um arquivo maior de áudio. O receptor deste arquivo, tendo posse das informações para encontrar a mensagem (por exemplo, reunir todas as sessões de 500 milésimos de segundo de duração intervaladas a cada catorze segundos), pode gerar um novo áudio com a mensagem originalmente escondida. Outra forma é baseada na gravação reversa da mensagem, que após ser misturada com outros sons do ambiente de gravação direta, podem resultar em um áudio aparentemente inofensivo de captação sobre os ruídos de uma floresta, porém, que eclipsam uma mensagem secreta.

A esteganografia molecular

Em 1999, um trio de pesquisadores norte-americanos demonstrou, pela primeira vez, que é possível usar moléculas de DNA para ocultar mensagens por meio de um sistema híbrido de estegano e criptografia.

Para isso, eles sintetizaram moléculas de DNA com uma sequência nucleotídica específica, que não é encontrada nos seres vivos. A sequência era AGT CTG TCT GGC TTA ATA ATG TCT CCT CGA ACG ATG GGA TCT GCT TCT GGA TCA TCCT CGA TCT TTG AAA. De acordo com a **chave**, que apenas o remetente e o destinatário conhecem (Figura 6), a mensagem (original em inglês) é JUNE 6 INVASION : NORMANDY, referindo-se à famosa incursão dos países aliados na região francesa da Normandia, então tomada pelos nazistas, durante a II Guerra Mundial.

Chave - instrução que especifica a transformação do texto puro em texto cifrado (codificação) ou vice-versa (decodificação). A chave apresenta, por exemplo, a correlação entre os símbolos presentes na mensagem codificada e as letras do nosso alfabeto. Assim, a chave permite compreender a mensagem criptografada.

Genética na Escola – ISSN: 1980-3540

Como essas moléculas são extremamente pequenas, da ordem de **nanômetros**, uma quantidade mínima delas, 100 cópias, por exemplo, é infimamente menor que o sinal de ponto final. Desta forma, pode-se pingar uma solução desses DNAs no ponto final da carta, deixar secar e enviar ao destinatário. Este, por sua vez, tendo conhecimento que há uma mensagem cripto e esteganografada em um **microponto de DNA**, pode realizar a extração de DNA desta parte do papel.

Nanômetro - a bilionésima parte do metro; equivalente à milionésima parte do milímetro.

Microponto de DNA - porção diminuta de espaço dentro de um suporte físico (por exemplo, um sinal de pontuação de uma carta), no qual moléculas de DNA com uma mensagem encoberta são depositadas.

A = CGA	K = AAG	U = CTG	U = ACT
B = CCA	L = TGC	V = CCT	1 = ACC
C = GTT	M = TCC	W = CCG	2 = TAG
D = TTG	N = TCT	X = CTA	3 = GCA
O = GGA	E = GGC	Y = AAA	4 = GAG
G = TTT	Q = AAC	= ATA	5 = AGA
H = CGC	R = TCA	, = TCG	6 = TTA
I = ATG	S = ACG	. = GAT	7 = ACA
J = AGT	T = TTC	: = GCT	8 = AGG

Figura 6.
Chave criptográfica para mensagem codificada em DNA. Ao se utilizar o DNA para codificar uma informação, é necessário criar uma correspondência (chave) entre trincas de nucleotídeos e as diferentes letras do alfabeto e símbolos, gerando um "código genético criptográfico". Neste caso, o uso da chave permite ler a mensagem JUNE 6 INVASION : NORMANDY.

Em seguida, para conseguir ler a mensagem, inicialmente é necessário aumentar o número de cópias dela, por meio da técnica de Reação em Cadeia da Polimerase (vide o artigo "A Reação em Cadeia da Polimerase" (PCR) – https://doi.org/10.55838/1980-3540.ge.2019.318). Este é outro ponto muito forte na segurança do processo de esteganografia com DNA, pois é necessário saber as sequências nucleotídicas dos iniciadores (*primers*) para realizar a PCR (compartilhado entre remetente e destinatário). Sem essa informação, o processo equivale a procurar uma agulha no palheiro.

Após ampliar milhões de vezes essa mensagem via PCR, ela poderá ser lida com o uso de um sequenciador de DNA, que é um aparelho capaz de identificar a sequência de bases nitrogenadas de um fragmento de ácido nucleico. Após obter essa informação, ela pode ser decodificada de acordo com a chave compartilhada, novamente entre remetente e destinatário (Figura 6).

Genética na Escola – ISSN: 1980-3540

Métodos para tentar identificar o DNA e contramedidas

O princípio da esteganografia com DNA emergiu no fim do século passado, e de lá até os dias atuais, grandes foram os avanços nas tecnologias de análise de DNA. Nesse sentido, atualmente, caso o inimigo saiba que há uma mensagem em DNA na carta, ele poderia usar algumas dessas técnicas recentes para tentar encontrar a informação. Por exemplo, ele poderia realizar o sequenciamento maciço de todo o DNA que venha a ser extraído da carta, isto é, tanto os DNAs das pessoas que manipularam a carta, quanto as moléculas sintéticas portadoras da mensagem em si. O uso dessa estratégia torna a amplificação do DNA via PCR um processo dispensável, não obrigatório.

Entretanto, o trio de pesquisadores também propôs que o DNA sintético fosse misturado com DNA humano ou com DNAs de diferentes espécies, de tal forma a dificultar ainda mais a identificação da molécula-alvo. Contudo, atualmente, é possível, por métodos computacionais e busca em bancos de dados, saber de qual espécie é determinada sequência de DNA. Portanto, o sequenciamento maciço de todo o DNA presente da carta, excluindo-se em seguida todos os DNAs de outros organismos, permitiria a identificação da molécula alvo.

Assim sendo, de acordo com a tecnologia atual, a esteganografia com DNA demandaria misturar a molécula (contendo a informação sigilosa e uma **etiqueta molecular**, conhecida apenas pelo remetente e destinatário) com milhares (ou milhões) de outras moléculas sintéticas de DNA (desprovidas de qualquer informação) e que, por serem artificiais, não estão disponíveis nos bancos de dados para pesquisa. Portanto, caso o inimigo viesse a sequenciar massivamente, ele obteria milhares ou milhões de sequências sintéticas sem saber qual delas contém a informação.

Conclusões

Assim como no passado e no presente, sem dúvidas no futuro a ocultação de informações sensíveis – tal como movimentações bancárias, informações militares, de agências de inteligência, de empresas de tecnologia, dentre muitos exemplos – continuará sendo essencial. Curiosamente, o advento dos computadores resultou em uma **corrida da rainha vermelha**. Isto é, ao mesmo tempo em que avanços na capacidade computacional permitiram quebrar códigos anteriormente considerados inquebráveis – tal como aqueles gerados pela famosa máquina alemã ENIGMA (tema do filme norte-americano de 2014, *Imitation game*, O jogo da imitação), o desenvolvimento de algoritmos criptográficos de última geração permite gerar sistemas cada vez mais seguros – até que ele venha a ser quebrado, e novos algoritmos sejam criados. Ou seja, estamos em uma perpétua perseguição, em que codificação e decodificação computacionais andam em ritmos semelhantes.

Corrida da rainha vermelha - metáfora que ilustra a situação de dois grupos, por exemplo, a presa e o caçador, que correm um atrás do outro. Entretanto, se ambos correm no mesmo ritmo, ou aceleram suas velocidades na mesma intensidade, eles permanecem parados um em relação ao outro. Diante disso, na metáfora, é impossível alcançar a rainha vermelha, pois ela está fugindo à mesma velocidade de seus perseguidores.

Adicionalmente, uma das promessas dos computadores quânticos é que eles sejam capazes de fazer cálculos de maneira paralela, avaliando múltiplas possibilidades simultaneamente, de forma a conseguirem quebrar qualquer tipo de cripto ou esteganografia computacional que venha a ser desenvolvida. Tomados em conjunto, o status de corrida da rainha vermelha na computação convencional, assim como a aproximação dos computadores quânticos, podem ser sugestivos de que a era da ocultação de informações em sistemas computacionais clássicos possa ter um fim em breve. Diante desse possível cenário, a preservação de dados sigilosos por meio de outras técnicas envolvendo o DNA é uma alternativa factível e com um potencial ainda pouco explorado.

Etiqueta molecular - uma sequência nucleotídica curta, de aproximadamente 50 bases, conhecida apenas entre remetente e destinatário, e presente em uma das extremidades da molécula de DNA onde a informação sigilosa se encontra.

Para saber mais

CLELLAND CT, RISCA V, BANCROFT C. Hiding messages in DNA microdots. *Nature*. 1999 Jun 10;399(6736):533-4.

Na D. DNA steganography: hiding undetectable secret messages within the single nucleotide polymorphisms of a genome and detecting mutation-induced errors. *Microb Cell Fact*. 2020 Jun 11;19(1):128.