

CAPA

# A NOVA ERA DO DNA





## Sequenciamento completo de genomas busca transformar o entendimento sobre doenças, diversidade genética humana, produção de alimentos e processos evolutivos

Maria Guimarães | ILUSTRAÇÃO Felipe Mayerle

Imagine que, ao sair com alguém que não conhecia, você queira verificar, por exemplo, se é verdade que aquela pessoa tem ascendência italiana. Basta aproveitar uma distração, passar um cotonete na marca da beira do copo, pôr a amostra de DNA no sequenciador que você leva no bolso, conectar ao telefone celular e, em minutos, ter a sua resposta. Parece ficção científica? Nem tanto: já existem aparelhos miniaturizados que fazem a leitura do material genético em tempo real e podem ser usados para identificar patógenos causadores de doenças. O dispositivo faz parte da revolução genômica que agora permite sequenciar o material genético completo cada vez em menos tempo, com menos erros e a um custo cada vez menor.

Quem era adulto há 20 anos, provavelmente se lembra da era dos genomas, quando foram divulgadas as sequências completas de uma série de organismos, com destaque para o Projeto Genoma Humano (PGH), cuja sequência quase completa foi anunciada em 2001. Grandes iniciativas – no Brasil, algumas delas financiadas pela FAPESP (ver Pesquisa FAPESP nº 174) – desvendaram o material genético de organismos como a bactéria *Xylella fastidiosa*, causadora da praga do amarelinho, que ataca laranjais, e as diversas formas de câncer, impulsionando laboratórios de genética e desenvolvendo potencial em bioinformática.

De lá para cá, esses dados estiveram à disposição da comunidade científica e pesquisas permitiram identificar uma infinidade de genes ligados a problemas de saúde, mas com impacto ainda restrito na vida dos pacientes (ver Pesquisa FAPESP nº 284) – sobretudo para as doenças complexas, como diabetes e câncer, que envolvem a ação de

muitos desses trechos do material genético responsáveis pela produção de proteínas.

“O Projeto Genoma Humano inaugurou a genética moderna”, avalia o bioinformata Pedro Galante, coordenador do laboratório de bioinformática do Instituto de Ensino e Pesquisa do Hospital Sírio-Libanês (IEP-HSL), em São Paulo. “Descobrimos muitos genes associados a uma diversidade de características.” O biólogo molecular Carlos Menck, do Instituto de Ciências Biomédicas da Universidade de São Paulo (ICB-USP), relata que o uso do sequenciamento de nova geração, há cerca de 10 anos, fez uma enorme diferença para o diagnóstico da doença xeroderma pigmentosa, cujos pacientes não podem ser expostos ao sol (ver Pesquisa FAPESP nº 323 e box na página 15). “Vamos em breve publicar um artigo com as mutações detectadas em 150 pacientes”, conta. Ele também vê aproximar-se a possibilidade de que esse conhecimento dê origem a tratamentos com base em edição ou regulação gênica por meio de ferramentas como Crispr-Cas9 e interferência por RNA.

Agora, a maneira de olhar o genoma está se alterando radicalmente, voltando-se não só para uma porcentagem muito pequena de genes que eram considerados essenciais, mas também para a sequência completa do DNA e comparando um grande número de indivíduos e até espécies. É um momento empolgante. Em maio, Galante esteve no congresso The Biology of Genomes, no laboratório Cold Spring Harbor, nos Estados Unidos, o mais importante da área. Ele relata que na palestra de encerramento o geneticista norte-americano Evan Eichler, da Universidade de Washington, declarou ter esperado anos pelo momento em que seu campo está agora. A expectativa é pelo salto que a pesquisa genômica deverá propiciar, especialmente na área da saúde.

Grande parte dos genes é semelhante entre organismos; cerca de 0,01% do genoma diferencia seres humanos uns dos outros

Uma das limitações do PGH foi que, apesar de ter colhido material de cerca de 20 voluntários, por volta de 70% do genoma sequenciado partiu de apenas uma pessoa, um norte-americano de ascendência europeia e africana da cidade de Buffalo, no estado de Nova York. Isso significa que variantes genéticas que diferem de uma população para outra podem não estar representadas nesse que é ainda hoje o genoma de referência. É a ele que os pesquisadores cotejam seus resultados para identificar a função e a variação de genes detectados em situações diferentes.

Um consórcio internacional, o Pangenoma Humano, pretende agora resolver essa questão realizando o sequenciamento completo de ao menos 350 pessoas de todos os cantos do planeta, deixando de ter apenas uma referência. “Se tivéssemos referências de genomas da população brasileira para comparar ao que vemos em nossos pacientes, seria muito melhor, porque variantes comuns aqui podem não estar presentes na população norte-americana ou vice-versa”, explica Galante, que está pleiteando a adesão ao consórcio para contribuir com conhecimento bioinformático.

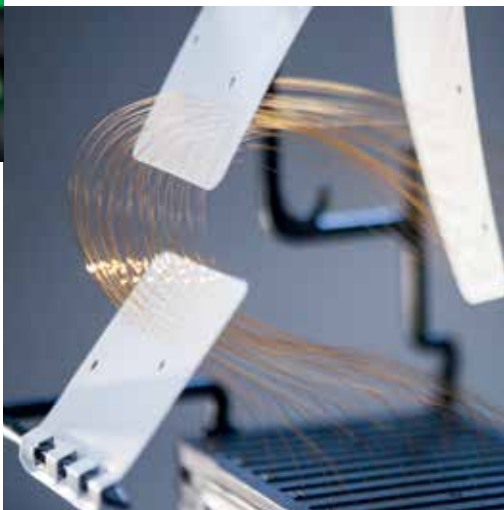
A promessa é fazer o chamado sequenciamento telômero a telômero (T2T), em referência às pontas dos cromossomos, cujas sequências muito repetitivas os tornam difíceis de destrinchar com base nos métodos anteriores. Para fazer isso, a estratégia tem sido combinar técnicas. Os aparelhos da empresa britânica Oxford Nanopore e da norte-americana Pacific Biosciences conseguem

ler trechos bem longos de DNA, com até milhões de bases. É eficiente para a montagem do genoma. “As leituras longas permitem determinar o arco-bouço do genoma”, explica a geneticista Glória Franco, da Universidade Federal de Minas Gerais (UFMG). Não à toa, a leitura longa foi destacada pela revista *Nature* como o método do ano de 2022, em editorial publicado em janeiro deste ano. Como a taxa de erro é maior, para refinar as sequências os geneticistas usam as leituras mais curtas feitas pelas máquinas da empresa californiana Illumina, que permitem olhar trechos menores com maior precisão.

**A**té recentemente, o sequenciamento vinha em fragmentos pequenos que precisavam ser montados, como um quebra-cabeça que pode ter partes sobrepostas (*ver infográfico abaixo*). Trechos repetitivos, abundantes no genoma, são especialmente difíceis de incluir na montagem e acabavam ficando de fora – na representação das bases nitrogenadas que compõem o DNA como A, T, C e G, poderia ser uma longa sequência de ATATATATATAT, por exemplo.

As iniciativas vão muito além dos genomas humanos, como destacado pela revista *Science* em janeiro entre os assuntos que seriam notícia neste ano. O Earth Biogenome Project (EBP), um

Máquinas de sequenciamento: *flow cell* na qual as amostras são inseridas na Illumina NovaSeq (*abaixo*); no método Sanger, o DNA passa por finos fios (*à dir.*)

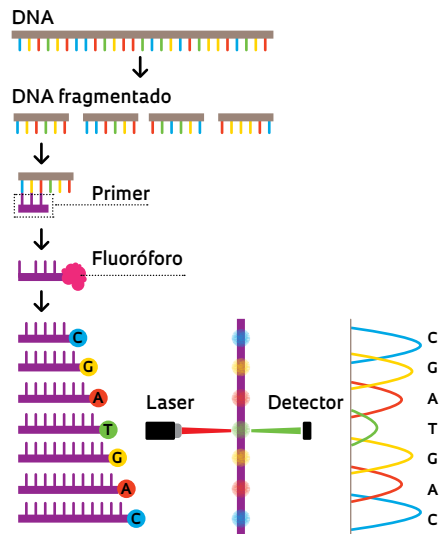


## OS GRANDES MARCOS DO SEQUENCIAMENTO

Avanços técnicos permitem análises cada vez mais extensas, rápidas e baratas

### 1977 SANGER

Sequência inicialmente determinada pela distância entre bases e marcas radioativas, depois substituídas por fluoróforos (extensões fluorescentes). Fragmentos de DNA clonados em bactérias são sequenciados e montados, com base nas sobreposições entre eles





consórcio lançado em 2022 que reúne pesquisadores de vários países, propõe sequenciar todas as espécies – inclusive a grande proporção ainda desconhecida pela ciência, sobretudo de seres unicelulares e pequenos invertebrados – ao longo de 10 anos. O objetivo é “conhecer, utilizar e conservar a biodiversidade”. Debaixo desse guarda-chuva estão outros projetos, como o sequenciamento de todos os mamíferos (Zoonomia, *ver* Pesquisa FAPESP nº 328) e de primatas (*ver* Pesquisa FAPESP nº 329).

A bioinformata brasileira Marcela Uliano-Silva, do Instituto Wellcome Sanger, no Reino Unido, integra o Darwin Tree of Life, que pretende desvendar o genoma de todos os organismos do Reino Unido. Ela justifica a necessidade desse olhar amplo: “Tudo na biologia é investigado na base da comparação”. São cerca de 70 mil espécies, das quais mil já estão com o sequenciamento quase pronto. “Temos ao todo por volta de 2.700 genomas em algum ponto do processo”, calcula. Enquanto são gerados, os dados já ficam disponíveis e podem ser analisados por quem tiver interesse. “O primeiro grande bloco de sequenciamentos foi de lepidópteros, a família das borboletas e mariposas”, conta a bioinformata.

Em busca da diversidade para além da genética, Uliano-Silva também integra a comissão de Justiça, Equidade, Diversidade e Inclusão do EBPP, que busca, entre outras coisas, evitar uma “neocolonização científica” em consequência da preponderância dos países mais abastados nesses

## A NOVA ONCOLOGIA

Pedro Galante, do IEP-HSL, avalia que a oncologia está em um patamar muito diferente em relação a poucos anos atrás e que os avanços agora são muito rápidos. Já é possível direcionar tratamentos agressivos – e extremamente caros – a pacientes que tenham mais chances de serem beneficiados, por características do perfil de mutações dos tumores e do sistema imunológico do paciente, por exemplo. “Isso graças a descobertas dos últimos cinco anos.”

Um exemplo, publicado este ano na revista científica *Cancers*, foi encontrar nas células de leucemia mieloide aguda – um tipo de câncer ainda sem tratamento satisfatório – uma grande atividade de 19 genes de receptores olfatórios, normalmente envolvidos com a detecção de odores. Foi um achado surpreendente, que pode auxiliar no diagnóstico precoce da doença e, quem sabe, revelar potenciais alvos terapêuticos.

Os tumores de pele em portadores da doença xeroderma pigmentosa têm um aumento de inserção de elementos de transposição, os retrotransposons, de acordo com artigo do grupo de Carlos Menck, do ICB-USP, publicado em maio na revista *Carcinogenesis*. “Fomos os primeiros a mostrar isso, graças a uma ferramenta bioinformática desenvolvida por Pedro Galante”, afirma Menck. O estudo indica que essas inserções causam instabilidade no genoma e podem estar na base do desenvolvimento de tumores e identifica uma enzima com papel importante na manutenção da estabilidade. “Estamos agora investigando o mecanismo de ação.”

1990

Início do Projeto Genoma Humano

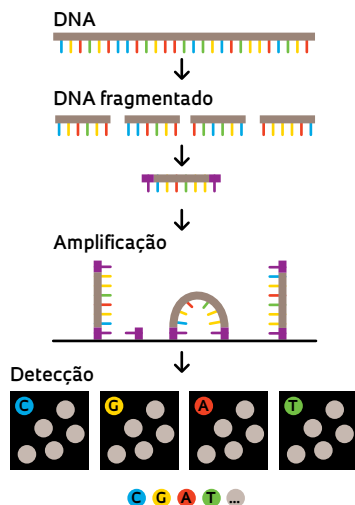
2001

Anúncio da sequência preliminar completa

2005

### NOVA GERAÇÃO (NGS)

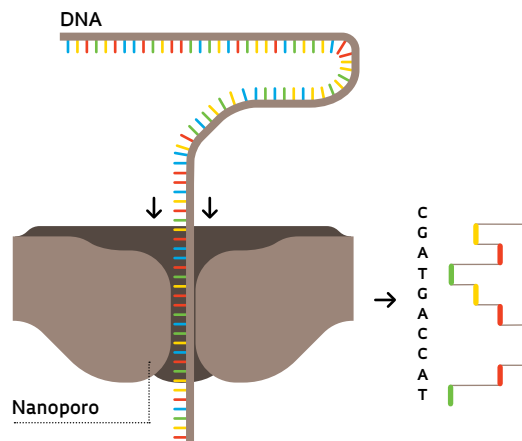
Depois de amplificação *in vitro*, milhões ou bilhões de fragmentos de DNA são imobilizados em uma superfície bidimensional, onde estão expostos ao mesmo tempo ao reagente que faz a leitura por marcações fluorescentes



2011

### TERCEIRA GERAÇÃO

Leitura de trecho longo de DNA em tempo real, distinguindo grupos de bases por diferenças de voltagem, permite o sequenciamento rápido e com custo menor, facilitando a montagem do genoma





Jaguar, lobo-guará e protozoário: organismos muito diferentes podem ter genes semelhantes



## AUTISMO

Estudos genômicos em quadros de autismo vêm indicando variantes de genes associadas ao distúrbio e a atrasos no desenvolvimento, de acordo com estudo publicado em 2022 na revista *Nature Genetics* com a participação de Maria Rita Passos-Bueno, do Centro de Estudos do Genoma Humano. A estratégia parece promissora para compreender melhor as bases neurobiológicas do transtorno do espectro autista (TEA).

Em busca de entender os fatores hereditários, ou os efeitos da idade dos pais no nascimento de crianças com TEA, parece ser promissora uma análise que considera três gerações de cada família, em vez de estudos populacionais mais amplos. Um artigo realizado no CEGH-USP relatando a análise de 33 famílias, publicado este ano na revista *Genetics*, favorece a abordagem – embora a amostra pequena ainda seja insuficiente para grandes conclusões.

grandes consórcios internacionais. Nesse papel, ela tem recebido contato de pesquisadores brasileiros com interesse em integrar esse tipo de projeto e defende que se organizem iniciativas semelhantes por aqui.

**U**m projeto em estágio inicial no Brasil planeja sequenciar todos os tetrápodes (vertebrados com quatro membros) do país, uma iniciativa do primatólogo Jean Boulbi, da Universidade de Salford, no Reino Unido, em parceria com o grupo do Centro de Estudos do Genoma Humano da USP (CEGH-USP). “Trabalhamos com uma sequenciadora Illumina NovaSeq 6000”, explica a geneticista Maria Rita Passos-Bueno, referindo-se a um dos aparelhos mais modernos. Embora o foco do CEGH esteja em doenças genéticas raras, ela está empolgada com a nova empreitada. “Estamos organizando a coleta de amostras com o Museu de Zoologia da USP, onde vão conferir a qualidade das amostras antes de nos enviar.” Um teste inicial sequenciou 70 amostras de aves, com sucesso. “Me interessa usar dados de primatas não humanos para entender o genoma humano”, diz ela, voltando a seu interesse principal. “Se encontramos em seres humanos uma variante que é comum em outros animais, a chance é grande de ela não estar por trás de problemas”, explica.

Conseguir uma infinidade de genomas completos nem é o salto mais importante da nova genô-

mica. O crucial foi perceber que o foco nos genes que produzem proteínas restringe a atenção a apenas 1,2% do DNA humano. No que diz respeito a eles, as pessoas são 99,9% iguais. A maior parte do genoma consiste em sequências que controlam e orquestram a ação dos genes e que são os verdadeiros responsáveis pelas diferenças entre organismos, em grande parte por ação de moléculas de RNA com funções regulatórias. Esse é um dos conceitos da área conhecida como “evo-devo” (evolução e desenvolvimento, *ver* Pesquisa FAPESP nº 152): ao longo do desenvolvimento, ajustes na construção das mesmas estruturas podem gerar resultados muito distintos, como a mão de uma pessoa e a asa de um morcego. “O hardware é o mesmo, o que muda é o software”, compara o geneticista molecular Paulo Amaral, do Insper. “A esponja-do-mar, que usamos de bucha, tem genes bem semelhantes aos nossos.”

O entendimento de que pequenas moléculas de RNA desligam genes (*ver* Pesquisa FAPESP nº 133) rendeu em 2006 o prêmio Nobel aos norte-americanos Andrew Fire e Craig Mello. Com as novas técnicas que surgiram desde então, a sua ação está cada vez mais posta em evidência. “Pela primeira vez, estamos produzindo genomas completos, incluindo as partes não codificantes, bem como os RNA produzidos por eles”, afirma Amaral.

Ele publicou neste ano, em parceria com o geneticista australiano John Mattick, da Universidade de Nova Gales do Sul, na Austrália, o livro *RNA – The epicenter of genetic information* (CRC Press). A base para o volume, depois ampliada e



Mico, abelha, boto, morcego, marsupial, esponja-do-mar: regulação no desenvolvimento gera parte das diferenças

## QUANTOS GENES?

O número não tem relação com a complexidade do organismo

Ser humano  
(*Homo sapiens*) ~20.000

Camundongo  
(*Mus musculus*) ~22.000

Verme  
(*Caenorhabditis elegans*) ~20.000

Mosca  
(*Drosophila melanogaster*) ~13.600

Planta  
(*Arabidopsis thaliana*) ~25.500

FONTE: MATTICK E AMARAL, 2023

atualizada, foi a revisão que lhe valeu três meses enfiado na biblioteca durante o doutorado, defendido em 2011 na Universidade de Queensland, também na Austrália, no então laboratório de Mattick. O volume repassa a história da biologia molecular desde o século XIX até agora, com foco no dilema que ainda na graduação atraiu Amaral: a função considerada principal para o RNA era traduzir genes em proteínas. Por que, então, a maior parte do genoma humano representa instruções para moléculas de RNA que não produzem proteínas? Exemplos marcantes são o gene *Xist*, que produz longos RNA reponsáveis por silenciar um dos dois cromossomos X nas células femininas de mamíferos, e o envolvimento em doenças como câncer. Junto com o grupo do imunologista Helder Nakaya, do Hospital Israelita Albert Einstein, em São Paulo, Amaral tem investigado a ação de RNA não codificadores na produção de resposta imunológica a vacinas ou em doenças cardiovasculares e neurológicas. Ao fim, o livro alça o RNA ao posto de “motor computacional da célula, do desenvolvimento, da cognição e da evolução”. “Já estamos escrevendo atualizações para uma segunda edição”, ele revela. As descobertas científicas estão avançando mais depressa que o processo editorial.

Assim, também ganham cada vez mais destaque os elementos de transposição (ver Pesquisa FAPESP nº 246), ou transposons, descritos inicialmente nos anos 1940 pela geneticista norte-americana Barbara McClintock como os responsáveis pela variação de cor nos grãos de espigas de milho – e que tendem a ter sequências repetitivas.

## DIABETES E METFORMINA

O principal tratamento oral para controlar a glicemia em portadores de diabetes do tipo 2 é a metformina, mas seu modo de ação ainda guarda mistérios e só genes codificadores de proteínas tinham sido estudados nesse contexto. O grupo de Glória Franco, da UFMG, buscou formas das moléculas de RNA longas não codificadoras (lncRNA), de acordo com artigo publicado em 2022 na revista científica *Non-coding RNA*. A conclusão, ainda exploratória, de que o medicamento regula formas críticas de lncRNA que podem afetar a resposta ao tratamento, assim como a proliferação celular e o metabolismo de energia das células podem ser promissores na investigação de como ele atua.

Franco participou dos primeiros projetos genoma trabalhando com o DNA do verme *Schistosoma mansoni*, causador da esquistossomose, e agora se interessa pela relação entre lncRNA e câncer, e resposta a fármacos.

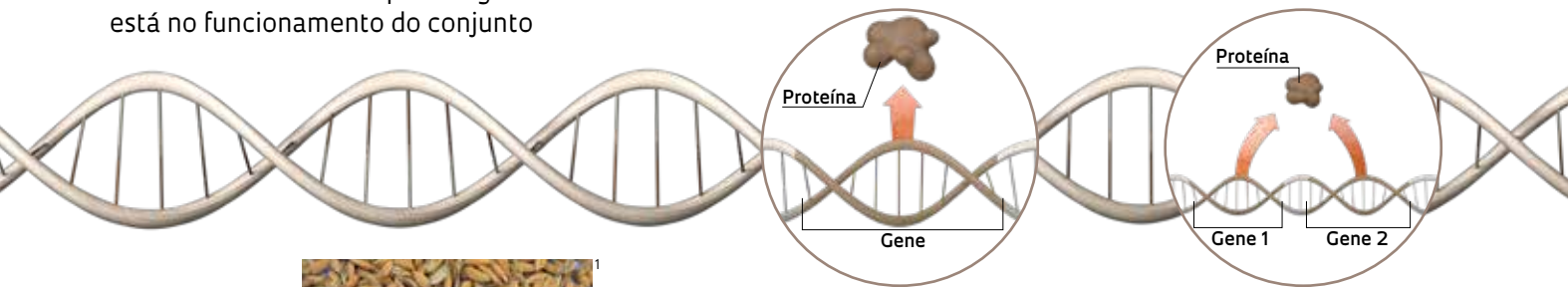


## DO GENE AO GENOMA

De meados do século XX para cá, a capacidade de enxergar em maior detalhe revelou que o segredo está no funcionamento do conjunto

Na visão mais antiga, um gene trazia o código para uma proteína; mutações podem ter efeitos graves

Mais dados indicaram que muitas características e doenças são influenciadas por vários genes atuando juntos



Arroz, tomate e borboleta: elementos de transposição contribuem para variações

São trechos replicados que podem saltar e se inserir em outra parte da fita de DNA, passando a influenciar os genes vizinhos com agilidade para sofrer alterações e gerar novidades.

Um artigo recente do grupo de Galante, depositado em fevereiro no repositório bioRxiv, investigou a origem de RNA regulatórios a partir de elementos de transposição, as retrocópias, em primatas. Cinco das 17 moléculas estudadas se mantêm iguais em todos os primatas, enquanto duas são específicas do genoma humano e podem estar envolvidas em processos biológicos essenciais, como metabolismo, comunicação entre células e desenvolvimento de vários tipos de câncer. “A linha que separa os elementos de transposição dos genes está cada vez mais tênue”, celebra a botânica Marie-Anne van Sluys, que via o tema como pouco valorizado, apesar de concentrar-se nele há décadas e atualmente liderar um estudo sobre genes, genomas e elementos de transposição em cana-de-açúcar e sua associação na interação com patógenos. Um estudo de seu grupo depositado em 2020 no repositório bioRxiv identificou um elemento de transposição capaz de modular o desenvolvimento e a expressão gênica em plantas de tabaco em resposta a estresse. “Em mamíferos, já ficou claro que eles se destacam como motores de diversificação.”

Os resultados também incluem as marcas epigenéticas (como padrões de metilação) que regulam o funcionamento do DNA sem alterar sua sequência e indicam como os genes interagem com o ambiente (*ver infográfico acima*). Eles podem estar por trás de certos tipos de tumores e sua detecção em DNA circulante na corrente sanguínea tem se revelado uma ferramenta diagnóstica promissora

## LAVOURA GENÔMICA

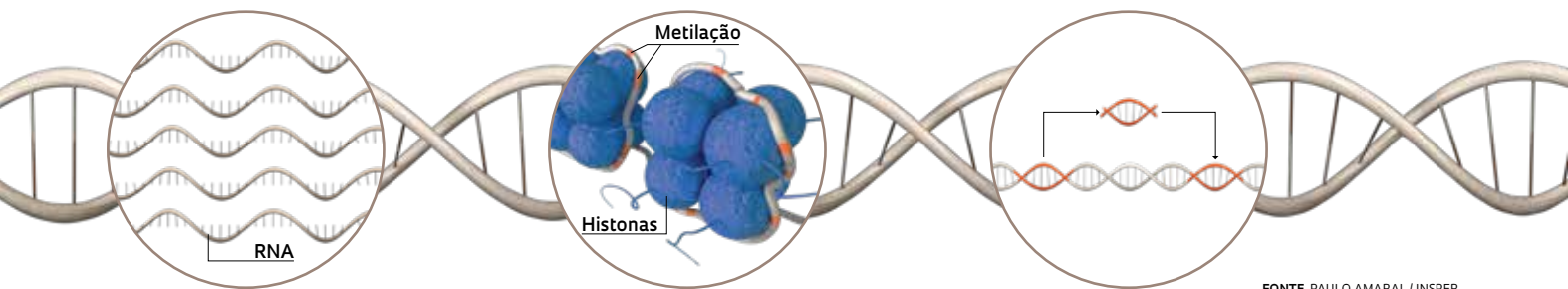
A pangenômica já chegou aos principais cultivos responsáveis por alimentar a humanidade. Sabe-se que algumas plantas passam naturalmente por duplicações do genoma inteiro, ficando com várias cópias do DNA em suas células. Esse processo aumenta a chance de gerar elementos de transposição – e, portanto, novas respostas a condições ambientais –, de acordo com artigo publicado em 2021 na revista *Genome Biology* pelo biólogo brasileiro Rafael Della Coletta, estudante de doutorado na Universidade de Minnesota, nos Estados Unidos.

Os transposons estão por trás de mudanças como reprimir o funcionamento do gene que promove o florescimento de milho quando os dias estão mais longos, promover tolerância a alumínio em arroz e gerar a forma oval dos tomates do tipo roma. De acordo com a publicação, identificar a relação causal entre mudanças no fenótipo (características observáveis de um organismo) e novidades genômicas, que podem ser alteradas por edição gênica conforme a necessidade de produção agrícola, representaria uma nova era na domesticação de plantas comestíveis.

O RNA, antes visto como uma ferramenta para fazer proteínas, emergiu como importante elemento regulador

Mecanismos epigenéticos, como metilação e modificação de histonas, têm papel crucial na orquestração do genoma

Elementos de transposição geram novidades genéticas e permitem adaptação mais rápida do que outros processos



FONTE PAULO AMARAL / INSPER

e prática, a chamada biópsia líquida desenvolvida em anos recentes (ver Pesquisa FAPESP nº 253).

### A URGÊNCIA DA BIOINFORMÁTICA

Processar e interpretar a imensidão de dados gerados requer a formação de profissionais muito especializados, os bioinformatas. Marcela Uliano-Silva está imersa nessa necessidade e acaba de desenvolver um programa de acesso aberto para a montagem de genomas mitocondriais, algo que ainda não existia, de acordo com artigo publicado em julho na revista científica *BMC Bioinformatics*. É um tipo de DNA especial: resulta de uma antiga simbiose com bactéria e é circular e transmitido geralmente de mães para filhos. Ela avisa: é impreterível que biólogos aprendam programação porque há muita necessidade de desenvolvimento de software para analisar e visualizar os dados.

Galante completa que a formação não costuma estar nos cursos de graduação brasileiros. Para tornar-se bioinformata, avisa, muitas vezes é necessário voltar aos estudos para aprender estatística e matemática, algo que a maior parte das pessoas com formação em ciências biológicas não está disposta a fazer. “As demandas em estatística e computação para analisar dados genômicos estão cada vez mais complexas”, diz. Enquanto isso, a necessidade surge e aumenta na iniciativa privada, segundo ele, e os estudantes formados em seu laboratório que dominam essa área terminam o doutorado com emprego garantido em laboratórios de análises clínicas e de oncologia, ou empresas que buscam o desenvolvimento de tecnologias computacionais. “Em cinco anos se formou um ecossistema dessas indústrias que demandam bioinformatas no país.”

Os avanços tecnológicos fazem sentido, e atingem seu potencial máximo, por meio de perguntas científicas sólidas. Mas essa é uma relação de mão dupla, ressalta Carlos Menck. “Quando buscamos uma ferramenta para resolver um problema, às ve-

## GENÔMICA NACIONAL

Iniciativas brasileiras, com apoio da FAPESP, mantiveram o país em linha com o que era feito no exterior. Iniciado em 1997, o sequenciamento da bactéria *Xyllela fastidiosa* uniu 35 laboratórios e 191 pesquisadores de instituições paulistas no âmbito da Rede de Organização para Sequenciamento e Análise de Nucleotídeos (Onsa). Foi o primeiro sequenciamento de um organismo causador de doença em plantas, com relevância comercial, que ganhou a capa da revista *Nature* em 13 de julho de 2000.

Nesse começo do século, foram também finalizados o Genoma Cana, o Genoma do Câncer e o da bactéria *Xanthomonas citri*, causadora do cancro cítrico.

Os resultados, fundamentais para o estabelecimento de uma cultura científica contemporânea, envolveram a instalação de dezenas de laboratórios, a articulação entre grupos de pesquisa, o estabelecimento de bancos de dados compartilhados e o desenvolvimento da bioinformática. Também inspirou a criação de grupos de pesquisa em todo o país.

“A formação da rede Onsa foi uma iniciativa muito ousada e de risco, que fluiu muito bem e gerou grandes resultados. Construir uma metodologia capaz de entregar uma sequência foi importantíssimo e formou muita gente nessa área”, disse o geneticista Marcio de Castro, diretor científico da FAPESP, em entrevista (ver Pesquisa FAPESP nº 328).

zes aprendemos algo que muda a questão científica.” Muitos pesquisadores de áreas ligadas à saúde que passam a maior parte do tempo no laboratório fazendo pesquisa básica, como é seu caso, têm um objetivo claro: resolver o problema dos pacientes. Enquanto buscam soluções para dificuldades práticas sobre diagnóstico ou como fazer o fármaco chegar aonde precisa, por exemplo, desvendam modos de ação moleculares; e enquanto investigam o funcionamento do DNA, podem encontrar inesperadas soluções aplicadas. ■

Os projetos e os artigos científicos consultados para esta reportagem estão listados na versão on-line.