

# GENÔMICA NO BRASIL UMA NOVA ERA NA BIOLOGIA

---

*Anamaria Aranha Camargo*

*N*o dia 22 de julho de 2002, uma reportagem na renomada revista norte-americana *The Economist* anunciava: “Samba, futebol e... genômica. A lista de aspectos pelos quais o Brasil é reconhecido foi recentemente ampliada”. O destaque foi dado à conclusão dos dois primeiros projetos genoma executados no Brasil, frutos do trabalho de centenas de pesquisadores paulistas que integravam a Rede ONSA (Organization for Nucleotide Sequencing and Analysis). Trata-se de um Instituto Virtual de pesquisa, sem sede física, criado com o apoio da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) e formado por mais de 30 laboratórios interligados pela Internet. Essa idéia inusitada, que tem servido de modelo para outros países que se iniciam na área, propiciou a estréia do Brasil no cenário da genômica e, conseqüentemente, da bioinformática ou biologia computacional.

## Da genética ao genoma

O termo *genoma* é utilizado para designar o conjunto de genes e seqüências regulatórias de um dado organismo. Os genes, por sua vez, carregam as informações genéticas que determinam todas as características de um organismo e a sua existência foi inicialmente inferida nos experimentos realizados por Gregor Mendel em 1865. Através de estudos de cruzamentos entre diferentes tipos de ervilhas, Mendel verificou que certas características físicas dessas plantas eram transmitidas de geração para geração, propondo então a existência de “fatores” que seriam responsáveis por essa transmissão. Por volta de 1902, Walter Sutton e Theodor Boveri verificaram que o padrão de herança dos “fatores” descritos por Mendel acompanhava a segregação dos cromossomos de células em divisão. Já em 1915, Thomas Morgan concluiu que estes “fatores” estavam organizados de maneira linear nos cromossomos e, assim, propôs, pela primeira vez, a correlação entre um gene (*genótipo*) e uma característica física (*fenótipo*).

A natureza química dos genes, entretanto, só foi revelada na década de 40 e 50 através de experimentos que demonstraram que os genes estavam contidos na molécula de DNA. Em 1953, James Watson e Francis Crick determinaram a estrutura física do DNA e o modelo proposto da dupla fita foi fundamental para a compreensão do mecanismo de transmissão e execução da informação genética. Após quatro anos, Crick e George Gamow estabeleceram o dogma central da biologia molecular (DNA-RNA-proteína) e, assim, propuseram um modelo através do qual a informação genética contida no DNA seria transmitida e executada pela célula. Já em 1966, Marshall Nirenberg, Heinrich Mathaei e Severo Ochoa demonstraram que seqüências sucessivas de três nucleotídeos do DNA (códon) determinavam a seqüência de aminoácidos de uma proteína. Estava, então, desvendado o código genético.

Com o desenvolvimento das técnicas de manipulação do DNA (técnicas do DNA recombinante) e, em especial, da técnica de seqüenciamento, tornou-se possível isolar e determinar a seqüência dos genes. Conseqüentemente, durante as décadas de 70, 80 e 90, as seqüências de milhares de genes de diferentes organismos foram determinadas. Em 1979, biólogos e matemáticos reuniram-se na Rockefeller University e propuseram a construção de um banco de dados para armazenar as seqüências de DNA produzidas pelos diversos grupos. O objetivo da construção deste banco era

que os pesquisadores pudessem disponibilizar e comparar suas seqüências com dados de outros grupos. Em 1981, foi criado, então, o EMBL Database (European Molecular Biology Laboratory Database) e, em 1982, o GenBank. Já no final de 1990 o GenBank armazenava mais de 50 milhões de nucleotídeos contidos em 40 mil seqüências de DNA.

Com os novos recursos disponíveis, foram dissecados vários mecanismos biológicos, como por exemplo a replicação do DNA, a divisão celular e o desenvolvimento embrionário. Cientistas das mais diversas áreas começaram a utilizar as ferramentas da biologia molecular para responder perguntas específicas dentro de suas linhas de pesquisa e passaram, desta forma, vários anos estudando um conjunto particular e restrito de genes.

### **O nascimento da genômica e da bioinformática**

No início da década de 80, isolar e caracterizar o conjunto completo de genes de um organismo parecia um objetivo muito distante. A forma manual e limitada com que os dados de seqüenciamento eram gerados e analisados apresentava-se como o fator limitante. A situação só foi revertida com o desenvolvimento dos primeiros seqüenciadores semi-automáticos de DNA. A preparação do DNA continuava sendo feita de forma manual, mas a aquisição e o processamento dos dados de seqüenciamento passaram a ser automatizados e informatizados.

Em 1995, um grupo de 30 pesquisadores do Instituto para Pesquisa Genômica (The Institute for Genomic Research – TIGR) liderados por Craig Venter e em colaboração com pesquisadores da Universidade John Hopkins, da Universidade Estadual de Nova Iorque, em Buffalo, e do Instituto Nacional de Padrões e Tecnologia (National Institute for Standards and Technology) publicaram a seqüência completa dos 1.8 milhões de nucleotídeos que compunham o material genético da bactéria *Haemophilus influenzae*. Nascia, assim, a genômica, uma ciência voltada para a produção e análise de seqüências de DNA de genomas completos.

Alguns meses depois, um grupo de somente 5 pesquisadores concluiu, em apenas 8 semanas, o seqüenciamento do genoma de 580kb do *Mycoplasma genitalium*. Desde então, o número de genomas completamente seqüenciados não pára de crescer. Com o desenvolvimento da tecnologia de seqüenciamento e o aperfeiçoamento dos programas para análise de dados, tem sido possível seqüenciar genomas cada vez maiores e mais complexos. Em 1997, mais um

marco foi alcançado com a publicação da seqüência completa do primeiro organismo eucarionte, a levedura. Desde então, cresce a lista de genomas eucariontes completamente seqüenciados: *C. elegans*, *Drosophila*, *Arabidopsis*, *Plasmodium* e, finalmente, o genoma humano.

A automatização e a informatização do processo de seqüenciamento foram essenciais, não só para aumentar a quantidade dos dados disponíveis, mas, principalmente, para viabilizar o desenvolvimento de ferramentas computacionais que permitissem a manipulação e a análise sistemática dos dados gerados. Nascia, assim, a irmã gêmea da genômica: a bioinformática.

A bioinformática é uma área multidisciplinar que envolve biólogos, físicos, técnicos em computação e é comumente dividida em duas escolas: a de desenvolvimento de novos algoritmos e ferramentas e a de análise e interpretação de dados. O termo bioinformática apareceu na literatura pela primeira vez em 1991, mas, hoje em dia, parece difícil imaginar a época em que bancos de dados de seqüências e ferramentas de análise não estavam disponíveis através da Internet.

## **A genômica no Brasil**

Em 1998, a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) criou a Rede ONSA (Organization for Nucleotide Sequencing and Analysis) com o objetivo ambicioso de formar uma rede de cientistas brasileiros especializados em uma das áreas mais promissoras e inovadoras da Biologia: a genômica. Ao contrário dos demais centros de seqüenciamento do mundo, a Rede ONSA é um Instituto Virtual de seqüenciamento, sem sede física, e formada por mais de 30 laboratórios de diferentes instituições de pesquisa do Estado de São Paulo que estão interligados pela Internet.

A idéia de criar um Instituto Virtual, proposta pelo Diretor Científico da FAPESP, professor José Fernando Perez, foi inovadora e vem sendo utilizada como modelo para outros países que começam a se aventurar na área da genômica. Por não haver necessidade de gastos com a construção de prédios e contratação de cientistas, o conceito de Instituto Virtual permitiu que todos os recursos destinados ao projeto fossem gastos em equipamentos de última geração, reagentes e treinamento de pessoal. A criação da rede permitiu, ainda, que o projeto fosse agilizado e executado em tempo menor que o previsto. Mais importante do que isso, no entanto, foi o estímulo oferecido aos diferentes

grupos de pesquisa para atuarem de forma colaborativa e integrada. Este espírito colaborativo entre os pesquisadores foi intenso e elogiado pelo comitê estrangeiro que fazia a avaliação do projeto.

O primeiro desafio da Rede ONSA foi realizar o seqüenciamento completo do genoma da bactéria *Xylella fastidiosa*. Essa bactéria vive no interior do xilema da laranja e é o agente causador da doença popularmente conhecida como amarelinho. A escolha do organismo foi justificada pela associação da bactéria com uma doença de impacto significativo na economia do Estado de São Paulo, um dos maiores produtores de laranjas do mundo. O projeto foi coordenado pelo Dr. Andrew J. G. Simpson do Instituto Ludwig de Pesquisa sobre o Câncer e durou, aproximadamente, dois anos de intenso trabalho e aprendizado. O trabalho resultou na publicação de um artigo científico na revista *Nature* e relatava, pela primeira vez, o seqüenciamento completo do genoma de um fitopatógeno. Características importantes da bactéria foram inferidas a partir do conjunto de genes identificados no projeto, servindo, desta forma, de base para futuros estudos funcionais aplicados voltados para o combate à doença.

O projeto genoma da *Xylella* marcou a estréia do Brasil no cenário da genômica. O próximo desafio era manter o sucesso obtido no primeiro projeto e, se possível, ampliar ainda mais os conhecimentos e os méritos alcançados na área. Esse objetivo não foi difícil, uma vez que na seqüência três grandes projetos foram iniciados: o Projeto Genoma da *Xanthomonas campestris*, o Projeto Genoma Humano do Câncer e o Projeto Genoma da Cana de Açúcar. Novos grupos foram incorporados à Rede ONSA, os projetos acabaram superando as expectativas iniciais e, conseqüentemente, o reconhecimento nacional e internacional foi obtido. Outros projetos foram concluídos ou ainda estão em andamento: Projeto Genoma de uma segunda cepa de *Xylella fastidiosa* causadora da doença de Pierce, Projeto Genoma de *Leifsonia xyli* e Projeto Genoma de *Schistossoma mansoni*. A genômica definitivamente se estabeleceu no Estado de São Paulo e a tecnologia e treinamento adquiridos passaram a ser aplicados em outras áreas não diretamente relacionadas aos projetos genomas, estendendo-se a um número cada vez maior de pesquisadores.

Como não poderia deixar de ser, o estabelecimento da genômica trouxe, também, o domínio da bioinformática. Laboratórios pioneiros liderados pelos pesquisadores João Setúbal, João Meidanis, João Paulo Kitajima e Sandro J. de

Souza coordenaram a bioinformática dos primeiros projetos genomas. Considerando-se que a disponibilidade de pessoas treinadas nessa área era limitada, o desafio ganhava ainda maior relevância. No entanto, esses laboratórios são hoje centros de referência no Brasil e no exterior e a formação em bioinformática passou a ser uma das prioridades dos cursos de graduação e pós-graduação.

Reconhecendo a importância da genômica e incentivado pelo sucesso dos projetos em São Paulo, o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e o MCT (Ministério da Ciência e Tecnologia) lançaram o Projeto Genoma Nacional. A criação de uma rede nacional de laboratórios de seqüenciamento era um projeto ambicioso dadas as grandes extensões territoriais. O alvo seria novamente um genoma bacteriano, mas, dessa vez, de importância ambiental. Como organismo alvo foi escolhida a bactéria *Chromobacterium violaceum* que vive nas águas de rios amazônicos e tem como característica principal a capacidade de produção de violaceína – um pigmento com atividade bactericida. A rede de seqüenciamento nacional foi estabelecida rapidamente cobrindo o país de norte a sul. Prevalecendo mais uma vez o espírito colaborativo e o trabalho dedicado, o sucesso foi merecidamente alcançado. Paralelamente à rede nacional, foram criadas as redes regionais de seqüenciamento, de modo que podemos definitivamente afirmar que o Brasil é, também, o país da genômica. A exemplo do futebol, já podemos nos considerar pentacampeões e tudo isso em menos de cinco anos. E ainda temos uma seleção formada por centenas de pesquisadores e técnicos altamente qualificados, prontos para encarar novos desafios.

### A genômica e a formulação de hipóteses

Ao contrário das demais áreas da biologia, a genômica não está atrelada à forma convencional de se fazer ciência, que consiste em formular hipóteses para solucionar problemas pontuais. Na área da genômica, a formulação de hipóteses é, em um primeiro momento, irrelevante, pois não é possível formular hipóteses com base no desconhecido, da mesma forma que não podemos interpretar um livro sem ler todos os seus capítulos. A formulação de hipóteses passa a ser, então, uma segunda etapa, dependente da descoberta dos genes, da caracterização do genoma e do processamento dos dados com o auxílio de recursos computacionais. Desta forma, a genômica não pode ser entendida como um produto final, mas sim como um meio para formular hipóteses e resolver problemas mais abrangentes e complexos.

A genômica está mudando de forma definitiva os horizontes da biologia e da medicina; o grande desafio das próximas décadas será o desenvolvimento de métodos experimentais em larga escala voltados para a validação das hipóteses formuladas *in silico*. Já é possível, por exemplo, identificar alterações no funcionamento de mais de mil genes da levedura quando esta é cultivada em um meio nutricional específico. Da mesma forma, pode-se determinar o perfil genético de uma célula tumoral e, com base no padrão de genes expressos, obter informações sobre a agressividade do tumor e a resposta ao tratamento. Além disso, já é possível escolher, de forma mais eficiente, genes candidatos à vacina contra patógenos humanos analisando diretamente o conjunto de genes que codificam proteínas de superfície em bactérias patogênicas como, por exemplo, a *Streptococcus pneumoniae*. Em um futuro não muito distante, acredita-se que os formuladores de hipóteses revolucionárias serão os cientistas que dominam e promovem o desenvolvimento da genômica e da biologia computacional. Que bom que o Brasil é também o país da genômica!

Anamaria Aranha Camargo é bióloga, pesquisadora-assistente do Laboratório de Genética do Câncer do Instituto Ludwig, em São Paulo e coordenadora do projeto Transcript Finishing Initiative Ludwig/FAPESP.  
anamaria@compbio.ludwig.org.br