

REPORTAGENS



A origem e o sentido da bioinformática

João Carlos Setubal

Um dos desdobramentos da descoberta de Watson e Crick em 1953 de que o DNA é estruturado como hélice dupla é a **bioinformática**, esse casamento entre o computador e a biologia molecular. É um desdobramento **novo**, recente. Há 10 anos, o termo nem sequer existia. Apesar de nova, era possível prever que a bioinformática iria acontecer, desde 1953. Para entender isso preciso contar um pouco de história.

A história começa na década de 1940, quando foi inventado o moderno computador digital. O primeiro computador moderno começou a funcionar por volta de 1946. Ele se chama digital, pois os dados são armazenados com um alfabeto binário, os **dígitos binários**, zeros e uns. A operação também é digital, toda em cima da lógica do liga/desliga. Um pouco antes, em 1944, Avery e colaboradores descobriram que o DNA era a substância que carregava a informação genética. Vemos então que o nascimento do moderno computador digital e da moderna biologia molecular (usando a descoberta de Avery como referência) se deram mais ou menos ao mesmo tempo.

A descoberta da hélice dupla, em 1953, mostrou que a informação genética também é armazenada de forma digital. Ou seja, a informação genética também é escrita com um alfabeto, só que ele é **quaternário** e não binário, pois são usadas 4 letras, as famosas A, C, G, e T. Mais tarde se descobriu que a forma dos genes operarem também é, até um certo ponto, digital: os genes podem ser "ligados" ou "desligados". Apenas esta observação já seria suficiente para prever, na década de 1950, que um dia informática e biologia molecular iriam juntas fazer nascer uma nova área do conhecimento. Esse nascimento entretanto teve que esperar muito tempo para realmente acontecer, e é essa a razão da bioinformática ser uma aparente novidade. Não existe uma data precisa, mas algumas pessoas consideram que a bioinformática passou a ser reconhecida como importante pelo mundo científico por volta de 1995 (é o ano em que o primeiro genoma de uma bactéria foi publicado). Por que tão longa demora?

Do lado da biologia molecular o motivo é muito simples: apesar da estrutura do DNA ter sido desvendada em 1953, a informação nela contida não podia ser "lida". Foi como se tivéssemos descoberto o alfabeto utilizado para escrever "o livro da vida", mas as "palavras" desse livro estavam com letrinhas tão pequenas que não conseguíamos lê-las. Foi preciso esperar até fins da década de 1980 para que aparecesse uma "lente de aumento" suficientemente boa (e automática -- uma máquina) que permitisse a leitura dessas letrinhas em grandes quantidades. Em 1995 uma única máquina dessas já conseguia ler milhares de letrinhas por dia.

Editorial
Bioinformática, genes e inovação Carlos Vogt
Reportagens
Matematização da biologia não é novidade da ciência moderna
Software para bioinformática ganha o mercado
Será a bioinformática uma área já constituída?
Comunidade realiza primeiro encontro da área
Computadores simulam vida artificial
Concentração da bioinformática decorre do Genoma
Capacitação ainda é o principal obstáculo para a área
Biotecnologia: aplicações e tendências
Artigos
A origem e o sentido da bioinformática João Carlos Setubal
Neurobioinformática: quando mentes e genes se encontram Luciano da F. Costa
Bioinformática: manual de utilização João Paulo Kitajima
As tecnologias de informação digital e a pós-organicidade Rosemary Segurado

Bioinformática na indústria João Meidanis
Rede de pesquisa e desenvolvimento em bioinformática do centro-oeste Georgios Pappas Jr.
Sting Millennium Suite: ferramentas para análise estrutural de proteínas Paula Kuser Falcão, Michel Yamagishi, Roberto Higa e Goran Neshich
Poema
A medida do tempo Carlos Vogt
Créditos

Do lado da computação foi também preciso um amadurecimento. Esse amadurecimento é a tão falada revolução da informática, com computadores sendo capazes de armazenar cada vez mais informação, de processá-la de modo cada vez mais rápido, a um custo cada vez menor. É interessante observar que se o seqüenciamento automático do DNA tivesse amadurecido mais rapidamente, digamos com 20 anos de antecedência, não haveria computadores com poder suficiente para dar conta dos dados gerados. Na década de 1970 a unidade básica de armazenamento de informação era o kilobyte -- 1000 bytes, aproximadamente 1000 letras. Um computador de grande porte daquela época tinha alguns kbytes de memória. Com tal memória um computador desses não seria capaz de processar nem sequer o genoma de um vírus, que pode chegar a 20 kilobases, ou 20 mil lettrinhas; que dirá o genoma humano, com seus 3 bilhões de lettrinhas.

Então, através de uma evolução que parece mais ou menos sincronizada, desembocamos em 1995, quando os computadores já estavam suficientemente poderosos para poder processar os milhões e milhões de lettrinhas que passaram a vir à luz. E assim nasceu a bioinformática, com a missão de ajudar-nos a entender a história que está escrita nesse livro da vida.

Estamos em 2003, portanto a bioinformática já está aí com uns 8 anos, e posso testemunhar que vai indo em geral muito bem. Do meu ponto de vista a biologia molecular e o computador "nasceram um para o outro". Bem, não é bem assim, visto que o computador é meio polígamo. Outra coisa a mencionar é que existe um gargalo: é a falta de gente com qualificação para trabalhar nessa área interdisciplinar. De onde vem esse gargalo? Apesar desse longo tempo de maturação que eu delinee, o aparecimento da bioinformática pegou quase todo mundo de surpresa. O motivo disso, a meu ver, é uma insistência da moderna vida intelectual, pelo menos desde o século XIX, se não for de antes, em compartimentalizar o conhecimento. Criamos essas "gavetas" chamadas biologia, química, física, matemática, etc, e disciplinamos nossas vidas acadêmicas ao redor delas. Quando aparece algo misturado, como é o caso da bioinformática, essas gavetas tendem mais a atrapalhar do que ajudar. Digo isto porque um bioinformata no sentido pleno do termo precisa entender tanto de biologia quanto de informática e é difícil conseguir uma tal formação com os cursos acadêmicos com sua estrutura atual.

Voltando para a poligamia da computação, é válido colocar a seguinte pergunta: até que ponto essa onda em torno de bioinformática é justificada? Afinal de contas, hoje quase toda atividade científica depende do computador. Então poderíamos falar em física-informática, astronomia-informática, arqueologia-informática, etc. Será que há algo de especial na bioinformática? Eu acho que sim e vou tentar explicar por quê.

Em primeiro lugar quero distingüir dois tipos de problemas em que atua a bioinformática. O primeiro tipo de problema é o que eu chamo de **problema biotecnológico**. O exemplo clássico é o da **montagem de DNA**. Um genoma de bactéria tem em geral 3 ou 4 milhões de lettrinhas ou bases. As máquinas seqüenciadoras conseguem ler apenas pedaços de cerca de 1000 bases. Então como é possível ler um livro de 3 ou 4 milhões de letras se só conseguimos ler fragmentos de 1000 letras? A solução é gerar uma enorme quantidade de fragmentos que tenham sobreposição entre si. Para ler

3 ou 4 milhões são necessários cerca de 100 mil desses fragmentos. Aí, obviamente, é necessário um programa de computador para montar esse quebra cabeça.

Mas esse problema é fruto de uma limitação tecnológica atual. É bem possível que daqui a 5 anos apareça um novo tipo de seqüenciador que consiga ler **diretamente** as 3 ou 4 milhões de bases. E quando essa tecnologia aparecer, o problema da montagem do DNA deixa de existir. É isso que eu chamo de problema biotecnológico. E, tal como esse, existem dezenas ou centenas de outros problemas na bioinformática, cada um deles motivado por uma particular tecnologia. Mas problemas desse tipo existem em outras ciências também. Certamente os telescópios modernos geram grandes quantidades de dados de um jeito e formato que exigem programas de computador sofisticados para coleta e interpretação; e quando mudam os telescópios mudam os programas.

O meu argumento é de que em bioinformática existe uma segunda classe de problemas que têm um interesse que vai além de tecnologias específicas, que transcende qualquer tecnologia, e diz respeito à natureza mesmo da biologia molecular. E, na minha opinião são esses problemas que dão um charme todo especial à bioinformática. Que problemas são esses? Eu diria que são basicamente de 2 tipos, embora isto certamente seja uma simplificação exagerada.

Primeiro, temos a interpretação do DNA como uma **linguagem**, a linguagem dos genes. Os genomas contém informação. Queremos saber que informação está contida neles. E sempre vamos querer saber isso. Vamos querer saber para diferentes espécies de organismos e para diferentes indivíduos de uma espécie, particularmente a nossa.

Essa interpretação requer métodos, técnicas, algoritmos que vêm principalmente da informática, pois afinal ela é a ciência da informação. Sem o uso dessas técnicas as seqüências de DNA produzidas pelas máquinas e montadas pelos programas não passam de uma inútil sopa de letrinhas. Dar sentido a essa sopa é tarefa dos bioinformatas. Nesta linha é interessante observar que existe um paralelo entre o uso da informática para decifrar a informação genética e uma das primeiras grandes aplicações dos computadores (ou mesmo motivador da sua criação), que foi o deciframento, pelos ingleses, do código secreto usado pelos alemães na 2ª. guerra mundial. Um artigo recente que faz uma apanhado geral sobre as técnicas das ciências da informação usadas no deciframento da linguagem genética é D. Searls, *The Language of Genes*, *Nature*, 420:211-217, 2002.

O segundo tipo de problema é o de entender os **efeitos** da informação genética. Temos aqui efeitos de **gênese, de criação**, ou seja, como a partir de um ovo ou de uma semente se chega a um indivíduo; e também efeitos de **manutenção da vida**, ou seja, como um indivíduo já formado responde ao ambiente em que vive. O entendimento dos efeitos da informação genética começa com o entendimento de como uma célula funciona. Até agora a biologia celular estudava isso com experimentos: perturba-se a célula de uma certa forma, e observa-se o resultado. Perturba-se a célula de outra forma, e observa-se o resultado, e assim por diante. Agora temos a possibilidade de tentar descobrir qual é o efeito que uma perturbação vai ter sem ter que efetivamente (em bancada de laboratório) realizar

essa perturbação. Para isso bastará "ler as instruções", "ler as especificações" e fazer as deduções necessárias. Isto é, ler o genoma e entender a cadeia de dependências entre os genes. Essas cadeias são assim: um efeito externo faz um gene produzir uma proteína, que ativa outro gene, que produz outra proteína, que ativa outro gene e assim por diante. Novamente aqui essa é uma tarefa que vai precisar de muita bioinformática. E novamente aqui já acumulamos vasta experiência com redes de circuitos elétricos e eletrônicos usados por computadores, e essa experiência será fundamental para o entendimento das **redes gênicas**.

Então este é o fascínio da bioinformática: ela está abrindo caminho para a compreensão desse novo mundo que é a biologia molecular baseada na informação genética. Para finalizar é interessante notar que a biologia molecular está causando influências na própria computação. Alguns exemplos são a utilização da palavra "vírus" para descrever certos tipos de programas, os algoritmos genéticos, a "vida artificial", e a computação com DNA. Mas esta é uma outra história.

João Carlos Setubal é coordenador do Laboratório de Bioinformática da Unicamp.



Atualizado em 10/08/2003

<http://www.comciencia.br>
contato@comciencia.br

© 2003
SBPC/Labjor
Brasil

Contador de acessos:

