

O seqüenciamento do genoma humano atesta as semelhanças científicas entre animais, plantas e bactérias e alimenta a idéia de que somos apenas uma espécie a mais sobre a Terra

Golpe no

CARLOS FIORAVANTI E MARCOS PIVETTA

Por um momento, o homem se sentiu pequeno. Exatamente no mesmo dia, 12 de fevereiro, os dois grupos que se digladiavam há anos pelo feito de terminar mais rápido o seqüenciamento de todo o genoma humano – o consórcio público internacional e a empresa privada norte-americana Celera Genomics – divulgaram, de forma separada e em publicações distintas, a mesma e surpreendente notícia. Depois de mapearem cerca de 95% do código genético humano, estimaram que o homem tem cerca de 30 mil genes, três a quatro vezes menos do que imaginavam. No emaranhado de dados, análises e opiniões que recheavam as duas radiografias iniciais do nosso genoma, impressas nas páginas da britânica *Nature* (consórcio público) e da norte-americana *Science* (Celera), esse número chamou a atenção de todos. Apenas 30 mil genes! A espécie que domina o planeta, o ser capaz de cravar sua bandeira na Lua e voltar à Terra, abriga em cada célula pouco mais que o dobro do número de genes de vermes e moscas. A reação da sociedade foi imediata. Espanto geral e piadinhas comparando o *Homo sapiens* a pequenos seres alados e rastejantes. Então seria essa a principal conclusão de um dos mais badalados e caros programas científicos já realizados pela humanidade?

Além dessa constatação, os dados do genoma mostraram que os genes são distribuídos de forma irregular pelos 23 pares de cromossomos, que formam o genoma humano. Há cromossomos com alta incidência de genes e outros com pouquíssimos (*ver pág. 35*). Também se notou a predominância de seqüências repetidas, o chamado DNA-lixo, cuja função ainda é pouco conhecida. Por enquanto, é interpretado como uma evidência de que nosso código genético incorporou seqüências de outro seres (bactérias, por exemplo) e ainda não se livrou desse material de utilidade duvidosa. O fim de um grande trabalho científico pode frequentemente produzir

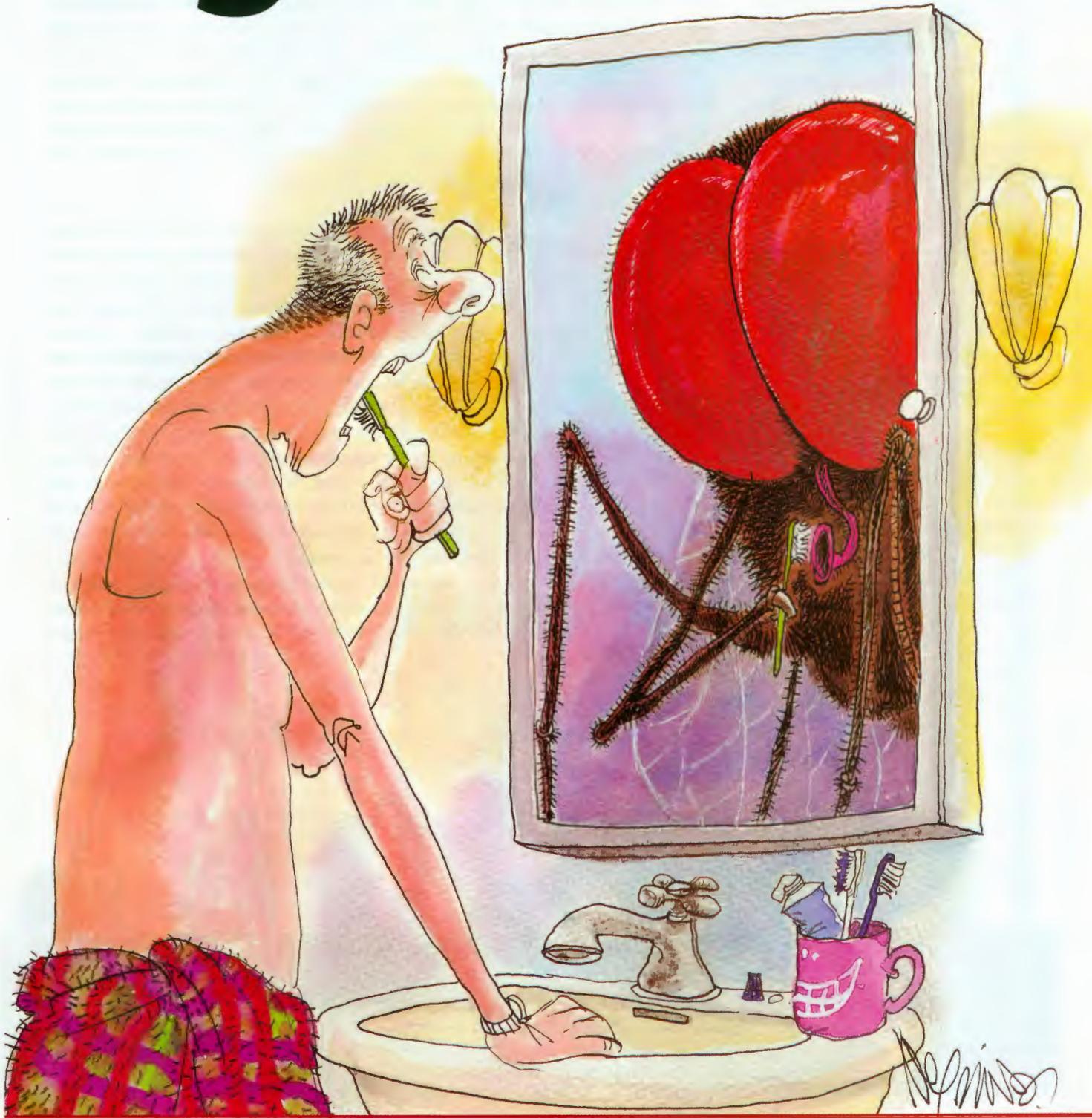
mais dúvidas do que certezas. Foi o que se deu com o genoma humano, disseminando um efeito aparentemente – apenas aparentemente – oposto ao que inicialmente se esperava. À medida que se analisam as informações dos dois rascunhos do nosso DNA, uma série de novas – e velhas – questões entra na ordem do dia. A seguir, algumas delas:

O fim do começo ainda não terminou

É preciso deixar bem claro: o mapa do seqüenciamento, tal qual ganhou as páginas da *Nature* e da *Science*, é ainda um rascunho (o segundo) do nosso DNA, embora já exibindo contornos muito próximos da forma final. É como se a humanidade tivesse recebido uma enorme biblioteca com um catálogo ainda precário, que não permite saber quantos livros existem em todas as estantes nem separar as obras importantes das medianas e das quase sem valor. Nenhum dos resultados apontados pelos estudos publicados nas duas revistas é definitivo e inquestionável. Os números são provisórios e precisam ser mais bem calculados, as análises ainda carecem de refinamento e há uma série de questões em aberto. “Gerada por programas de computador, a atual configuração do genoma, com esse reduzido número de genes, é uma ótima hipótese de como deve ser o nosso DNA, mas ainda é uma hipótese”, pondera Marcelo Briones, professor de biologia e evolução molecular da Universidade Federal de São Paulo (Unifesp).

Esses senões são necessários por um motivo simples. A rigor, os pesquisadores dos dois grupos ainda não botaram ponto final no gigantesco trabalho a que se propuseram: estabelecer a ordem correta dos 3,2 bilhões de bases nitrogenadas – adenina, citosina, guanina e timina, representadas respectivamente pelas letras A, C, G e T – dispersas pelos cromossomos. Em junho do ano passado, já haviam anunciado a conclusão de mais de 80% do seqüenciamento do genoma humano, mas, na ocasião, não botaram no papel os seus achados. Agora, deram um

orgulho vãõ

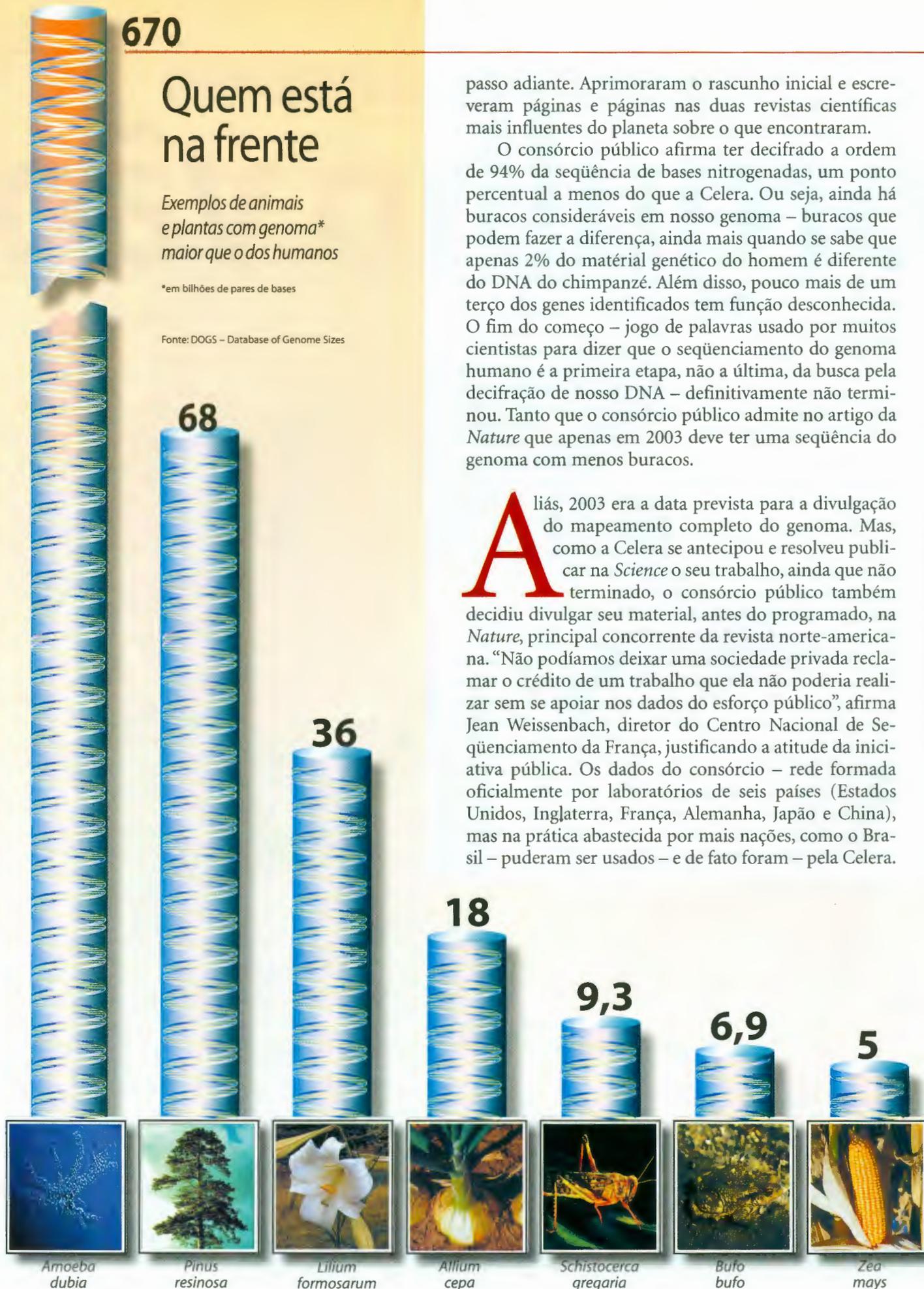


Quem está na frente

Exemplos de animais e plantas com genoma* maior que o dos humanos

*em bilhões de pares de bases

Fonte: DOGS – Database of Genome Sizes



passo adiante. Aprimoraram o rascunho inicial e escreveram páginas e páginas nas duas revistas científicas mais influentes do planeta sobre o que encontraram.

O consórcio público afirma ter decifrado a ordem de 94% da seqüência de bases nitrogenadas, um ponto percentual a menos do que a Celera. Ou seja, ainda há buracos consideráveis em nosso genoma – buracos que podem fazer a diferença, ainda mais quando se sabe que apenas 2% do material genético do homem é diferente do DNA do chimpanzé. Além disso, pouco mais de um terço dos genes identificados tem função desconhecida. O fim do começo – jogo de palavras usado por muitos cientistas para dizer que o seqüenciamento do genoma humano é a primeira etapa, não a última, da busca pela decifração de nosso DNA – definitivamente não terminou. Tanto que o consórcio público admite no artigo da *Nature* que apenas em 2003 deve ter uma seqüência do genoma com menos buracos.

Aliás, 2003 era a data prevista para a divulgação do mapeamento completo do genoma. Mas, como a Celera se antecipou e resolveu publicar na *Science* o seu trabalho, ainda que não terminado, o consórcio público também decidiu divulgar seu material, antes do programado, na *Nature*, principal concorrente da revista norte-americana. “Não podíamos deixar uma sociedade privada reclamar o crédito de um trabalho que ela não poderia realizar sem se apoiar nos dados do esforço público”, afirma Jean Weissenbach, diretor do Centro Nacional de Seqüenciamento da França, justificando a atitude da iniciativa pública. Os dados do consórcio – rede formada oficialmente por laboratórios de seis países (Estados Unidos, Inglaterra, França, Alemanha, Japão e China), mas na prática abastecida por mais nações, como o Brasil – puderam ser usados – e de fato foram – pela Celera.

Afinal, quantos genes temos?

Os dois artigos, da *Nature* e *Science*, situam o número de genes do *Homo sapiens* entre 26 mil e 40 mil. A tão comentada cifra de 30 mil genes humanos é uma espécie de média de consenso, que parece ter agradado tanto aos cientistas do consórcio público quanto aos da Celera. Antes da publicação, estimava-se que nossa espécie tinha cerca de 100 mil genes. Algumas previsões falavam em até 120 mil, 140 mil genes.

No entanto, há quem diga que a atual previsão de 30 mil genes é tão cristalina e indiscutível quanto o resultado da última eleição para presidente dos Estados Unidos. Eles apostam que – mais dia, menos dia – vai haver uma recontagem. Cientistas que participam de projetos genômicos no Brasil acreditam que podem existir mais genes ainda não detectados pelos modelos matemático-computacionais da Celera e do consórcio público. O total, dizem, pode chegar a 50 mil, se forem considerados os chamados genes transcritos, que formam as moléculas de ácido ribonucléico (RNA), a base da síntese das proteínas.

Andrew Simpson, coordenador do Genoma Humano do Câncer (GHC), projeto financiado pela FAPESP e pelo Instituto Ludwig, é um dos que sustentam essa projeção. No ano passado, sua equipe aplicou no cromossomo 22, um dos menores e dos primeiros a ser decifrado, a técnica usada pelo GHC, as ESTs ou etiquetas de seqüências expressas, que indica apenas os trechos atuantes da molécula de ácido desoxirribonucléico (DNA). Resultado: o grupo paulista encontrou 219 novas regiões transcritas, que parecem corresponder a cerca de 100 genes que ainda não haviam sido descritos. De tão relevante, a descoberta ganhou as páginas da edição de 7 de novembro de 2000 da *Proceedings of the National Academy of Sciences*, dos Estados Unidos. “Nosso

trabalho, se continuar bem feito, pode contribuir para determinar com precisão o número de genes do genoma humano”, comenta Simpson.

Outra indicação de que o número de genes humanos pode estar sujeito a ajustes vem da equipe paulista do recém-concluído genoma da cana-de-açúcar, que mapeou parcialmente o DNA dessa planta e rastreou cerca de 80 mil genes (ver Pesquisa FAPESP nº 59). Cruzando as informações obtidas no seqüenciamento da cana-de-açúcar com as disponíveis no GenBank, banco de dados sobre todos os genomas concluídos ou em andamento, os pesquisadores ligados ao projeto da FAPESP encontraram – quem diria – de 200 a 1.000 genes ainda não identificados na *Arabidopsis thaliana*, a primeira planta inteiramente seqüenciada, no final do ano passado. Ainda que sujeito a valores mais precisos, o resultado da comparação pode aumentar em até 5% o total de genes (25 mil) previstos na *Arabidopsis*. Se foram encontrados novos genes nessa planta, por que o mesmo não pode acontecer com o ser humano?

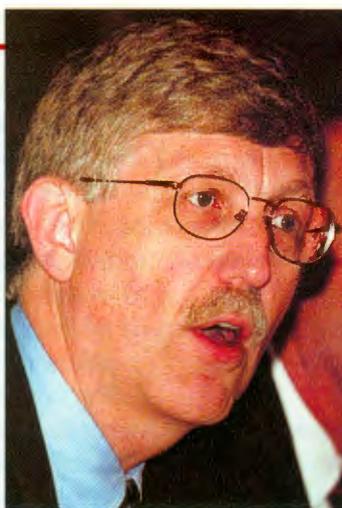
Determinismo genético versus fatores ambientais

A hipótese de o ser humano ter apenas 30 mil genes reavivou esse velho debate. Os críticos do determinismo genético, em geral pessoas e cientistas da área de humanas, mas também ilustres biólogos como Richard Lewontin, da Universidade de Harvard, nos Estados Unidos, ganharam alguns reforços com a notícia da suposta escassez de genes do *Homo sapiens*. Com tão poucos genes, como podemos creditar tudo que somos – aparência física, propensão a doenças e gostos pessoais – apenas ao DNA e relegar a segundo plano o papel do ambiente? Críticos mais incisivos decretaram a morte do conceito de gene, como fez a *Folha de S. Paulo* em um editorial logo após a publicação dos dados do consórcio público e da Celera.



No que parece ser uma mudança de postura, Craig Venter, dono e principal cientista da Celera, certamente um dos homens que mais sonham em ganhar dinheiro com o estudo dos genes, começou a proferir uma série de afirmações antibombásticas sobre o peso das seqüências de As, Cs, Gs e Ts em nossa existência, após a publicação do artigo de sua companhia. “A montagem da seqüência do genoma humano é apenas o primeiro e hesitante passo de uma longa e excitante jornada na direção do entendimento do papel do genoma na biologia humana”, disse. Ou: “Duas falácias devem ser evitadas: o determinismo, a idéia de que todas as características de um ser são ditadas pelo genoma; e o reducionismo, (acreditar que) agora que a seqüência humana é totalmente conhecida é apenas uma questão de tempo entendermos as funções e interações dos genes que darão uma completa descrição causal da variabilidade humana” (ver entrevista abaixo).

Francis Collins, principal coordenador do trabalho do consórcio público do genoma humano, não acompanhou Venter no discurso moderado. Dias após a publi-



Collins: concorrência e resultados antecipados

cação do artigo de sua equipe, Collins participou da reunião anual da Associação Americana para o Progresso da Ciência, em São Francisco, nos Estados Unidos, e insistiu em difundir a idéia-clichê de que o DNA é o “livro da vida”, metáfora que Venter, também presente no encontro científico, fez questão de rechaçar. Os dois, que sempre tiveram tantas divergências – acerca do patenteamento e dos métodos de seqüenciamento de genes –, conseguiram acrescentar mais um item em sua lista de diferenças.

Genômica comparativa

Uma área que sai em alta com a publicação das seqüências presentes em nosso DNA, por mais que as analogias tendam a jogar por terra o antropocentrismo humano, é a genômica comparativa. Já se sabia que o tamanho do genoma – a quantidade de pares de bases – não guarda relação com o *status* evolutivo de um organismo. Um protozoário, a *Amoeba dubia*, tem 670 bilhões de pares de bases em seu genoma – 220 vezes maior que o humano. E não é só esse ser que bate

ENTREVISTA

CRAIG VENTER

Forças equivalentes

O ambiente pode ser tão determinante quanto os genes

Pode-se não gostar do irrequieto Craig Venter, o fundador da Celera Genomics, a empresa norte-americana que publicou no mês passado sua versão do genoma humano na revista *Nature* enquanto o consórcio público fazia o mesmo nas páginas da *Science*. Mas é impossível ignorar sua influência nos destinos da pesquisa genômica nos últimos dez anos. Em 1992, inconformado com a burocracia dos NIH (*National Institutes of Health*) dos Estados Unidos, ele decidiu fundar um instituto de pesquisa genômica. Nascia em Rockville, no Estado de Maryland, a meia hora da capital, Washington, o TIGR (*The Institute for Genomic Research*), que

se tornou um produtivo centro de estudos e seqüenciamento.

Em 1998, Venter começou a crescer definitivamente seu nome na história da genômica. Nessa data, fundou a Celera, também em Rockville, com um propósito claro: estabelecer a ordem correta dos 3 bilhões de pares de bases que formam o genoma humano num prazo máximo de três anos, bem menos tempo do que o inicialmente proposto pelo consórcio público internacional, que planejava atingir esse nível apenas em 2005. A entrada da Celera – uma companhia disposta a impor um ritmo frenético nas pesquisas e, assim, ganhar dinheiro com suas descober-

tas – causou muita polêmica e acirrou o debate ético. Mas, sem dúvida, fez o consórcio público rever o seu cronograma de trabalho. De 2005, o consórcio passou a prometer o seqüenciamento completo para 2003.

Com a chegada de 2001, pode-se dizer que a Celera cumpriu a sua promessa, ainda que não de forma integral. A empresa publicou uma versão do genoma humano, que ainda contém buracos e não é a definitiva – a exemplo do que ocorreu com o trabalho do consórcio público. Duas semanas após divulgar parcialmente os dados de seu rascunho do genoma – só tem acesso integral às informações da Celera quem virar assinante do banco de dados da empresa –, Venter deu a seguinte entrevista a Marcos Pivetta, por telefone:

■ *A revelação de que o genoma humano tem menos genes do que o esperado é uma notícia boa ou ruim?*

— Bem, não é necessariamente nem uma coisa nem outra. Mas diria que é uma boa notícia porque agora sabe-

o *Homo sapiens*. Até o cachorro, nosso melhor amigo, deixa o homem para trás: o *Canis familiares* deve ter uns 100 milhões de pares a mais que seus donos.

Nossa quantidade de genes, depois das novas projeções rebaixada para 30 mil, também deixou de ser, por si só, motivo de orgulho para a espécie. A popular mosca-da-fruta (*Drosophila melanogaster*) tem mais de 13 mil genes; o verme *Caenorhabditis elegans*, 19 mil; e a *Arabidopsis*, 25 mil. Sem falar na cana-de-açúcar, mapeada parcialmente em São Paulo e já com cerca de 80 mil genes. Para quem acredita que o homem é um ser único, as más notícias brotam de todos os lados. Segundo o consórcio público, os cromossomos humano e do camundongo apresentam muitas semelhanças: há no mínimo 200 segmentos com pelo menos dois genes comuns e na mesma ordem. “As comparações com o camundongo vão ajudar a identificar novos genes humanos”, diz Sandro José de Souza, coordenador de bioinformática do Genoma Humano do Câncer.

Na genômica comparativa, o DNA do ser humano é colocado lado a lado com o de outros organis-

mos qual é a resposta (para o número de genes). É uma ótima base para os pesquisadores seguirem adiante.

▪ *Mas isso quer dizer que vai ser mais fácil ou difícil entender o genoma?*

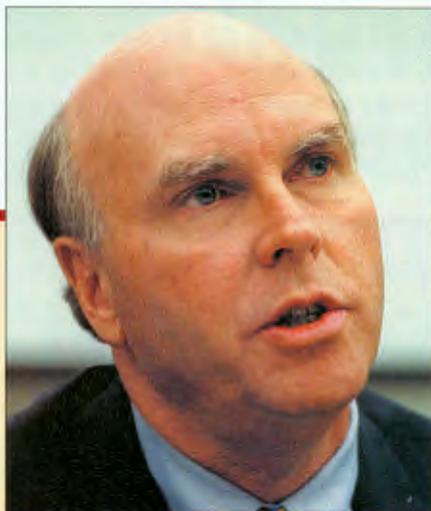
— Ironicamente, o baixo número de genes mostra que a biologia é mais complicada do que muitas pessoas gostariam que fosse. Muita gente pensava na base de “um gene, uma proteína, uma função biológica, uma doença”. Mas as coisas não funcionam dessa forma. Acho que o seqüenciamento do genoma ajuda a demonstrar algo que muitos de nós já sabíamos, mas não tínhamos a noção da extensão dessas complicações. A biologia ainda é a mesma de antes de seqüenciarmos o genoma. Só que agora temos uma melhor avaliação de sua complexidade.

▪ *Menos genes no ser humano quer dizer menos genes para serem patenteados. Como a Celera vê essa questão?*

— Isso é irrelevante. Patentear genes não é uma parte-chave do que a Cele-

mos. As conclusões podem ser surpreendentes. “Geneticamente, nós nos parecemos mais com as plantas do que com os fungos, diferentemente da filogenia, que põe as plantas e os fungos juntos”, comenta Carlos Frederico Martins Menck, do Instituto de Ciências Biomédicas (ICB) da Universidade de São Paulo (USP). Após analisar 120 genes de reparo de DNA – os guardiões do genoma, que consertam os danos que ocorrem nessa molécula –, ele e sua equipe encontraram notáveis semelhanças e, ao mesmo tempo, nítidas diferenças entre os genomas do homem, animais, leveduras (fungos), bactérias e plantas (cana-de-açúcar e *Arabidopsis*).

Um mesmo gene pode ser encontrado em diversas espécies, mas as plantas, por exemplo, podem ter genes só encontrados no homem ou típicos de bactéria – ou não ter genes indispensáveis a outros organismos. “O pior é quando a gente não encontra nada nessas comparações de genoma. Nesses casos, temos de procurar novamente até ter a certeza de que não existe nada em comum mesmo”, comenta Valéria Rodrigues de Oliveira, da equipe de Menck. No final do ano passado, ela encontrou pela primeira vez um gene de reparo de



APP PHOTO/LESLIE KOSOFF

“As pessoas que só enxergam os genes ou apenas o ambiente saem perdendo”

ra está fazendo. Nunca foi. Nosso objetivo é ajudar o resto da comunidade científica, farmacológica e farmacêutica a entender como usar o genoma humano para desenvolver melhores diagnósticos e terapias. Esse é nosso único objetivo, desenvolver novos tratamentos para as doenças.

▪ *O genoma ainda pode ser considerado o “livro da vida”? O senhor concorda com os pesquisadores que dizem que as informações do genoma são um duro golpe no determinismo genético?*
— Sou uma das pessoas que você está

citando. Espero que eu conde com isso (*rindo*). Há muitos termos diferentes usados para qualificar o genoma: livro da vida, fotografia da humanidade. O código genético não é o retrato de um ser humano, não é um dicionário da vida. Ele tem importantes partes de nossa história, importantes instruções para nossas células, sobre como modificá-las. Mas você não pode ir em um cromossomo e encontrar lá as instruções para fazer o coração, o cérebro. Essa discussão tem a ver com a questão anterior, sobre a complexidade do ser humano. A informação está nos níveis seguintes, nas interações entre as proteínas, entre as estruturas das células. Tudo isso não está diretamente codificado no nosso DNA.

▪ *Daqui para a frente, a influência do ambiente vai ser reavaliada no aparecimento de doenças?*

— Nossos estudos mostram que, de uma forma geral, os genes e o ambiente têm provavelmente a mesma importância. Em cada doença, em cada

bactéria em humanos – o mesmo que, depois se viu, funciona em cloroplastos, compartimentos das células vegetais em que se realiza a fotossíntese. “A troca de genes entre organismos é muito mais intensa do que pensávamos”, afirma Menck. “Genomas são misturas de genomas.”

A questão das raças

As duas seqüências quase completas do genoma humano trouxeram novas evidências de que, pelo menos do ponto de vista genético, não há diferenças significativas que justifiquem a noção de raça para qualificar seres humanos. O genoma de uma pessoa é igual em 99,99% de sua composição quando confrontado com o DNA de qualquer outro indivíduo na face da Terra – branco, preto, amarelo ou de origem indígena. A Celera estima que as diferenças entre o genoma de duas pessoas se resumam a 1.250 pares de bases com “letras” distintas. O artigo da empresa afirma que as diferenças genéticas entre pessoas que pertencem a uma mesma etnia podem até ser



EDUARDO CESAR

Menck: “Genomas são misturas de genomas”

maiores do que entre dois indivíduos de raças distintas.

A notícia, obviamente, é boa e contribui – espera-se – para diminuir o preconceito racial. Mas não se pode interpretá-la de forma errada. Sim, somos todos extremamente parecidos no interior de nosso DNA. Mas isso não quer dizer que cada grupo étnico não tenha predisposições genéticas específicas, o que pode aumentar ou diminuir a ocorrência de certas moléstias nessas populações. Causada por uma modificação na hemoglobina, resultante de alterações em um gene, a anemia falciforme tem, por exemplo, maior ocorrência na população negra. Já o outro tipo de anemia, a talassemia,

também provocada por alterações genéticas beta, apresenta maior incidência em indivíduos de origem mediterrânea. “Em alguns casos, pessoas de diferentes etnias podem ter respostas distintas para uma mesma droga. Isso tem de ser levado em conta na hora de se desenvolver medicamentos”, diz Mayana Zatz, coordenadora do Centro de Estudos do Genoma Humano e pesquisadora do Instituto de Biociências da USP.

ENTREVISTA

condição humana, há um equilíbrio diferente da influência desses dois fatores. A biologia molecular provou que o ambiente é realmente uma parte essencial da vida, da biologia. Eles não são separáveis. As pessoas que só enxergam os genes ou apenas o ambiente saem perdendo. Tem de ser os dois juntos por definição.

■ *O senhor acha que essa idéia é aceita pela maioria dos pesquisadores?*

— Isso é algo que eu não tenho como assegurar. Eu esperaria que sim. Mas suspeito que ainda seja necessário muito tempo para as pessoas absorverem toda essa informação. Eu tive a oportunidade de pensar sobre isso muito mais do que a maioria das pessoas.

■ *O que vem depois do genoma?*

— Vem todo o resto da biologia. Não há um passo único que vai resolver tudo. Temos de integrar toda a informação para entender a biologia.

■ *Muita gente tem falado do proteoma*

como o passo seguinte da genômica?

— Temos grandes pesquisas em desenvolvimento na separação e seqüenciamento de proteínas. Estamos construindo a maior unidade do mundo de seqüenciamento de proteínas. Temos feito pesquisas na área de proteômica há um bom tempo.

■ *Como o senhor compara os dados da Celera e do consórcio público?*

— Acharmos que o consórcio público fez um grande trabalho. Mas o banco de dados da Celera é substancialmente mais acurado. Nossa seqüência está numa ordem mais correta, algo que o consórcio público ainda vai demorar dois ou três anos para fazer. Uma coisa é ter as partes da seqüência, outra é colocar tudo na ordem certa e obter a seqüência correta dos genes. É por isso que o banco de dados da Celera é tão popular entre os cientistas. Eles sabem, de longe, que temos o melhor dado.

■ *Quantos assinantes tem o banco de dados da Celera?*

— Até ontem (1º de março), tínhamos 37 assinantes, incluindo grandes escolas de medicina e instituições acadêmicas, como o Rockefeller Institute. Cientistas de todo o mundo, do Instituto Karolinska (da Suécia), do Instituto Max Planck, confiam na nossa informação.

■ *Esse número de assinantes já é suficiente para o senhor?*

— A Celera tem apenas dois anos e meio. Nesse tempo, seqüenciamos a *Drosophila melanogaster* (mosca-da-fruta), o genoma humano e agora o do camundongo. Tudo isso em dois anos e meio. Já temos um número substancial de assinantes. Estamos apenas no começo. Estamos esperando o Brasil assinar nosso serviço.

■ *Pesquisadores brasileiros, a maioria financiada pela FAPESP, produziram uma grande quantidade de dados sobre o genoma humano, um milhão de ESTs de regiões codificadoras de genes. Como o senhor avalia a qualidade dessas informações?*

Matéria-prima para as ciências

Embora não sejam a versão final do genoma humano, os dois rascunhos do nosso DNA contêm dados em quantidade e qualidade suficientes para impulsionar pesquisas nas mais variadas áreas por anos e anos. “Agora vamos poder fazer a medicina do século 21”, diz Sérgio Danilo Pena, professor da Imunologia da Universidade Federal de Minas Gerais (UFMG) e diretor do Centro de Análise e Tipagem de Genomas do Hospital do Câncer A.C. Camargo, em São Paulo.

É uma matéria-prima abundante também para engenheiros, especialistas em computação, matemáticos e filósofos. “Nos próximos meses, vamos começar a sentir o impacto do genoma, talvez de forma diluída”, prevê Marco Antonio Zago, da Faculdade de Medicina da USP de



Briones: “Uma ótima hipótese sobre nosso DNA”



Anamaria: “Ainda temos muitos genes para encontrar”

FOTOS: MIGUEL BOYMAN

Ribeirão Preto. “Grupos do mundo inteiro vão divulgar trabalhos usando as informações das sequências.”

Assim que saíram os artigos da *Science* e *Nature*, o físico Murilo da Silva Baptista, do Instituto de Física da USP, ligou para os biólogos que conhecia. Não os deixou em paz até conseguir a sequência completa do cromossomo X, um dos menores. Seu plano: estudar padrões de organização da molécula de DNA e ci-

elos de repetição de códons, conjunto de três bases que forma aminoácidos, componentes das proteínas.

A comprovação da existência de uma regra que controla a ocorrência de códons pode ajudar a prever quando um deles deve aparecer novamente. “O genoma humano é um terreno extremamente fértil e só precisamos pôr as sementes”, diz o físico, que, no ano passado, encontrou esses padrões de repetição ao analisar os geno-

— Não tenho conhecimento específico sobre as ESTs humanas produzidas no Brasil. Mas sei que outras sequências de ESTs, as primeiras que foram conseguidas em seus país e que foram produzidas em colaboração com o meu instituto (TIGR, *The Institute for Genomic Research*), eram de alta qualidade. Os cientistas brasileiros fizeram um grande trabalho no seqüenciamento do primeiro patógeno de plantas (a *Xylella fastidiosa*).

▪ *No Brasil, os pesquisadores têm falado mais no estudo do transcriptoma do que no do proteoma nos próximos anos, diferentemente do que ocorre no exterior. Por que isso está acontecendo e qual a importância do transcriptoma?* — É um termo que ninguém reconhece, até onde eu saiba, como um termo válido do ponto de vista científico. É por isso que não se ouve muito falar dele. Acho que nunca usei essa palavra nesse sentido. As pessoas estão estudando o RNA todo o tempo. É um intermediário que pode ajudar com algumas indicações sobre

“O baixo número de genes mostra que a biologia é mais complicada do que muitas pessoas gostariam que fosse”

as variações nas proteínas, mas ele não pode dizer com grande precisão o que acontece com as proteínas.

▪ *O senhor considera então muito mais importante o estudo do proteoma do que do transcriptoma?*

— Minha posição é adotar uma visão holística da biologia. Tento não fazer julgamentos... Você não teria nenhuma proteína no corpo sem o RNA e nenhum RNA sem o genoma. Tudo é importante e precisa ser estudado e entendido. A maior parte da biologia ocorre no nível das proteínas. As pessoas medem RNA para geralmente tentar adivinhar quanta proteína existe. Agora, temos novas técnicas

para medir as proteínas diretamente. Mas ainda há interesse acadêmico em medir o RNA em alguns lugares. As duas técnicas são muito importantes.

▪ *O senhor acha que os softwares usados para seqüenciar o genoma humano também podem ser usados no estudo das proteínas?*

— Estamos desenvolvendo novos softwares para lidar com proteínas, mas entendê-las está totalmente baseado em entender o genoma. Quando se lida com um sistema integrado, todos os componentes são necessários. Mas definitivamente vamos precisar de novos e melhores softwares e computadores.

▪ *Tudo o que foi publicado na Nature e na Science sobre o genoma humano fez o senhor mudar alguma de suas crenças mais íntimas, como em Deus, no destino ou na evolução humana?*

— Não. Nenhuma das minhas crenças mudou. Ainda sou a mesma pessoa. Apenas me dei ainda mais conta das complexidades da biologia.

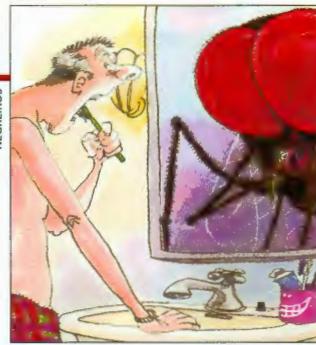
mas da drosófila e da bactéria *Mycoplasma genitalium*. São sistemas com regras próprias, mas que, sabe-se lá por quê, apresentam características matemáticas em comum com a oscilação dos índices econômicos das bolsas de valores e com o comportamento de partículas atômicas carregadas eletricamente, o chamado plasma.

A era pós-genômica: transcriptoma e proteoma

Com o trabalho de determinação da ordem dos 3 bilhões de pares de bases já bem encaminhado, a nova onda na área genômica – ou pós-genômica – é o estudo do proteoma, o conjunto de proteínas de um organismo. Como se sabe, os genes produzem proteínas, as moléculas que formam as células e os tecidos, cujo excesso ou falta pode causar doenças. Muitas moléculas conhecidas são proteínas: hemoglobina, insulina, hormônios e neurotransmissores como a dopamina e a serotonina – sem mencionar as enzimas, indispensáveis para as reações químicas. As proteínas representam cerca de 90% do peso seco do sangue, 80% dos músculos e 70% da pele.

No homem, o estudo da interação das proteínas promete ser uma tarefa ainda mais complicada do que o deciframento do genoma. Um dos motivos: ninguém tem uma idéia muito clara do tamanho do nosso proteoma. Ao contrário do DNA, idêntico em qualquer parte do corpo, as proteínas produzidas num tipo de célula não são as mesmas encontradas em outras. Como o homem tem cerca de 100 trilhões de células, desta vez será difícil chegar logo a um número de consenso. Desde já, as estimativas vão de 100 mil a um milhão de proteínas. Tanto o consórcio público quanto a Celera já começaram a pesquisar nesse campo. No Brasil, há poucos grupos especializados. Fora de São Paulo, um dos únicos é o Centro Brasileiro de Serviços e Pesquisas em Proteínas da Universidade de Brasília (UnB).

Para alguns cientistas, a corrida rumo ao proteoma é inevitável, mas está ocorrendo de forma precipitada. “Deveríamos fazer diferente”, afirma Simpson. Antes de se debruçar sobre as proteínas, dizem, convém entender o transcriptoma, o conjunto de genes que por serem expressos geram moléculas de RNA, necessárias para a síntese de proteínas. Quando conseguem demonstrar que uma região do DNA é transcrita, os pesquisadores comprovam que ali há um ou mais genes. A Iniciativa para Validação do Transcriptoma Humano, um recém-iniciado projeto conjunto da FAPESP e do Instituto Ludwig, pretende encontrar 4 mil genes transcritos nos próximos dois anos. Trinta e um laboratórios participam do projeto, orçado em US\$ 1 milhão. “Para nós, o fato de a Celera e o consórcio público não terem terminado todo o trabalho foi uma ótima notícia”, afirma Anamaria Aranha Camargo, do Ludwig, uma das coordenadoras do Transcriptoma Humano. “Ainda temos muitos genes para encontrar e validar.”

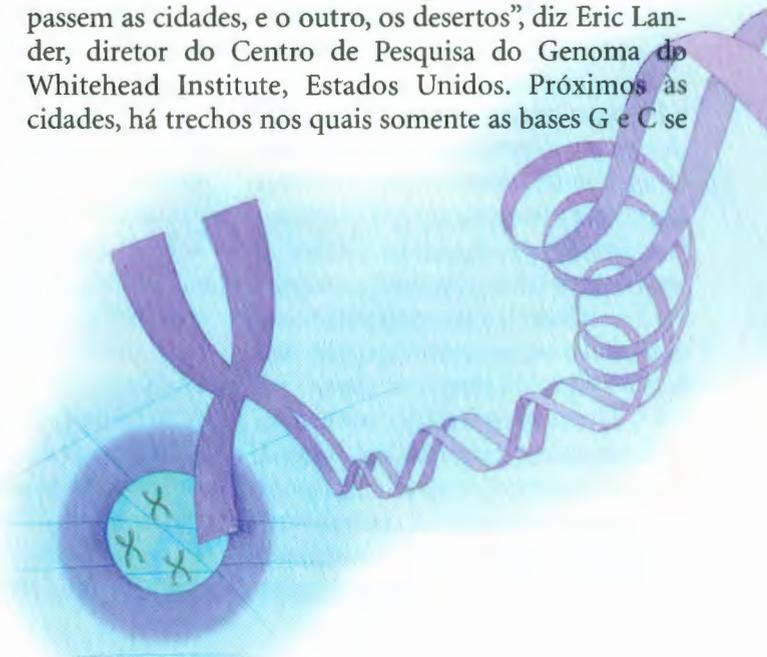


Entre cidades e desertos

Nem todos aderiram aos gráficos, tabelas e longos relatos científicos para entender o genoma humano. Bob Waterson, diretor do Centro de Genoma da Universidade de Washington em St. Louis, Estados Unidos, não hesitou em lançar mão de metáforas com o propósito de deixar clara a irregularidade com que os genes se distribuem ao longo dos cromossomos humanos (*ver tabela*). “Em algumas regiões, os genes estão bastante amontoados, como os prédios nas cidades”, disse ele. “Há também grandes desertos, onde o DNA-lixo pode ser encontrado, e cada região contém informações únicas sobre a história de nossa espécie”. Esse cenário contrasta fortemente com o genoma de outras espécies, como a *Arabidopsis*, o *C. elegans* ou a *Drosophila* – bem mais uniformes, esparramando-se em subúrbios, com uma distribuição relativamente regular de genes nos cromossomos.

Os centros urbanos, densos em genes, são constituídos predominantemente por blocos de duas bases nitrogenadas, guanina e citosina, G e C. Já os desertos ou DNA-lixo são ricos em adeninas e timinas, A e T. Em cada cromossomo, há longos trechos de GC, um com uma densidade de 60% e outro somente com 30%, por exemplo. Nunca ocorrem de modo padronizado e constituem o que Waterson chama de vizinhanças, com soques distintos.

“É como se as regiões de genes e o DNA-lixo tivessem feito um acordo, de modo que os primeiros ocupassem as cidades, e o outro, os desertos”, diz Eric Lander, diretor do Centro de Pesquisa do Genoma do Whitehead Institute, Estados Unidos. Próximos às cidades, há trechos nos quais somente as bases G e C se



repetem 30 mil vezes ou mais. São as ilhas CpG, pouco representadas ao longo do genoma, que ajudam a regular as funções dos genes.

Outra peculiaridade: cada gene humano pode originar, na média, três proteínas, mais que as dos vermes e das moscas. É uma decorrência do chamado *splicing* alternativo, no qual as partes de uma proteína podem ser rearranjadas de modo diferente – ABC, CBA ou BAC, no caso hipotético de apenas três elementos –, como peças de um brinquedo de montar. Esse processo é possível porque os genes estão espalhados ao longo do DNA e as regiões que codificam proteínas não são necessariamente contínuas.

A espécie humana também expandiu as famílias de proteínas. Calcula-se que cerca de 60% das famílias de proteínas humanas contenham mais elementos que em qualquer outra espécie. E a maioria dos grupos de proteínas está associada a funções fisiológicas mais desenvolvidas nos vertebrados. O artigo da *Science* lista 247 genes reguladores do desenvolvimento ou associados ao sistema nervoso, à interação de proteínas, a moléculas de sinalização ou a respostas do sistema imunológico, no homem, na *Drosophila*, no *C. elegans*, na levedura e na *Arabidopsis* – com soberba vantagem para nós.

O refinamento é evidente: só o organismo humano, entre os estudados, produz os genes da interleucina, um tipo de anticorpo. Também temos cerca de três vezes mais que as drosófilas e os vermes um grupo de proteínas que regula a resposta a infecções, as imunoglobulinas, ausentes em fungos e plantas. E temos pelo menos dez genes que pertencem a quatro famílias de proteínas envolvidas na produção de mielina, o revestimento dos nervos; as drosófilas têm apenas um desses genes, e o *C. elegans*, nenhum.

A análise do genoma detectou a presença de remanescentes de uma migração que ocorreu nos primeiros ancestrais vertebrados. Como tinham poucas defesas contra parasitas invasores, as bactérias podiam tornar-se resistentes no interior dos organismos. O resultado dessa coexistência é que o genoma humano abriga cerca de 200 genes que parecem provir de bactérias ou de genomas intermediários de vírus, embora não se descarte inteiramente a hipótese que as bactérias possam também ter furtado genes de ancestrais vertebrados. De acordo com o artigo da *Nature*, cerca de metade do genoma deriva dos chamados elementos transponíveis ou transposons – genes que saltam de um ponto a outro do cromossomo ou mesmo de um cromossomo para outro e regulam a função de outros genes.

Outro ponto interessante é o que os cientistas estão chamando de “mania de colecionar quinquilharias”, em contraste com outras espécies. A quantidade de lixo acumulado em nosso genoma excede o de espécies mais

antigas, com exceção da ameba. Há repetições em meta-de de nosso genoma, muito mais que na *Arabidopsis* (11%), no *C. elegans* (7%) e na drosófila (3%). “Esse fato sugere que fomos muito lerdos ao fazer a limpeza de casa”, compara Arian Smit, bioinformata do Instituto de Biologia de Sistemas. Calcula-se que a drosófila tenha limpado a casa há cerca de 12 milhões de anos, enquanto os mamíferos há 800 milhões. •

Lógica peculiar

Nem sempre os maiores cromossomos contêm mais genes, nem os menores, menos

CROMOSSOMO	TAMANHO* (EM MILHÕES DE PARES DE BASES)	NÚMERO DE GENES**	DENSIDADE (GENES POR MILHÃO DE PARES DE BASES)
1	220	2.453	11
2	240	1816	7
3	200	1611	8
4	186	1145	6
5	182	1366	7
6	172	1467	8
7	146	1219	8
8	146	940	6
9	113	1018	8
10	130	1027	7
11	132	1586	12
12	134	1342	10
13	99	582	5
14	87	873	10
15	80	804	10
16	75	995	12
17	78	1210	15
18	79	472	6
19	58	1409	23
20	61	697	11
21	33	286	8
22	36	641	17
X	128	992	7
Y	19	104	5
Total	2907***	26.383	9

*Estimativa.

** Considerando a estimativa mais baixa de genes humanos, na qual se incluem 328 genes que a Celera ainda não conseguiu associar a nenhum dos cromossomos humanos.

*** Incluindo os 75 milhões de pares de bases que a Celera ainda não conseguiu associar a nenhum dos cromossomos humanos.

Fonte: *Science*