ISSN 1980-3540

Genética NA ESCOLA

06.01, 53-56 (2011)
www.sbg.org.br

GENES, GENOMAS, RNAS NÃO-CODIFICADORES E A COMPLEXIDADE BIOLÓGICA

Francis de Morais Franco Nunes

Pesquisador associado ao Laboratório de Biologia do Desenvolvimento de Abelhas, Universidade de São Paulo, *campus* Ribeirão Preto
E-mail: francis@rge.fmrp.usp.br

O Projeto Genoma Humano iniciou-se em 1990, quando se estimava que o nosso material genético possuísse cerca de 100 mil genes. Em 2001 foram publicados dois artigos que apresentaram o primeiro esboço referente ao sequenciamento de quase todo o genoma humano, e tornou-se possível uma previsão mais realista do número de genes, calculada entre 30 a 40 mil (Lander et al., 2001; Venter et al., 2001). Dez anos se passaram e muitas pesquisas sobre o tema foram realizadas nesse período. Revisamos informações da literatura científica e de bancos de dados, e encontramos que o número mais aceito atualmente corresponde a 22.333 genes humanos (Pertea e Salzberg, 2010) espalhados num universo de 3,2 bilhões de pares de bases (Lander et al., 2001; Venter et al., 2001).

Nesta revisão, estaremos usando o termo gene* para nos referirmos apenas às sequências de DNA que são transcritas em moléculas de RNAs mensageiros (mRNAs) e, por consequência, traduzidas em proteínas.

Curiosamente, o verme *Caenorhabditis elegans* possui um genoma cerca de 30 vezes menor que o humano (aproximadamente 100 milhões de pares de bases) e possui 19.735 genes (Chen et al. 2005). Por que organismos com graus de complexidade tão distintos apresentam similaridade na quantidade de genes e diferença no tamanho do genoma?

Essa é uma questão de ordem evolutiva, e nos remete a dois índices: o valor-C e o valor-G. O valor-C refere-se à quantidade de DNA em um genoma nuclear haplóide (Swift, 1950). Geralmente a quantidade de DNA é calculada em picogramas (pg) ou em milhões de pares de bases (Mb). Assim, os termos "quantidade de DNA" e "tamanho do genoma" são equivalentes e ambos refletem o valor-C de uma espécie qualquer. Para conversão, é amplamente aceito que 1 pg ≈ 978 Mb (Dolezel et al., 2003). Já o valor-G diz respeito ao número de genes codificadores de proteínas presentes num genoma nuclear haplóide (Hahn e Wray, 2002).

Ambos os índices são intrigantes, pois não são coerentes com o conhecimento que temos hoje sobre a história evolutiva dos grupos de seres vivos. As contradições provenientes do debate sobre esses índices culminaram no estabelecimento de dois enigmas biológicos: o "paradoxo do valor-C" (Thomas, 1971) e o "paradoxo do valor-G" (Hahn e Wray, 2002).

***Nota:** *Uma das razões da discrepância nas estimativas do número de genes humanos deve-se ao uso de diferentes conceitos sobre "genes". A problemática conceitual interfere nas predições gênicas apenas do ponto de vista técnico. Os métodos computacionais utilizados atualmente para predição de genes não são completamente precisos, e potenciais erros são constantemente corrigidos por meio de validações experimentais e reanotações manuais dos genes, um a um. Ainda que seja importante definir o número exato de genes de uma dada espécie, os esforços científicos têm centrado muito na evolução dos genomas e dos organismos, e na regulação e funções gênicas. Informações evolutivas, reguladoras e funcionais servem como parâmetros para a compreensão dos diferentes processos de desenvolvimento orgânico e, consequentemente, graus de complexidade biológica. Essas informações vão desde a organização genômica e molecular à diversidade celular (número, formas, funções).*

Tais paradoxos estão ilustrados na Figura 1, onde se observa que os genomas de mamíferos (como o homem, camundongo, gato, cavalo e cão) são, em geral, maiores que os de outros vertebrados (como galinha, rã e o peixe), invertebrados (como o ouriço-do-mar, mosquito, mosca e verme), eucariotos unicelulares (levedura), plantas (arroz) e procariotos (bactéria). No entanto, é equivocado concluir que organismos mais derivados

possuem genomas maiores do que os organismos mais primitivos. Basta notar que o invertebrado transmissor da dengue, *Aedes aegypti*, tem o genoma ligeiramente maior do que o da galinha (*Gallus gallus*), porém um número menor de genes.

Similarmente, o potencial codificador desses genomas, indicado pelo número de genes, também não reflete o grau de complexidade evolutiva. Na Figura 1, observamos que organismos completamente distintos como ouriço-do-mar, peixe e arroz são os que se destacam com relação à abundância de genes, superando todos os mamíferos ilustrados.

Biologicamente, o ser humano é a espécie mais complexa vivente nos dias atuais, sobretudo pela sua extraordinária capacidade cognitiva e de comunicação. Esses atributos são, sem dúvida, resultantes da expressão de elementos genéticos (sendo os genes apenas um entre vários elementos) que controlam nosso desenvolvimento, nossa fisiologia, nosso comportamento e criatividade. Cabe ressaltar que, na natureza, encontramos organismos menos complexos que o homem, e que possuem genomas extraordinariamente grandes. São os casos da ameba *Amoeba dubia*, do peixe *Protopterus aethiopicus*, da cebola *Allium cepa* e do sapo *Bufo bufo*, os quais possuem genomas com aproximadamente 209, 43, seis e duas vezes o tamanho em humanos, respectivamente.

Alguns estudos conseguiram detectar conexões entre o tamanho de genomas e certas características de alguns grupos. Sabe-se que, para muitos organismos, o tamanho do genoma correlaciona-se positivamente com o tamanho das células e negativamente com a taxa de divisão celular (revisado por Gregory, 2002). Para exemplificar, podemos comparar os procariotos, que possuem células com comprimento médio entre 1-5 micrômetros e genomas compactos, com os eucariotos, que possuem células com comprimento médio entre 5-100 micrômetros e genomas maiores. Por outro lado, os genomas menores dos procariotos replicam-se mais rapidamente e a divisão celular é, portanto, mais veloz que nos eucariotos. Esses estudos com enfoque na relação do valor-C *versus* caracteres morfo-fisiológicos são pontuais e não se constituem em leis biológicas.

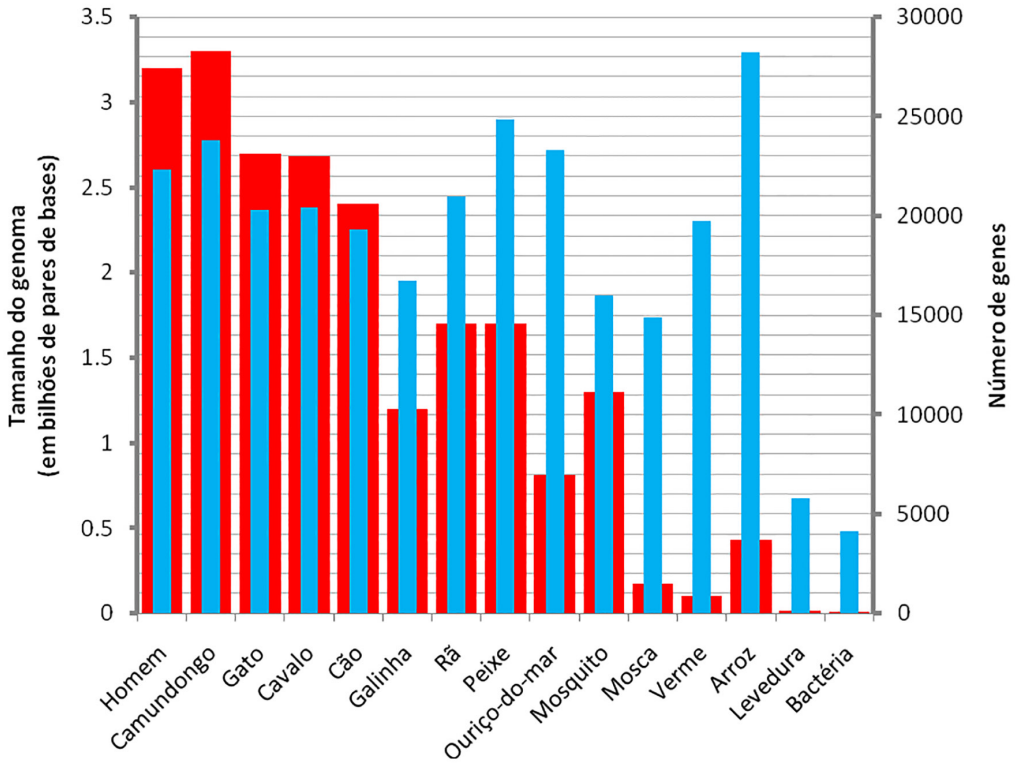

FIGURA 1: *Distribuição comparativa do tamanho dos genomas (barras vermelhas e valores no eixo vertical esquerdo) versus o número de genes (barras azuis e valores no eixo vertical direito) em diferentes organismos. Dados compilados de informações do NCBI-GenBank (http://www.ncbi.nlm.nih.gov/genbank) e literatura científica (buscas em http://www.ncbi.nlm.nih.gov/pubmed) referente aos genomas sequenciados de cada espécie: homem (*Homo sapiens*), camundongo (*Mus musculus*), gato (*Felis catus*), cavalo (*Equus caballus*), cão (*Canis lupus familiaris*), galinha (*Gallus gallus*), rã (*Xenopus tropicalis*), peixe (*Danio rerio*), ouriço-do-mar (*Strongylocentrotus purpuratus*), mosquito (*Aedes aegypti*), mosca (*Drosophila melanogaster*), verme (*Caenorhabditis elegans*), arroz (*Oryza sativa*), levedura (*Saccharomyces cerevisiae*) e bactéria (*Escherichia coli*).*

Do ponto de vista genético, existe algum parâmetro capaz de explicar o aumento da complexidade biológica ao longo da história evolutiva? A resposta é **sim** e discutiremos a seguir.

Até poucos anos atrás, as proteínas foram o centro das atenções científicas. A relevância das proteínas se deve aos diferentes papéis celulares que elas desempenham como atividades enzimáticas, estruturais e reguladoras. Os eventos de processamento alternativo (*splicing alternativo*), por exemplo, podem gerar dois ou mais tipos de proteínas a partir de um único mRNA. É correto dizer que o *splicing alternativo* aumenta o potencial codificador nos organismos e, assim, gera complexidade. Mas, sozinho, o *splicing* alternativo não é capaz de explicar toda a diversidade de vida no planeta.

A resposta à questão acima está aqui: a grande quantidade de informações de sequências genômicas disponíveis nos bancos de dados (especialmente no GenBank) nos revela que houve uma expansão e acúmulo de DNA não-codificador de proteínas durante o curso da evolução das espécies. Também, descobriu-se que existe uma forte correlação entre a complexidade biológica e o aumento na quantidade de sequências não-codificadoras (Taft e Mattick, 2003; Taft et al., 2007). Ao se dividir o número de sequências não-codificadoras (**nc**) de um organismo pelo seu valor-C (**tg**, tamanho do genoma) e, em seguida, comparar as razões **nc/tg** entre diferentes espécies, observa-se uma escala de valores com coerência evolutiva. As razões **nc/tg** descritas por Taft e Mattick (2003) foram capazes de discriminar diferentes grupos de seres vivos, do mais basal ao mais derivado (Tabela1).

TABELA 1: *Razões entre o número de sequências de DNA não-codificador (**nc**) pelo tamanho do genoma (**tg**). As razões foram calculadas para diferentes grupos de seres vivos por Taft e Mattick (2003).*

Grupo		**Razões**
Procariotos		< 0,25
Eucariotos unicelulares		entre 0,26 e 0,52
Eucariotos multicelulares	Plantas	entre 0,71 e 0,80
	Invertebrados	entre 0,74 e 0,93
	Vertebrados	entre 0,89 e 0,98

Hoje sabemos que a informação genética vai muito além da transcrição de RNAs mensageiros (mRNAs), transportadores (tRNAs) e ribossômicos (rRNAs). Esses três tipos de RNAs são clássicos e funcionam em sintonia para a ocorrência de síntese proteica. Nos últimos anos tecnologias para análise de expressão gênica e sequenciamento em larga-escala foram aprimoradas (ex: *tiling arrays*, *next-generation RNA-Seq*), as quais permitiram uma nova compreensão sobre o potencial de transcrição das informações genéticas em diversos genomas.

As regiões DNA não-codificadoras foram consideradas por muito tempo como "lixo genético", dada a ausência de informação para síntese de proteínas. Em leveduras, estimou-se que 74,5% do genoma é transcrito (excluindo-se sequências repetitivas), sendo que a maioria destes não gera proteínas (Nagalakshmi et al., 2008). Assim, os resultados experimentais têm apontado que regiões de DNA não-codificador em diferentes espécies são alvos de transcrição, ou seja, produzem moléculas de RNAs não-codificadores.

As recentes evidências científicas apontam que, juntamente com as proteínas, os RNAs não-codificadores são capazes de coordenar o metabolismo celular e definir os fenótipos dos organismos. A partir de 2005 tornou-se claro que os genomas de animais e plantas produzem uma abundância de RNAs não-codificadores (revisado por Mattick, 2009). Em humanos e outros mamíferos, por exemplo, sabe-se que RNAs não-codificadores são transcritos em ambas as fitas do DNA, inclusive sobrepondo regiões promotoras e de mRNAs (Carninci et al., 2005; Kapranov et al., 2007).

Os genomas são mais dinâmicos e complexos do que se podia imaginar, e a "soberania proteica" está em queda. Baseado no exposto, propomos um novo índice, que chamaremos de valor-R (de RNA), relativo a todo o potencial de transcrição de diversos tipos de RNA (codificadores e não-codificadores de proteínas) de um genoma nuclear haplóide. Já podemos também nos antecipar e propor outro índice, o valor-N (N de *networks* = redes), cujas inferências basear-se-ão no grau de interações mo-

leculares e dinamismo metabólico em diversos organismos. A expectativa é que tais índices sejam robustos o suficiente para hierarquizar a complexidade biológica sem que surjam novos paradoxos.

REFERÊNCIAS BIBLIOGRÁFICAS

CARNINCI P et al. 2005. The transcriptional landscape of the mammalian genome. *Science* 309(5740):1559-1563.

CHEN N et al. 2005. WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Research* 33(Database issue):D383-389.

DOLEZEL J et al. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry* A 51(2):127-128.

GREGORY TR. 2002. Genome size and developmental complexity. *Genetica* 115(1):131-146.

HAHN MW, WRAY GA. 2002. The G-value paradox. *Evolution and Development* 4 (2):73-75.

KAPRANOV P et al. 2007. Genome-wide transcription and the implications for genomic organization. *Nature Review Genetics* 8(6):413-423.

LANDER ES et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921.

MATTICK JS. 2009. The genetic signatures of noncoding RNAs. *PLoS Genetics* 5(4):e1000459.

NAGALAKSHMI U et al. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320(5881):1344-1349.

PERTEA M, SALZBERG SL. 2010. Between a chicken and a grape: estimating the number of human genes. *Genome Biology* 11(5):206.

SWIFT H. 1950. The constancy of desoxyribose nucleic acid in plant nuclei. *Proceedings of the National Academy of Sciences of the United States of America* 36:643-654.

TAFT RJ, MATTICK JS. 2003. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biology*, Preprint Depository http://genomebiology.com/2003/5/1/ P1.

TAFT RJ, PHEASANT M, MATTICK JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29(3):288-299.

THOMAS CA JR. 1971. The genetic organization of chromosomes. *Annual Review of Genetics* 5:237–256.

VENTER JC et al. 2001. The sequence of the human genome. *Science* 291(5507):1304-51.