



Universidade Federal da Bahia
Universidade Estadual de Feira de Santana



Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências

GENES, INFORMAÇÃO E SEMIOSE
DO CONHECIMENTO DE REFERÊNCIA AO ENSINO DE BIOLOGIA

Vanessa Carvalho dos Santos

Salvador, Bahia
Agosto
2008

Universidade Federal da Bahia
Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências

GENES, INFORMAÇÃO E SEMIOSE
DO CONHECIMENTO DE REFERÊNCIA AO ENSINO DE BIOLOGIA

Vanessa Carvalho dos Santos

Dissertação apresentada ao Programa
de Pós-Graduação em Ensino,
Filosofia e História das Ciências, da
Universidade Federal da Bahia como
requisito parcial para obtenção do
título de Mestre

Orientador: Prof. Dr. Álvaro João Magalhães de Queiroz
Co-orientador: Prof. Dr. Charbel Niño El-Hani

Salvador, Bahia
Agosto
2008

Universidade Federal da Bahia
Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências

**GENES, INFORMAÇÃO E SEMIOSE
DO CONHECIMENTO DE REFERÊNCIA AO ENSINO DE BIOLOGIA**

Vanessa Carvalho dos Santos

Prof. Dr. José Luis de Paula Barros Silva
Coordenador do Curso

Salvador, Bahia
Agosto
2008

Universidade Federal da Bahia
Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências

BANCA EXAMINADORA:

Prof. Dr. Charbel Niño El-Hani (co-orientador)
Universidade Federal da Bahia - UFBA

Prof. Dr. Waldomiro José da Silva Filho (membro interno)
Universidade Federal da Bahia - UFBA

Prof.^a Dr.^a Ana Maria de Andrade Caldeira (membro externo)
Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP

Salvador, Bahia
Agosto
2008

Agradecimentos

Aos meus orientadores, Charbel El-Hani e João Queiroz, pela orientação e socorro nos momentos mais críticos e pela inacreditável paciência ao aceitar todos os prazos não cumpridos. Tenho profunda admiração por vocês e trago uma enorme satisfação por fazer parte desse grupo de pesquisa.

Aos professores Ileana Greca, Diogo Meyer e Ângela Freire, pelas críticas e excelentes contribuições no exame de Qualificação.

Aos professores Waldomiro Silva Filho e Ana Maria de Andrade Caldeira, por terem aceitado gentilmente a participação na banca de defesa. Agradeço também aos professores Diogo Meyer e José Luis Barros Silva por aceitarem participar na banca como suplentes.

Aos colegas do Grupo de Pesquisa em História, Filosofia e Ensino de Ciências Biológicas (GPHFECB), sobretudo Leyla Mariane, pelo auxílio na análise e nas discussões dos ‘relatórios do mal’.

Aos colegas do mestrado, pela companhia durante os últimos anos.

À FAPESB, pela concessão da bolsa, que permitiu a realização deste trabalho.

Aos secretários do Programa, Lene e Orlando, pelo auxílio em questões burocráticas.

A meus pais, por todo o apoio e carinho que me dedicaram.

Aos meus amigos Fabiano Timm, Mônica Calil, Francismare, Catarina, Eagles, Milena, Cíntia, Jéssica e Thiago. Amo vocês.

Ao Tiago, o peixe.

Ao Mario Malpica, por ter chegado na hora exata. Serás sempre ‘minha doce lembrança’.

À Rodamagia: ‘O seu olhar melhora o meu!’

Ao André (Wall-E). ‘O amor espelhado infinitas vezes dança a maior poesia: leve, uva, doce lábios. Ei! Eu te amo!’

Para André

e

para a Rodabiogigante

*... And in this crazy life, and through these crazy times
It's you, it's you, You make me sing
You're every line, you're every word, you're everything*

Resumo

Introdução – Informação como processo semiótico.....	1
Abordagem quantitativa da informação biológica.....	5
Signos, sim; mas sem linguagens e metáforas.....	10
Plano de vôo.....	14
Capítulo Um – O conceito de gene na era pós-genômica.....	18
1. Introdução: gene e informação genética na berlinda.....	18
1.1. Breve histórico do conceito de gene.....	19
1.2. Desafios ao gene molecular clássico.....	25
2. A importância do processamento de RNA: emenda alternativa.....	30
3. Gene, informação e a semiotização espontânea da biologia.....	51
4. Considerações finais.....	55
5. Referências.....	57
Capítulo Dois – Emenda alternativa de RNA, semiose e realização gênica.....	63
1. Introdução.....	63
2. Função e realização gênica.....	69
2.1. Função do gene, análise funcional da célula.....	70
2.2. Soluções múltiplas? Bricolagens, remendos e emenda alternativa	74
2.3. Estudo de caso: emenda alternativa do gene <i>Dscam</i> de <i>Drosophila melanogaster</i>	79
3. Análise semiótica da emenda alternativa de RNA.....	90
3.1. Algumas idéias básicas na teoria dos signos de Peirce.....	93
3.2. Níveis de Semiose.....	100
3.3. Análise semiótica de processos de realização gênica: em cena, a emenda de RNA.....	101
3.4. Posição da emenda alternativa na escala do estocástico ao altamente regulado.....	110

4. Considerações Finais.....	115
Referências.....	121

Capítulo Três – Explicações sobre genes em livros didáticos de biologia do ensino médio

publicados no Brasil.....

1. Introdução.....	127
2. Breve retrospectiva: Bases históricas dos modelos de função gênica.....	130
3. Retórica do ‘gene-para’ um traço: ferramenta conceitual ou fonte de determinismo genético.....	134
4. Metodologia.....	136
4.1. Fase explanatória.....	137
4.2. Categorias para análise.....	137
4.3. Validação interna.....	141
5. Resultados e discussão.....	142
5.1. Frequência de ocorrência dos conceitos de gene.....	142
5.2. Modelos de função gênica.....	149
5.2.1. Frequência de ocorrência de funções atribuídas ao gene.....	150
5.2.2. Modelos de função gênica por categoria de contexto.....	152
5.3. O papel do ambiente: visões deterministas e interacionistas sobre a origem das características.....	159
5.4. Desafios ao conceito molecular clássico.....	163
6. Considerações finais.....	169
Referências.....	171

Anexo

Considerações Finais.....	175
---------------------------	-----

Referências.....	183
------------------	-----

Lista de Figuras:

Capítulo 1

- Figura 1. Modelos estruturais para um gene codificante de proteínas.....29
- Figura 2. Padrões de emenda alternativa..... 32

Capítulo 2

- Figura 1. Definição de éxon e íntron na reunião do spliceossomo..... 82
- Figura 2. Organização do gene *Dscam* de *D. melanogaster* *Dscam*..... 83
- Figura 3. Modelo para o mecanismo de emenda mutuamente
excludente do éxon 6 de *Dscam*.....85
- Figura 4. Estabelecimento de padrões de emenda tecido-específicos
a partir de poucos fatores regulatórios..... 88
- Figura 5. Signo como meio de transmissão de uma forma
(uma regularidade ou invariância) do objeto para o interpretante.....96
- Figura 6. A relação triádica S-O-I forma uma cadeia de tríades..... 98
- Figura 7. Visão geral dos processos semióticos envolvidos na realização
de signos potenciais no DNA..... 104
- Figura 8. Integração do processo de emenda de RNA na realização
de signos potenciais no DNA.....109

Capítulo 3

- Figura 1. Frequência de ocorrência de conceitos de gene nas
obras de biologia de ensino médio analisadas..... 145
- Figura 2. Frequência de ocorrência de idéias sobre função gênica
nas obras de ensino médio analisadas..... 151
- Figura 3. Normas de Reação e diferentes modelos da relação genótipo-ambiente..... 162

TABELAS

Capítulo 3

Tabela 1. Características epistemológicas e suas variantes contidas nas unidades de registro.....	140
Tabela 2. Variantes de características epistemológicas dos modelos históricos de função gênica.....	141
Tabela 3. Percentual de sobreposições do gene molecular clássico e da concepção informacional a outros conceitos de gene.....	148
Tabela 4. Modelo de função gênica encontrado no contexto ‘glossário’.....	155
Tabela 5. Modelo de função gênica encontrado no contexto ‘caracterização da vida’.....	155
Tabela 6. Modelo de função gênica encontrado no contexto ‘biologia celular e molecular’.....	156
Tabela 7. Modelo de função gênica encontrado no contexto de ‘genética’.....	156
Tabela 8. Modelo de função gênica encontrado no contexto ‘evolução’.....	157
Tabela 9. Percentual de ocorrência de conceitos de gene por categoria de contexto.....	157
Tabela 10. Obras analisadas que abordam genes interrompidos, splicing alternativo e as implicações do splicing alternativo para o conceito molecular clássico de gene.....	166
Tabela 11. Idéias e termos utilizados para descrever éxons, íntrons e splicing.....	168

Resumo:

Na era pós-genômica, co-existe uma diversidade de conceitos e modelos de gene. Frente à demanda de se assimilar a complexidade e a diversidade da arquitetura genética, as quais dificultam as individualizações moleculares de gene, tem se argumentado a favor do abandono da busca por um conceito unificado de gene para adotarmos um pluralismo científico e uma visão instrumentalista dos mesmos. Em suma, podemos inferir que a necessidade de acomodar os novos resultados empíricos à idéia do gene como unidade força os programas de pesquisa a enfatizarem diferentes características em situações variadas ou para distintos propósitos. Opiniões como estas foram emitidas por muitos cientistas e filósofos da biologia, para os quais é desnecessária a adoção de um conceito de gene único e abrangente, que inclua toda a diversidade de significados e funções epistêmicas conectada a esse termo e que, contanto que os contextos de aplicação sejam bem definidos, esta diversidade será frutífera em termos explicativos e heurísticos. O *status quo* na genética e na biologia molecular torna patente a subordinação dos padrões de expressão genética ao complexo e dinâmico aparato regulatório celular. Nesta conjuntura, a depender dos propósitos de um domínio de investigação, é necessário prescindir não apenas da concepção de gene como unidade de estrutura, função e informação, mas ainda das propriedades de permanência entre gerações e de estabilidade tradicionalmente atribuídas a ele. Decerto que estas duas últimas propriedades são salvaguardadas quando conferidas ao DNA. Todavia, há muito se sabe que esta molécula não é um agente autônomo que encerra em si mesmo a decisão de quando e onde se expressar e duplicar, muito menos é a sua estabilidade um atributo inabalável. Assim, não há nada de contra-intuitivo em atentar que ser estável e permanente não é algo peculiar do DNA em si mesmo, que estas são características inapropriadamente imputáveis a ele, se não como *propriedades relacionais*. Se esta ressalva for estendível ao geral das propriedades biológicas significantes atribuídas convencionalmente aos ácidos nucleicos (DNA ou RNA), torna-se premente a necessidade de se ressaltar a relevância de suas interações dinâmicas com a miríade de processadores enzimáticos que convergem sobre eles, subordinados ao contexto sistêmico em que estão imersos. A semiótica do lógico pragmatista Charles Sanders Peirce constitui uma ferramenta analítica heurísticamente poderosa para acomodar os desafios atuais aos conceitos de gene e de informação na biologia. No modelo semiótico que empregamos, as maquinarias de transcrição (e.g., RNA polimerase), de tradução (e.g., o ribossomo), da emenda alternativa (e.g., o spliceossomo) e, na sinalização celular, um receptor de membrana que reconhece um dado sinal extracelular são considerados como sistemas interpretativos subordinados ao intérprete global, a célula. Cremos que, dentro desta perspectiva, podemos refutar de maneira contundente as visões genocêntricas que atribuem aos genes um papel privilegiado na hereditariedade e no desenvolvimento com base na suposição de que os mesmos, e apenas eles, carregariam informação – mais precisamente, em sua seqüência de bases. Nossos resultados sugerem que há, nos livros didáticos de ensino médio, um equivalente à chamada ‘semiotização espontânea’ da biologia, que se encontra, porém, intrinsecamente associada à controversa concepção seqüencial da informação biológica. Será interessante averiguar, em trabalhos futuros, se a transposição do modelo alternativo semiótico que aqui aplicamos para compreensão dos sistemas genéticos de informação pode ser conduzida em termos de uma ‘semiotização estimulada’ no ensino médio, dentro das possibilidades de transposição didática da própria semiótica.

Palavras-chave: Gene, Informação genética, Emenda alternativa, Signos, Semiótica, Peirce, Processos, Livros didáticos, Ensino Médio.

Introdução

INFORMAÇÃO COMO PROCESSO SEMIÓTICO

A questão central desta dissertação é: *Que caminhos se apresentam como promissores e se insinuam como ferramentas fundamentais, ao lado das explicações mecanicistas, para penetrarmos os enigmas da complexidade biológica?*

Dito de outra maneira, nós colocaremos em xeque a crença de que sistemas complexos como os organismos podem ser compreendidos por meio de explicações reducionistas, baseadas na mera investigação de seus constituintes mais fundamentais – a saber, as biomoléculas e suas reações físico-químicas. Em prol disso, argumentaremos que uma miríade de atributos dos sistemas vivos só adquire significância quando se faz referência a níveis de organização superiores ao molecular. Para fins do argumento, adotamos a perspectiva da biossemiótica, um campo que busca analisar sistemas biológicos como sistemas semióticos, e operacionalizamos a idéia de emergência de significados em *processos de interpretação de signos* pela célula. O sentido preciso em que processos biológicos específicos podem ser concebidos como envolvendo a ação de signos será fundamentado, aqui, na semiótica do filósofo pragmatista norte-americano Charles Sanders Peirce (1839-1914) e conterà elementos essenciais de uma recente abordagem peirceana da informação, desenvolvida por Queiroz, Emmeche & El-Hani (2005) e El-Hani, Queiroz e Emmeche (2006).¹

É neste íterim que realçamos o discurso informacional na biologia como um instrumento imprescindível à captura de características básicas dos organismos que são negligenciadas num arcabouço teórico alicerçado meramente em reações bioquímicas. Todavia, rechaçamos a

¹ Ver também El-Hani, Arnellos e Queiroz (2007) e Queiroz e El-Hani (2007).

maneira usual como informação vem sendo entendida nas ciências biológicas – i.e., como informação seqüencial em DNA, RNA ou proteínas –, acomodando esta concepção mais apropriadamente numa abordagem orientada a processos.

O que pretendemos, acima de tudo, é reiterar sugestões advindas tanto do estado da arte da biologia molecular na era genômica, quanto da reflexão de cientistas e filósofos, no esforço de refinar o arcabouço teórico da biologia para acomodar as crescentes descobertas sobre a complexidade e a natureza dinâmica de sistemas vivos. Nesse contexto, vale indagar: ‘E se informação não for nada que a célula contém de antemão, mas algo que ela desempenha, pratica, exerce como um sistema dinâmico complexo interagindo com o ambiente?’; ‘E por que não considerar que apenas dessa maneira ácidos nucleicos e proteínas podem engendrar, porém não encerrar em si mesmos, algum significado?’; ‘E se este significado for tão inexoravelmente conectado a um processo situado (dependente de contexto) e dialógico (dependente de intérprete), de maneira a demandar um modelo semiótico pragmático de informação?’. E, por fim, a questão mais crucial: ‘E se este processo apresentar a mesma forma geral independentemente de envolver estímulos internos – evoluídos e herdados genética- ou epigeneticamente –, ou estímulos do ambiente externo, de modo que nenhum estatuto informacional privilegiado possa ser imputado a genes?’.

A nosso ver, questões como estas podem ser explicadas se ‘informação’ for concebida como processo semiótico. Neste caso, ela *emergiria* da complexa ‘dança’ molecular, como uma ‘performance’ de ações de signos orquestradas pela célula, entendida como um sistema semiótico autocorretivo capaz de interpretar seu ambiente. A informação genética, em particular, não poderia ser nada inerente a seqüências de nucleotídeos ou aminoácidos que, ao fluir a partir do DNA, conferiria ao genoma o poder de programar o desenvolvimento e as funções celulares; tampouco seria a informação genética dedutível do comportamento das moléculas, em isolado ou

em reações bioquímicas em si mesmas, sem considerar o contexto celular em que estão situadas. Torrentes de objeções às diferentes formas de reducionismos na genética e na biologia molecular poderão ser fortalecidas com uma abordagem processual da informação biológica.

Este é, porventura, o entendimento pelo qual atualmente clamam os pesquisadores dos sistemas vivos, ávidos por compreender o significado funcional da grande quantidade de dados produzidos no seqüenciamento dos genomas. A cada seqüência finalizada, um misto de júbilo e frustração: pouco do que se encontra se assemelha a uma mensagem distinta, clara e precisa do que deve ser feito para se construir uma dada molécula e, menos ainda, para se construir um organismo; o máximo que se poderia encontrar é algum tipo de sintaxe – ainda assim, a propriedade sintática do genoma é questionável (cf. Kay 2000). A suposta propriedade semântica dos genes, que seria intrínseca à informação contida em sua seqüência de bases, revela-se infundada.

Nesta conjuntura, constata-se que muito ainda resta a se compreender sobre os sistemas biológicos e é provável que sua inteligibilidade resida em algo ainda mais difícil de se ‘capturar’ do que seqüências moleculares que lembrem um sistema de linguagem digital, discreto, estável e rico em sintaxe; é provável que a significância biológica resida em algo muito mais rico em detalhes e sujeito a contingências, mas que parece ser a principal fonte de semântica num organismo: a rede contínua e analógica de processos, interações e dinâmicas que compõem o contexto celular e supracelular, que se desenrola numa história de desenvolvimento, de reconstrução de organismos a cada geração. Na biologia pós-genômica, se recorrermos à charada sobre o ovo e a galinha, convém afirmar que é necessária uma ênfase maior na fase da ‘galinha’, o código para ação, a fase analógica dos organismos, que interfere com o entorno físico e compartilha com este as propriedades de extensão e continuidade. É cada vez mais patente a indispensabilidade desta fase para a significância da fase digital do ovo, o código de memória,

protegido da esfera da ação e que alegadamente geraria mensagens com sinais discretos de modo linear, porém, segmentado e à semelhança de elementos lingüísticos – nomeadamente, o DNA (sobre a dualidade de código dos organismos, cf. Hoffmeyer e Emmeche [1991]2005). A relevância da dualidade de código reside no reconhecimento da valia de uma memória localizada e digital, porém, sem superestimá-la: as vantagens da posse de um sistema de sintaxe, do caráter discreto e de ‘entidade’, da estabilidade e permanência, e dos potenciais pertinentes ao DNA têm como *pedra de toque* uma memória distribuída e analógica, de natureza processual, contínua, transitória e mais suscetível a contingências, que é peculiar à célula como um todo.

Este é o ensejo para a introdução do arcabouço conceitual semiótico: significados surgem na ação dos signos, ou semiose, que pode ser concebida como um *processo sistêmico emergente*, i.e., uma propriedade encontrada somente no nível do sistema como um todo, e não no nível de suas partes (Queiroz e El-Hani 2006). Se a célula for concebida como sistema semiótico, a construção de um conceito semântico de informação deverá levar em conta que o significado biológico é relacionado às propriedades de todo o contexto celular, bem como irreduzível a meras reações bioquímicas. Uma abordagem semiótica peirceana, contudo, não toma a distinção ‘analógico-digital’ como tão crucial e, na verdade, argumenta-se que ela também não o é na semiótica em geral (ver Nöth 1995, p. 208): pode-se considerar que qualquer ato de semiose envolve uma interconversão, a interface destas duas fases. Se extrapolarmos o sentido desta asserção, poderemos constatar que não é oportuno focalizar apenas um dos dois lados da moeda; talvez seja preciso incorporar uma perspectiva mais dinâmica e complexa do que a linearidade que correlaciona, por exemplo, um genótipo digital a um fenótipo analógico e que fomenta metáforas instrutivas de programas, projeto, livro da vida etc. (ver Nijhout 1990). É oportuna a consideração daquilo que não só converte digital em analógico e vice-versa, mas que também *interpreta* uma mesma coisa ora como digital, ora como analógica.

Urge a busca pelo terceiro lado da moeda, a noção do organismo como intérprete de signos internos e externos. Para este intuito, acreditamos em avanços singularmente promissores no ponto em que as perspectivas processuais e biossemióticas convergem e esta constitui nossa melhor conjectura para a pergunta que iniciou esta seção: um *pensamento orientado a processos* e uma *abordagem semiótica* dos mesmos apontam um caminho frutífero para o entendimento da complexidade biológica e para a construção de uma teoria semântica e pragmática da informação na biologia.

Abordagem quantitativa da informação biológica

Numa noção tradicional, na qual a palavra ‘informação’ significava o processo ou ação de informar ou ser informado (Kay 2000; Hoffmeyer e Emmeche [1991]2005), havia uma dimensão mental e subjetiva implícita: alguém é informado, instruído, treinado, educado.² Trata-se de um sentido em que o termo ‘informação’ incorpora os três níveis semióticos, provavelmente mais familiares na nomenclatura de Charles Morris (1964): o sintático, que estuda as relações entre signos; o semântico, que se ocupa das relações entre signos e referentes; e o pragmático, que estuda as relações entre signos e seus intérpretes. Nesta acepção, informação é um conceito qualitativo – visto que está necessariamente associado ao significado –, assim como também o é num uso cotidiano do termo, em que:

‘Informação é uma propriedade semântica de uma mensagem porque somente mensagens significativas podem ser informativas. Informatividade depende de fatores pragmáticos da comunicação, a saber, um emissor, que informa, e um receptor, que nota certo grau de novidade na mensagem’ (Nöth 1995, p.134).

² Neste sentido, informação significa inteligência e informar, portanto, é uma ação cognitiva, é dar forma a uma mente, e não a uma entidade material. E mesmo neste último caso, a palavra forma (no sentido aristotélico) não é algo derivado da matéria, mas algo induzido sobre ela por uma mente.

Num movimento gradual de reificação, ao invés de denotar a ação de informar – um verbo –, informação se tornou sinônimo daquilo que é comunicado, da notícia, da novidade: pessoas recebem informação, algo cada vez mais próximo da idéia de entidade, de nomes, e independente de sujeito, emissor ou receptor. Este movimento pode culminar numa extrema ‘atomização do conhecimento’ (Hoffmeyer e Emmeche [1991]2005, p. 39), em fragmentos que podem chegar a unidades de ‘bits’, como ocorre na teoria da informação. O advento de modelos derivados desta teoria, como também se chamou a teoria matemática da comunicação, de Shannon e Weaver (1949), fomentou a introdução do pensamento informacional na biologia nas décadas de 1950 e 1960. Todavia, neste caso, informação não é precisamente uma entidade, mas uma quantidade, desprovida de conteúdo semântico e de pragmática: quantidade de informação é a medida da improbabilidade de um sinal ocorrer num dado ponto de uma mensagem, a medida da raridade estatística de sinais. A propriedade de frequência ou raridade concerne à relação entre a ocorrência textual de um sinal e as regras sintáticas subjacentes aos arranjos entre os símbolos de um código.

A propagação do vocabulário informacional na biologia prosseguiu por um caminho cada vez mais divergente do sentido técnico matemático da teoria da informação, dada a dificuldade em se determinar a medida probabilística da informação alegadamente contida no DNA. E ainda que esta medida pudesse ser determinada (em *bits*), ela seria apenas uma medida da sintaxe. Assim como uma coleção aleatória de letras pode ter a mesma *quantidade de informação* que a contida num soneto (Kay 2000), seqüências de DNA com mesmo comprimento poderiam conter a mesma quantidade de informação, fosse uma seqüência não-codificante ou, conforme argumenta Jablonka (2002), fosse codificante de uma enzima funcional ou codificante de uma enzima não funcional. Nenhuma semântica seria obtida até que um devido contexto (celular, orgânico, ambiental) fosse especificado. Por se restringir ao aspecto sintático e ser indiferente

ao semântico, a teoria da informação não poderia legitimar a visão do DNA como o livro da vida, como a fonte do significado biológico. A quantificação da informação com base apenas em aspectos sintáticos deixa em aberto a maneira como a informação foi criada e o que deve se entender por sua significância.

Entretanto, à revelia de todos os paradoxos e aporias, convencionou-se tratar informação como a determinação precisa da seqüência de bases de ácidos nucléicos ou de resíduos de aminoácidos numa proteína, i.e., como 'informação seqüencial'. O discurso da informação foi empregado no meio científico, popularizado e utilizado como ferramenta de retórica para persuasão, coloquialmente e de maneira não-sistemática, com base numa suposta propriedade semântica dos genes, mas sem clarificar se uma semântica genuína estaria envolvida.

Deste ponto em diante, tornou-se manifesto o emprego concomitante de (i) uma retórica reducionista e mecanicista estimulada pela *tour de force* despendido em investigações de estruturas moleculares, como as que resultaram no modelo da dupla-hélice do DNA e, anos mais tarde, na decifração do código genético e (ii) um discurso da informação que, paradoxalmente, emergia a partir dos esforços biomoleculares, a despeito de denotar atributos imateriais como cognição, potencialidade e fim (*telos*) nas explicações desta ciência.

No início, não houve incompatibilidade entre os dois tipos de discurso. O vocabulário informacional veio referendar a fé no poder preditivo das seqüências de DNA, com base na idéia de um suposto programa genético, de uma presumida correspondência direta entre genes, estruturas e funções. Este tipo de visão remonta à característica da especificidade, i.e., à complementaridade de estruturas biológicas altamente ordenadas, um dos aspectos estruturais mais sedutores para a biologia molecular. O pareamento de bases, em especial, ganhou destaque por sugerir uma interpretação e caracterização simbólica, e não apenas físico-química, da relação entre as fitas de DNA e, possivelmente, entre DNA e outras moléculas da célula (ver Sarkar

1998). Esta peculiaridade foi um dos fatores a contribuir para a difusão do vocabulário informacional na biologia molecular, o qual impregnou o imaginário tecnocientífico e as práticas discursivas, em particular, com representações textuais e lingüísticas do genoma. Seu poder retórico e persuasivo perdura no âmbito popular e midiático atual e estimulou, há poucos anos, as iniciativas de seqüenciamento dos projetos genoma.

De acordo com Lily Kay (2000, p.41), o termo ‘informação’ foi inicialmente empregado como metáfora para a especificidade molecular, ao lado de representações escriturais – como ‘mensagem’, ‘alfabeto’, ‘instrução’, ‘código’, ‘texto’, ‘programa’ etc. Claramente, o nascente e mesclado discurso da biologia molecular refletia um dualismo Cartesiano: o léxico oriundo da esfera cultural, com seus atributos mentais e intencionais, foi empregado devido às suas qualidades essencialmente metafóricas, que os tornavam dispositivos convenientes para propor hipóteses e para expor, de forma mais simples, um conhecimento físico-químico mais preciso. Não havia, no início, a disposição para uma interpretação literal desse vocabulário num campo que pertencia à esfera da ‘natureza’. Esta postura, no entanto, não durou muito. Em pouco tempo, a noção de informação foi concebida em termos ontológicos e se seguiu uma grande empreitada para compreender as regras da suposta ‘linguagem natural’ do DNA. Hoffmeyer e Emmeche ([1991]2005) ressaltam que o conceito de informação parecia dissolver o antigo dilema entre substância e forma, já que o DNA (uma substância) e o programa genético (informação/ou forma potencial) eram tratados como a mesma entidade. Entretanto, mesmo nesta fase, ainda se encontravam vestígios de dualismo: havia uma bipartição estrita dos organismos numa parte lingüística e atemporal – o DNA – e noutra temporal e operante dinamicamente – o sistema de proteínas. Em outras palavras, havia uma separação entre o que era entendido como componente semiótico e um componente não-semiótico dos sistemas biológicos.

Entendido dentro dessa perspectiva, o conceito de informação seqüencial parece encerrar pequeno poder preditivo e heurístico, sobretudo por ter herdado, a despeito de todas as divergências, algumas deficiências apontadas na crítica à própria teoria da informação, mais precisamente, à sua visão mecanicista e idealizada sobre o processamento de textos. O modelo quantitativo de Shannon e Weaver para a comunicação – entendida como a transferência de sinais numa interação bilateral entre dois sistemas – não é heurísticamente interessante no ambiente de modelagem de processos de comunicação, devido ao seu caráter linear, que sugere um modo de causalidade simples: o emissor do sinal aparece como uma causa que tem um efeito calculado no receptor. A cadeia linear de transmissão de sinais pode ser vista como uma série mecanicista de relações diádicas, em que cada evento depende, única e exclusivamente, do evento anterior. Esta perspectiva sugere a interação de um participante ativo com outro passivo, um potencial quase total para a manipulação do receptor (Nöth 1995).

Este aspecto nos remete à visão do DNA como a fonte de onde flui toda informação do organismo, e do genoma como um programa para o desenvolvimento. O único fator que parece ameaçar esse processo é o erro na leitura desse programa, i.e., o erro no processo de transcrição e tradução, o que, na teoria da informação, é entendido como ruído – distorção indesejada na mensagem transmitida. Neste modelo, portanto, a célula é um ente passivo que processa a informação recebida do DNA, mas sem afetá-lo. Todavia, à medida que avança o conhecimento sobre a complexidade estrutural e dinâmica do genoma, torna-se mais difícil argumentar pela existência de um programa genético. Hoje, há uma miríade de desafios que torna plástica, contingente e dependente de contexto a relação do DNA com qualquer efeito que ele possa ter na célula ou no organismo. Diante deste quadro e da falta de precisão na atribuição de sentido ao termo ‘informação’ na biologia, podemos encontrar posições céticas em relação ao discurso informacional, nas quais se argumenta pela sua eliminação (Stuart 1985, Sarkar 1996): este

discurso não passaria de um conjunto de metáforas inadequadas que podem conduzir a visões distorcidas, em que genes são vistos como causas determinantes de fenótipos, a despeito da importância de fatores epigenéticos e da complexidade do desenvolvimento, em organismos multicelulares.

Signos, sim: mas sem linguagens e metáforas

Uma posição otimista quanto ao discurso da informação reconhece que este se apóia numa série de metáforas ainda pouco elucidadas, que se encontram à espera de uma teoria da informação biológica que possa atribuir-lhes sentido mais preciso (Griffiths 2001; Stotz et al. 2004). Aqui, podemos nos valer da crítica à visão mecanicista de fluxo de informação, que é procedente da semiótica do texto. Este é um dos campos que mais têm enfatizado a autonomia do receptor ou leitor no processo de comunicação. Em vez de focar no autor como a chave hermenêutica para o trabalho de arte, a semiótica do texto se volta para o intérprete no processo de interpretação (Eco 1979; Eco e Sebeok 1983). Num de seus modelos de comunicação, é possível notar uma inversão no sentido das setas que tradicionalmente indicam o fluxo de informação: ‘texto → emissor → receptor’ é substituído por ‘emissor → texto ← receptor’ (cf. Koch 1971, citado por Nöth 1995). Neste novo modelo, as setas não representam fluxo de informação, mas o foco dos participantes, sua atenção e seu comportamento seletivo em vista de um potencial *a priori* indeterminado de estruturas semióticas no texto.

Esta visão sobre a natureza do texto pode ser entendida como um *insight* essencial obtido do arcabouço semiótico peirceano, o qual, na descrição de Köller (1977, p.73, citado por Nöth 1995, p.46), assume que: “signos de linguagem não têm uma estrutura estática, mas formam um evento dinâmico”, e que a ‘linguagem não pode ser adequadamente estudada da perspectiva de sistema,

mas apenas da perspectiva de processo”.³ Nöth (ibid.) ressalta que o lema de pesquisa da semiótica do texto no espírito peirceano se resume assim, nos termos do próprio Peirce:

“Símbolos crescem. Eles vêm a ser pelo desenvolvimento a partir de outros signos, particularmente de ícones, ou de signos misturados compartilhando a natureza de ícones e símbolos [...] Um símbolo, uma vez na existência, se difunde [...] No uso e na experiência, seu significado cresce” (CP 2.302).⁴

Com este exemplo da semiótica do texto, não queremos ratificar a visão do DNA como o livro da vida, repleto de instruções para criar um organismo. É bastante controversa a visão do código genético como um código ou sistema de linguagem natural. Este código pode ser visto, antes, como uma simples tabela de correlações, embora nem de longe tão sistemática e preditiva como a tabela periódica, por exemplo, devido às contingências, degenerações e ambigüidades (Kay 2000). Contudo, nosso intento é ressaltar que, a despeito disso, o genoma pode conter signos e estes, porventura, podem vir a ser concebidos como símbolos. O ponto crucial deste argumento é que, na semiótica peirceana, o signo é tratado em suas características formais, não importando qualquer modo particular de sua produção. Disto segue que, para um signo ser signo, é irrelevante se é produzido como um fenômeno mental, social, ou biológico. Qualquer coisa pode se tornar um signo, tão logo satisfaça às condições formais dos signos como tais (Liszka 1996).

Por conseguinte, nesta perspectiva, não há coisas que sejam signos e coisas que não o sejam – e.g., signos não são entidades mentais convencionais – e.g., linguagem – em contraposição aos

³ Em sua divisão mais fundamental dos signos, Peirce caracteriza ícones, índices e símbolos, correspondendo, respectivamente, a relações de similaridade, contigüidade e lei – seja convenção ou regularidade natural – entre o Signo e o Objeto que ele representa na tríade Signo-Objeto-Interpretante (S-O-I). Signos de linguagem são símbolos e, segundo a semiótica do texto, a emergência de seu significado durante o processo de interpretação justifica o motivo pelo qual, dentro do arcabouço peirceano, não se pode classificar sem ambigüidades um dado signo como pertencendo a uma única classe; isso se deve ao fato de que seu comportamento como signo em si mesmo, i.e., sua natureza, pode mudar com sua função, história, perspectiva e o processo de sua interpretação.

⁴ Manteremos a maneira tradicional de citar o *Collected Papers of Charles Sanders Peirce* (Peirce 1931-35, 1958) por CP, seguido do número do volume, ponto decimal e número(s) do parágrafo.

objetos naturais. Não há este dualismo na semiótica de Peirce. Tendo isto em conta, vale adiantar que não será o caso, na análise aqui empreendida, de se empregar conceitos semióticos na análise de processos naturais apenas em termos semióticos— como seria o caso se estivéssemos utilizando o sentido lingüístico da semiologia saussuriana. Sequer almejamos argumentar que sistemas biológicos têm uma ‘linguagem natural’ – o que poderia até justificar o emprego literal de termos semióticos, como ‘informação’. Antes, uma abordagem semiótica peirceana implica não se restringir ao arcabouço conceitual lingüístico na biologia, mas, sim, naturalizar o vocabulário informacional e semiótico, com base no entendimento de que signos não são necessariamente símbolos, nem estes são necessariamente de linguagens. Na perspectiva da biossemiótica, organismos são concebidos como sendo profusos em signos e informação – legítimos como tais, sem se restringirem inevitavelmente ao DNA e se configurarem como textos.

Não há, contudo, uma classe de coisas com características intrínsecas de signos, dado que estes não são tipos naturais – e.g., não podemos manipular signos no ambiente ou isolar um signo celular no laboratório. Qualquer característica convencional, inerente, ou natural de um objeto pode contribuir apenas para discernir um dado tipo de signo, e mesmo assim, somente depois de este objeto adquirir as características formais que qualquer signo deve ter. Este aspecto torna o escopo de aplicação da semiótica peirceana bastante amplo. Fenômenos biológicos constituem um modo particular genuíno de produção de signos dentre tantos possíveis. Para Peirce, desde que satisfaça as condições gerais, qualquer coisa pode ser um signo, e é neste sentido que se conciliam suas idéias de que o universo é permeado de signos e de que nada é essencialmente um signo em si mesmo (Liszka 1996).

O *insight* peirceano identificado na semiótica do texto e que, por analogia, acomodaremos na análise do genoma é que signos não devem ser estudados como elementos de um sistema autônomo de linguagem, independente do contexto no qual são empregados; antes, eles devem

ser vistos da perspectiva dos *processos* em que são interpretados num dado contexto e nos quais emergem seus significados a partir da rede de conexões com outros signos. Adicionalmente, o aspecto dialógico e situado desse processo permite que classifiquemos diferencialmente signos biológicos, a depender dos propósitos de nossa investigação, sobretudo porque sua natureza de símbolo, índice ou ícone – sendo esta a tríade mais básica da classificação peirceana dos signos – está à mercê de sua significância para a célula, o intérprete primordial na vida⁵, em cada contexto.

O discurso da informação e a pleora de termos relacionados compreendem o conjunto de conceitos (proto-) semióticos mais familiares na biologia. Plausivelmente, a consolidação deste discurso, hoje alvo de contenda, está intrinsecamente atrelada à mudança do foco em entidades discretas para um foco em processos, sobretudo na contenda em torno do conceito de gene (Neumann-Held 2001; Keller 2005) e dentro deste último, à integração da abordagem de processos físico-químicos à abordagem de processos semióticos. É bastante provável que, neste último passo, o conceito de informação possa ser dissociado das versões fortes de redução que marcaram, nos termos de Sarkar (1998, p. 141), a ‘molecularização da biologia’, sendo legitimado, em vez de eliminado, no contexto de abordagens mais sistêmicas da hereditariedade, do desenvolvimento e da evolução, que se fazem tendência nas ciências biológicas (Griffiths 2001; Oyama et al. 2001).

⁵ Somos a favor da idéia de que os sistemas vivos foram o primeiro sistema semiótico genuíno.

Plano de vôo

Para fins do argumento de que sistemas biológicos podem ser concebidos como sistemas semióticos e para o entendimento de como eles instanciam processos sígnicos, nosso ponto de partida será a discussão dos problemas que cercam os conceitos de gene e informação genética, na estrutura do conhecimento biológico. Assim, no primeiro capítulo, abordamos como o conhecimento atual da estrutura e dinâmica dos sistemas genéticos de informação está compelindo as explicações na biologia para uma abordagem que enfoca processos celulares e desenvolvimentais, ao invés de apenas entidades (substâncias) dentro das células. Parte das respostas de biólogos teóricos e filósofos da biologia aos desafios ao conceito de gene podem ser tratados como indícios de que uma abordagem biossemiótica, calcada na metafísica de processos peirceana, é deveras oportuna.

No segundo capítulo, utilizamos uma abordagem semiótica peirceana para tratar dos conceitos de gene e informação, dando seqüência a trabalhos anteriores no mesmo domínio (Queiroz et al., 2005; El-Hani et al., 2006; Queiroz & El-Hani 2007). No contexto de tal abordagem, considera-se que estruturas genéticas, e suas ações, não podem ser compreendidas isoladamente, mas apenas no contexto de um sistema mais amplo por meio do qual são interpretadas. Em sintonia com a moldura teórica estabelecida pela teoria pragmática do signo de peirce, e conforme desenvolveremos neste trabalho, processos semióticos são ‘logicamente dependentes de contextos’. Dessa perspectiva, o significado de um gene para seu intérprete, a célula, ou a significância biológica de um gene, não pode ser encontrada nas seqüências de DNA de um cromossomo. Esta tese se contrapõe à versão mais corrente e usual, relativamente ao modo como a informação genética é abordada no conhecimento biológico, seja de referência ou escolar, a saber, à compreensão da ‘mensagem’ de um gene como informação seqüencial no DNA – refletida em informação seqüencial em proteínas (Sarkar 1998). A informação genética não

poderia estar restrita, nos termos de uma análise peirceana, a uma seqüência de nucleotídeos no DNA, mas deveria incluir também o efeito do ‘gene-come-um-signo’ na célula ou, no caso dos organismos multicelulares, no organismo como um todo, e, em última análise, a própria função dos subsistemas celulares como intérpretes de trechos de DNA.

Aplicando o sistema triádico básico de Salthe (1985) e a teoria dos signos de Peirce para analisar genes e informação, podemos dizer, por ora sem maiores detalhes, que um trecho de DNA que codifica uma proteína pode ser considerado um Signo, que representa uma seqüência específica de aminoácidos de um polipeptídeo (Objeto, na tríade semiótica, como explicaremos mais abaixo), determinando a reconstrução (Interpretante) da seqüência polipeptídica. Esta reconstrução consiste, por sua vez, na instanciação de uma regularidade estabelecida ao longo da evolução dos sistemas genéticos de informação.

Adicionalmente, de acordo com este modelo, uma seqüência de nucleotídeos no DNA não possui qualquer significado intrínseco, na ausência de um aparato celular para interpretá-la. Ela somente pode possuir um significado potencial. Nesses termos, as seqüências de nucleotídeos no genoma funcionam como condições iniciais – no que podemos denominar ‘nível micro-semiótico’, como veremos adiante – que apenas delimitam quais cadeias de tríades são possíveis, i.e., podem ser realizadas no nível focal – o nível em que um processo semiótico é observado; mas o que regula as cadeias de tríades que serão efetivamente concretizadas neste nível são as condições de contorno (estabelecidas pelo que denominamos ‘nível macro-semiótico’), constituídas por redes de cadeias de tríades ou semióticas encontradas no sistema celular. Estas redes correspondem a sistemas integrados de vias de sinalização celular. Aplicaremos este modelo para analisar um processo biológico fundamental no sistema genético de informação, nomeadamente, a emenda alternativa de RNA e sua regulação por vias de sinalização celular. Conduziremos esta análise tendo como estudo de caso o gene da proteína de membrana *Dscam*

em *Drosophila melanogaster*, bem como as vias de sinalização que regulam sua expressão. Numa última etapa deste capítulo, fornecemos uma visão global de nossa análise, dialogando com os outros trabalhos nessa linha já desenvolvidos em nosso grupo de pesquisa.

No terceiro e último capítulo, passamos a tratar dos conceitos de gene e informação no conhecimento escolar, na medida em que este é a fonte das concepções que a maioria das pessoas tem sobre estes conceitos tão cercados por controvérsias, que podem ter seus significados elucidados no contexto de uma teoria da informação biológica que faça uso do referencial peirceano. Para tanto, analisamos qualitativa e quantitativamente como livros didáticos de ensino médio abordam idéias sobre genes e informação genética. Com essa investigação, pretendemos mostrar *como* o discurso sobre genes, bem como a chamada semiotização espontânea da biologia – entendida como a implementação de um vocabulário com conotações semióticas na ciência, sem de fato se dar conta de sua natureza semiótica (ver Emmeche 1999) – se propagam nesse estágio da transposição didática do conhecimento biológico e alcançam o nível médio de escolaridade. A abordagem dos genes e da informação genética no ensino médio pode ser uma fonte importante de equívocos das pessoas quanto à compreensão dos sistemas biológicos, com sérias conseqüências para o exercício de sua cidadania, sobretudo em relação à promoção de visões deterministas das relações entre genótipo, desenvolvimento e fenótipo. Neste capítulo, argumentamos, ainda, a favor da necessidade de incluir na educação científica de nível médio uma abordagem mais rica, atualizada e epistemologicamente bem fundamentada do sistema genético e de suas relações com outros sistemas biológicos, bem como dos fenômenos que desafiam idéias típicas ao seu respeito.

Esta dissertação está organizada na forma de artigos. Cada capítulo corresponde, assim, a um artigo que deverá estar preparado, em sua versão final, para submissão a periódicos. Em vista disso, cada capítulo segue as normas das revistas para as quais serão submetidos. O formato em

artigos implicará necessariamente alguma repetição de conteúdo entre os capítulos, bem como a existência de uma lista de referências para cada um destes, além da lista completa de referências colocada ao final da dissertação⁶. A organização em artigos tem a vantagem de tornar mais expedito o processo de dar publicidade aos resultados da investigação, passo essencial na construção do conhecimento científico, mas implica que nos afastaremos do modelo tradicional das dissertações e teses como um documento único, com um argumento desenvolvido continuamente ao longo de todo o texto. Haverá, naturalmente, algum grau de quebra dentro do argumento, em virtude da própria organização em artigos. Consideramos, contudo, que as vantagens deste modo de organizar a dissertação superam, em muito, as limitações dele decorrentes. Outrossim, fizemos um esforço substancial para manter o argumento bem conectado ao longo dos três artigos que compõem este texto, como o plano de vôo apresentado acima permite entrever.

⁶ As referências utilizadas nesta introdução se encontram na lista completa de referências ao final da presente dissertação.

Capítulo Um

O CONCEITO DE GENE NA ERA PÓS-GENÔMICA

1. Introdução: gene e informação genética na berlinda

A compreensão dos genes como portadores de informação inerente a uma seqüência de DNA e, por sua vez, da informação como correspondente a instruções para a construção de fenótipos é tão freqüente em livros didáticos do ensino médio e do ensino superior, nos discursos da imprensa leiga e especializada, e, inclusive, na literatura científica, que parece invocar asserções não-problemáticas. Nessa abordagem, a informação genética tem sido empregada para explicar, através de uma perspectiva molecular, tanto a hereditariedade quanto o desenvolvimento. A hereditariedade é restringida, assim, à transmissão de moléculas de DNA de uma geração a outra, não obstante o crescente reconhecimento da importância de mecanismos epigenéticos (Sarkar 1996). O desenvolvimento, por sua vez, é visto freqüentemente como a leitura de um programa genético, que, por assim dizer, forneceria instruções de comando para a função celular e para a realização de todos os estágios desenvolvimentais. Essas idéias estão intimamente relacionadas ao conceito molecular clássico de gene e ao dogma central da biologia molecular.

O gene, de acordo com o ‘conceito molecular clássico’ (Griffiths & Neumann-Held, 1999; Stotz, Griffiths & Knight, 2004), que foi estabelecido na primeira década de construção da estrutura da biologia molecular, na esteira do modelo da dupla hélice, é uma unidade discreta no DNA, cuja seqüência contínua de nucleotídeos codifica um produto funcional, seja uma cadeia polipeptídica ou uma molécula de RNA (El-Hani, 2007). Mas o conceito de gene sempre esteve em contínua evolução e seu significado tem sido objeto de controvérsias praticamente desde sua criação, por Johannsen, em 1909 (Keller 2002; Portin, 1993; Carlson, 1991; Falk, 1986). Os avanços da biologia molecular e da genômica nos últimos trinta anos, em particular, tornaram

difícil até mesmo explicar e localizar o que é um gene. Alguns filósofos da biologia têm uma visão pessimista quanto ao futuro do gene, sugerindo sua substituição por um novo conceito, que se mostre mais adequado às novas descobertas experimentais (Keller 2002)⁷, ou a invenção de uma nova terminologia para quando maior precisão for necessária (Portin 1993). Mas é importante destacar que, ao alardearem a ‘morte’ do gene, a chegada ao limite de seu poder explicativo, os autores estão claramente se referindo ao gene molecular clássico (Solha, 2005; Solha e Waizbort, 2007). Este conceito está realmente em apuros, mas o mesmo não pode ser dito do conceito de gene em geral: parece haver alternativas no fim do túnel.

1.1. Breve histórico do conceito de gene

A idéia de uma unidade hereditária e a polêmica por ela gerada perpassam a história do conceito de gene, aparecendo antes mesmo da criação desse termo. Já no final do século XIX, precedendo o mendelismo, firmou-se a noção geral de herança particulada, presente nos conceitos de pangene de De Vries e determinantes de Weismann. Essa noção de herança veio a contribuir substancialmente para incutir o conceito de unidade na genética. Em 1900, o resgate dos trabalhos de Mendel inaugurou o que Keller (2002) chamou de ‘o século do gene’. Na primeira década do século XX, o termo ‘*unit-character*’ era amplamente utilizado para se referir tanto ao hipotético ‘fator unitário’ encontrado nas células germinativas que determinaria características fenotípicas, quanto ao seu efeito, o caráter manifesto em um organismo, que se comportaria como uma unidade indivisível de herança mendeliana (Falk 1986). Até então, não havia, pois, uma diferença clara entre o potencial para um traço – o ‘fator unitário’ de Mendel – e o próprio traço.

⁷ Em artigo posterior, Evelyn Fox Keller (2005) não se compromete com a idéia de que o termo ‘gene’ deveria ser abandonado, como o fez em *O Século do Gene*.

Visando desfazer essa ambigüidade, já que se tratava de coisas distintas, o geneticista dinamarquês Wilhelm Johannsen cunhou o termo ‘fenótipo’ para designar o conjunto de características de um organismo e o termo ‘genótipo’, para indicar os correspondentes a essas características que se encontrariam nas células germinativas ou, nas suas próprias palavras, ‘algo que alcançamos por inferência’ (Falk 1986).

Em 1909, Johannsen criou a palavra ‘gene’ para indicar os próprios fatores, as ‘condições especiais’, os ‘fundadores’ e ‘determinantes’ das características dos organismos, que estariam presentes de maneira única, separada e, portanto, independente na linhagem germinativa, ou seja, seriam as unidades de herança particulada (Keller 2002; Portin 1993; Falk 1986). É evidente que o conceito de gene foi formulado a partir da conjunção dos conceitos de genótipo, criado por Johannsen, e de unidade. Os genes seriam as unidades do genótipo. Apesar de Johannsen se referir a algo supostamente presente nos gametas e no zigoto, ele foi cauteloso ao formular o termo ‘gene’ de maneira instrumentalista, buscando não carregá-lo com qualquer hipótese quanto à natureza material de seu referente, mesmo porque era praticamente impossível definir materialmente o que era, então, apenas inferido (Moss 2003).

Conseqüentemente, nos primeiros anos da genética clássica, ou seja, por grande parte da primeira metade do século XX, prevaleceu uma visão instrumentalista sobre a condição do gene como um conceito teórico; ele foi considerado uma abstração, útil como unidade de cálculo nos cruzamentos mendelianos, para expressar regularidades nos traços fenotípicos, mas sem qualquer contrapartida material clara (El-Hani 2005; Falk 1986). Durante muito tempo, a única forma de se identificar um gene era através do traço visível, seu ‘marcador’ ou ‘representante’. Sem qualquer suposição sobre a natureza dos genes, procediam-se os cálculos dos cruzamentos, aplicando-se procedimentos heurísticos ‘como se existissem genes’ (Falk 1986, p. 135). Pode parecer uma idéia circular dizer que o gene era determinado através de seu fenótipo. Mas esta era

apenas uma forma de defini-lo, em termos de seu efeito. Usando a comparação feita por Sterelny e Griffiths (1999), podemos identificar algo como veneno mesmo não sabendo do que se trata, mas apenas pelo seu efeito letal. Da mesma forma, o gene era identificado por um traço, ou mais especificamente, pela aparência alternativa de um traço, ainda que não se soubesse a que ele precisamente se referia.

Não tardou a aparecer, contudo, uma visão realista acerca do gene, cujo maior defensor foi inicialmente Herman J. Müller. De acordo com esta visão, genes não eram meras unidades operacionais, mas deviam ser assumidos, antes, como unidades materiais. Como salienta Falk (1986), o gene foi então reificado, o que representou o primeiro passo para sua identificação material. Até a quarta década do século XX, entretanto, os pesquisadores não haviam chegado a um consenso sobre a natureza dos genes, ainda que as pesquisas buscassem com ímpeto esse objetivo.

Muitos mendelianos especulavam sobre a natureza dos genes, mas o que eles buscavam era algo análogo aos átomos da química (no sentido original de uma entidade não-divisível), ou seja, elementos únicos, separados (discretos), independentes (em suas ações e nos efeitos da seleção sobre eles) e determinantes das características de um organismo. Mayr (1998) ironicamente apelidou essa forma de conceber os genes de ‘visão do saquinho de feijão’ (*bean-bag view*), numa alusão às práticas encontradas em manuais de genética das décadas de 1940 e 1950, em que os genes eram representados por feijões coloridos, reagrupados a cada geração, sem qualquer interação entre eles. Devido à influência da noção de herança particulada e, sobretudo, aos esforços do grupo de pesquisa liderado por Thomas Hunt Morgan, o conceito de unidade foi plenamente assimilado pela genética clássica. O gene foi considerado uma unidade indivisível de recombinação, mutação e função.

Adicionalmente, os genes, como unidades estruturais, seriam os blocos elementares a partir dos quais o estudo da genética deveria começar ou, em outras palavras, as *menores partes* de todo um patrimônio genético, o que já correspondia a uma idéia de continuidade ou indivisibilidade do gene (Portin, 1993). A unidade de função, por sua vez, pode ser considerada a menor entidade que tem significado único, uma única função, e o gene, identificado a partir de um fenótipo, poderia ser considerado como tal unidade. Assim, genes seriam as menores porções do genoma que sofrem recombinação e mutação, e estão relacionadas a uma função. Estes aspectos foram adicionados à idéia de genes como elementos *únicos, discretos* (separados) e *independentes*, influenciada pela noção geral de herança particulada.

De acordo com Falk (1986, p. 169), o conceito de gene evoluiu através de uma ‘confrontação dialética’ entre a abordagem instrumentalista e a realista, seguindo um padrão que ele denomina de ‘bonecas russas’⁸; neste padrão, descobertas sobre a natureza química do gene levavam à elaboração de novas definições funcionais, que, por sua vez, geravam significados estruturais mais profundos, levando a significados funcionais ainda mais profundos, e assim sucessivamente. Como conseqüência, o gene permaneceu como *unidade de função*, mas não mais como unidade de mutação e recombinação, visto que estas duas últimas propriedades, devido ao então mais refinado conhecimento da natureza química do gene, passaram a ser mais apropriadamente atribuíveis a entidades, respectivamente, menores e maiores do que os genes.

Benzer, de acordo com sua interpretação, mostrou que as unidades de função, que ele chamou de ‘cístrons’ (termo bastante usado na literatura primária no lugar de ‘gene’), são tipicamente maiores do que as unidades de recombinação e mutação (que ele chamou de ‘récon’ e ‘múton’, respectivamente)⁹ (Benzer [1957]. *cf.* El-Hani 2005). Kitcher (1982) argumenta que, embora Benzer tenha sugerido a substituição do termo ‘gene’ pelas palavras cístron, récon e

⁸ Falk se refere àquelas bonecas russas (*matryoshka*) que se encontram dentro de outra, que, por sua vez, se encontram dentro de outra, e assim por diante.

múton, ele de fato refinou o conceito de gene. Os geneticistas moleculares precisavam de uma unidade ‘significativa’ no DNA e, portanto, depois de um curto eclipse do uso do termo ‘gene’ (porque não estava claro se o referente era a unidade de função, de recombinação ou de mutação), o ‘cístron’ terminou por ser identificado com o ‘gene’ (Falk 1986).⁹

A partir da década de 1940, estudos com o fungo *Neurospora crassa* e com a bactéria *Escherichia coli* (ver Beadle e Tatum 1941; Horowitz e Leupold 1951) forneceram a base para uma formulação mais precisa da função gênica, levando à chamada hipótese ‘um gene - uma enzima’, segundo a qual cada gene teria como função a produção de uma enzima específica. Pouco tempo depois, em 1944, o DNA foi identificado como molécula estruturalmente importante e funcionalmente ativa na determinação das atividades bioquímicas e das características específicas das células de pneumococos. Com a revelação da composição de bases do DNA, o “átomo genético foi clivado em suas partes constituintes” (Portin 1993, p. 175).

O modelo da dupla hélice, proposto por Watson e Crick em 1953, consolidou a *visão realista* dos genes e estabeleceu o DNA como base material da herança. Mas foi mantida a tensão dialética entre as visões funcional e estrutural do gene, que atravessou a genética mendeliana: o gene poderia ser identificado tanto operacionalmente por seu produto (e pela função deste), quanto como um segmento discreto de DNA em seu próprio direito, independentemente de sua transcrição ou da tradução de seu transcrito (Falk 1986). Adicionalmente, com o florescimento da biologia molecular, estabeleceu-se outra tensão, relacionada ao problema da segmentação (Kitcher 1982). Desde o início da genética clássica, o gene fora identificado com algum tipo de segmento cromossômico, o que levantou o problema de como dividir apropriadamente

⁹ Com a descoberta da composição do DNA, os nucleotídeos se tornaram unidades fundamentais – mas não significativas da genética molecular, visto que apenas uma trinca deles encerra a capacidade de representação de um aminoácido – e, por algum tempo, acreditou-se que os nucleotídeos individuais seriam as unidades de mutação e recombinação, tornando desnecessários os termos múton e récon, que foram eliminados; mas essa atitude foi prematura, visto que nem a unidade de recombinação, nem a de função corresponde a um único nucleotídeo (Falk 1986, p. 158).

cromossomos em genes, denominado por Kitcher (1982) o ‘problema da segmentação’: os cromossomos poderiam ser divididos em genes identificados a partir de sua função na produção de um fenótipo ou a partir de sua ação imediata, a produção de transcritos primários de RNA. De acordo com Kitcher (ibid., p. 349), a primeira abordagem foi característica da genética clássica e a segunda, da genética molecular.

Uma solução aparente para estas tensões foi fornecida pelo conceito molecular clássico de gene, construído na esteira do modelo da dupla hélice. Este conceito combinou numa só unidade as definições estrutural e funcional do gene, mostrando-se poderoso o suficiente para desfazer, por um lado, a tensão entre a definição operacional de genes através de seus produtos e funções e a visão de genes como segmentos cromossômicos e, por outro, a tensão entre identificar genes através de suas funções ou por suas ações imediatas (El-Hani 2005).

Com o advento do conceito molecular clássico, além de serem unidades de função, os genes passaram a ser vistos também como *unidades de estrutura*. Estas corresponderiam à forma mais elementar em que um polipeptídeo ou RNA estaria representado no DNA, ou, mais precisamente, seriam seqüências ininterruptas, colineares a um produto gênico final. Desta forma, a concepção mendeliana de genes como *unidades* foi ‘atualizada’ na compreensão molecular clássica do que é um gene (Fogle 1990). Finalmente, sobretudo a partir da década de 1960, a biologia molecular passou a utilizar termos derivados da cibernética e da teoria da informação, resultando no que tem sido chamado o ‘discurso da informação’ (*information talk*) nas ciências biológicas (Griffiths 2001; Oyama et al. 2001; Jablonka 2002). O gene veio a ser considerado, então, também uma unidade de informação.

Devido ao notável poder heurístico do modelo da dupla hélice, na década que se seguiu à sua proposição, todos os problemas que eram também considerados fundamentais para a nascente biologia molecular se renderam rapidamente, sem dificuldade ou surpresa (Keller 2002). O

biólogo molecular Gunther Stent (1968), em seu artigo da *Science* intitulado ‘*That was the molecular biology that was*’ (‘Esta era a biologia molecular que já era’), previu o declínio da biologia molecular, que, segundo ele, já havia entrado em sua ‘fase acadêmica’ e não traria mais nenhum paradoxo; faltaria apenas polir os detalhes do conhecimento sobre o material genético e seu funcionamento. Ao afirmar isso, Stent e a maioria dos biólogos moleculares não contavam com o advento da tecnologia do DNA recombinante, muito menos com a série de descobertas que desestabilizariam o próprio conceito molecular clássico de gene.

1.2. Desafios ao gene molecular clássico

Nas últimas três décadas, uma série de descobertas sobre a organização e a dinâmica do genoma tem tornado claro que genes não são discretos – há genes superpostos e nidados –, nem contínuos – há íntrons dentro dos genes; que eles não têm necessariamente uma localização constante – há transposons –, nem sempre uma função claramente definida – há pseudogenes; que eles não são unidades nem de função – há genes que sofrem emenda alternativa, genes que codificam proteínas multifuncionais, e uma forte dependência da ação gênica em relação aos contextos celular e supracelular –, nem de estrutura – há muitos tipos de seqüências *cis*-atuantes¹⁰ que influenciam a transcrição, genes interrompidos etc. (El-Hani, 2007; Falk, 1986; Griffiths e Neumann-Held 1999; Keller 2002; Fogle 1990, 2000).

El-Hani (2007) sumariza os problemas enfrentados pelo conceito de gene – em particular, pelo conceito molecular clássico – por meio de três características observadas nos genomas: (i) correspondências de um-para-vários entre segmentos de DNA e RNAs/polipeptídeos (como, por

¹⁰ ‘*Cis*’ e ‘*trans*’ são termos usados na genética para identificar elementos localizados na mesma molécula e em moléculas diferentes, respectivamente. Sítios no DNA ou RNA (e.g., promotores, terminadores, intensificadores – *enhancers* –, silenciadores) são *cis*-atuantes quando afetam apenas as atividades de seqüências em sua própria molécula, o que geralmente implica que não codificam, mas são alvos de reconhecimento por elementos *trans*-atuantes. Estes, por sua vez, são produtos gênicos (moléculas de RNA ou proteínas) difusíveis que agem sobre qualquer outra molécula.

exemplo, na emenda alternativa de RNA); (ii) correspondências de vários-para-um entre segmentos de DNA e RNAs/polipeptídeos (como no caso dos rearranjos genômicos, como aqueles que originam os genes de receptores de linfócitos T); e (iii) ausência de correspondência entre segmentos de DNA e RNAs/polipeptídeos (como na edição de mRNA).

Diante dessas características, podemos dizer, em termos gerais, que o colapso do conceito molecular clássico de gene se fundamenta na inadequação da idéia de que há alguma unidade no sistema genético, seja uma estrutural, funcional ou informacional. Apenas as descobertas que não afetaram a noção de unidade no material genético puderam ser acomodadas ao conceito molecular clássico de gene, como o fato de que alguns segmentos de DNA são transcritos, mas não traduzidos (genes de rRNAs, tRNAs etc.), e outros não são sequer transcritos, apesar de guardarem grande semelhança com regiões transcritas ('pseudogenes'). Da mesma forma, não se constituiu um problema a substituição gradual da hipótese 'um gene-uma enzima' por 'um gene-um polipeptídeo' (Falk, 1986).

Foi o modelo do operon – o primeiro modelo de regulação gênica – proposto por Jacob e Monod em 1961, que suscitou os primeiros problemas com a concepção do gene como unidade, particularmente, como unidade estrutural (Falk 1986; Fogle 1990; Keller 2002). Por um lado, com a proposição do modelo do operon, postulou-se uma diferença entre genes estruturais e regulatórios. Esta pôde ser, contudo, facilmente assimilada, visto que a ação de genes regulatórios envolve a síntese de proteínas e não há, portanto, maiores dificuldades em tratar genes regulatórios como unidades estruturais e funcionais. Por outro lado, problemas mais sérios foram levantados pelo fato de que diferentes polipeptídeos podem ser seqüencialmente traduzidos a partir de um único mRNA policistrônico, transcrito de um operon bacteriano (El-Hani *et al.*, 2006). Até aquele momento, o conceito molecular clássico tinha se mostrado satisfatório para resolver a tensão entre delimitar genes nos cromossomos a partir de sua ação imediata e a partir

de sua função. Mas só não havia maiores dificuldades até então porque as expressões ‘unidade de transcrição’, ‘unidade de tradução’, ‘unidade estrutural, funcional e informacional’ possuíam o mesmo referente: um segmento no DNA colinear a um RNA que, sendo um mRNA, também era colinear à seqüência de aminoácidos do polipeptídeo final. O modelo do operon começou a dissolver essa simetria e ressuscitou a tensão supracitada: o mRNA (produto intermediário ou imediato) já não era colinear à seqüência de aminoácidos de um só polipeptídeo, mas à de três. Qual segmento agora seria o gene estrutural? No operon *lac*, por exemplo, tomando como referência o *produto imediato*, podemos considerar que o gene estrutural é a unidade de transcrição, que encerra três unidades de tradução e função. Mas, considerando *os polipeptídeos*, a unidade de transcrição encerra três genes estruturais que são também unidades de função e tradução. A situação é ainda pior, porque os três genes estruturais e o gene regulatório podem ser vistos como uma mensagem unitária (uma unidade de informação), que resulta em um produto.¹¹

Este foi um primeiro desafio do modelo do operon à idéia de unidade, mas um problema ainda maior diz respeito ao estatuto ontológico das regiões operadora e promotora (Falk 1986; Fogle 1990; Keller 2002): apesar de compartilharem muitas características com as unidades genéticas (têm funções, são herdáveis, sofrem mutações e influenciam o fenótipo), não são consideradas genes apenas por não serem transcritas (El-Hani, 2007). Diante do compartilhamento de tantas características, a distinção entre ‘unidades’ (genes) e ‘não-unidades’ (seqüências regulatórias *cis*-atuantes não-transcritas, como, por exemplo, a região operadora e a promotora, DNA repetitivo etc.) parece bastante arbitrária.

É interessante notar que a distinção entre genes estruturais e elementos regulatórios, que podem ser genes regulatórios ou seqüências *cis*-atuantes, provocou uma outra mudança importante na estrutura teórica da genética: a transição da concepção de ação gênica para a de

¹¹ Como veremos adiante, esse problema aparece não somente em procariotos, nos quais encontramos o operon *lac*, mas também em eucariotos, nos chamados genes de poliproteínas.

ativação gênica (Keller, 2002). Neste último modelo, genes não agem simplesmente – como concebia a visão predominante nos primeiros anos da genética clássica –, mas precisam ser ligados ou desligados, a depender do contexto celular, para ter algum efeito.

A tentativa de preservar o modelo de gene como uma unidade no material genético tem se revelado infrutífera. Fogle (1990) aponta quatro modelos possíveis para a compreensão de genes como unidades estruturais (Figura 1), mas argumenta que todos se defrontam com sérias dificuldades. O modelo A, o mais inclusivo, abrange a região transcrita e todas as seqüências vizinhas *cis*-atuantes (promotores, intensificadores – *enhancers* –, terminadores, reguladores etc.) que têm alguma influência detectável sobre sua transcrição. No caso do modelo do operon, o gene seria, nesse caso, o próprio operon.¹² A principal vantagem desse modelo é romper com a distinção arbitrária entre unidades e não-unidades no genoma. Mas as desvantagens superam os benefícios. Por exemplo, alguns elementos *cis*-atuantes, como intensificadores (*enhancers*) – e silenciadores, podem afetar a transcrição independentemente da distância da região codificante. Logo, seria difícil demarcar empiricamente os limites do gene no modelo A. Além disso, há muitos tipos de elementos regulatórios que podem operar simultaneamente sobre diferentes alvos genéticos, específicos ou não (neste caso, influenciam qualquer promotor compatível dentro de seu alcance), e em combinações complexas e variadas. Assim, o gene não poderia ser uma unidade discreta, visto que este modelo conduziria a uma superposição considerável entre genes cuja transcrição é regulada pelas mesmas seqüências *cis*-atuantes.

Em nossa visão, o que torna o modelo A mais inadequado, entre os muitos problemas que apresenta, é o fato de que há diferentes tipos de elementos regulatórios, em geral operando em combinações complexas e variadas. Seria como se cada gene se ‘diluísse’ aos poucos no genoma.

¹² Mas, nesse caso, o que seriam os atualmente chamados genes estruturais, que não formam *clusters* (grupos) e não são regulados? Outro tipo de gene? Fica claro que o modelo A não resolveria o problema de referentes múltiplos para o termo ‘gene’.

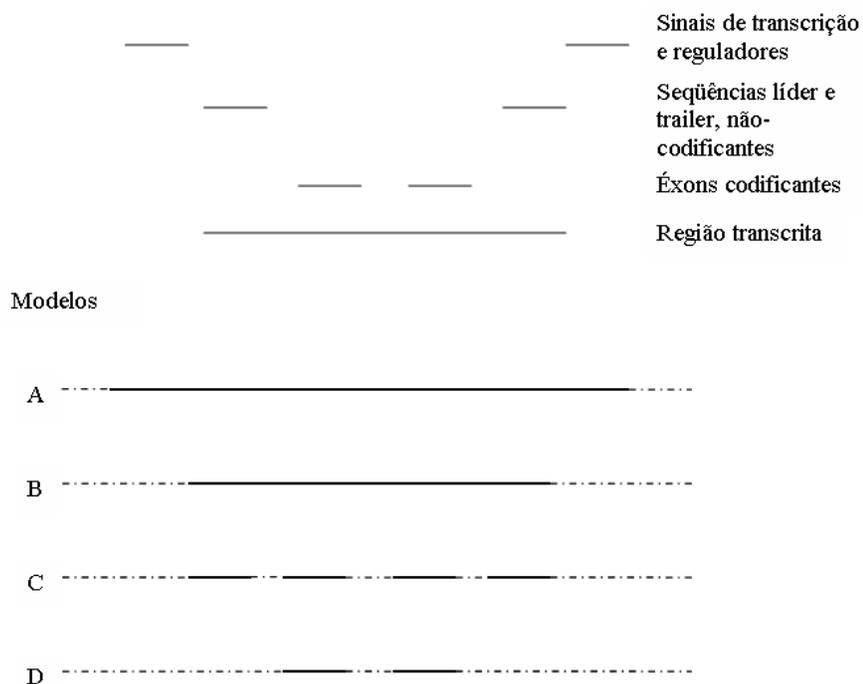


Figura 1. Modelos estruturais para um gene codificante de proteínas (Fogle, 1990). As linhas cheias representam as áreas incluídas em cada modelo.

Limitar o gene à região transcrita, conforme o modelo B, seria a melhor saída para manter a idéia de unidade de estrutura (ainda que permanecesse o problema da distinção arbitrária entre unidades e não-unidades no genoma). Esse modelo possui as vantagens de basear-se nos limites claros que a transcrição, em princípio, parece estabelecer e de comportar uma relação interessante entre a unidade de transcrição e as seqüências necessárias para produzir um polipeptídeo. Mas, por mais interessante que seja, esse modelo é desafiado de maneira incontornável pelos genes interrompidos e pela emenda alternativa, como veremos a seguir. A discussão sobre estas dificuldades da idéia do gene como unidade estrutural nos levará naturalmente ao tratamento dos modelos C e D propostos por Fogle (1990), que discutiremos na seção seguinte.

2. A importância do processamento de RNA: emenda alternativa

Nos eucariotos, há um conjunto de mecanismos de controle pós-transcricional da expressão genética, operando entre a transcrição do DNA e a exportação do mRNA final do núcleo para o citoplasma. Esse conjunto é conhecido como processamento do pré-mRNA (ou mRNA precursor) e envolve uma série de eventos comuns aos genes eucarióticos, como os eventos que conferem estabilidade ao novo transcrito, os eventos de emenda – que seguem à transcrição de genes interrompidos – e a edição de mRNA.

Os genes eucarióticos têm, por assim dizer, um aspecto de mosaico, com seqüências codificantes de polipeptídeos contidas numa matriz de DNA ‘silencioso’ – i.e., aparentemente sem sentido. Mais precisamente, as regiões expressas de uma unidade de transcrição são interrompidas por seqüências intervenientes que não aparecem no RNA maduro – ou seja, que não serão traduzidas. Gilbert (1978) propôs para tais seqüências as denominações ‘éxons’ (de *expressed regions*) e ‘íntrons’ (de *intragenic regions*), respectivamente. Somente após a remoção dos íntrons e a emenda dos éxons remanescentes, o mRNA estará apto a ser traduzido, o que não quer dizer que ele tenha alcançado necessariamente sua versão final. Algumas moléculas de mRNA só ficam maduras após passarem por um processo adicional, chamado de edição de mRNA, no qual nucleotídeos específicos são quimicamente modificados, com substituição de uma base individual, adição ou deleção de bases (Graveley 2001; Hanson 1996).

Fenômenos como a existência de íntrons e os eventos de processamento do transcrito primário estão entre os maiores desafios à idéia de genes como unidades no DNA e como portadores de informação em suas seqüências de bases. Uma forma particular de emenda, a emenda alternativa, traz desafios adicionais. Esse processo envolve a excisão de íntrons e a subsequente emenda dos éxons de um transcrito primário (pré-mRNA) de maneira diferencial, a

dependem, por exemplo, do genótipo sexual, do estado de desenvolvimento/diferenciação celular, da idade da célula e/ou da ativação de uma via particular de sinalização celular.

Black (2003) sumariza os diferentes padrões de emenda alternativa, ilustrados na Figura 2. Em (A), temos um éxon alternativo que pode ser incluído ou excluído do mRNA; em (B), dois ou mais éxons alternativos dispostos em série são mutuamente exclusivos, de forma que a emenda sempre inclui apenas uma das várias escolhas possíveis; em (C, D), um éxon particular pode ser alongado ou encurtado por sítios alternativos de emenda 5' ou 3'; em (E, F), promotores alternativos e sítios alternativos de poliadenilação podem mudar o primeiro e o último éxon de um transcrito, respectivamente; em (G), um íntron, normalmente excisado, é retido no mRNA, no padrão chamado de retenção de íntron, que regula negativamente a expressão de uma sequência transcrita do DNA, ou seja, impede a tradução de seu produto polipeptídico.

Os padrões de emenda expostos na Figura 2A e 2B são do tipo *exon-skipping* (salto de éxon), envolvendo éxons alternativos conhecidos como 'cassete', os quais são ou completamente incluídos ou completamente excluídos do mRNA (Stamm *et al.*, 2005; Magen e Ast, 2005; Black 2003). De acordo com Thanaraj e colaboradores (2004), os padrões que envolvem éxons cassetes são os mais frequentes nos genomas eucarióticos, sobretudo em vertebrados, sendo que 10% destes éxons existem aos pares (sendo apenas um dos dois incluídos num mRNA maduro, como no padrão B da figura 2).

Uma vez que um domínio funcional de uma proteína pode estar inteiramente, ou em grande parte, codificado em um único éxon, a emenda alternativa de éxons cassetes pode criar mRNAs que codificam variantes protéicas (isoformas), com diferenças sutis na função ou mesmo com funções biológicas completamente distintas. Este último caso é chamado de 'conversão de genes' (*gene switching*), podendo ser interpretado, de um lado, como compartilhamento de genes e, de outro, como sobreposição de genes (Downes 2004).

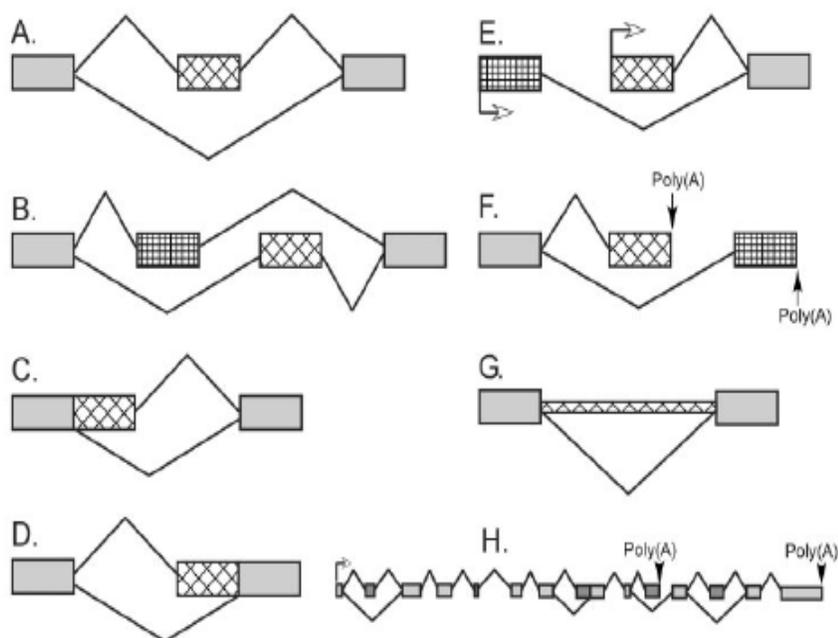


Figura 2. Padrões de emenda alternativa. Éxons constitutivos, presentes em todos os mRNAs finais, estão em cinza. Éxons cassette estão em xadrez. As linhas ligando os éxons representam as diferentes formas de emenda. Em G, um íntron é representado retido entre dois éxons. Diferentes combinações de padrões de emenda alternativa (envolvendo vários introns) podem gerar uma família de mRNAs relacionados (H). [Retirada de Black 2003].

Podemos ilustrar o impacto da existência de íntrons e da emenda alternativa sobre a compreensão acerca de genes e informação genética utilizando uma analogia. Se pudéssemos traduzir a seqüência do gene da β -globina humana como uma sentença, ela se pareceria com a seguinte: ‘Essa seqüência *desahsjkdhinão* expressa *fia* a informação *kjkgosdhuhsj* no gene da β -globina humana’. Obviamente, lida do início ao fim, esta representa uma mensagem truncada. Mas, retirando-se as regiões italicizadas (nossos supostos íntrons), entenderíamos que ‘Essa seqüência expressa a informação do gene da β -globina humana’. Porém, devido à emenda alternativa, esta não seria a única sentença final possível. Se incluíssemos na sentença a sílaba ‘*desa*’, no início da primeira região italicizada, bem como a segunda região italicizada ‘*fia*’, e retirássemos a palavra ‘expressa’, teríamos uma mensagem completamente diferente: ‘Essa

seqüência desafia a informação no gene da β -globina humana'. Alternativamente, poderíamos dizer que 'Essa seqüência não é a informação do gene da β -globina humana', se incluíssemos apenas o *não* na primeira região italicizada, i.e., no fim do primeiro 'íntron', e apenas o *e* da palavra 'expressa'. E poderíamos, claro, ainda obter a sentença 'Essa seqüência não expressa a informação do gene da β -globina humana', se incluíssemos somente a palavra 'não' da primeira região italicizada. Os diferentes padrões de emenda alternativa, portanto, mostram que uma mesma região codificante, ora será removida como íntron, ora emendada como éxon. Esta analogia fornece indícios de por que a presença de íntrons e a emenda alternativa desafiam a definição de informação nos sistemas genéticos e, em particular, a idéia de que esta última estaria contida numa seqüência de bases no DNA que funcionaria como uma unidade de estrutura, função e informação.

Retomando a análise da Figura 1 da seção anterior, o gene, como unidade de transcrição, seria um mosaico, em que regiões codificantes são interrompidas por íntrons. Nesse caso, perde-se a unidade de estrutura, a colinearidade entre o gene e seu produto final, que seria a principal vantagem do modelo B. Sendo os íntrons removidos durante a emenda de éxons, a unidade de transcrição não é a seqüência que será traduzida num polipeptídeo. A unidade de tradução e, portanto, a proteína que ela codifica existem no cromossomo apenas como possibilidades. E nem mesmo a unidade de função poderia ser atribuída ao gene, porque devido à emenda alternativa, múltiplas proteínas, que poderão ser produzidas por diferentes padrões de emenda, existem potencialmente no genoma.¹³

A emenda alternativa é, portanto, um fenômeno que desafia não só o conceito molecular clássico de gene, mas a interpretação geral de genes como unidades, que, como vimos, antecede a

¹³ Diante desse fato, alguns biólogos moleculares sugerem deslocar o gene para o mRNA maduro (ver, e.g., Kampa et al. 2004). Isto aparentemente resgataria o gene como unidade de estrutura e função (mas veremos que este não é necessariamente o caso). Além disso, o próprio gene (e não somente seus possíveis polipeptídeos) existiria no zigoto apenas como possibilidade. Discutiremos essa questão na próxima seção.

própria genética, o que mostra a extensão de sua influência sobre nossa compreensão da herança. Contudo, os desafios atuais a esta idéia são tão contundentes que não é mais possível manter a sobreposição do gene molecular e da idéia de unidade genética, que é parte integrante do conceito molecular clássico e permitiu a sobrevivência da idéia de unidade até os nossos dias. Nossa compreensão crescente da complexidade da estrutura e função gênicas demanda que abandonemos tal idéia. A emenda alternativa é um dos exemplos mais contundentes entre os desafios a esta idéia. Este processo suprime a idéia de unidade de estrutura, mesmo quando esta é atribuída a uma entidade menor do que o gene. Devido aos diferentes padrões de emenda, a unidade funcional, por sua vez, é claramente incompatível com qualquer explicação do gene como se estivesse situado no cromossomo, quer ela inclua ou não seqüências além das que podem especificar um produto gênico.

Para fins do argumento, consideremos que uma forma de confrontar a descontinuidade da região codificante e resgatar a idéia de unidade seria definir o gene sem incluir os íntrons. Poderia ser adotado, portanto, o modelo C da Figura 1, em que o gene representa a unidade formada pelo conjunto de éxons que compartilham um mesmo transcrito.¹⁴ Nesse modelo, admite-se que genes não são, mas apenas encerram unidades de estrutura, os éxons, as menores seqüências codificantes e ininterruptas. Ele acomoda, em princípio, a emenda alternativa, reconhecendo que múltiplos produtos podem ser produzidos pela combinação diferencial dos éxons. Entretanto, surge um novo problema: se o critério para a definição de gene for incluir apenas as seqüências codificantes, por que considerar parte de um gene os éxons das seqüências líder e *trailer*, que podem ser incluídas ou não no mRNA por diferentes padrões de emenda, mas que, no final das

¹⁴ Venter e colaboradores (2001, p. 1317) empregam esta compreensão do gene em um dos artigos que relatam o esboço da seqüência do genoma humano: “Um gene é um locus de éxons co-transcritos. Um único gene pode originar múltiplos transcritos e, portanto, múltiplas proteínas distintas com múltiplas funções, por meio da emenda alternativa e dos sítios alternativos de iniciação e terminação da transcrição”.

contas, não serão traduzidas? Uma saída pode ser adotar o modelo D, que inclui apenas os éxons codificantes.

Mas, nesse ponto, deve-se considerar que a emenda alternativa desafia também a própria idéia de éxons como unidades de estrutura. Como ressalta Fogle (1990), o embaralhamento de éxons para formar isoformas protéicas múltiplas constitui um dos mecanismos da emenda alternativa, mas não o único. Outros padrões de emenda podem alterar até mesmo a extensão do éxon. Uma vez que esse processo pode afetar o tamanho e a região codificante dos éxons, estes segmentos codificariam, na verdade, várias seqüências superpostas, que diferem em extensão e são expressas de forma mutuamente exclusiva em padrões de emenda alternativos. O que é assumido como parte de um éxon, em um padrão, seria parte de um íntron em outro. Assim, opcionalmente, seria possível afirmar que os éxons, como unidades estruturais, são definidos apenas após a emenda, existindo no DNA apenas como possibilidades. A consequência deste fato para o conceito de gene, que já não poderia ser considerado unidade de função, seria a completa perda de qualquer aspecto de unidade. Se as unidades estruturais (éxons) que compõem um conjunto (o gene) podem estar superpostas no DNA, na verdade, não poderemos manter a idéia de unidade, pelo menos não no sentido de entidades discretas e contínuas. Enfim, nenhum dos modelos estruturais do gene apresentados na Figura 1 se sustenta. El-Hani (2007) argumenta que a emenda alternativa apenas mostra que o modelo D não é absolutamente geral e, levando-se em conta que na biologia não há modelos inteiramente gerais, não há motivos para que demandemos tal generalidade desse gene estrutural – assim, ele não deveria ser simplesmente eliminado devido a um tipo particular de emenda alternativa, uma vez que esse modelo permanece útil a despeito de possíveis exceções.

Outros desafios ao conceito de gene, em particular, ao gene molecular clássico, decorrentes de fenômenos como genes superpostos, genes dentro de genes, edição de mRNA, trans-emenda

(*trans-splicing*), pequenas moléculas de RNA com capacidades regulatórias etc., também desafiam a idéia de unidade. A edição de mRNA merece atenção especial aqui, porque constitui exemplo de ausência de correspondência entre RNAs/polipeptídeos e segmentos de DNA (El-Hani 2005, 2007). A adição ou deleção de bases individuais do mRNA pré-editado pode causar alterações no quadro de leitura, de modo que a seqüência de aminoácidos codificada pela versão final do mRNA poderá não ter correspondente no DNA (mesmo levando em conta a emenda). Neste caso, a informação correspondente à versão final do mRNA não poderá ser transmitida (pelo menos não diretamente) para a próxima geração (Keller 2005).

Outro fenômeno notável são os genes nidados. Dois genes que compartilham seqüências no DNA são ditos nidados ou superpostos (Lewin, 2004). No primeiro caso, são encontrados genes dentro de genes. Já os genes superpostos podem ser encontrados numa mesma fita ou cada qual em uma das fitas do DNA. Pode acontecer também de um trecho ser compartilhado de forma mais sutil, sendo lido em diferentes quadros de leitura. Assim como a emenda alternativa e a edição de mRNA, os genes nidados e superpostos também tornam problemáticos os modelos estruturais examinados por Fogle (1990).

No caso do *trans-splicing*, em que são emendadas duas moléculas de RNA codificadas em diferentes *loci* do genoma, há dúvidas se o RNA produzido por esse processo é produto da fusão de duas unidades ou se ele próprio é a unidade codificante, a despeito de ter sido emendado. Este constitui um forte desafio a idéia de gene como unidade estrutural.

Entre os mais recentes desafios ao conceito de gene temos a descoberta de pequenas moléculas de RNA (micro-RNAs) com funções regulatórias. Micro-RNAs são codificados por seqüências espalhadas no genoma, contidas, em grande parte, nas regiões anteriormente descritas como 'DNA lixo' (o que, aliás, mostra a inadequação de tal descrição). Conforme ressalta Keller

(2005), cerca de 98,5% do genoma humano consiste de regiões que não codificam proteínas, sendo que algumas destas regiões codificam micro-RNAs com funções regulatórias.

A conclusão que tiramos do exame de todos estes desafios ao gene molecular clássico é que a interpretação de genes como unidades não pode ser reconciliada com o conhecimento atual sobre a organização estrutural e funcional dos genomas, o que, de certa forma, abre uma porta para a tentativa de salvar o conceito de gene, em vez de abandoná-lo (El-Hani 2005; Keller 2005).

De acordo com Fogle (2000), a existência de produtos múltiplos para um *locus* conspira para forçar decisões arbitrárias sobre se um ou mais genes estão representados naquele *locus*. Segundo este autor, há resistência entre os geneticistas quanto à subdivisão de uma região em múltiplos genes, quando vários produtos compartilham relação funcional e estão relacionados ao mesmo *locus*. Assim, a saída mais corrente é centralizar a delimitação de gene na localização no DNA e considerar que cada unidade de transcrição possui um único gene com várias funções. É fundamental destacar que, como implicação de tal reformulação, o gene é destituído dos poderes de agência e especificação (Keller 2002; Hall 2001): a complexa dinâmica regulatória da emenda alternativa evidencia que cada uma das múltiplas isoformas protéicas não é escolhida nem produzida *por* um gene, mas *a partir dele* (ou melhor, utilizando-o como um dos recursos) pela maquinaria enzimática celular, que, por sua vez, é regulada por sistemas celulares e organizmicos.

A opção de se admitir vários genes para um mesmo *locus* do DNA suscita modificações mais drásticas, relacionadas à própria localização do gene. Se não há como sustentar a idéia de unidade de estrutura e função no DNA, localizar o gene no mRNA maduro (após emenda e edição) parece em princípio ser uma boa saída para manter sua unicidade – um locus no DNA não conteria vários genes superpostos, mas representaria, por assim dizer, a fonte destes, que seriam realizados em um transcrito completamente processado. Podemos encontrar essa sugestão inclusive em

artigos empíricos, como, por exemplo, o de Kampa et al. (2004, p. 331), que argumentam a favor de ‘uma reavaliação do número total de genes humanos e um termo alternativo a “gene” para incluir essas crescentes, e novas classes de transcritos de RNA no genoma humano’. Eles argumentam que a observação de que 49% dos nucleotídeos transcritos nos cromossomos humanos 21 e 22 correspondem a uma nova classe de transcritos de RNA, enquanto que 31,4% corresponde aos já bem-caracterizados genes. Isto, segundo eles “... suporta fortemente o argumento para uma reavaliação do número total de genes humanos e um termo alternativo para ‘gene’ que abarque esta nova classe crescente de transcritos de RNA no genoma humano (ibid., p. 331). Eles não sugerem que devemos abandonar o termo gene, mas observam que “[...] o uso do termo ‘gene’ para identificar todas as unidades transcritas no genoma pode precisar de reconsideração, dado o fato de que este é um termo que foi cunhado para denotar um conceito genético e não necessariamente uma entidade física e mensurável. Com relação ao esforço para se enumerar todas as unidades transcritas, pode ser útil considerar o uso do termo ‘transcrito(s)’ no lugar de gene” (ibid., p.341).

Esse artifício traz um novo problema, porque tais genes existiriam no zigoto recém-formado apenas como possibilidades, delineadas somente posteriormente (*post-factum*) (Keller 2002). Além disso, os genes não teriam a permanência e a estabilidade tipicamente atribuídas a eles e talvez não fossem encontrados nos cromossomos e, muitas vezes, sequer no núcleo, visto que a versão final do transcrito pode ser concluída apenas no citoplasma (El-Hani 2007). É claro que poderíamos propor a eliminação das idéias de estabilidade e permanência, ou de existência nos cromossomos ou no núcleo, do conceito de gene. Mas um conceito de gene assim redefinido poderia cumprir os papéis explicativos que lhe cabem na genética e biologia celular e molecular? Apenas para ilustrar as dificuldades que seguiriam de tal redefinição, como poderíamos pensar nos genes como entidades que mediam a herança de características de uma geração a outra se eles

não possuíssem suficiente estabilidade e permanência, e existissem no zigoto apenas como possibilidades? Torna-se aparente que a tentativa de manter o conceito de gene como unidade poderia terminar, assim, tendo um preço alto demais, afastando nossa compreensão dos genes dos requisitos explicativos que nos fizeram construir tal conceito em primeiro lugar, como um meio de entender a estabilidade das características dos organismos de uma geração a outra (Pitombo *et al.*, 2008a, b).

Na verdade, o ‘cístron’, ou unidade de função, pode não se correlacionar nem mesmo à unidade de tradução, o mRNA maduro, já que existem proteínas multifuncionais. Outro desafio seriam as poliproteínas: após a tradução, alguns polipeptídeos precursores, particularmente de neuropeptídeos e hormônios, se subdividem por clivagem proteolítica em partes idênticas ou divergentes (nesse caso, cada parte seria uma unidade de função independente) (Fogle, 2000; Portin 1993). O gene seria, então, colinear a um longo produto de tradução que abriga vários polipeptídeos menores, um caso semelhante ao do RNAm policistrônico no operon *lac* (ver acima). Dessa forma, nem mesmo o processo de tradução pode ser considerado como uma condição de contorno¹⁵ – a função protéica só pode ser especificada pelo contexto celular, e não por qualquer seqüência de nucleotídeos ou de aminoácidos.

De acordo com Fogle (2000), a alternativa de se admitir vários genes para o mesmo *locus* representa uma inversão da fórmula usual ‘um gene-um polipeptídeo’ para ‘um polipeptídeo-um gene’. Este autor vai além, afirmando que esta decisão seria como aplicar uma espécie de curativo ao problema da ameaça à noção de unidade, e não uma tentativa séria de resolvê-lo. Realmente, esta solução desconsideraria conseqüências metodológicas ou ontológicas importantes. Em termos metodológicos, por exemplo, seria necessário fazer uma revisão do

¹⁵ No caso específico que estamos analisando, a expressão gênica, a ‘condição de contorno’ se refere ao contexto celular em que o gene é expresso e que define o que de fato será expresso. Entende-se como ‘condição de contorno’ as restrições seletivas estabelecidas por um nível hierárquico acima do nível em que se observa uma interação de entidades ou processos, e que determina de maneira descendente o que, no DNA, seria expresso (para uma discussão mais detalhada, ver El-Hani et al., 2006).

sistema de nomenclatura de milhares de *loci*. Ontologicamente, as estimativas do número de genes para humanos e para a maioria dos eucariotos seriam profundamente afetadas.

Uma extensão dessa proposta é localizar o gene, a depender do caso, no DNA ou numa molécula de RNA. Esta sugestão foi apresentada por Kitcher (1982), motivado pelo fato de que catalogar os usos correntes do termo ‘gene’ nos leva a um conceito altamente nebuloso e ambíguo (Kitcher 1982, p.130). A necessidade de acomodar os novos resultados empíricos à idéia do gene como unidade força os programas de pesquisa a enfatizarem diferentes características em situações variadas ou para distintos propósitos. O gene molecular se transforma, então, num *objeto epistêmico* (*sensu* Rheinberger 2000), um ‘gene consenso’ (Fogle 2000), o que constitui, de certa forma, um retorno a uma visão instrumentalista. Isso porque o gene passa a ser entendido como um construto idealizado, ‘aquilo que se espera que ele seja’, um padrão geral de arquitetura bioquímica para dar conta da explosão de estruturas e processos recém-conhecidos. Tal flexibilidade fornece uma falsa impressão de coerência ao gene molecular, mas, ao fim e ao cabo, seu correlato material se torna tão indefinido quanto o era nos primeiros dias da genética. Desta perspectiva, pode até parecer de fato mais sensato abandonar o termo ‘gene’, como propuseram Portin (1993), Gelbart (1998) e Keller (2002), em vez de forçar fenômenos moleculares tão diversos dentro de um mesmo conceito.

Mas há autores mais otimistas, que não propõem o abandono do termo ‘gene’, mas a compreensão de sua diversidade de significados e sua conseqüente reformulação. Muitos apresentam, para este fim, conceitos derivados desde perspectivas pragmáticas. Um conceito com forte dimensão pragmática, relacionado ao de Kitcher, foi proposto por Waters. Ele afirma que a definição subjacente às diferentes aplicações do termo ‘gene’ é a de ‘*um gene para uma seqüência linear em um produto em algum estágio da expressão genética*’ (Waters 1994, p. 178). Segundo o autor, esse conceito seria aplicado, ‘num dado contexto investigativo’, para designar

uma seqüência específica no DNA, contínua ou descontínua, codificando um segmento linear num produto, o qual poderia ser uma seqüência de nucleotídeos no transcrito primário, no RNA processado, maduro, ou até uma seqüência de aminoácidos num polipeptídeo. Então, para cada diferente estágio da expressão genética, haveria um respectivo ‘gene’ no DNA. Esse conceito permite, por exemplo, que um íntron seja ou incluído ou excluído do gene, a depender de qual seqüência linear num estágio da expressão genética esteja sendo considerada. O íntron fará parte do gene quando se focalizar o processo de transcrição no estágio do transcrito primário, mas não fará parte se o foco estiver, por exemplo, na cadeia polipeptídica. Como apontam Griffiths e Neumann-Held (1999), se uma variedade de mRNAs maduros for produzida por meio de emenda alternativa, então, pela definição de Waters, vários ‘genes’ resultarão de um ‘gene’ único. Isso significa que a palavra ‘gene’ não seria utilizada como o nome de uma estrutura molecular específica, mas antes ‘como um rótulo flutuante, cuja referência é fixada pelo contexto local de uso’ (Sterelny e Griffiths 1999, p. 133).

A proposta de Waters (1994) leva a uma sucessão de genes dentro de um mesmo locus no DNA, com o decorrer do processo de expressão gênica e conforme o objetivo da pesquisa. Embora esse conceito esteja próximo do uso corrente do termo ‘gene’ e possa, assim, satisfazer os requisitos pragmáticos dos geneticistas e biólogos moleculares, ele não parece ser suficientemente elucidativo, abrigando muitos problemas que não podem ser ignorados, da perspectiva de uma tentativa de avançar na resolução do problema do conceito de gene. Entre estes problemas, temos, por exemplo, a indefinição do número de genes num genoma e as oscilações na contagem de genes num mesmo *locus*, a depender da etapa da expressão gênica e dos objetivos da pesquisa.

Downes (2004), por sua vez, apresenta o ‘conceito pragmático de gene’, segundo o qual o gene pode ser definido como *qualquer região de ácido nucléico de interesse*. Para este autor,

conceitos pragmáticos de gene seriam como um tipo intermediário entre dois extremos: conceitos que localizam o referente do termo gene no DNA, que ele chama de ‘conceitos de gene na seqüência de DNA’, e conceitos que dispersam o referente do termo gene sobre partes variadas da maquinaria celular, chamados por ele de ‘conceitos inclusivos ou amplos de gene’ (ibid., p. 97).

Consideramos que o elo entre estes três conceitos de gene é o fato de: (i) terem surgido como resposta à problemática levantada, sobretudo, pela emenda alternativa; e (ii) terem como alicerce a compreensão de que a delimitação de um polipeptídeo e, portanto, da função codificada pelo gene está sob o jugo do contexto celular, e mesmo organizmico, no qual ele é expresso. Assim, só faz sentido falar ‘das’ funções possíveis de um *locus* no DNA, ou ‘da’ função desse *locus* numa célula, estágio desenvolvimental ou circunstância particular. Qualquer definição de gene que venha a ser cunhada deve levar esse aspecto em conta.

Griffiths e Neumann-Held (1999) concordam que diferentes conceitos de gene podem ser úteis em áreas distintas da biologia, mas ressaltam que é preciso distinguir um do outro claramente e usar cada qual em seu próprio domínio. Similarmente, El-Hani (2007) argumenta que não há necessidade de adoção de um único e abrangente modelo ou conceito de gene, que inclua toda a diversidade de significados e funções epistêmicas conectada a esse termo, mas também destaca necessidade de demarcar conceitos ou modelos e seus domínios de aplicação. Adicionalmente, a coexistência da diversidade de conceitos e modelos de ‘gene’ poderá até mesmo propiciar maior poder explicativo e heurístico, desde que seus domínios de aplicação sejam bem delimitados.

Os conceitos pragmáticos propostos por Waters e Downes tentam acomodar os desafios ao conceito molecular clássico, localizando, a depender do interesse da investigação, o referente do termo ‘gene’ em seqüências de bases no DNA ou no RNA; ou ainda, localizando o gene apenas

no DNA, mas assumindo que, para um mesmo *locus*, são identificados vários genes que se sucedem no decorrer do processo de expressão genética. Enfim, estes conceitos explicitam que referente do termo ‘gene’ pode ser inferido a partir do contexto da pesquisa. Neumann-Held (2001, pp. 71 e 74) argumenta, entretanto, que, quando o maior interesse é entender a produção regulada de polipeptídeos (como no estudo do desenvolvimento), ao invés de focar na identificação de uma seqüência de ácido nucléico que participe causalmente na síntese de um polipeptídeo, a questão deve ser saber ‘como as seqüências de DNA são usadas no processo de produção de um polipeptídeo’ e se não há ‘outras influências causais que sejam dependentes de um contexto específico, em adição ao DNA’. Estas são idéias relacionadas ao seu ‘conceito molecular processual de gene’ (*process molecular gene concept*).

Frente aos desafios à compreensão do que são os genes, Griffiths e Neumann-Held (1999, p. 2) abandonam a tendência de localizar o gene no DNA, em favor da manutenção do gene como parte central das explicações desenvolvimentais dos traços fenotípicos, ou, em outras palavras, postulam ‘a correspondência de um-para-um entre gene e uma unidade desenvolvimental significativa’. Eles caracterizam o gene como ‘o processo recorrente que conduz à expressão espacial e temporalmente regulada de um produto polipeptídico particular’ (ibid., 1999, p. 4), ou seja, ao proporem esse entendimento sobre genes, eles argumentam em favor da manutenção do gene como unidade de função. Para isso, eles incluem as condições epigenéticas no seu conceito molecular processual de gene, segundo o qual o gene é ‘o processo (i.e., o curso dos eventos) que une o DNA e todas as outras entidades relevantes que não sejam o DNA na produção de um polipeptídeo particular’ (Neumann-Held 2001, p. 74; ver também Griffiths e Neumann-Held 1999).

Essa definição permite dizer que um segmento de DNA não é, em si mesmo, um gene, mas poderá fazer parte de um ou mais genes, a depender de sua contribuição para a produção de certo

polipeptídeo. No caso da emenda alternativa, a atuação do sistema desenvolvimental na escolha diferencial de sítios de emenda acomodaria o fato de que um trecho de DNA poderá, em um gene, fazer parte de um éxon, em outro, de um íntron. Logo, o que é éxon ou íntron só será definido no momento da seleção do par de sítios de emenda, ou, em outras palavras, no ‘decorrer de um gene processual’. Como elementos regulatórios no íntron são necessários à definição de um padrão de emenda e, no final das contas, à produção de um polipeptídeo, os íntrons também seriam incluídos no gene molecular processual.

O conceito molecular processual de gene suscita alguns problemas, como aponta Moss (2001): (i) o número de genes em eucariotos aumentaria substancialmente, considerando-se as múltiplas isoformas produzidas em processos como a emenda alternativa¹⁶; (ii) com a inclusão de sistemas multimoleculares que participam da transcrição e da emenda, o gene passaria para um nível superior na hierarquia biológica; e (iii) por causa da incorporação de processos contingentes, i.e., que podem ou não acontecer, a depender do contexto de regulação, seria difícil individualizar um gene processual. Contudo, este é um conceito que merece, sem dúvida, investigação adicional, visando buscar soluções para esses problemas. Um campo promissor para essa busca ser bem sucedida reside na análise destes problemas dentro de uma perspectiva dos críticos da metafísica de substâncias, os quais incluem vários pensadores como Whitehead, Peirce, Weiss, Alexander, Lloyd Morgan, entre outros (ver Rescher, 1996). Esses pensadores defendem uma filosofia de processos, em que processos são considerados mais fundamentais em termos ontológicos do que entidades. Como salienta El-Hani (2007), a partir dessa perspectiva, o problema de individualizar genes processuais seria visto como algo natural, uma vez que processos são, peculiarmente, mais difíceis de se individualizar, por serem altamente dependentes

¹⁶ Deve-se considerar, contudo, que o problema de contar genes é um problema geral no que tange ao conceito de gene. Não é, pois, exclusivo do gene molecular processual (ver Keller, 2002, 2005). Essa dificuldade levou biólogos moleculares e geneticistas a contarem transcritos, como discutimos mais acima, ao tratarmos da sugestão de Kampa e colaboradores (2004) de trocar ‘genes’ por ‘transcritos’.

de contexto. E acrescenta que, ‘se chegarmos à conclusão de que genes são mais propriamente conceituados como processos, nós não devemos nos esquivar da tarefa de individualizá-los, ainda que seja uma tarefa enorme’ (ibid., p. 304).

Moss (2003) argumenta ainda que a manutenção da relação entre o gene e um fenótipo específico está subjacente ao conceito processual de gene proposto por Griffiths e Neumann-Held. Para ele (ibid., p. 45), entretanto, falar de um gene para um fenótipo é ‘falar como se, mas *apenas* como se, ele determinasse diretamente o fenótipo’. A esse tipo de gene, ele dá o nome de Gene-P (P, de preformacionista), que não traz a idéia de determinação, já no nascimento, de todos os traços de um organismo maduro, mas funciona, antes, apenas como uma ferramenta instrumental. Conforme Moss (id. ibid.), o Gene-P é indeterminado em relação a uma seqüência no DNA, sendo que o fenótipo ou diferença fenotípica que ele determina não se correlaciona à presença de uma seqüência específica, mas pode se relacionar à ausência de um recurso que é uma seqüência molecular ‘normal’, que pode estar ausente de muitas formas diferentes. Por exemplo, possuir um Gene-P para olhos azuis é não ter herdado um ou mais recursos para produzir olhos castanhos. Moss reconhece a utilidade desse conceito em alguns contextos teóricos e empíricos. Afinal, ele captura de modo adequado o uso do termo gene na análise de heredogramas, seja na genética médica, seja no contexto escolar, como um meio de predizer efeitos patológicos, como um câncer de mama ou a fibrose cística.

Este autor distingue, contudo, outro tipo de gene referido na literatura, o Gene-D, que, contrariamente ao Gene-P, é definido por sua seqüência molecular e indeterminado em relação ao fenótipo. O Gene-D é um recurso desenvolvimental (por isso, sua nomeação usando a letra D), definido como uma unidade transcricional no cromossomo (que se estende da trinca de nucleotídeos de iniciação à trinca de terminação), que fornece moldes para a produção direta de RNA ou produção indireta de polipeptídeos relacionados (Moss 2003, p. 46; Moss 2001, p. 88).

Essa abordagem realista dos genes, proposta por Moss, assim como a de Griffiths e Neumann-Held, acomoda processos de regulação da expressão gênica, retira a primazia causal de um segmento de DNA em relação a outras biomoléculas e reconhece que ele pode contribuir para inúmeros efeitos fenotípicos diferentes, até mesmo opostos. Mas ao localizar o gene no cromossomo, ao invés de concebê-lo como um processo, Moss adota a tendência de manter o gene no DNA, abdicando da idéia de gene como unidade de função. Ele afirma que, como um recurso desenvolvimental ao nível molecular, ‘o Gene-D está ontologicamente no mesmo plano’ que outras moléculas como proteínas, RNAs, oligossacarídeos etc. (ibid., p. 47). Moss evita introduzir no conceito de gene os eventos que levam à especificação de um polipeptídeo, porque, em sua visão, o conceito molecular processual conduz a uma explosão de complexidade e contingência que impossibilita a construção de uma taxonomia genética apropriada, em decorrência dos problemas que discutimos acima (Moss 2001).

Por fim, Moss ressalta que não há nada que seja simultaneamente um Gene-D e um Gene-P (Moss, 2003, p. 47), um gene utilizado instrumentalmente como determinante de diferenças fenotípicas e um gene que teria um referente material no genoma e atuaria como um recurso para o desenvolvimento, sem determinar características. A confusão entre Gene-P e Gene-D, em parte produzida pela visão de que um gene no DNA possui informação para um fenótipo em sua seqüência de bases, seria uma das fontes do determinismo genético e ocultaria a enorme diferença, apontada por Keller (2002, p. 89), entre a mera acumulação de proteínas e um organismo, ou, em outros termos, a lacuna de conhecimento que nos resta quando confinamos o desenvolvimento a uma caixa preta. O Gene-P prediz um fenótipo ou uma diferença fenotípica, que é o resultado altamente complexo de um processo de desenvolvimento. O Gene-D, por sua vez, é um recurso, par a par com vários outros que participam dos processos de desenvolvimento, que fornece moldes para proteínas, representando fenótipos somente em níveis muito basais na

hierarquia biológica. Uma característica fenotípica, como a presença de olhos azuis ou a fibrose cística, é sempre o produto de interações complexas entre muitos fatores, cujas funções são contingentes, dependentes de um contexto para o qual também contribuem. Assim, não faz sentido dizer que características em níveis da organização biológica que se situam acima do proteoma sejam determinadas por um recurso do desenvolvimento apenas, como os genes. Até mesmo no caso das proteínas, são comuns casos em que genes não determinam fenótipos, como aqueles em que o dobramento espacial da proteína é dependente da ação de chaperonas.

O Gene-D, diferentemente do modelo estrutural D analisado por Fogle (1990), não é apenas um conjunto de éxons, mas a unidade transcricional que os contém e, portanto, parece incluir os íntrons – logo, corresponderia à unidade de transcrição do modelo B proposto por aquele autor. Para ambos, são colocados desafios à idéia de gene como unidade estrutural, sejam os problemas suscitados pela distinção arbitrária entre unidades e não-unidades no genoma, pelos genes interrompidos ou pela emenda alternativo. Mas diferentemente do modelo B, a proposta do gene-D não se alicerça na idéia de unidade de estrutura, muito menos na de unidade de função, visto que acomoda a possibilidade da produção de muitos polipeptídios a partir da mesma unidade de transcrição. Por um lado, sua relevância reside no fornecimento de recursos de moldes moleculares. Por outro, este conceito não enfatiza a importância dos íntrons na compreensão do que servirá de molde num dado contexto, nem se refere à questão da indefinição das seqüências *cis*-atuantes, como promotores e terminadores.

Fogle (1990), por sua vez, define gene como um conjunto de domínios para uma transcrição ativa (*Domain Set for Active Transcription – DSAT*), sendo um domínio cada um dos tipos de seqüências que se combinam para formar o gene (ibid., pp. 367-368), i.e., seqüências de nucleotídeos que podem ser distinguidas de outras seqüências com base em suas propriedades estruturais e/ou atividades – como, por exemplo, éxons, íntrons, promotores, intensificadores

(*enhancers*), operadores etc. Seu conceito de gene acomoda o fato de que conjuntos de domínios são necessários para influenciar um fenótipo, combinando-se para formar genes. Esse modelo elimina a idéia de unidade do gene molecular e permite a superposição de genes, visto que um mesmo domínio pode fazer parte de mais de um DSAT. Fogle ressalta que nem todos os domínios que influenciam a expressão gênica precisam ser especificados, mas só uma relação parcial suficiente para prescrever um conjunto de interesse a uma comunidade de pesquisadores em suas investigações teóricas e empíricas.¹⁷ De acordo com esta definição, portanto, o que existem no DNA são domínios, e não genes, sendo estes últimos melhor entendidos como objetos teóricos que não devem ser vistos de maneira realista, mas instrumentalista. Os domínios, por sua vez, seriam vistos de maneira realista.

Em contrapartida, o modelo de Fogle, em relação aos modelos que buscam tratar o gene como uma unidade molecular real no DNA, tem a vantagem de não simplificar excessivamente a organização genômica. Fogle esclarece que nem todos os domínios podem ser encapsulados dentro de genes, não tendo qualquer relação empiricamente estabelecida com qualquer gene (é o caso, por exemplo, de muitas seqüências de DNA repetitivo). Desta forma, saber o número de genes seria um parâmetro menos útil do que a medida da densidade, do tipo e do número de domínios presentes num dado organismo. Como o número de DSATs no genoma excede bastante o número de genes definidos como unidades, ele forneceria uma medida mais acurada, na visão de Fogle, do proteoma dos eucariotos.

Um conceito de gene similar ao de Fogle, mas que se aplica também a seqüências de RNA, é o ‘conceito sistêmico de gene’, proposto por Pardini e Guimarães (1992). De acordo com este conceito, um gene é uma combinação de uma ou mais seqüências de ácidos nucleicos (DNA ou RNA) que corresponde a um produto (polipeptídeo ou RNA), mas que só é definida num

¹⁷ Assim, a construção de um conjunto de domínios parece ser feita pela comunidade científica, como um objeto epistêmico, e não pela célula, como proposto no conceito sistêmico de gene de Pardini e Guimarães (1992), que discutimos logo adiante.

determinado contexto de um sistema, que pode ser a célula em interação com o ambiente, ou o próprio ambiente, no caso de sistemas subcelulares ou pré-celulares. Assim, uma seqüência no DNA pode ter múltiplos significados, cada um respectivo a um contexto específico no qual ele se insere. Esse conceito, portanto, enfatiza a relevância dos processos que mediam a relação entre uma seqüência no DNA e um dado produto gênico, mas sem adotar uma visão processual do gene, como fazem Griffiths e Neumann-Held (1999). E, diferentemente dos DSATs de Fogle (1990), para Pardini e Guimarães, o gene é construído pela célula, e não pela comunidade científica, como um objeto epistêmico.

Similarmente a Fogle (1990), Keller (2005) também admite que seqüências de DNA, e não genes, existem independentemente de qualquer outra coisa, sendo um recurso importante para os pesquisadores em busca da compreensão da função biológica. Mas, como Pardini e Guimarães (1992), ela destaca as seqüências de DNA como recursos, sobretudo para a célula, na construção das complexas redes de interações intra- e extra-celulares. Para Keller (ibid.), a célula é um sistema produtor de significado que converte seqüências de nucleotídeos em genes.

Keller (2005), portanto, compartilha com Neumann-Held (2001) uma visão processual e realista de gene. Também é a realista a visão de genes de Pardini e Guimarães (1992). Nas três visões, genes não se localizam no DNA, mas apenas os potenciais para genes. Conforme Falk (1986) argumenta, o genoma representa a possibilidade de realização material de genes e, portanto, não deve haver compromisso com a existência de qualquer entidade, além de sua seqüência de bases. Nos termos de Pardini e Guimarães (1992), é a célula ou o sistema interpretativo (a célula interagindo com o ambiente ou o próprio ambiente, no caso de sistemas subcelulares ou pré-celulares) que irá *construir* o gene que, no caso do ‘conceito sistêmico de gene’, será uma combinação específica de seqüências de DNA ou RNA. Na interpretação de

Neumman-Held (2001) e Keller (2005), qualquer combinação desse tipo será apenas uma parte do gene, entendido como todo o processo interpretativo conduzido pela célula.

Alternativamente, os genes podem ser entendidos como entidades sempre existentes no genoma, como acontece se adotarmos o gene-D de Moss, que também permite acomodar a especificação, pelo sistema interpretativo, de uma seqüência no DNA. Neste caso, o que impediria que as diferentes isoformas protéicas produzidas por emenda alternativa a partir de um mesmo trecho de DNA fossem todas vistas como o mesmo gene (fornecedor de moldes moleculares, caso definido como um gene-D)? Hall (2001, p. 225) faz essa observação e recusa o abandono do termo 'gene', mostrando maior tolerância com a miríade de aspectos estruturais e funcionais que podem ser atribuídos aos genes. Ele argumenta que a heterogeneidade poderia ser também uma característica dos genes, assim como é das células. Para Hall (ibid.), o gene não está morto, mas apenas órfão e sem lar, por ser ainda visto fora de seu contexto, a célula. Ele ressalta que o maior desafio não é simplesmente situar os genes dentro de uma célula ou um núcleo celular, mas sim encontrar seu significado dentro de um contexto temporal e espacial em que eles sejam *capacitados*, por condições do núcleo, citoplasma, célula, matriz extracelular, tecido, e assim por diante, a exercer suas funções no desenvolvimento e na evolução.

No século XXI, é patente a necessidade de substituir ou, pelo menos, complementar o paradigma da biologia molecular, cujo foco são as partes na construção do todo, para acomodar a construção histórica e simultânea de partes e todos ao longo da evolução. Em substituição ao genocentrismo e ao DNA-centrismo (e inclusive ao RNA-centrismo que vem surgindo timidamente, como uma consequência da crise do gene), a tendência, na biologia molecular e genômica, é a emergência de uma abordagem que enfatize a análise do genoma como componente de uma complexa rede interativa na célula.

A interpretação fundamentada na semiótica peirceana de genes como *signos* no DNA, ou ainda de genes como todo o *processo* interpretativo da célula ou do sistema supracelular, que conecta uma seqüência no genoma a um dado produto polipeptídico, se apresenta como uma possibilidade promissora para uma compreensão mais apropriada dos sistemas genéticos. Esta abordagem pode tornar possível conferir a devida importância ao contexto interativo do qual o gene faz parte.

3. Gene, informação e a semiotização espontânea da biologia

Desde o nascimento da biologia molecular, pode-se identificar o crescimento, em grande parte de uma forma espontânea, da tendência de adoção de um novo vocabulário, importado sob o estímulo da concomitante ascensão da cibernética. O problema é que a maneira como o termo ‘informação’ estava sendo entendido na biologia tinha pouca coisa a ver, se alguma, com o sentido original atribuído a ele na teoria da informação. Após cinquenta anos de discurso da informação, encontramos na biologia atual um intrincado conjunto de conceitos e termos, tais como ‘expressão gênica’, ‘transferência ou transmissão de informação’, ‘reconhecimento’, ‘fatores de sinalização’, ‘decodificação’, ‘*cross-talk*’, ‘tradução’, ‘transcrição’, ‘informação genética’ etc. Todos esses termos implicam a aceitação de uma causalidade que tem pouca conformidade com a visão da natureza estabelecida na física clássica, porque a informação implicada no contexto da biologia requer um processo semiótico. A implementação de um vocabulário com conotações semióticas pelos cientistas, sem se darem conta de sua natureza semiótica, foi chamada por Emmeche (1999, p. 274) de ‘semiotização espontânea’ da biologia molecular.

O uso de termos como ‘informação’, ‘codificação’, ‘transcrição’, ‘tradução’ etc. teve grande valor heurístico na genética molecular dos anos 1960, marcando uma distinção crucial desta nova

ciência em relação à genética clássica (Sarkar 2000; Maynard-Smith 2000). À medida que aumentava o entendimento da natureza do DNA e do RNA, a antiga idéia de ação gênica, segundo a qual o gene causava, por atividade direta, a produção de uma proteína (Keller 2002), foi suplantada por uma visão de genes como portadores de códigos, nos quais residiriam ‘informações’ para a produção de traços fenotípicos. Estas informações, por sua vez, deveriam ser processadas pela maquinaria enzimática da expressão gênica. Modelos de regulação e interação gênica foram desenvolvidos de acordo com essa visão, embasando uma maneira de compreender o desenvolvimento como a execução de um programa contido inteiramente no genoma. Tal programa seria levado a cabo satisfazendo-se três regras, apontadas por Sarkar (1996): (i) a informação hereditária reside na seqüência específica de bases no DNA; (ii) a informação passa do DNA ao RNA, por transcrição, e do RNA para uma proteína, por tradução; (iii) a informação nunca é transferida de uma proteína para um ácido nucléico. Este conjunto de regras constitui o chamado ‘dogma central da biologia molecular’, que, conforme explicado por Crick (1958), pode ser interpretado como a transferência da informação para uma proteína através de um caminho sem volta. Ou seja, uma vez que assume a forma de seqüência de resíduos de aminoácidos, a informação não pode retornar ao estado de seqüência específica de bases nos ácidos nucléicos, nem pode ser repassada a outra proteína.

A partir dessa interpretação, podemos entender que, na transmissão do DNA através das gerações, o que é relevantemente herdado é a ‘informação’ contida em sua seqüência de bases. Como a informação não pode passar da proteína para o DNA, está excluída, em princípio, a possibilidade de herança de características adquiridas. Nisso consiste o chamado weissmanismo molecular (Sterelny e Griffiths, 1999). A ‘informação’ relevante para o desenvolvimento das gerações futuras estaria completamente nos genes, sendo os outros fatores causais reduzidos a papéis secundários no desenvolvimento, sendo tipicamente vistos como disparadores de

processos específicos de expressão gênica. Suspeitamos que talvez não seja anacrônico falar que essa noção de programa já podia ser, na década de 1960, considerada equivocada; afinal, já se reconhecia a participação do aparato celular na duplicação, transcrição e tradução do DNA. O problema é que não lhe era dada a devida importância: o pensamento na biologia molecular estava arraigado à concepção do DNA como uma ‘molécula mestra’, a ponto de se superestimar seu valor e de se atribuir à maquinaria celular a função de mera coadjuvante.

Conforme El-Hani, Queiroz e Emmeche (2005), ‘quando a informação é concebida dessa forma [de acordo com o dogma central da biologia molecular], o DNA se torna um tipo de reservatório do qual toda ‘informação’ numa célula flui e ao qual ela deve ser, em última análise, reduzida’ (p. 4). Esse modo de compreender a informação genética pode estar relacionada a formas mais brandas de determinismo genético, de acordo com o qual a emergência de um traço em um organismo é explicada pela presença do gene correto e de condições ambientais ‘normais’, que seriam aquelas ‘adequadas à produção de organismos viáveis’ de um dado tipo, de modo que somente mudanças extremas afetariam o traço de modo a produzir ‘efeitos catastróficos’ (Sterelny e Griffiths, 1999, pp. 14-15). Assim, a maquinaria celular, como componente dessas ‘condições normais’, seria apenas uma ferramenta necessária para ‘rodar’ o programa genético, afetando-o, grosso modo, somente por sua presença ou ausência. O aparato enzimático celular seria mera matéria, em oposição à informação contida no DNA.

Se seguirmos a Susan Oyama ([1985]2000), contudo, poderemos concluir que não se deve diminuir a importância da conexão entre esse modo usual de explicar a noção de ‘informação genética’ e o determinismo. Ela argumenta que o determinismo genético é inerente à maneira como representamos genes e sua função em sistemas biológicos. Na visão dessa filósofa da biologia, enquanto genes forem representados como se carreassem informação sobre como um organismo se desenvolverá, eles continuarão a ser vistos como causas determinantes, não importa

quanta evidência exista contra essa idéia. Em oposição ao determinismo genético, Oyama propôs sua ‘teoria dos sistemas de desenvolvimento’, na qual a paridade causal entre genes e outros recursos desenvolvimentais é um dos postulados básicos (Oyama, [1985]2000; Oyama *et al.*, 2001). Essa perspectiva busca destacar um elemento que estaria ausente nas explicações deterministas da relação genótipo-fenótipo, a saber, o desenvolvimento, ao longo do qual, genes, organismos e ambientes interagem uns com os outros de tal maneira que cada um deles é tanto causa quanto efeito, de uma maneira complexa (Lewontin, 1983, 2000).

Não faltam motivos, contudo, para ressaltarmos a ausência de primazia do DNA e o aspecto dinâmico do desenvolvimento: genes, além de não se autoduplicarem, autotranscreverem e autotraduzirem, precisam ser ativados, inclusive os regulatórios, por fatores não-genéticos, que, no final das contas, determinam como o ‘programa genético’ será lido em cada contexto celular. Diante de padrões diferentes de metilação, emenda alternativa, edição de mRNA, informação posicional no citoplasma do óvulo, por exemplo, não seria mais razoável – caso queiramos manter a metáfora do ‘programa’ – considerar o DNA como um sub-programa ou componente do programa desenvolvimental principal, que estaria distribuído no zigoto? Como em muitos casos a estrutura primária de uma proteína não é especificada pela seqüência de bases no DNA, talvez o genoma não seja sequer uma espécie de sub-rotina, mas antes um conjunto de dados¹⁸ (Atlan e Koppel 1990) a serem utilizados pela célula numa combinação própria e numa seqüência de eventos especificada pelo contexto celular (e, em muitos casos, extracelular). Além disso, a expressão gênica temporal ou tecido-específica, que está na base da diferenciação celular, evidencia que, mesmo se fôssemos capazes de estabelecer uma correspondência simples entre um gene e uma proteína, ainda seria preciso preencher a lacuna entre proteínas e organismos; afinal, como resalta Keller (2002, p. 89), como poderia o organismo ser feito de mera acumulação de

¹⁸ A partir de uma perspectiva semiótica, é preferível pensar num sistema de signos potenciais em vez de um conjunto de dados.

diferentes proteínas? As interações celulares e a influência da matriz extracelular são de importância crucial para a compreensão do desenvolvimento dos organismos.

4. Considerações finais

Em nossa visão, a asserção de que o sistema genético poderia ‘especificar qualquer polipeptídeo concebível’ suscita uma grande controvérsia: a natureza ilimitada, atribuída *a priori* ao sistema genético de herança, não deveria ser mais corretamente vista como uma propriedade do sistema desenvolvimental como um todo, ao invés do genoma isolado? Diante de processos como a emenda alternativa, na qual a especificação de um produto gênico e a regulação de sua expressão são determinadas por condições de contorno, num contexto celular específico, talvez seja melhor, como sugerem Atlan e Koppel (1990), pensar o sistema genético como um conjunto de dados para a maquinaria dinâmica da célula, e não como um programa. Esta maquinaria seria uma espécie de ‘rede de computação paralela’, cujos elementos ‘são reações bioquímicas e transportes’ (ibid., p. 338). Referindo-se às expectativas do PGH, Atlan e Koppel advertem que as análises apresentadas em seu artigo

“[...] tornam extremamente improvável que o seqüenciamento do genoma de um organismo humano na forma de uma lista de pares de bases de DNA possa produzir *por si próprio* um entendimento de seu significado, como se fosse a listagem de um programa real feito por instruções dirigidas para a execução de um dado objetivo. [...] Essa tarefa não pode ser separada da decifração da estrutura de conexão das redes multicamadas de genes e proteínas interagindo ao menos parcialmente em paralelo” (ibid., 346).

A questão é que uma nova dimensão semiótica, além da semântica e da sintaxe, precisa ser considerada: a pragmática. A dimensão pragmática ressalta que um significado só pode ser entendido no contexto da situação do agente, que aqui identificamos como uma célula ou mesmo

um organismo, em um ambiente. O ambiente de um sistema genético corresponde, aqui, não apenas aos fatores externos ao organismo, mas ao próprio contexto celular no qual aquele sistema está imerso. Parece-nos mais razoável a posição de que, se existe um programa para o desenvolvimento, ou mesmo se esta for apenas uma metáfora com valor heurístico na biologia molecular e genômica, esse programa não deve estar situado ou ser metaforicamente atribuído a locais específicos; ele deve ser, antes, distribuído na célula como um todo (Keller, 2002) e, no caso dos sistemas de desenvolvimento, possivelmente no organismo como um todo.

Concordamos com a afirmação de Sarkar (1996, p. 187) de que ainda “não há uma noção técnica clara de informação na biologia molecular”, sobretudo porque a atribuição de um aspecto seqüencial à informação, alegadamente contida no DNA, tem se revelado incongruente, como discutimos acima. Uma saída, claro, pode ser considerar simplesmente que não existe algo que possa ser chamado de ‘informação’ genética, como argumenta Sarkar (ibid.).

É controverso, portanto, se a informação genética pode ser de fato entendida apenas em termos de seqüências de nucleotídeos no DNA (El-Hani et al 2006) e encontramos na literatura, inclusive, um reconhecimento de que o discurso da informação na biologia não passa de uma série de metáforas à espera de uma teoria da informação biológica que possa atribuir-lhes sentido preciso (Griffiths 2001; Stotz et al 2004). Entretanto, esta teoria não foi ainda construída.

A controvérsia a respeito do conceito de gene torna imprescindível a busca por novos modelos de explicação do sistema genético. Adicionalmente, o significado do conceito de ‘informação genética’ não está claro na biologia e demanda uma formulação mais apropriada no contexto de uma teoria da informação biológica que ainda não foi construída (Stotz et al., 2004; Griffiths, 2001). Uma possível base teórica para a construção de tais modelos se encontra na teoria dos signos de C.S. Peirce. Acreditamos que a aplicação de uma abordagem biossemiótica na discussão sobre o conceito de gene poderá ajudar na elaboração de uma definição mais precisa

de informação genética, considerando tanto sua dimensão semântica quanto pragmática. Essa conciliação é plausível, porque, numa perspectiva biossemiótica, a informação pode ser entendida como um processo (El-Hani et al., 2006), ressaltando-se a importância dos sistemas interpretativos de signos, e, logo, sem sustentar um determinismo genético.

Com o intuito de construir tal teoria da informação, estamos propondo um caminho para avançar na direção que Keller (2005), de se forjar uma biologia que enfatize processos, em vez de entidades. Conduziremos nossa discussão na direção de uma perspectiva que contraria a forte tendência da metafísica ocidental, que remonta a época dos filósofos pré-socráticos, a favor de ‘entidades’, de estabilidades fixadas, ao invés de processos, de modos de mudança. A visão processual nos convida a enfatizar a contingência, a emergência, a novidade e a criatividade, a aceitar o mundo em sua inteira complexidade e a enxergar que os itens que classificamos como ‘coisas’ podem ser mais bem compreendidos como instanciações de certos tipos de processos ou complexos de processos. Acreditamos que uma abordagem processual dos sistemas genéticos de informação nos conduzirá do determinismo genético para uma perspectiva mais sistêmica, que considere organismos como sistemas integrados de processos coordenados.

5. Referências

- ATLAN, H. & KOPPEL, M. The Cellular Computer DNA: Program or Data? *Bulletin of Mathematical Biology*, vol. 52 (3), pp. 335-348, 1990.
- BEADLE, G.W. & TATUM, E.L. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA*, vol. 27, pp. 499-506, 1941.
- BLACK, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, vol. 72, pp. 291-336, 2003.

- CARLSON, E.A.C. Defining the Gene: An Evolving Concept. *Am. J. Hum. Genet.*, vol. 49, pp. 475-487, 1991.
- CRICK, F. H. (1958). On protein synthesis. *Symposium of the Society of Experimental Biology* 12:138-163.
- DOWNES, S.M. Alternative splicing, the gene concept, and evolution. *History and Philosophy of the Life Sciences*, vol. 26, pp. 91-104, 2004.
- EL-HANI, C. N. Controvérsias sobre o Conceito de Gene e suas Implicações para o Ensino de Genética. Atas do V Encontro de Pesquisa em Educação em Ciências. 2005.
- EL-HANI C (2007) Between the cross and the sword: the crisis of the gene concept. *Genet Molec Biol* 30(2):297-307
- EL-HANI C, QUEIROZ J, EMMECHE C (2006) A semiotic analysis of the genetic information system. *Semiotica* 60:1-68
- EMMECHE, C. (1999) The Sarkar challenge to biosemiotics: is there any information in a cell? *Semiotica* 127 - 1/4: 273-293.
- FALK, R. What is a gene? Studies in the *History and Philosophy of Science*, vol. 17, pp. 133-173. 1986.
- FOGLE, T. Are genes units of inheritance? *Biology and Philosophy*, vol. 5, pp. 349-371. 1990.
- _____. The Dissolution of Protein Coding Genes, in: Beurton, P.; Falk, R. & Rheinberger, H.-J. *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press., pp. 3-25, 2000.
- GELBART, WILLIAM (1998). Databases in genomic research. *Science* 282, 659-661.
- GILBERT, W. Why genes in pieces? *Nature*, vol. 271, p. 501, 1978.
- GRAVELEY, B.R. Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics*, vol. 17 (2), pp. 100-107. 2001.

- GRAVELEY, B.R. Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell*, vol. 123 (1), pp. 65-73. 2005.
- GRIFFITHS, P. Genetic Information: A metaphor in Search of a Theory. *Philosophy of Science*, vol. 68, pp. 394-405. 2001.
- GRIFFITHS, P.E. & NEUMANN-HELD, E. The many faces of the gene. *BioScience*, vol. 49 (8), pp. 656-662. 1999.
- GUIMARÃES, R.C. & MOREIRA, C.H.C. *O conceito sistêmico de gene –uma década depois. In: Auto-Organização – estudos interdisciplinares 2*, org. por I.M.L.D’Ottaviano & M.E.Q. Gonzales, Coleção CLE (Centro de Lógica e Epistemologia) 30, pp. 249-280, UNICAMP, Campinas, SP, Brasil, 2000.
- HALL, B. K. The gene is not dead, merely orphaned and seeking a home. *Evolution and Development* vol. 3, pp. 225-228. 2001.
- HANSON, M.R. Protein products of incompletely edited transcripts are detected in plant mitochondria. *The Plant Cell*, vol. 8, pp. 1-3. 1996.
- HOROWITZ, N.H. & Leupold, U. Some recent studies bearing on the one gene–one enzyme hypothesis. *Cold Spring Harbor Symp. Quant. Biol.* vol. 51, pp. 65-74, 1951.
- JOHANNSEN, W. *Elemente der Exakten Erblchkeitslehre*. Jena: Gustav Fischer. 1909.
- KAMPA, D.; CHENG, J.; KAPRANOV, P.; YAMANAKA, M.; et al.. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Research*, vol. 14, pp. 331-342. 2004.
- KELLER, Evelyn Fox. *O Século do Gene*. Belo Horizonte: Editora Crisálida, 2002.
- KELLER, E.F. The century beyond the gene. *Journal of Bioscience*, vol. 30 (1), pp. 101-108. 2005.

- KITCHER, P. Genes. *British Journal for the Philosophy of Science* vol. 33, pp. 337-359. 1982.
- LEWIN, B. *Genes VIII*. Upper Saddle River-NJ: Pearson Education. 2004.
- LEWIS, B.P.; GREEN, R.E. & BRENNER, S.E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci.*, vol. 100, pp.189–192, 2003.
- LEWONTIN, R.C. The organism as the subject and object of evolution. *Scientia*, vol. 118, pp. 63-83, 1983.
- _____. (2000). *The Triple Helix: Gene, Organism, and Environment*. Cambridge-MA: Harvard University Press.
- MAGEN, A. & AST, G. The importance of being divisible by three in alternative splicing. *Nucleic Acids Research*, vol. 33 (17), pp. 5574-5582, 2005.
- MAYNARD SMITH, J. The concept of information in biology. *Philosophy of Science*, vol. 67, pp. 177-194, 2000.
- MAYR, E. *O Desenvolvimento do Pensamento Biológico*. Brasília: Editora Universidade de Brasília, 1998.
- MOSS, L. (2001). Deconstructing the gene and reconstructing molecular developmental systems. In *Cycles of Contingency: Developmental Systems and Evolution*, Susan Oyama, Paul E. Griffiths and Russell D. Gray (eds.), 85-97. Cambridge-MA: MIT Press.
- _____. *What Genes Can't Do*. Cambridge-MA: MIT Press. 2003.
- NEUMANN-HELD, E. Let's talk about genes: The process molecular gene concept and its context. In: *Cycles of Contingency: Developmental Systems and Evolution*, Susan Oyama, Paul E. Griffiths and Russell D. Gray (eds.), pp. 69-84. Cambridge-MA: MIT Press, 2001.
- OYAMA, S. 1985. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd Ed. Cambridge: Cambridge University Press. 1st Ed.

- OYAMA, S.; GRIFFITHS, P.E. & GRAY, R.D. (Eds.). *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge-MA: MIT Press. (2001).
- PARDINI, M.I.M.C. & GUIMARÃES, R.C. A systemic concept of the gene. *Rev. Brasil. Genet.*, vol. 15 (3), pp. 713-721, 1992.
- PEIRCE, C.S. (1931–1935). *The Collected Papers of Charles Sanders Peirce*. Electronic edition reproducing Vols. I–VI [C. Hartshorne & P. Weiss (Eds.), Cambridge-MA: Harvard University Press, 1931–1935], Vols. VII–VIII [A. W. Burks (Ed.), same publisher, 1958]. Charlottesville: InteleX Corporation. (Aqui referido como CP).
- PITOMBO, M. A., ALMEIDA, A. M. R. & EL-HANI, C. N. (2008). Conceitos de gene e idéias sobre função gênica em livros didáticos de biologia celular e molecular do ensino superior. *Contexto & Educação (Brasil)* 77: 81-110.
- PITOMBO, M. A., ALEMIDA, A. M. R. & EL-HANI, C. N. (2008). Gene concepts in higher education cell and molecular biology textbooks. *Science Education International* 19(2): 219-234.
- PORTIN, P. The concept of the gene: short history and present status. *Quarterly Review of Biology*, vol. 56, pp. 173-223. 1993.
- RESCHER, N. 1996. *Process Metaphysics: An Introduction to Process Philosophy*. New York: State University of New York Press.
- RHEINBERGER, H.J. 2000. Gene concepts: Fragments from the perspective of molecular biology. In: Peter Beurton, Raphael Falk, and Hans-Jörg Rheinberger (eds.), *The Concept of the Gene in Development and Evolution. Historical and Epistemological Perspectives*. Cambridge University Press, Cambridge: 219-239.
- SARKAR, S.. Biological information: A skeptical look at some central dogmas of molecular biology. In: *The Philosophy and History of Molecular Biology. New Perspectives*, Sahotra Sarkar (ed.), pp. 187-231. Dordrecht: Kluwer, 1996.

- _____. Information in genetics and developmental biology: Comments on Maynard Smith, *Philosophy of Science*, vol. 67(2), pp. 208-213, 2000.
- SOLHA, G. C. F. *Os Genes Interrompidos: Introdução Histórica ao Impacto da descoberta dos introns (1977) na Controvérsia Sobre a Definição de Gene Molecular Clássico (1960)*. Dissertação (Mestrado em História das Ciências da Saúde) - Casa de Oswaldo Cruz/Fiocruz, Rio de Janeiro, 2005.
- STAMM, S.; BEN-ARI, S.; RAFALSKA, I. et al. Function of alternative splicing. *Gene*, vol. 344, pp. 1–20, 2005.
- STENT, G.S. “That Was the Molecular Biology that Was”, *Science*, vol. 160, pp. 390-395, 1968.
- STERELNY, K. & GRIFFITHS, P. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press, 1999.
- STOTZ, K, GRIFFITHS, P, KNIGHT, R (2004) How biologists conceptualize genes: An empirical study. *Stud Hist Philos Biol Biomed Sci* 35:647-673
- THANARAJ, T.A.; STAMM, S.; CLARK, F. et al. ASD: the alternative splicing database. *Nucleic Acids Res.*, vol. 32, pp. D64–D69, 2004.
- VENTER, J.C. et al. 2001. The sequence of the human genome. *Science*, 291 (5507): 1304-1351.
- WATERS, C. K. Genes made molecular. *Philosophy of Science* vol. 61, pp. 163-185. 1994.

EMENDA ALTERNATIVA DE RNA, SEMIOSE E REALIZAÇÃO GÊNICA

1. Introdução

A idéia central deste texto é simples e visa questionar: ‘a mera descrição físico-química de estruturas e mecanismos biomoleculares pode elucidar a função dos mesmos nos sistemas biológicos dos quais fazem parte?’ Supomos que não. O indício mais notório a nos impelir para esta via de pensamento é a surpreendente constatação, na era pós-genômica, de que a despeito de quão numerosos possam ser os genomas¹⁹ seqüenciados, estes não serão suficientes para prover o significado funcional da torrente de dados obtida. Também há dúvidas de que o escrutínio de mecanismos moleculares envolvidos na expressão genética seja capaz de fazê-lo; e esta hesitação concerne não apenas à genômica funcional, mas se estende a qualquer função biológica, sobretudo dos próprios mecanismos subjacentes aos processos de replicação do DNA, transcrição, tradução e das entidades moleculares a eles inerentes, bem como das vias de transdução de sinais que regulam a expressão gênica.

Se concebermos função em termos do papel específico que muitos componentes de um sistema desempenham, contribuindo para a estabilidade do próprio sistema, e se considerarmos que cada função resulta da relação entre as partes quando restringidas dentro de uma estrutura de nível superior, caberá desdobrarmos a resposta à nossa pergunta inicial numa faceta pouco trivial, porém promissora. Aludimos ao que será a pedra fundamental deste artigo, a saber, que uma

¹⁹ O genoma é tradicionalmente entendido como a identidade e localização de todos os genes no DNA; o transcriptoma e o proteoma, por sua vez, representam as concentrações individuais de todos os mRNAs (RNAs mensageiros) e proteínas, respectivamente (Bruggeman et al. 2002).

modelagem semiótica de processos biológicos é requerida como contrapartida não apenas da abordagem mecanicista dos mesmos, mas também de sua análise funcional.

Nesta linha argumentativa, salientaremos a proficuidade da integração de uma abordagem biossemiótica sobre os sistemas biológicos às discussões sobre atribuições e explicações funcionais na filosofia da biologia, mais precisamente, ao modelo de análise funcional proposto por Robert Cummins ([1975]1998). Este modelo tem como diferencial a abordagem de função em termos de *disposições e capacidades complexas*, bem como o foco sobre as complexidades e relações entre propriedades das partes e do todo, em um sistema complexo.²⁰ Ao seguirmos o caminho sugerido por Cummins, a natureza extremamente complexa e dinâmica dos organismos, que tem sido revelada pelos recentes avanços nas ciências biológicas, poderá ser explicada mediante uma análise funcional que não depende de qualquer avaliação de adaptatividade, i.e., de como uma capacidade complexa se relaciona à capacidade do organismo de manter a espécie (cf. Cummins [1975]1998, p.182; Cummins 2002, p.167).²¹ Examinaremos estas idéias em maiores detalhes mais adiante. Por ora, vale ressaltar que na era pós-genômica, a genética – i.e., o estudo do processamento do DNA na construção de características fenotípicas (Keller 2005) – pode se beneficiar da incorporação de uma abordagem semiótica a uma análise funcional desta natureza, sobretudo, no que tange à visão dos fenótipos como sendo controlados por redes auto-

²⁰ Este modelo se contrapõe, portanto, às abordagens que recorrem às considerações evolutivas – a saber, as abordagens etiológicas seletcionistas, denominadas por Cummins (2002) de neo-teleologia, as quais consideram função como algo que explica a existência ou permanência do item organizmico sob consideração. Diferentemente da neo-teleologia, que busca responder à questão ‘por que-ele-está-ali’ respondendo à questão prévia ‘o-que-ele-é-para’ (*what-is-it-for*), a abordagem de Cummins (2002, p. 158) sobre a função de um item não coloca a questão ‘por que-ele-está-ali’, mas sim, a questão ‘como-ele-funciona’.

²¹ Uma análise funcional sistêmica deste tipo não leva em conta o caráter histórico dos sistemas sob estudo. No entanto, o apelo à seleção natural não é tão espúrio como sugere Cummins, visto que ela fornece um mecanismo consistente para explicar por que traços funcionais se tornam mais comuns na população, ainda que não possa explicar a origem de traços funcionais (Nunes-Neto 2008). A explicação da origem de um traço biológico é dada por sua história desenvolvimental, a qual precede qualquer função que ele venha a ter, i.e., o que quer que conte como a função de um item, deverá ser uma atividade que o item executa após seu desenvolvimento ter ocorrido, como bem apontou Cummins (2002, p. 162). Mas a seleção natural pode explicar por que constituições genotípicas e fenotípicas se tornam mais comuns numa população, já que o aumento de sua frequência pode ser simultâneo ao exercício de sua função.

organizadas que apresentam uma dinâmica por todo o sistema (cf. Strohman 2002). Um resultado natural desta abordagem será tornar explícito o caráter ‘grosseiramente ingênuo’, nos termos de Strohman (ibid.), da visão do determinismo genético; em outras palavras, da visão de que características complexas são causadas por um único gene ou conjunto de genes.

Para investigar a compatibilidade de um modelo semiótico com esta forma sistêmica de se pensar a função biológica, nós seguiremos pistas como aquelas fornecidas pela filósofa e historiadora da ciência Evelyn Fox Keller (2005), de que é preciso buscar indícios desta função não em genes particulares, ou na estrutura do DNA, ou sequer em seus produtos protéicos, mas antes nas redes de comunicação das quais o DNA e seus produtos fazem parte. A sutileza do argumento de Keller está na ressalva de que (i) urge irmos não apenas além do gene (como sugere explicitamente o título de seu artigo de 2005, *The century beyond the gene*), mas também ultrapassarmos a compreensão mecanicista de como as partes funcionam em interação umas com as outras, para visar o entendimento sobre: (ii) sua interação com entidades maiores nas quais estão imersas; (iii) os sistemas de regulação gênica extraordinariamente complexos e versáteis; (iv) os sinais que mediam as relações entre diferentes níveis de organização; (v) a variedade de mecanismos epigenéticos de herança, cuja existência e relevância têm sido claramente estabelecidas nos últimos anos; e (vi) o *feedback* evolutivo entre estes diferentes mecanismos, i.e., a maneira como a evolução de cada um destes mecanismos afeta os mecanismos previamente evoluídos. Seguindo estes passos, Keller acrescenta, nós seremos naturalmente compelidos, de um lado, para a necessidade de um novo léxico – ‘um que tenha a capacidade de representar a interatividade dinâmica dos sistemas vivos’ – e de outro, ‘para descrever os tipos de entidades inerentemente relacionais que podem emergir desta dinâmica’ (ibid., p.107).

Tais sugestões se insinuam como um *déjà vu* aos biossemiotistas, os quais concebem sistemas biológicos como sistemas semióticos. Dentro desta perspectiva, a linha condutora deste

artigo reconhece processos de interpretação de signos – i.e., processos semióticos – como necessários para que sistemas vivos, como a célula, levem a cabo sua capacidade auto-corretiva frente ao ambiente em que estão situados e considera que estes processos são subordinados a padrões de condições restritivas estabelecidos pelo sistema celular como um todo, i.e., pela estrutura de nível superior em que entidades e mecanismos biomoleculares estão imersos. Nestes termos, contemplaremos objeções à visão de que sistemas complexos, como os sistemas vivos, podem ser satisfatoriamente compreendidos investigando-se apenas seus constituintes mais fundamentais – e.g., pelo escrutínio de seqüências de DNA e de estruturas e interações moleculares em geral. Estaremos em ressonância, assim, com alguns alertas contra as abordagens reducionistas, como a do filósofo Thomas Nagel que, num simpósio sobre reducionismo em 1997, argumentou que ‘princípios adicionais, não evidentes nas leis que governam constituintes básicos, são necessários para explicar fenômenos de ordem superior’, ou como a do biólogo celular Paul Nurse, de que talvez ‘não seja possível ou mesmo necessário explicar todos os fenômenos celulares em termos de interações moleculares precisas’ (ambos citados por Williams 1997, p. 476-477). Enfatizaremos que a busca por catalogar sistematicamente todas as moléculas e suas interações dentro de uma célula viva (Barabási & Oltvai 2004, p. 101) pode constituir apenas uma dentre as múltiplas abordagens necessárias para o entendimento da função biológica. Seguindo este caminho, estaremos próximos de uma visão pluralista que combina métodos sintéticos e analíticos, abordagens mais globais e abordagens mecanicistas dos sistemas vivos (Bruggeman et al. 2002).

No que tange ao *tour de force* que deverá ser despendido para pintarmos um retrato mais fiel do que se pode chamar de um ‘genoma vivente e respiratório’²² (Rheinberger & Müller-Wille 2008, p.13), esperamos contribuir, em particular, para clarificar por que só em referência a níveis

²² Um genoma com uma configuração dinâmica e flexível, constituído de entidades e mecanismos moleculares que lhe conferem uma característica modular, como um corpo de pedaços remendados (*‘a dynamic body of ancestrally tinkered pieces’*) (Rheinberger e Müller-Wille 2008, p.13).

de organização superiores ao molecular um dado trecho de DNA poderá funcionar como um gene. Para este fim, teremos como pedra de toque o escrutínio da *emenda alternativa de RNA*, um dos principais mecanismos de controle pós-transcricional da expressão gênica. Este é um processo singularmente versátil de regulação, que pode estar integrado a outros mecanismos regulatórios para a modulação de respostas celulares a sinais fisiológicos e desenvolvimentais, bem como para delimitar a função de fatores regulatórios e para diversificar o inventário bioquímico dentro de e entre células, produzindo polipeptídeos variados a partir de uma mesma seqüência de DNA (Lopez 1998; Black 2000, 2003; Cáceres e Kornblihtt 2002). A ocorrência da emenda alternativa em vários domínios biológicos representa, por si só, um desafio à visão de que há correspondência simples – i.e., uma relação de ‘um para um’ – entre genes e fenótipos e demanda o entendimento de que o contexto celular e o organizmico influirão na manifestação de um fenótipo. Adicionalmente, a emenda alternativa pode ser vista como um caso particular de uma classe de processos que devem ser compreendidos para que o mapeamento entre genótipo e fenótipo seja feito, o que constitui um dos grandes desafios da biologia, ao lado da necessidade de explicar o que são genes.

O caminho em que analisamos a função de regulação genética da emenda alternativa segue alguns dos passos sugeridos por Keller (2005) – citados acima – como necessários para ‘trazer o genoma à vida’ (Keller 2005, p.102). Começaremos a (i) ir além dos genes ao elucidarmos noções gerais do mecanismo da emenda alternativa, ressaltando como este processo contribui para que se mostrem infrutíferos os esforços de individuações moleculares dos mesmos – i.e., as divisões em genes feitas pelas delimitações de seqüências de nucleotídeos (Rosenberg e McShea 2008). Em seguida, iremos além desta descrição mecanicista para examinar como o acoplamento da emenda alternativa a outros passos da expressão gênica²³ sugere, de maneira singular, que o

²³ Por exemplo, à transcrição, à poliadenilação, à exportação e aos mecanismos de supervisão do RNA.

entendimento da estrutura e da dinâmica de entidades e processos em sistemas biológicos demanda que eles sejam localizados em vias e redes informacionais complexas (Ideker et al. 2001), i.e., – retomando as dicas de Keller – em (ii) entidades maiores que constituem (iii) sistemas de regulação gênica extraordinariamente complexos e versáteis.

Enfim, tomaremos o quarto passo sugerido – i.e., (iv) a investigação dos sinais que mediam os diferentes níveis de organização²⁴ – como o ensejo para introduzirmos a análise semiótica da emenda alternativa, seguindo o modelo peirceano de análise de processos semióticos desenvolvido por Queiroz, Emmeche e El-Hani (2005; ver também El-Hani, Queiroz e Emmeche 2006).²⁵ Esta abordagem permite a construção de um conceito semântico e pragmático de informação – um problema em aberto na filosofia da biologia (Küppers 1990; Jablonka 2002) – entendida como processo de significação ou semiose, i.e., como a relação irreduzivelmente triádica e processual entre um *Signo* (S), o *Objeto* (O) que o signo representa e o *Interpretante* (I), que é o efeito do signo num intérprete. Nesta perspectiva, informação não é algo contido em alguma estrutura – e sequer inerente apenas à estrutura de nucleotídeos. Informação genética é, assim, apenas um caso de um modelo de informação biológica geral e processual que se aplica a qualquer entidade ou processo nos sistemas vivos e que permite, portanto, suprimir a tradicional primazia que se atribui ao DNA. De acordo com este modelo, que também se fundamenta no estruturalismo hierárquico de Stanley Salthe (1985), a *emergência* de semioses de diferentes tipos pode ser entendida como resultante de interações fundamentais num processo hierárquico triadicamente organizado: a semiose que se observa num dado nível de observação emerge da interação das potencialidades estabelecidas por um nível micro-semiótico (condições iniciais) e a

²⁴ Foge ao escopo do presente trabalho dar conta dos outros passos sugeridos por Keller, a saber, (v) da variedade de mecanismos epigenéticos de herança e do (vi) *feedback* evolutivo entre eles.

²⁵ Sugerimos ainda a consulta de El-Hani, Arnellos e Queiroz (2007) e Queiroz e El-Hani (2007).

influência seletiva de um nível macro-semiótico (condições de contorno), conforme discutiremos mais adiante.

Este arcabouço analítico implica a influência de processos de nível superior de organização sobre processos moleculares, como a transcrição, a emenda alternativa, a tradução etc. É por meio desta regulação que os sistemas vivos instanciam um tipo de potencial criativo em que seus componentes adquirem papéis funcionais que não teriam se independentes desse sistema (cf. Juarrero 2000). Contudo, uma vez que estes papéis podem ser preenchidos por quaisquer componentes que satisfaçam os requisitos do nível superior, múltiplos caminhos físico-químicos produzidos na evolução biológica podem ser classificados como ajustados a uma dada função. Se o foco da investigação de um sistema como a célula é uma capacidade c que ele apresenta e que pode ser analisada na série de capacidades c_1, c_2, \dots, c_n apresentadas por componentes desse sistema, o que mais importa é a ocorrência desta série de capacidades, a despeito de quais estruturas moleculares precisas e minúcias de reações biomoleculares irão se acomodar a elas (cf. Cummins ([1975]1998; ver também a discussão mais adiante).

Os organismos engendram, portanto, um padrão de *soluções múltiplas* e também uma freqüente indiferença à maneira exata como elas são efetuadas, desde que satisfaçam certas condições gerais. Provavelmente, o nível molecular é o de maior espectro de estruturas alternativas que têm efeitos indistinguíveis em níveis superiores de organização (Rosenberg & McShea 2008).

2. Função e realização gênica

O que propomos, deste ponto em diante, é um convite à reflexão sobre possíveis significados para o termo ‘genear’ – um neologismo que aqui forjamos como norte num caminho frutífero a ser seguido para a construção de um arcabouço na biologia que seja mais compatível com a

natureza extremamente complexa e dinâmica dos sistemas biológicos. Se na era pós-genômica podemos vislumbrar os vindouros frutos de uma biologia molecular do material genético, mas não tão claramente os frutos de uma biologia molecular do gene, talvez este seja o momento para uma mudança de *gestalt* que enfatize o ‘tornar-se’, o contínuo ‘vir-a-ser’ dos organismos. Nesta nova perspectiva, conceitos estruturantes do conhecimento biológico como gene e informação podem aparecer como um tipo de processos fisiológicos ou ontogenéticos ou, inexoravelmente, como resultados destes, sendo mais bem descritos por nomes que denotam ação ou mesmo por verbos, como Keller (2005) salienta oportunamente. Tomar esta trilha pode requerer soluções inusitadas como deslocar do DNA o que implementa a função gênica.

2.1. Função do gene, análise funcional da célula

De acordo com Cummins ([1975]1998), atribuir função a algo é, ao menos em parte, atribuir uma disposição a este algo. Nesta perspectiva, podemos explicar uma dada disposição de uma seqüência de DNA para a produção de RNA ou proteína, vinculando-a a uma *regularidade disposicional legiforme* – a qual, segundo Cummins (ibid.), é uma regularidade peculiar do comportamento de certo tipo de objeto e que ocorre em virtude de alguns fatos sobre aquele tipo de objeto. Nesta estratégia, pode se aplicar a regularidade legiforme relativa ao código genético, que subsume um caso particular de codificação em questão, associada a informações sobre este caso, por exemplo, a respeito: da seqüência exata de nucleotídeos codificantes; da continuidade ou da interrupção desta seqüência por trechos não-codificantes; da presença ou não de mutações pontuais; e da ocorrência de condições iniciais particulares que propiciam a fidelidade dos mecanismos de transcrição, processamento e tradução de ácidos nucléicos.

Por sua vez, a explicação das regularidades disposicionais de todo o *sistema genético de informação* demandaria, mais adequadamente, uma estratégia mais sofisticada do modelo de

Cummins (ibid.), a estratégia analítica, em que não se recorre aos fatos da instanciação desta regularidade, mas se apela às capacidades de cada componente desse sistema de realizar tarefas específicas de modo organizado; a capacidade do sistema como um todo seria resultante da manifestação das capacidades dos componentes.²⁶ A mesma estratégia analítica valeria para explicarmos as capacidades de níveis superiores de organização, como o celular, histológico, de órgãos e de todo o organismo.

Para Cummins (ibid.), uma abordagem analítica será tanto mais adequada quanto maior for a diferença de sofisticação entre as capacidades analisadas e as capacidades analisadoras ou, de outro ângulo, quanto mais esta lacuna demandar um papel relevante da organização das partes e processos componentes do sistema. Tal observação é propícia aos argumentos que desenvolveremos no presente trabalho. Consideremos o interesse de compreender a capacidade de reconhecimento de um sistema celular, como um neurônio. Levando-se em conta os diferentes níveis hierárquicos em uma célula, constataremos que a atribuição de função se mostrará apropriada em diferentes graus. Considerando o próprio nível celular, na busca de funções no contexto sistêmico, a estratégia analítica é menos adequada, pois a manifestação da capacidade de reconhecimento numa célula particular será uma mera instanciação da regularidade disposicional legiforme, do tipo de célula em questão, relativa ao reconhecimento, que foi possibilitada por determinadas condições iniciais precipitantes. Neste caso, em que se enfoca o próprio nível celular, a estratégia analítica não é apropriada e sequer faz sentido falar sobre função. No extremo oposto, como nas abordagens reducionistas em que se investigam funções no nível molecular, como a função de genes no DNA ou de outras biomoléculas, a distância de sofisticação entre as capacidades desses componentes e a capacidade analisada de reconhecimento celular evidencia a

²⁶ Note-se que, na estratégia analítica, Cummins ([1975]1998, p. 187) substitui o termo disposição pelo termo capacidade (ou habilidade), que é mais freqüentemente explicado por meio de análise.

relevância da *complexidade da organização* dos referidos componentes e a adequação de uma estratégia analítica.

Se enfatizarmos subsistemas celulares progressivamente mais complexos – e.g., partindo de maquinarias enzimáticas macromoleculares para um mecanismo de processamento de RNA, e deste para todo o sistema genético de informação e, por fim, para uma rede regulatória que conecta vários sistemas como o genético e as vias de sinalização – veremos as capacidades analisadoras se aproximarem, em termos de sofisticação, da capacidade de reconhecimento da célula. Disto segue que à medida que se sobe na hierarquia dos níveis de organização que constituem a célula, a estratégia analítica se torna cada vez mais simples e a atribuição de função cada vez mais sem sentido, pois a partir do foco nestes subsistemas mais complexos como partes, torna-se mais imediata a compreensão da instanciação da referida capacidade da célula.

Esta nos parece uma perspectiva assaz vantajosa para iluminar a dependência de contexto da função gênica e o papel da regulação nos sistemas vivos, uma vez que, conforme discutido acima, a maneira como estão organizados os componentes moleculares e mecanismos físico-químicos é altamente relevante para entendermos a função dos mesmos no sistema celular. Por sua vez, o entendimento da rede regulatória que acopla e organiza estes componentes se revela imprescindível, sobretudo porque sua complexidade é mais próxima daquela do sistema celular que se pretende explicar. Vale destacar que o modelo de Cummins inova por salientar que enquanto se atribui uma função a um componente de um sistema complexo, o alvo da explicação é uma capacidade deste sistema continente, não a presença de certo item.²⁷ Este enquadramento do modelo de Cummins a uma perspectiva sistêmica do mundo ajuda a tornar mais inteligível o porquê de as diferenças na complexidade e no plano básico dos organismos estarem em grande medida relacionadas a mudanças na regulação temporal e espacial de padrões de expressão

²⁷ Este modelo se contrapõe, portanto, às abordagens etiológicas, em que coincidem os alvos da atribuição funcional e da explicação funcional, i.e., em que atribuir função a algo é explicá-lo funcionalmente (ver Cummins 2002).

genética (Carroll et al., 2005), e não tanto à evolução dos próprios genes como mostrado pelas comparações entre vários genomas animais.²⁸

Se adotarmos um modelo funcional do sistema, onde as capacidades dos componentes são tratadas como as funções destes e explicam, juntamente com a organização do sistema, a realização da capacidade sistêmica em questão, esteja ela no nível celular ou supracelular, parecemos frutífero em termos heurísticos somar a este quadro ferramentas de um arcabouço analítico que explique, em termos gerais: (1) de que maneira potencialidades – sejam elas disposições ou capacidades – se tornam realidades (*actualities*); e (2) como este processo de realização pode ser influenciado pela organização do sistema continente. A nosso ver, uma modelagem de processos biológicos segundo à teoria dos signos de C. S. Peirce satisfaz a estes requisitos, por exemplo, explicando a disposição de uma seqüência de DNA (um signo) para a produção de RNAs ou proteínas (objetos) como a incorporação representativa de um ‘potencial real’ (EP 2.388)²⁹ ou ‘forma’ – definida por Peirce como ‘um poder’, i.e., nada como um existente, mas ‘o fato de que alguma coisa aconteceria sobre certas condições’ (MS 793: 1-3), que é verdadeira do objeto que o signo representa, i.e., uma potencialidade que o próprio objeto apresentaria se pudesse estar presente. Nesta perspectiva, o que realiza as disposições ou os potenciais de seqüências genômicas é a transmissão de uma ‘forma’ para um efeito (interpretante) produzido num biosistema (um intérprete, como a célula ou um organismo) em um *processo de informação*. Ao

²⁸ A título de ilustração, citamos os resultados do Projeto Genoma Humano, em que, das predições iniciais de 50.000 a mais de 140.000 genes, só foram identificados 26.000-38.000 genes (Venter et al., 2001), de acordo com o projeto de seqüenciamento privado, e 30.000-40.000 genes, de acordo com o projeto público (Lander et al., 2001). O número de genes humanos está relativamente próximo dos números de genes da mosca da fruta *Drosophila melanogaster* (~14.000), do nemátode *Caenorhabditis elegans* (~ 19.000) (Modrek e Lee 2002) e da planta *Arabidopsis thaliana* (~25.498) (Szathmáry et al. 2001). O número estimado de proteínas humanas, por sua vez, é de cerca de 90.000 (Magen e Ast 2005), cerca de cinco vezes maior que o proteoma da mosca da fruta (Downes 2004). Vale ressaltar que, nestas investigações, genes são identificados com base numa numeração padrão, como unidades de transcrição (ver discussão mais adiante).

²⁹ Seguiremos aqui a prática de citar o *Collected Papers de Charles Sanders Peirce* (Peirce 1931-1935, 1958), pelo número do volume e número do parágrafo, precedido por ‘CP’; o *Essential Peirce* (Peirce 1992, 1998) por número do volume e número da página, precedidos por ‘EP’. Referências ao *Annotated Catalogue of the Papers of Charles Sanders Peirce* (1967) serão indicadas por ‘MS’ seguido pelos números do manuscrito e das páginas.

mesmo tempo, sob uma abordagem semiótica peirceana, a organização de entidades e mecanismos moleculares pode ser entendida como a emergência de significados biológicos que é instanciada quando, nos termos da abordagem de Cummins ([1975]1998), as possibilidades de relação entre as capacidades c_1, c_2, \dots, c_n de componentes físico-químicos – i.e., as capacidades analisadoras de uma capacidade c do sistema celular – são selecionadas pelas restrições colocadas por níveis superiores de organização. Assim, uma provável contribuição desta modelagem semiótica é iluminar como e por que as atribuições funcionais em cada nível de organização demandam o entendimento da influência de outros níveis; um modelo semiótico também contribui à análise funcional, ao explicar como as restrições impostas pelo todo sobre suas partes conferem novos papéis funcionais às mesmas, que elas não teriam se independentes do sistema que as contém (cf. Juarrero 2000).

Todavia, antes de desenvolvermos esses argumentos, prosseguiremos com o devido escrutínio de alguns aspectos do estado da arte da biologia molecular, dos quais nos valeremos em nossa análise semiótica, mais precisamente, os concernentes à emenda alternativa de RNA.

2.2. Soluções múltiplas? Bricolagens, remendos e emenda alternativa

Ao apresentarem uma das críticas ao reducionismo, Rosenberg e McShea (2008, p.106) iluminam o calcanhar-de-aquiles do projeto de se atrelar função à estrutura na investigação de genes, i.e., de se identificar a seqüência de DNA que ‘realiza’, ‘implementa’, ‘instancia’ ou constitui a coisa (RNA ou produto funcional) que realiza uma dada função. Com este intuito, eles recorrem à constatação de que, se pudéssemos encontrar agora uma generalização que fosse verdadeira de todos os organismos, ela seria eventualmente falsificada na busca interminável da natureza por vantagem adaptativa. Um ponto central de seu argumento é o de que os organismos lidam com problemas suficientemente gerais, de forma que, para tais problemas, sempre haverá

mais de uma solução disponível. Este aspecto da dinâmica dos sistemas vivos está entre as razões pelas quais as individualizações de genes com base na função não se alinham com qualquer maneira óbvia de individualá-los pela estrutura dos ácidos nucléicos.

Neste quesito, nosso foco serão as implicações trazidas pela emenda alternativa de RNA. Este processo constitui um caso típico da capacidade dos sistemas vivos de empregar soluções múltiplas em seus problemas de *design* funcional. Analisaremos o caso notável do gene *Dscam* (*Down syndrome cell adhesion molecule*) da mosca-da-fruta *Drosophila melanogaster*, cujo potencial de gerar uma quantidade descomunal de proteínas transmembrana servirá de exemplo do relevante papel de uma diversidade de estruturas no aumento de complexidade dos circuitos neurais durante a evolução. Este refinamento da complexidade neuronal é alcançado pela função desempenhada por proteínas de reconhecimento relacionadas, mas com diferentes seqüências de aminoácidos, bem como pelas diferentes seqüências de ácidos nucléicos que as originam, tanto em outras espécies, como também dentro de cada uma delas, nos diferentes tecidos de um organismo, e até mesmo entre as células individuais. Uma sugestão deste quadro é que a existência de um largo repertório de estruturas moleculares é, em geral, mais importante para o desenvolvimento e a função dos sistemas biológicos do que a seqüência exata de cada estrutura alternativa (para o caso específico das drosófilas, cf. Graveley et al. 2004).

A diversidade de estruturas no DNA que podem contribuir para um mesmo papel funcional na célula pode ser entendida como um reflexo de sua configuração dinâmica e flexível, produzida por um grande arsenal de remendos de DNA (*DNA tinkering*). Este genoma ‘vivente e respiratório’ (Rheinberger e Müller-Wille 2008, p.13) é constituído de entidades e mecanismos moleculares que lhe conferem uma característica modular. Entre os mecanismos que fazem do genoma um corpo de pedaços remendados (*‘a dynamic body of ancestrally tinkered pieces’*, *ibid.*), nós podemos citar: as mutações pontuais, adições, deleções e inversões de nucleotídeos; as

recombinações meióticas de seqüências de nucleotídeos; os remendos de gene e a emenda de DNA envolvida na produção potencial de milhões de anticorpos na resposta imune de organismos multicelulares; e a duplicação e o embaralhamento tanto de genes, quanto de éxons (*exon-shuffling*) (Boue et al. 2003), regiões codificantes numa unidade de transcrição interrompida por íntrons – regiões não-codificantes intragênicas. Entre as entidades que compõem este genoma modular, a mais notável são os *transposons* – elementos genéticos móveis que podem ser excisados e inseridos nos genomas bacterianos e eucarióticos.

Vemos, assim, que a extraordinária versatilidade dos genomas tem como fonte diferenças genômicas que podem ser, mas não precisam se tornar, ‘compartimentadas em genes’ durante a evolução, conforme ressalta Beurton (2000, p.303). O gene, entendido como unidade de transcrição no DNA, não deve ser visto como unidade de evolução, mas antes como seu produto tardio, como o resultado de uma longa história de condensação genômica (ver Rheinberger e Müller-Wille 2008). Nesta longa história, antes de e durante as eventuais compartimentalizações de diferenças genômicas, permeia um processo de geração de estruturas genéticas, que surgem, são refinadas e integradas em estruturas cada vez mais complexas pela combinação fortuita de módulos moleculares já existentes (ver Keller 2007).

Herbert Simon (1996) e Mattick (2004) apontam esta modularidade como uma via para o alcance do que chamam de complexidade organizada a qual, segundo Simon, consiste numa hierarquia de composição.³⁰ A complexidade organizada, para Mattick, ‘é autorizada pela *insignificância* das estruturas geradas pela combinatória absoluta de interações complexas’ (p.184; *itálico* nosso). Na continuação de seu argumento, ele afirma que a evolução e o desenvolvimento precisam percorrer estas possibilidades para encontrar as que são coerentes e

³⁰ O termo hierarquia é empregado por Simon se referindo a todos os sistemas complexos analisáveis em conjuntos sucessivos de subsistemas, diferindo de uma hierarquia formal, que envolve relações de autoridade (cf. Keller 2007). Nossos argumentos nas seções subseqüentes parecem estar mais relacionados a este segundo sentido de hierarquia na medida em que nos referimos às condições de controle, restrições gerais colocadas por entidades e processos de nível superior sobre a dinâmica de processos em níveis inferiores de organização (ver discussão mais adiante).

competitivas. A complexidade organizada condiz, portanto, com uma peculiaridade dos sistemas vivos de alcançar novas maneiras de persistir pela criação de novas estruturas por uma espécie de bricolagem, i.e., de criação e geração de novidades pelo aproveitamento e combinação de partes pré-existentes. A insignificância a que Mattick (ibid.) se refere acima pode estar relacionada à relevância secundária da maneira como será estruturado um módulo funcional: quaisquer componentes que satisfaçam os requisitos de níveis superiores de organização serão classificados como bem ajustados a uma dada função (cf. Juarrero 2000). Os benefícios do alcance de robustez, i.e., de estabilidade em relação a prováveis perturbações (Keller 2007, p.9) são muito mais prementes. Assim, é favorável evolutivamente a existência de estruturas e vias moleculares que distintas entre si, mas variando dentro de certos limites, de modo que compartilhem o mesmo espectro de possibilidades que implementa uma dada função biológica. Analisaremos, adiante, um exemplo notável desta estratégia bricolagem dos sistemas biológicos, nomeadamente, a emenda alternativa de RNA.

Ao comentar a visão de Simon (1996) e Mattick (2004), Keller (2007) ressalta que ambos deixam de capturar qual tipo de organização irá justificar, por exemplo, a atribuição da propriedade de função a uma dada estrutura ou padrão. Nosso palpite é o de que uma resposta para esta questão não está em nenhuma organização em si, mas na interação entre diferentes níveis de organização num sistema biológico. Recorremos novamente a Cummins ([1975]1998; 2002) para ressaltar que a função de remendos genômicos, das eventuais compartimentalizações em genes e, em geral, de qualquer componente de um sistema não pode explicar a origem dos mesmos, visto que qualquer atividade exercida pelo item biológico que pode contar como sua função deve ser uma atividade que o item executa *ex post facto*, i.e., após ele ter se desenvolvido.

Este pode ser um ensejo para não focarmos tanto na estrutura de ‘genes’ ou de qualquer elemento genômico com vistas a identificar a seqüência capaz de realizar, implementar,

instanciar um efeito selecionado. Ainda que possamos apelar à função para explicar porque um item selecionado é comum na população, não podemos dizer que sua função explica por que ele está presente *tal como é*. Numa abordagem funcional sistêmica, prevalece a idéia antirreducionista de que embora dois mundos idênticos fisicamente não possam diferir em sua biologia, uma mesma biologia pode ser obtida a partir de dois mundos fisicamente diferentes (Rosenberg e McShea 2008); em outras palavras, que o conjunto de fatos físicos que compõem um dado tipo, classe ou categoria biológica – e.g., termorregulado, camuflado, sistema digestivo, molécula catalítica, gene, proteína *Dscam* etc. – será muito heterogêneo fisicamente – sempre haverá alternativas para se realizar uma dada função – e, por isso, menos relevantes para a biologia, já que não são fonte de diferenças biológicas. A nosso ver, função está mais intrinsecamente relacionada àquilo que transforma estruturas e padrões em *self* (Keller 2007, p. 7), algo que, nos termos de Bateson (1972, p. 453), poderia ser descrito como ‘uma diferença que faz a diferença’ e, nós acrescentaríamos, faz a diferença para um sistema que o interpreta. Informação: como palavra-chave de nossos argumentos, é mais um verbo, como o ‘genear’.

A seguir examinaremos o mecanismo da emenda alternativa em prol de colher elementos que fundamente nosso argumento de que uma análise semiótica desse processo revela a plausibilidade de que a célula não contém genes, *a priori*, mas os constrói dentro de contextos específicos. Seguindo a sugestão de Keller (2005) de forjarmos uma biologia ao redor de verbos e processos, a célula, por assim dizer, ‘geneia’: disto pode ser compreendido que ela, apenas num processo semiótico situado, produz genes no DNA, ou nos transcritos de RNA, ou ainda, genes entendidos como o próprio processo de realização gênica, o que ressoa com o ‘conceito molecular processual de gene’ proposto por Eva Neumann-Held (2001): o gene como o processo recorrente que conduz à expressão regulada espacial- e temporalmente de um produto particular (ver também Griffiths e Neumann-Held 1999).

2.3. Estudo de caso: emenda alternativa do gene *Dscam* de *Drosophila melanogaster*

A emenda de RNA é um processo indispensável nos genomas eucarióticos, devido à natureza descontínua da vasta maioria de seus genes: nestes, bem como em seus transcritos primários, as regiões codificantes – éxons – são interrompidas por seqüências intervenientes não-codificantes – íntrons. Para a produção de um RNA funcional³¹, os íntrons precisam ser removidos e os éxons, emendados subsequenteiramente. Hoje, sabe-se que a emenda de RNA ocorre em todos os domínios da vida, ainda que as freqüências e os métodos de emenda empregados variem entre os organismos: em Eubacteria e Archaea, os íntrons são extremamente raros e excisados por auto-emenda, e não pela via spliceossômica³². Em eucariotos unicelulares, como protistas, leveduras e outros fungos, íntrons são incomuns e, quando presentes, sua densidade média varia de um a cinco por gene (com algumas notáveis exceções). Já nos eucariotos multicelulares, a emenda é um fenômeno corrente, sendo o número médio de íntrons por gene intermediário em plantas (cerca de quatro por gene) e elevado em animais (acima de sete em vertebrados) (Roy & Gilbert 2006; Jeffares et al. 2006; Hirschman et al. 2006).

A depender do genótipo sexual, do estágio desenvolvimental, do tipo celular ou da regulação por vias de sinalização (Black 2003), as seqüências que serão incluídas no RNA maduro podem ser reconhecidas de maneira diferencial pelo *spliceossomo*, o grande complexo molecular que catalisa a reação de emenda, de forma que um mesmo gene pode originar transcritos variados de mRNAs, cada um contendo uma seleção particular dos éxons disponíveis em sua molécula precursora. Visto de outro ângulo, alguns éxons podem ser alternativamente excisados juntamente com os íntrons – e adicionalmente, alguns íntrons podem ser alternativamente retidos

³¹ Apesar de diversos tipos de RNAs resultarem da transcrição do DNA – e.g., mRNA (RNA mensageiro), rRNA (RNA ribossômico), tRNA (RNA transportador), snRNA (RNA nuclear pequeno, do inglês *small nuclear RNA*) e scRNA (RNA citoplasmático pequeno, do inglês *small cytoplasmic RNA*) –, restringiremos nossa análise à emenda alternativa dos pré-RNAs que resultam em mRNAs funcionais, traduzíveis numa cadeia polipeptídica.

³² O spliceossomo (termo derivado da palavra *splicing*, i.e., ‘emenda’ em inglês) é o complexo enzimático que catalisa a reação de emenda em eucariotos.

no RNA final. Neste processo, conhecido em geral como *emenda alternativa de RNA*, múltiplas isoformas (variantes protéicas) são geradas a partir de um mesmo trecho de DNA, que então engendra o potencial para uma enorme plasticidade funcional, entre os diferentes tecidos que compõem um organismo. Estudos recentes mostram que também as células de um mesmo tecido podem gerar isoformas distintas (Matlin et al. 2005) e que a taxa e os padrões de emenda alternativa podem variar não apenas entre grupos supra-específicos, mas entre indivíduos de uma mesma espécie, de maneira herdável (Kwan et al. 2007).

O mecanismo básico de excisão de íntrons de pré-mRNAs pelo spliceossomo é ilustrado na Figura 1 (para revisão, ver Cáceres e Kornblihtt 2002; Black 2003; Stamm et al., 2005). A maquinaria de emenda compreende quatro partículas de ribonucleoproteínas (snRNPs), U1, U2, U4, U5 e U6, em adição a muitas proteínas não-snRNPs.³³ A formação do spliceossomo procede através de uma série de complexos discretos. O primeiro passo envolve a associação da snRNP U1 ao sítio de emenda 5' no chamado complexo E ('E' derivado das palavras 'early', inicial)³⁴. O reconhecimento do sítio da extremidade 3' do íntron no complexo E inclui a associação do fator auxiliar U2 (U2AF) com o trato de polipirimidina no sítio de emenda 3' e a associação da proteína fator de emenda 1 (SF1) com o ponto de ramificação, também no sítio de emenda 3'. Embora a snRNP U2 possa ser detectada no complexo E, sua associação estável com o pré-mRNA depende da formação do pré-spliceossomo ou complexo A, caracterizado pela dupla U2 e snRNA no ponto de ramificação. A associação da tri-snRNP U4-U5-U6 com o pré-mRNA forma o complexo B, seguido por rearranjos que levam ao complexo C, o spliceossomo maduro, no qual

³³ Associações de proteínas às chamadas pequenas moléculas de RNA (*small RNA molecules*) constituem as snRNPs (sigla para *small nuclear ribonucleoproteins*; lê-se 'snurps') que, juntamente com proteínas acessórias se agregam progressivamente sobre o pré-mRNA, compondo 'complexos pré-emenda' intermediários (Cáceres e Kornblihtt 2002; Black 2003). A reação de emenda só ocorre quando todos os componentes spliceossômicos tiverem sido reunidos.

³⁴ Os sítios de emenda são seqüências especiais na junção éxon/íntron, denominados de acordo com a posição relativa *ao íntron*. Assim, as extremidades 5' e 3' de um íntron correspondem aos sítios de emenda 5' (sítio da esquerda ou doador) e 3' (sítio da direita, ou receptor), respectivamente (Lewin 2004).

as duas etapas da emenda ocorrem, a saber: a definição de éxons e a definição de íntrons, como mostrado na Figura 1 (Schellenberg et al. 2008).

Na definição de íntrons, os sítios de emenda em ambas as extremidades intrônicas são reconhecidos como unidade. Este mecanismo predomina em organismos cujos escassos genes interrompidos possuem, em geral, dois grandes éxons separados por um pequeno íntron – e.g., fungos e a maioria dos protistas. Por sua vez, o reconhecimento de éxons mais curtos num mar de longos íntrons, típicos dos vertebrados, requer a definição de éxon, i.e., que os sítios de emenda nas extremidades exônicas sejam reconhecidos como unidade. Uma vez que apenas pares intrônicos de sítios de emenda são funcionais, a definição de éxon constitui uma rota indireta para a efetivação da reação de emenda. Após a formação do complexo C por definição de éxon, este spliceossomo maduro se rearranja sobre o íntron, como ilustra a Figura 1(b). Assim, a distinção entre definição de íntrons e definição de éxons se refere ao que é primariamente reconhecido pelo spliceossomo.

Deste mecanismo, típico dos eucariotos multicelulares³⁵, sobretudo dos metazoários, resulta predominantemente um padrão de emenda alternativa conhecido como éxons cassetes, em que éxons são completamente incluídos ou completamente excluídos do mRNA (Thanaraj et al. 2004). A produção de éxons cassetes pela definição de éxon constitui uma vantagem evolutiva que predomina em organismos multicelulares, visto que esta permite maior flexibilidade e complexidade combinatória de transcritos emendados alternativamente do que as alcançadas através de definição de íntron (McGuire et al. 2008). A emenda alternativa ilumina, portanto, o porquê da ausência de correlação entre o número de genes e a complexidade nos organismos em geral. Este processo, principalmente a abundância de éxons cassetes no proteoma – i.e., no

³⁵ Como vimos acima, apesar de sua ocorrência variar nos domínios biológicos, tanto o mecanismo de definição de íntron quanto o de definição de éxon atuam na emenda de todos os grupos filogenéticos, inclusive numa mesma espécie e até num mesmo gene (McGuire et al. 2008).

conjunto completo de proteínas expresso por todo um genoma –, pode ajudar a explicar por que o número de genes codificantes de proteínas no genoma humano em pouco excede os de vermes e moscas, apesar da percepção – ‘admitidamente enviesada’, como ressaltam MacGuire e colaboradores (2008) – de nossa complexidade orgânica muito maior.

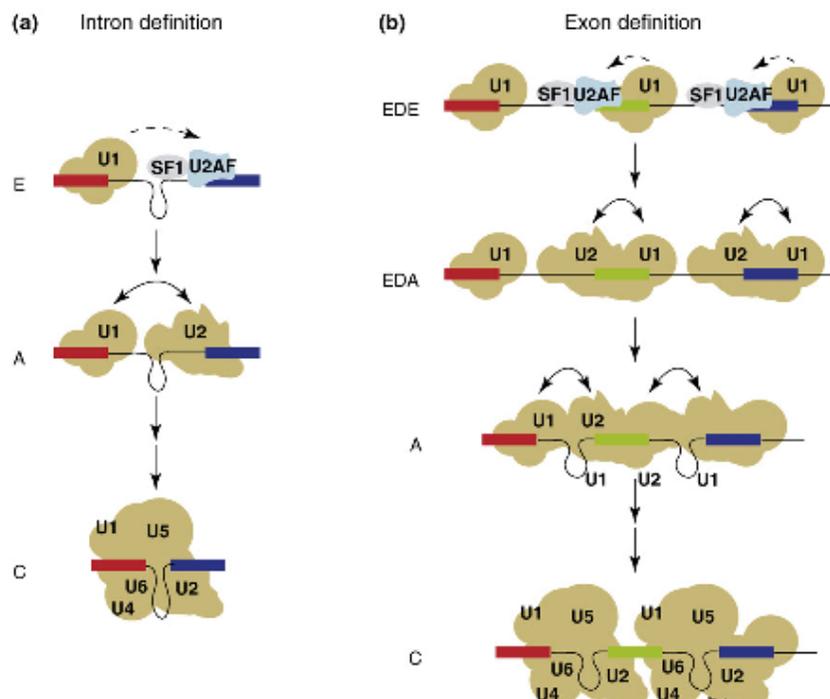


Figura 1. Definição de éxon e íntron na reunião do spliceossomo. (a) Definição de íntron: num pré-mRNA contendo um íntron (linha preta sólida) e dois éxons (vermelho e azul) interações entre fatores nos sítios de emenda no complexo E ocorre ao longo do íntron. (b) Definição de éxon: num pré-mRNA contendo múltiplos íntrons (linhas pretas sólidas) e éxons (vermelho, verde e azul), as interações iniciais entre fatores que se ligam nos sítios de emenda no complexo E ocorrem sobre o éxon num complexo inicial definido por éxon (*exon-defined early complex*; EDE). Após a formação do complexo A definido por éxon (*exon-defined A complex*, EDA), deve haver uma transição para um complexo definido por íntron no prosseguimento para a formação do spliceossomo maduro (complexo C) (Figura retirada de Schellenberg et al. 2008).

Embora somente 2-3 variantes protéicas sejam, em geral, codificadas por emenda alternativa, há vários exemplos de genes com um repertório muito maior de isoformas. Um dos casos mais fascinantes é o do gene da *Dscam* (*Down syndrome cell adhesion molecule*) da mosca das frutas *Drosophila melanogaster*, com o potencial de gerar, nos sistemas nervoso e imune, 38.016 proteínas transmembrana, intimamente relacionadas, da superfamília das imunoglobulinas – este

número é cerca de duas a três vezes o número de genes em todo o genoma desta espécie (Graveley 2005). O gene compreende 115 éxons, dos quais 95 são emendados alternativamente, estando organizados em quatro blocos de éxons alternativos arranjados em série; cada transcrito final diferente contém um éxon alternativo de cada bloco (Zipursky et al. 2006), como ilustrado na Figura 2. Cada isoforma protéica resultante compreende um dos 19.008 receptores extracelulares alternativos – obtidos por uma combinação única de três ectodomínios, i.e., três domínios externos à membrana, do tipo imunoglobulina – que são ligados a um de dois segmentos transmembrana alternativos (Schmucker et al. 2000).

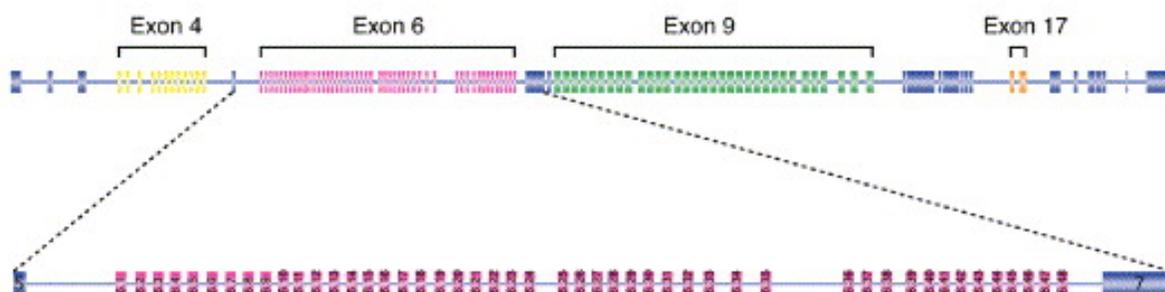


Figura 2. Organização do gene *Dscam* de *D. melanogaster*. *Dscam* contém 115 éxons, 95 dos quais são emendados alternativamente. Os 19.008 ectodomínios alternativos são obtidos por uma combinação única de domínios variáveis de imunoglobulina codificados por 12, 48 e 33 alternativas para os grupos de éxons 4, 6 e 9 respectivamente. O grupo de éxon 17 contém dois éxons que codificam versões alternativas do domínio transmembrana. Dentro de cada grupo, os éxons são emendados alternativamente de maneira mutuamente exclusiva [Figura adaptada de Graveley 2005].

De uma perspectiva mecanicista, um dos aspectos mais enigmáticos do gene *Dscam* é a maneira como a emenda dos seus grandes grupos de éxons ocorre num padrão mutuamente excludente de éxons cassetes (Olson et al. 2007). Os mecanismos conhecidos que asseguram que, *num par* de éxons alternativos, apenas um seja emendado – sendo este um dos arranjos mais comuns de éxons cassetes (Thanaraj et al. 2004) – não podem explicar a emenda mutuamente excludente dos múltiplos éxons cassetes dos grupos de éxons 4, 6 e 9 em mRNAs de *Dscam* (cf. Graveley 2005).

A partir de *insights* obtidos da genômica comparativa, Graveley (2005) propôs um modelo inovador para explicar como apenas um dos 48 éxons mutuamente excludentes do grupo de éxons 6 é incluído por vez, um fato intrigante, sabendo-se que cada um desses éxons possui sítios de emenda funcionais e guardam entre si distância suficiente e potencial para serem emendados de modo independente. Este modelo inclui duas classes de elementos conservados: cada éxon cassette seria precedido por (i) uma seqüência seletora (*selector sequence*), que seria complementar a um trecho diferente do chamado (ii) sítio de ancoragem (*docking site*), uma seqüência intrônica que antecede o grupo de éxons cassetes (Figura 3).

Dado que o pareamento desses elementos não pode explicar sozinho por que não são incluídas as outras variantes de éxons cujas seqüências seletoras não se ligam ao sítio de ancoragem, é imprescindível a atuação de fatores regulatórios que tenham a função de assegurar a fidelidade da emenda mutuamente excludente de um dos éxons 6. Assim, uma proteína-chave inicialmente reprimiria a emenda de todos estes éxons e, num determinado contexto celular, o éxon a ser emendado, e somente ele, parearia com o sítio de ancoragem no seu ponto específico, através de sua seqüência seletora, desfazendo a repressão de sua emenda. Um estudo recente (ver Olson et al. 2007) apóia e estende esse modelo, ao revelar um papel essencial da hrp36, uma ribonucleoproteína nuclear heterogênea (hnRNP): ao se associar a elementos regulatórios no pré-mRNA – sobretudo nos éxons –, ela impede proteínas SR (ricas em serina/arginina) de promoverem a inclusão ectópica de múltiplas variantes do éxon 6.³⁶

³⁶ As proteínas SR e as hnRNPs constituem dois grandes grupos de fatores acessórios que se ligam a elementos regulatórios no pré-mRNA – estes são chamados de intensificadores ou realçadores de emenda (*splicing enhancers*), se atuam positivamente para o reconhecimento de sítios de emenda, e de silenciadores ou repressores de emenda (*splicing silencers* ou *repressors*), se bloqueiam a reunião do spliceossomo em determinado sítio. Ambos os tipos de elementos regulatórios podem ser intrônicos ou exônicos (Black 2003; Lopez 1998). Algumas seqüências regulatórias criam estruturas secundárias que afetam o reconhecimento de um sítio de emenda, mas a maioria parece ser de seqüências ligantes a proteínas. Diferentes proteínas SR atuam em várias etapas da formação do spliceossomo, funcionando tanto como fatores essenciais de emenda quanto como fatores regulatórios (Staley & Guthrie 1998).

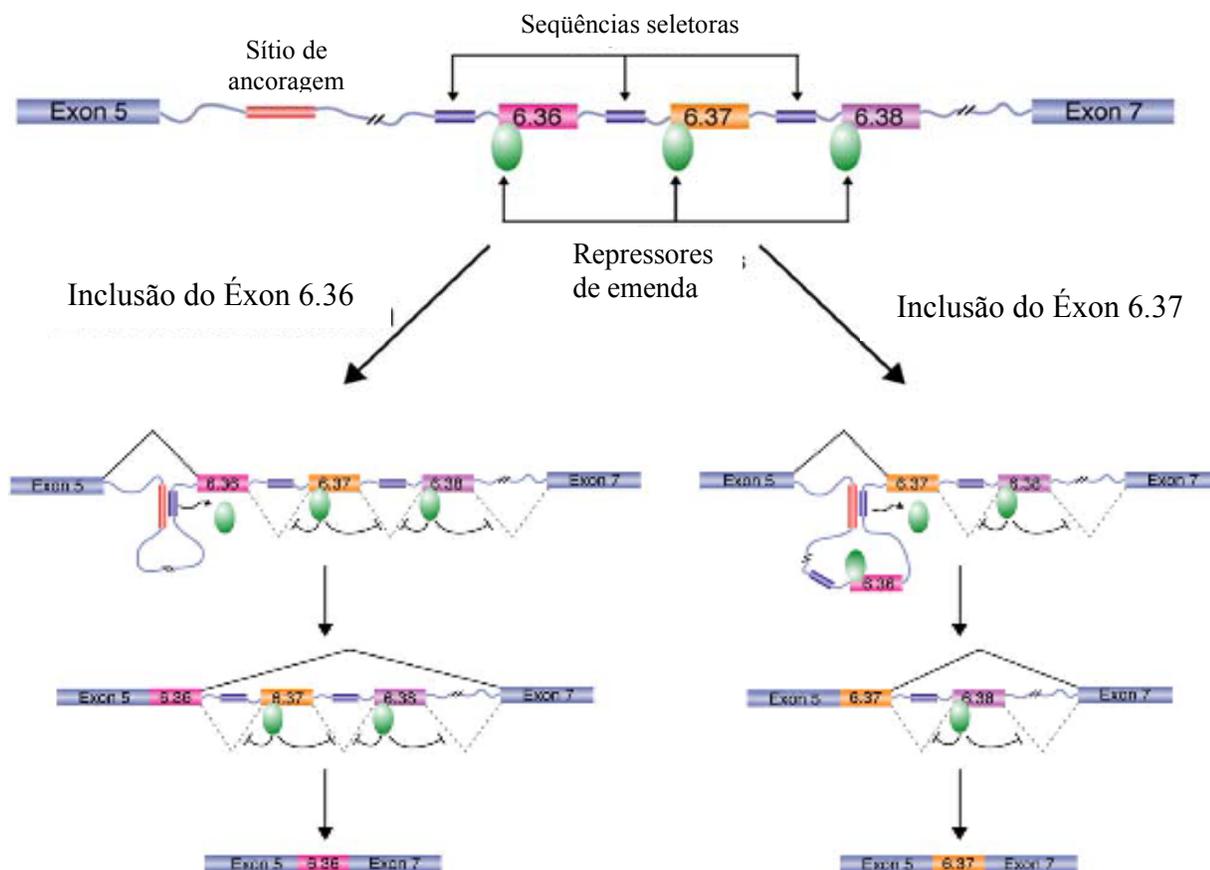


Figura 3. Modelo para o mecanismo de emenda mutuamente exclusiva do éxon 6 de *Dscam*. Apenas as variáveis 6.36, 6.37 e 6.38 estão representadas. Um componente-chave desse modelo é que um repressor de emenda (em verde) funciona impedindo as variantes do éxon 6 de serem emendadas ao mesmo tempo. Para que uma dessas variantes sejam incluídas no mRNA de *Dscam*, a sequência seletora anterior ao éxon deve interagir com o sítio de ancoragem. Por exemplo, se o éxon 6.36 deve ser incluído (à esquerda na figura), a sequência seletora anterior a ele irá interagir com o sítio de ancoragem; se o éxon 6.37 deve ser incluído, a sequência anterior a este éxon irá interagir com o sítio de ancoragem. Por algum mecanismo ainda desconhecido, a interação sítio de ancoragem-sequência seletora inativa o repressor de emenda no éxon imediatamente posterior a esta ligação e, conseqüentemente, ativa a emenda desta variante do éxon 6 ao éxon 5 constitutivo. Subseqüentemente, o éxon que é unido ao éxon 5 só pode ser emendado ao éxon constitutivo 7, porque as variantes restantes do éxon 6 estão ativamente reprimidas pelo repressor de emenda. Como resultado, apenas uma variante é incluída no mRNA (Figura adaptada de Graveley 2005).

Apesar de se conseguir, desta forma, explicar muitos aspectos de como a emenda no grupo do éxon 6 ocorre de maneira mutuamente exclusiva, muitas questões permanecem em aberto. Os próprios autores desse trabalho mencionam que não se sabe como a interação entre o sítio de ancoragem e uma sequência seletora libera, no éxon posterior a esta ligação, a repressão exercida

pela hrp36 e, sequer, o que determina qual das 48 seqüências seletoras interage com o sítio de ancoragem num dado pré-mRNA (ibid.).

Outra dúvida intrigante é sobre se este processo ocorre de maneira aleatória ou altamente regulada. A organização do grupo do éxon 6 – em que cada variante exônica é reprimida pela hrp36 e contém uma seqüência seletora que pode interagir com o sítio de ancoragem – leva a crer que, mesmo se a seleção de seus éxons cassetes fosse randômica, a emenda ainda ocorreria de modo mutuamente excludente. Este fato condiz com fortes indícios de que – tanto nos neurônios, nos quais a *Dscam* participa da conectividade neural, quanto nos hemócitos, em que ela funciona de maneira análoga aos anticorpos de vertebrados – é irrelevante a identidade real das isoformas que são expressas, desde que cada célula individual expresse um conjunto diferente de isoformas em relação às suas vizinhas (Olson et al, 2007; Hattori et al. 2007). A maneira mais simples de se alcançar esta condição é pela via de um mecanismo estocástico.

Vale notar, contudo, que o caráter aleatório da emenda desse grupo de éxons poderia ser facilmente anulado numa via desenvolvimental-, tecido- ou cito-específica pela expressão de fatores regulatórios intensificadores ou repressores de éxons determinados. Adicionalmente, é bem conhecido que muitas hnRNPs atuam antagonicamente às SR e que, em geral, é a taxa entre esses dois tipos de proteínas regulatórias que determina o padrão de emenda (Matlin et al. 2005). E sendo as interações desses fatores com o pré-mRNA geralmente fracas, a regulação de uma dada emenda por um determinado fator depende da presença de outras proteínas regulatórias que formam complexos específicos de pré-mRNA e proteínas. É devido a esta combinação que poucos tipos de fatores de emenda regulam tipos variados de transcritos primários que codificam isoformas com funções biológicas diferentes (cf. Stamm et al. 2005).

Num determinado contexto celular, estímulos externos à célula influenciam padrões de eventos de fosforilação de fatores de emenda – afetando a ligação destes ao RNA ou a outras

proteínas –, o que, por sua vez, aumenta o número de combinações das proteínas que se reúnem nos complexos regulatórios. Como os fatores parecem regular funções biológicas coerentes (Ule et al. 2003), a mudança na sua atividade resultará, muito provavelmente, numa resposta coordenada ao estímulo que disparou o sinal inicial de fosforilação. Em vista disso, o perfil molecular de uma célula num dado instante certamente incluirá várias proteínas, cuja expressão, apesar de parecer não-relacionada, pode ser traçada até um conjunto limitado de fatores regulatórios, desde que o mecanismo de regulação seja conhecido. A Figura 4 ilustra como padrões de emenda tecido-específicos podem ser estabelecidos a partir de poucos fatores regulatórios.

Na célula A, nenhuma das proteínas regulatórias é produzida e apenas AE1 é emendado, sendo que só são expressas isoformas protéicas intracelulares que não interagem entre si. Na célula B, somente a proteína regulatória 1 é expressa, resultando na inclusão de um trecho de AE1 que não foi incluído na célula A, e que agora resulta na expressão da proteína transmembrana (vermelho). A ligação ao éxon 2 não pode ativar a inclusão deste éxon, já que fatores adicionais são necessários (no caso, intensificadores não ocupados). O fator não influencia o gene 3, visto que ele não contém o intensificador apropriado. Em C, apenas a proteína regulatória 2 é expressa (verde). Ela não influencia o gene 1, porque ele não contém seu intensificador. Como resultado, o éxon AE1 não é incluído no mRNA maduro e uma proteína citoplasmática solúvel é produzida (vermelho). O fator não pode ativar AE2, porque são necessários fatores adicionais. Ele induz o éxon AE3, que codifica um sítio de fosforilação, resultando na ligação da proteína codificada pelo gene 2 àquela codificada pelo gene 3. Em D, ambos os fatores de emenda, 1 e 2, são expressos, levando à ativação de todos os éxons alternativos. AE2 é ativado, porque a ligação de ambos os fatores de emenda é estabilizada pela interação proteína-proteína (área vermelha entre os fatores de emenda). Visto que AE2 codifica

um códon de terminação prematuro, nenhuma proteína é sintetizada a partir deste gene (círculo tracejado).

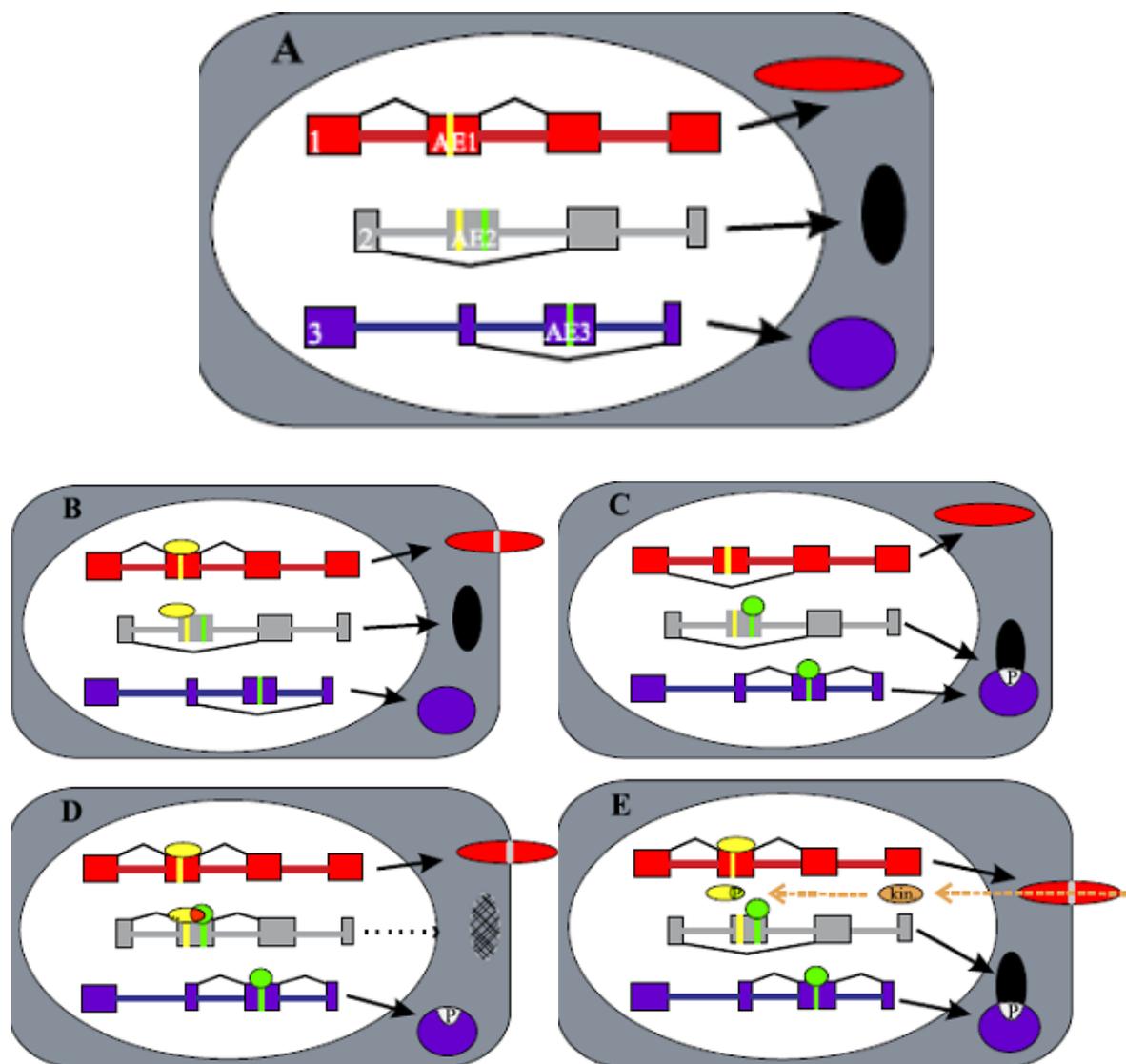


Figura 4. Estabelecimento de padrões de emenda tecido-específicos a partir de poucos fatores regulatórios. Os retângulos cinza representam células, sendo o núcleo, a região branca; três genes são mostrados no núcleo e indicados por diferentes cores. Proteínas codificadas por esses genes são mostradas na mesma cor, à direita de cada célula. Éxons são indicados por blocos e introns, por linhas horizontais. Pequenas elipses indicam proteínas regulatórias da emenda. Cada gene tem um éxon emendado alternativamente (AE1, AE2 e AE3) [Torne mais visíveis estas siglas, está difícil de ver]. Nesta simplificação, a regulação da emenda é realizada pelas proteínas regulatórias 1 (elipses amarelas) e 2 (círculos verdes), que ativam um éxon, após ligarem-se ao intensificador apropriado (linhas amarelas ou verdes no éxon). O éxon AE1 codifica um domínio transmembrana; AE2, um códon de terminação prematuro, e AE3, um sítio de fosforilação (Figura adaptada de Stamm et al., 2005).

Em E, sinais externos podem mudar a seleção do sítio de emenda. O estímulo externo resulta num influxo iônico para a célula (linha tracejada laranja), o que ativa uma quinase (kin), que fosforila o fator de emenda 1 (amarelo). Segue que o éxon 2 é pulado e uma isoforma funcional é feita (Stamm et al. 2005) (Figura 4).

Podemos inferir que o contexto da célula E, acima, representa de forma simplificada o modo como uma *rede biológica significativa de genes* é regulada por um pequeno número de fatores. Estes, se não são tecido-específicos, têm em sua maioria concentrações características em cada tecido (Hanamura et al. 1998). Genes diferentes, aparentemente não relacionados, podem ter sua expressão coordenada constituindo uma resposta biológica significativa em cada concentração específica de fatores; esta resposta regulada pela emenda alternativa pode compor um padrão de expressão que inclui atividades biológicas diferentes, porém coerentes, como a simultânea ligação de uma proteína à membrana, a geração de certa proteína funcional e a ligação desta a outra proteína (Figura 4E). Stamm e colaboradores (2005) nos alertam de que, para avançar no entendimento dessas redes genéticas mecanicista e funcionalmente intrigantes, será necessário determinar os genes que são alvos fisiológicos de mais fatores de emenda, criar sistemas-modelo que prevêm a regulação alternativa de éxons e acessar o papel fisiológico da mudança coordenada nas isoformas protéicas. E mais: ‘Uma vez que as vias de emenda alternativa mudam continuamente durante processos fisiológicos, será importante desenvolver as vias de transdução de sinal que conduzem ao spliceossomo e determinam seus mecanismos de ação’ (p. 14).

Nós acrescentaríamos que o devido escrutínio das vias de sinalização resultantes num dado padrão de emenda alternativa demanda um arcabouço analítico que explique como estímulos ambientais, sinais biológicos evoluídos – e.g., sinais bioquímicos de regulação – ou o material hereditário no DNA – i.e., um tipo muito especial de sinais biológicos evoluídos (cf. Jablonka 2002, p. 580) – se integram numa característica peculiar da célula, a saber: a dialética na qual este

sistema vivo mais elementar instancia sua habilidade de reconhecer e categorizar eventos em seu ambiente e de responder a eles expressando funcionalmente um ‘controle previamente categorizado’ – e.g., a produção de uma enzima que tenha algum efeito em seu entorno – conforme ressalta Rocha (1997, p.2). A nosso ver, uma abordagem semiótica Peirceana, como arcabouço analítico, preenche esse requisito; alguns indícios que fundamentam esta conjectura serão examinados a seguir.

3. Análise semiótica da emenda alternativa de RNA

Na presente seção, incorporamos a emenda alternativa de RNA numa recente modelagem semiótica dos sistemas genéticos de informação e de vias de sinalização celular (ver Queiroz et al. 2005; El-Hani et al. 2006; El-Hani et al. 2007), visando contribuir para a investigação da legitimidade e do poder heurístico de se conceber as redes funcionais dos seres vivos em um arcabouço teórico biossemiótico, em vez de um arcabouço puramente mecanicista (ver Emmeche 1999). Sistemas semióticos, de acordo com Fetzer (1988), podem ser entendidos como sistemas que *incorporam* a semiose, um processo que conecta irreduzivelmente três elementos interdependentes, um *Signo* (S), o *Objeto* (O) que o signo representa e o *Interpretante* (I), que é o efeito do signo num intérprete. Esta explicação recorre, portanto, ao modelo triádico de semiose concebido pelo lógico pragmatista Charles Sanders Peirce. Neste modelo, a semiose ou ‘ação do signo’ é um processo situado – dependente de contexto – e dialógico – dependente de intérprete – no qual se comunica uma forma ou hábito de O para I através de S, de modo a restringir, em geral, o interpretante como um signo ou, em sistemas semióticos, o comportamento do intérprete. A consequência direta da natureza da semiose é que sistemas semióticos são auto-corretivos.

Se sistemas biológicos são semióticos, eles incorporam a produção, transmissão, recepção e interpretação de signos de diferentes tipos como meio de assegurar sua estabilidade ao responder

a traços ambientais de maneira regular, habitual – as restrições ao sustento de sua estrutura ganham um caráter informativo dentro de um contexto auto-referencial e recursivo (El-Hani et al. 2007). Nesta dinâmica, mensagens externas sobre restrições ambientais relevantes precisam ser recriadas e armazenadas internamente por uma maquinaria interpretativa que, por sua vez, irá expressar funcionalmente essas restrições nas recorrências futuras de um dado contexto.

Este modelo tem como requisito a existência de sistemas interpretativos de signos, i.e., sistemas que estão constantemente respondendo a sinais selecionados em seu entorno – por exemplo, as maquinarias de transcrição (e.g., RNA polimerase), de tradução (e.g., o ribossomo), da emenda alternativa (e.g., o spliceossomo) e, na sinalização celular, um receptor de membrana que reconhece um dado sinal extracelular são considerados como sistemas interpretativos de signos. As ações desses sistemas, contudo, estão subordinadas ou reguladas pelo sistema maior do qual fazem parte, a saber, a célula ou intérprete ‘global’ (ver Jablonka 2002). De toda a discussão acima, segue que o argumento de que é legítimo e heurísticamente frutífero conceber as *redes funcionais* dos seres vivos num arcabouço biossemiótico (Emmeche 1999) parte da premissa de que significados biológicos, para serem compreensíveis, demandam o reconhecimento, pelos sistemas interpretativos, de estruturas e eventos bioquímicos como traços ou parâmetros de avaliação de um dado contexto (Hoffmeyer 2001). Esta condição, por sua vez, indica a necessidade de uma análise semiótica, e não meramente físico-química.

Dito de outra maneira, instanciados em sistemas biológicos, estruturas moleculares e processos físico-químicos – ou mecanismos, como são chamados na biologia molecular – podem ser dotados de um aspecto peculiar que inexiste quando eles ocorrem em sistemas mais simples, não-vivos. Esta peculiaridade consiste no fato de que, adicionalmente às interações reativas, eventos bioquímicos engendram outro tipo de relação – a relação sígnica triádica –, inerente ao processo comunicativo entre um ambiente e um sistema – a entidade viva – que para manter sua

independência parcial de controle ambiental externo, deve ser capaz de interpretar e compreender elementos de seu entorno como cruciais para sua manutenção. Os mecanismos pelos quais esse sistema reconhece e atua sobre aspectos do ambiente importantes à sua sobrevivência são, então, ditos funcionais em referência ao próprio sistema (Rocha 1997).

Assim, a função de cada componente é atribuída a partir da perspectiva de sua contribuição para a estabilidade dinâmica do sistema como um todo – o que é compatível com a análise funcional de Cummins ([1975]1998) discutida no início deste artigo: muitas das capacidades c_1, c_2, \dots, c_n das partes, necessárias para a manifestação de uma capacidade c do todo, são resultados das restrições colocadas pelo próprio todo e não capacidades que estas partes teriam quando independentes do mesmo. As restrições que o todo impõe sobre suas partes são, portanto, criativas num sentido funcional: quando partes previamente independentes se tornam componentes de um sistema maior, elas adquirem novos papéis funcionais e, com isso, todo o sistema recém-criado também tem maior potencial do que seus componentes independentes, não-relacionados (cf. Juarrero 2000, p.40). Entender como isto ocorre demanda o exame de como processos internos serão funcionalmente ajustados às condições ambientais, criados como uma organização dentro do sistema que irá se referir a algo fora do sistema – em prol disto, é essencial o processo de ‘referencialidade’, no qual um signo interno se refere ao mesmo objeto e da mesma maneira que um signo externo se refere (El-Hani et al. 2007, p. 27).

A teoria dos signos de Peirce contém a chave para explicar como esta referência a um mesmo objeto externo (coisa ou processo) é mantida, a saber, a transição interpretante-signo (I-S), na qual o mesmo elemento que numa relação triádica (S-O-I) exerce o papel de interpretante, na relação subsequente exercerá o papel de signo. Este processo estabelece uma *cadeia contínua de signos* que encerra uma referência recursiva ao mesmo aspecto de um dado objeto. Iremos examinar detalhadamente estas idéias mais adiante. Por ora, queremos apenas destacar um dos

pontos de partida da abordagem semiótica que será aqui empreendida, a saber: a idéia de que a referida cadeia de signos consta como relações dos componentes de um dado sistema – que instancia um particular (*token*) de um dado tipo (*type*) de estrutura – as quais, em conjunto, são restringidas pela influência organizacional e regulatória desta estrutura. Recorreremos, portanto, ao argumento de que há uma *relação determinativa lógica*³⁷ de processos particulares no nível inferior por princípios organizacionais do nível superior no qual os primeiros estão imersos. Esta relação inter-níveis pode ser expressa na forma de uma *tendência* ou *disposição* para instanciar um dado processo, atribuída a entidades de nível inferior – e.g., entidades biomoleculares – quando elas estão sob influência de um princípio geral de organização – e.g., um princípio inerente a contexto celular.

3.1. Algumas idéias básicas na teoria dos signos de Peirce

A semiótica ou ‘ciência formal dos signos’, como concebida por Peirce, descreve e analisa a estrutura de processos semióticos independentemente de suas bases materiais, ou das condições sob as quais eles podem ser observados – dentro de células (citossemiose), entre tecidos e populações de células, na comunicação animal (zoossemiose), ou nas atividades tipicamente humanas. Em outras palavras, a conceito peirceano de semiótica diz respeito a uma teoria do signo em que este é tratado em seu sentido mais geral, i.e, em suas características formais, não importando qualquer modo particular de sua produção. Disto segue que, para um signo ser signo, é irrelevante se é produzido como um fenômeno mental, social, ou biológico. Qualquer coisa pode se tornar um signo, tão logo satisfaça às condições formais dos signos como tais (Liszka 1996).

³⁷ Vale ressaltar que em seu sentido lógico, determinação deve ser entendida como restrição, em contraposição ao sentido causal, i.e., ao seu entendimento como um processo causalmente determinístico. O sentido lógico de determinação é empregado por Queiroz e colaboradores (2006), seguindo Hulswit (2001), ao conceberem ‘determinação descendente’ com base na semiótica Peirceana.

Peirce fundamenta formalmente a semiótica em sua teoria das categorias, que constitui um corpo altamente abstrato de idéias e é o centro de sua ‘filosofia arquitetônica’ (Parker 1998, p. 60) e de seu modelo de semiose (ação do signo) (Murphey 1993, p. 303-306). Esta teoria inclui uma lista de três categorias – Primeiridade, Secundidade e Terceiridade – que funciona como uma lista daquilo que não pode ser experimentado, pensado, imaginado, se não por meio de seus elementos (Queiroz 2004). As categorias também podem ser descritas logicamente como um sistema exaustivo de classes de relações organizadas hierarquicamente – relações monádicas, diádicas e triádicas (Houser 1997, p. 14; Brunning 1997).

Em poucas palavras, as categorias podem ser definidas como: (1) Primeiridade ou mônada: aquilo que é tal como é, sem referência a qualquer outra coisa; é a categoria da liberdade, originalidade, possibilidade (CP 1.25; CP 8.328); (2) Secundidade ou relação diádica (díade): aquilo que é tal como é em relação a alguma outra coisa, mas sem relação com qualquer outra terceira entidade; é a categoria da reação, oposição, ação bruta (CP 8.330); (3) Terceiridade ou relação triádica (tríade): aquilo que é tal como é em virtude de sua capacidade de colocar uma segunda coisa em relação com uma primeira da mesma maneira que ele coloca a si próprio em relação com a primeira e a segunda entidade; é a categoria do signo, da mediação, do hábito, generalidade, cognição (CP 1.304).

De acordo com Peirce, qualquer descrição de semiose envolve uma relação irreduzivelmente triádica entre três termos conectados, Signo-Objeto-Interpretante (S-O-I), seus três elementos constitutivos mínimos (MS 318: 81; CP 2.242, CP 2.274). A referência e a relevância atribuída ao intérprete - no qual se instancia o interpretante, que é efeito do signo – é um diferencial da semiótica peirceana em relação aos modelos clássicos diádicos de signo. Para Peirce, o intérprete não é necessariamente um ser consciente – ele concebe o intérprete como uma ‘*Quasi-mind*’ (CP 4.536), uma concepção que acomoda também organismos não-humanos, uma célula individual e

mesmo estruturas moleculares (ver Santaella-Braga 1994). Conforme Colapietro (1989, p.4), um signo ‘não apenas significa alguma coisa, mas significa para alguém – para uma mente’, e como implicação, ‘o signo é um fenômeno mais complexo do que a definição clássica indica’, i.e., não é equivalente ao conceito diádico de representação (significante/significado) na lingüística estrutural.

Peirce concebe um ‘Signo’ como um ‘Primeiro’ que está numa relação genuinamente triádica com um ‘Segundo’, o seu ‘Objeto’, de modo a ser capaz de determinar um ‘Terceiro’, seu ‘Interpretante, a assumir a mesma relação triádica com o seu Objeto na qual ele próprio está com seu Objeto (CP 2.274; CP 2.303; CP 2.92). A relação triádica é irreduzível no sentido de que não é decomponível em nenhuma relação mais simples (monádica ou diádica). Esta relação triádica foi chamada por Peirce de ‘semiose’ e corresponde à ação do signo³⁸.

Peirce também definiu um signo como um meio para a comunicação de uma forma ou um hábito, incorporado no objeto, para o interpretante (De Tienne 2003; Hulswit 2001), de modo a determinar o interpretante como um signo (Figura 5). Nas palavras de Peirce:

‘[...] um signo pode ser definido como um Meio para a comunicação de uma forma. [...] Como um meio, o Signo está essencialmente numa relação triádica, com seu Objeto que o determina, e com seu Interpretante que ele determina. [...] Aquilo que é comunicado do Objeto através do Signo para o Interpretante é uma Forma; isto quer dizer, não é nada como um existente, mas é um poder, é o fato de que alguma coisa aconteceria sob certas condições (MS 793: 1-3; EP 2.544, n.22).

³⁸ Uma das características mais notáveis da semiótica peirceana é a ênfase sobre relações dinâmicas e eventos, ao invés de coisas, conteúdo, essência, substância. O complexo S-O-I pode ser compreendido como o fator-focal de um processo dinâmico (Hausman 1993, p.72). Peirce era um verdadeiro adepto de um pensamento orientado a processos (ver Rescher 1996).

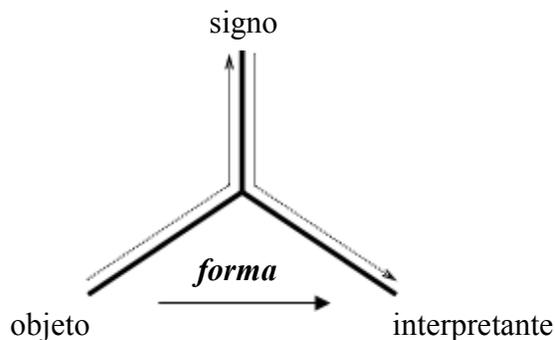


Figura 5. Signo como meio de transmissão de uma forma (uma regularidade ou invariância) do objeto para o interpretante.

Nos trabalhos de Peirce, forma é definida como tendo o ‘ser de predicado’ (EP 2.544) e também é formulada pragmaticamente como uma ‘condição proposicional’ que afirma que certas coisas aconteceriam sob certas circunstâncias (EP 2.388). A forma é algo que está incorporado no objeto como um hábito, uma ‘regra de ação’ (CP 5.397; CP 2.643), uma ‘disposição’ (CP 5.495; CP 2.170), uma potencialidade ou ‘potencial real’ (EP 2.388) ou, simplesmente, uma ‘permanência de alguma relação’ (CP 1.415). Forma tem, assim, tanto o caráter de primeiridade quanto de secundidade.

Vale ressaltar que a forma comunicada do objeto para o interpretante através do signo não é uma configuração particular de um objeto, ou algo parecido, mas uma regularidade, um hábito que permite um dado sistema semiótico interpretar aquela forma como indicadora de uma classe particular de entidades, processos, fenômenos, e, assim, responder a ele numa maneira regular. De outro modo, o sistema não seria realmente capaz de interpretar o objeto por meio de seu efeito nele (interpretante), mediado pelo signo.

A semiose implica necessariamente a instanciação de cadeias de relações triádicas (tríades), visto que um signo numa dada tríade levará à produção de um interpretante, que é, por sua vez, um novo signo. Peirce identifica, portanto, uma continuidade existente na transição da função de signo e de interpretante, a qual implica que nenhum signo existe sozinho, sem estar conectado a

outros signos: cada interpretante é tanto o terceiro termo de uma relação triádica quanto o primeiro termo (signo) da relação triádica subsequente, estabelecendo, assim, uma cadeia contínua de tríades³⁹ (Savan 1986), como ilustra a Figura 6.

Para fins de nosso argumento, é importante ressaltar algumas distinções feitas por Peirce concernentes às subdivisões das naturezas do Objeto. O Objeto Imediato (OI) é o objeto como representado no signo – é o Objeto dinâmico em sua forma semioticamente disponível. O Objeto Dinâmico (OD) é alguma coisa na realidade que determina o signo para sua representação e que, pela natureza das coisas, o Signo não pode expressar, mas apenas indicar e deixar o intérprete encontrar por *experiência colateral* (CP 8.314; EP 2.498), i.e., pela interpretação independente do processo semiótico, em que diversos signos são reconhecidos como sendo signos do mesmo objeto. O sistema que é causalmente afetado pelo signo deve estabelecer qual OD o Signo indica através de processos que foram selecionados na história evolutiva deste tipo de sistema. Na escala de tempo ontogenética, o sistema irá adquirir sua competência semiótica, i.e., sua competência como um intérprete de signo, através do desenvolvimento.

Numa cadeia de signos, quando a transição I-S ocorre, também há uma mudança no ocupante do papel funcional do objeto imediato (Figura 6). Quando o interpretante se torna um signo de outra tríade, a relação de referência ao mesmo objeto dinâmico depende do fato de que o novo ocupante do papel de objeto imediato indica o mesmo aspecto do objeto dinâmico que o objeto imediato da tríade anterior indicou. Via semiose, um objeto é um objeto plural.

³⁹ É importante salientar que há casos em que um processo semiótico termina na produção de interpretantes que não são signos, tais como, por exemplo, ações num sistema vivo, como o disparo de uma dada reação química ou o comportamento de fuga de uma presa em relação a um predador. Certamente, tal reação ou comportamento pode ser um signo para futuras interpretações, mas aquela cadeia de signos particular que estava ocorrendo chega ao fim, de fato, com tal ação e comportamento, e os processos interpretativos posteriores precisam ser modelados como uma nova cadeia de signos. O fato de que há interpretantes que não são signos – i.e., que não têm a natureza de signos – foi reconhecido pelo próprio Peirce na fase mais madura de sua semiótica, após 1907.

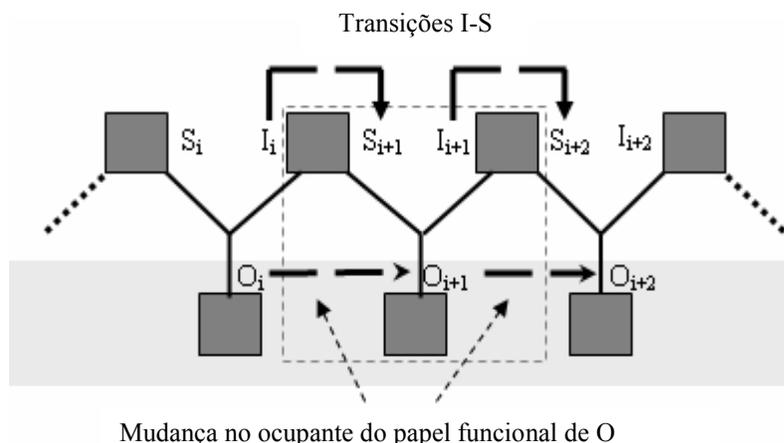


Figura 6 A relação triádica S-O-I forma uma cadeia de triádes. A área cinza ao fundo da figura mostra que todos os signos na cadeia de triádes se referem ao mesmo objeto dinâmico através de uma série de objetos imediatos. As setas mostram a transição interpretante-signo (I-S) e as mudanças no ocupante do papel funcional do objeto imediato (Figura retirada de Queiroz & El-Hani 2006).

Também devemos considerar a distinção entre dois tipos de interpretantes. O Interpretante Imediato (II) é o possível e pertinente efeito imediato do Signo, em sua inteireza primitiva não-analisada, o espectro de interpretabilidade do Signo – é o espectro de possíveis efeitos que um signo é capaz de produzir (ver Johansen 1993, pp. 166-167). Por sua vez, o Interpretante Dinâmico (ID) é o efeito real produzido sobre um intérprete numa dada ocasião e num dado estágio de consideração do signo, i.e., é a instanciação de um dos possíveis efeitos estabelecidos no II (MS 339d- 546-547).⁴⁰ Como o efeito do signo sobre o intérprete, o Interpretante Dinâmico pode ser tratado como sendo essencialmente igual à significância do Signo quando visto numa perspectiva dinâmica e orientada a processos.

Considerando que Peirce (CP 8.177) escreve que um signo determina um interpretante em alguma mente ‘real’ (*actual*) ou ‘potencial’, é possível estabelecer a diferença entre semiose ‘potencial’ e semiose ‘efetiva’ (Queiroz et al. 2005; El-Hani et al., 2006). A semiose potencial é

⁴⁰ Peirce distingue ainda o Interpretante Final (IF), que pode ser entendido, num processo semiótico, como o estado final deste processo, uma tendência sendo realizada quando uma dada cadeia de triádes é disparada, mas não determinada a acontecer, visto que outros estados finais podem seguir do processo semiótico, como no caso, por exemplo, de um erro de interpretação (Queiroz et al. 2008).

definida como um processo organizado triadicamente que não está ocorrendo, que está apenas em potência. A semiose efetiva, por sua vez, é um signo em ação efetiva, i.e., um signo que, por ser realizado (*actualized*), tem um efeito real (*actual*) no intérprete. Seguindo esta distinção, podemos adotar a distinção entre *informação potencial* e *informação efetiva*.

Queiroz e colaboradores (2008) sintetizam a concepção de informação desenvolvida em seus trabalhos prévios, a partir de sua interpretação das idéias de Peirce, que diverge da tendência de se substancializar informação. Segundo eles, informação tem uma natureza processual, é o processo de comunicação de uma *forma* – um hábito, uma regularidade incorporada – do Objeto para o Interpretante através do Signo, que opera como uma influência restritiva sobre os possíveis padrões de comportamento do sistema semiótico (i.e., do intérprete). Nesta perspectiva, informação não pode ser dissociada da noção de um agente situado. De acordo com esta abordagem, informação pode ser fortemente associada ao conceito de semiose⁴¹ e também ao de *significado*, uma vez que este último é definido por Peirce como algo que é comunicado na semiose – o que torna plausível também explicá-lo como estando associado ao interpretante, que, afinal, incorpora e retém uma invariância na forma reconstruída do Objeto.

Se a célula for concebida como um sistema semiótico, a construção de um conceito semântico-pragmático de informação deve levar em conta que o significado biológico é funcional e, portanto, relacionado às propriedades de todo o contexto celular, bem como irreduzível a meras reações bioquímicas. Na modelagem semiótica que serviu de base para esta seção e na qual acomodaremos a análise semiótica da emenda alternativa de RNA, a semiose é descrita como um processo sistêmico emergente: significados surgem, na ação dos signos, como uma propriedade encontrada somente no nível do sistema como um todo, e não no nível de suas partes (ver Queiroz e El-Hani 2006). Um esboço desta abordagem será fornecido na subseção seguinte.

⁴¹ Informação potencial e informação efetiva, portanto, estão associadas às noções de semiose potencial e efetiva, respectivamente.

3.2. *Níveis de semiose*

A concepção da semiose como processo dinâmico emergente está intrinsecamente conectada ao estabelecimento de uma cadeia de tríades, a qual foi descrita na seção anterior. A cadeia de tríades é concebida, na abordagem de semiose como processo emergente (ver Queiroz e El-Hani 2006), como o *nível semiótico focal*. A delimitação desse nível de análise é sugerida no contexto do sistema triádico básico de Stanley Salthe (1985), em que este autor enfatiza a necessidade de se descrever as interações fundamentais de uma dada entidade ou processo em três níveis: (i) o nível focal, em que as entidades ou processos são observados realmente; (ii) o nível inferior na hierarquia – em geral, mas não necessariamente, o nível imediatamente mais abaixo –, em que são investigadas as relações entre suas partes; (iii) e o nível superior na hierarquia – também, em geral, o mais imediatamente acima – no qual o nível focal está imerso. O que nos permite explicar a emergência de processos, e.g., o processo semiótico, no nível focal são restrições colocadas tanto pelo nível inferior, quanto pelo nível superior. No nível inferior, as condições restritivas concernem às possibilidades ou condições iniciais para o processo emergente, enquanto que as restrições no nível superior estão relacionadas a um papel de um ambiente seletivo exibido pelas entidades nesse nível, estabelecendo condições de contorno que coordenam ou regulam a dinâmica no nível focal.

Conforme este modelo, o *nível micro-semiótico* se refere às condições iniciais para o processo emergente ou possibilidades de relações (S-O-I) de determinação entre signos, objetos e interpretantes potenciais; este nível é o domínio dos Signos, Objetos e Interpretantes potenciais.⁴² A semiose do nível focal corresponde a uma dada cadeia de tríades; e o *nível macro-semiótico* se refere a uma rede de cadeia de tríades que influenciam umas às outras, a qual corresponde ao que

⁴² Um Signo potencial é algo que pode ser o signo de um objeto para um interpretante e que se torna um signo apenas quando submetido à relação mediadora de determinação entre Objeto e Interpretante. Por sua vez, um Objeto potencial é algo que pode ser o Objeto de um signo para um interpretante. Por fim, um Interpretante potencial é algo que pode ser o interpretante de um signo.

Peirce chama de experiência colateral, noção que nos remete a idéia de sistema. A semiose do nível focal emerge como um processo neste nível através da interação entre processos micro- e macro-semióticos, i.e., entre as relações de determinação dentro de cada tríade e a inserção de cada cadeia individual numa rede seletiva e regulatória de processos sígnicos. Por conseguinte, embora a semiose seja instanciada no nível focal, ela pode ser descrita como um processo sistêmico, já que a realização de tríades potenciais depende de condições de contorno estabelecidas pelo nível macro-semiótico.

Como uma propriedade sistêmica, a semiose é irreduzível no sentido de que ela depende de comportamento específico que componentes exibem num sistema de um dado tipo, e este comportamento, por sua vez, não segue do comportamento dos sistemas quando isolados ou quando em outros tipos de sistemas (mais simples) (Queiroz e El-Hani 2006).⁴³ Neste ponto, fica claro como uma abordagem semiótica desta natureza vem a contribuir para o modelo de análise funcional de Cummins ([1975]1998): quaisquer capacidades que componentes de um sistema semiótico exibem quando isolados em um sistema mais simples não serão suficientes para se deduzir certas capacidades, como a capacidade semiótica – i.e., de interpretação de signos – do sistema; isto ocorre porque muitos papéis funcionais, como os de S, O ou I, que são necessários à manifestação de uma capacidade sistêmica só se realizam nas partes através das restrições colocadas pelo sistema.

3.3. Análise semiótica de processos de realização gênica: em cena, a emenda de RNA

Um esboço da análise semiótica de sistemas genéticos de informação proposta por El-Hani e colaboradores (2006) é fornecido na Figura 7. De baixo para cima, são ilustradas três cadeias de

⁴³ Para uma discussão mais aprofundada sobre como a semiose pode ser concebida como processo emergente, recomendamos a consulta do trabalho original (Queiroz e El-Hani 2006). Para as discussões sobre como a irreduzibilidade pela não-dedutibilidade do comportamento das partes do sistema se relaciona à noção de causação descendente (DC) – ou ‘determinação descendente’, como reelaborada por estes autores – ver El-Hani e Queiroz (2007).

tríades (sentido horizontal) que correspondem aos processos semióticos envolvidos na realização de signos potenciais no DNA, respectivamente: (i) a transcrição de RNAs a partir de seqüências de DNA; (ii) a tradução de mRNA em seqüências de aminoácidos; e a (iii) síntese da aminoacil-tRNA.

Na cadeia da transcrição, cada tríade corresponde à semiose que realiza como Signo uma trinca de nucleotídeos de DNA (descrita como um *Signo Simples* em relação à seqüência codificante completa que será transcrita, entendida como *Signo Composto*).⁴⁴ Este Signo Simples ($^N S$) é relacionado triadicamente ao Objeto Imediato Simples ($^N O_i$) – um códon no transcrito primário de RNA⁴⁵ – e ao Interpretante Imediato ($^N I_i$), que é o espectro de interpretabilidade estabelecido pelas regras de pareamento de bases pelas quais nucleotídeos específicos no DNA determinam nucleotídeos específicos no RNA. A *enzima RNA polimerase é o sistema interpretativo* que estabelece uma cadeia de tríades ao se mover ao longo das séries de Signos Simples – i.e., trincas de nucleotídeos de DNA.

De modo similar, na cadeia da tradução na figura 3, Signos Simples ($^T S$) – os mesmos códons de mRNA que na cadeia anterior eram Objetos Imediatos Simples – são relacionados triadicamente a seus Objetos Imediatos Simples ($^T O_i$) – que são tRNAs com anticódonos específicos – e a seus Interpretantes Imediatos ($^T I_i$) na tradução – i.e., o espectro de interpretabilidade estabelecido pelas regras de pareamento de bases pelas quais nucleotídeos específicos em mRNA são pareados com nucleotídeos específicos em tRNA.

O *ribossomo é o sistema interpretativo* que realiza cada tríade da cadeia ao se mover ao longo de cada códon do mRNA. Por fim, na síntese de aminoacil-tRNA, etapa também necessária à produção do polipeptídeo, cada tRNA disponível na célula para o processo de tradução é um

⁴⁴ As noções de Signos e Objetos Simples e Compostos foram derivadas para adequar a aplicação do arcabouço semiótico em sistemas biológicos. Para maiores detalhes desta análise, consultar os originais (ver Queiroz et al., 2005; El-Hani et al. 2006).

⁴⁵ Este é parte do Objeto Imediato Composto que resulta da transcrição, i.e., a molécula de RNA transcrita.

Signo Simples ($^H S_i$), reconhecido pelo *sistema interpretativo que consiste na enzima aminoacil-tRNA* sintetase, a qual também reconhece um aminoácido específico – o Objeto Imediato ($^H O_i$) – ao qual o tRNA é relacionado triadicamente, bem como ao seu Interpretante Imediato ($^H I_i$), que são as regras do código genético.

Deste quadro que acabamos de descrever, podemos concluir que a passagem da potencialidade para a realização de um ‘gene’⁴⁶ resulta na produção de um Objeto Imediato composto – a seqüência de aminoácidos traduzida – a partir da realização seqüenciada de pareamentos específicos – Interpretantes Dinâmicos Simples – dentre os possíveis pareamentos estabelecidos pelos Interpretantes Imediatos, em cada ação de Signo Simples. O Objeto Dinâmico é a proteína funcional que a célula terá que encontrar por *experiência colateral*, após outros processos como seu dobramento tridimensional e alteração química de polipeptídeos. No entanto, o OD está disponível semioticamente na estrutura primária protéica, i.e, na seqüência de aminoácidos que é o Objeto Imediato Composto. Este processo semiótico integrado, que depende de sistemas interpretativos subordinados às condições de contorno colocadas pelas redes de cadeias de tríades que compõem o contexto celular, é concebido como o *processo de informação genética* que transforma um signo potencial no DNA em um signo efetivo, real (*actual*). Neste processo, a forma de um Objeto Dinâmico – certo tipo de proteína que tem sido favorável na história evolutiva de um organismo – é comunicada ao Signo – uma seqüência de DNA – que é determinado (no sentido de ser restringido) por esse objeto.

⁴⁶ Mais precisamente, de uma unidade de transcrição no DNA.

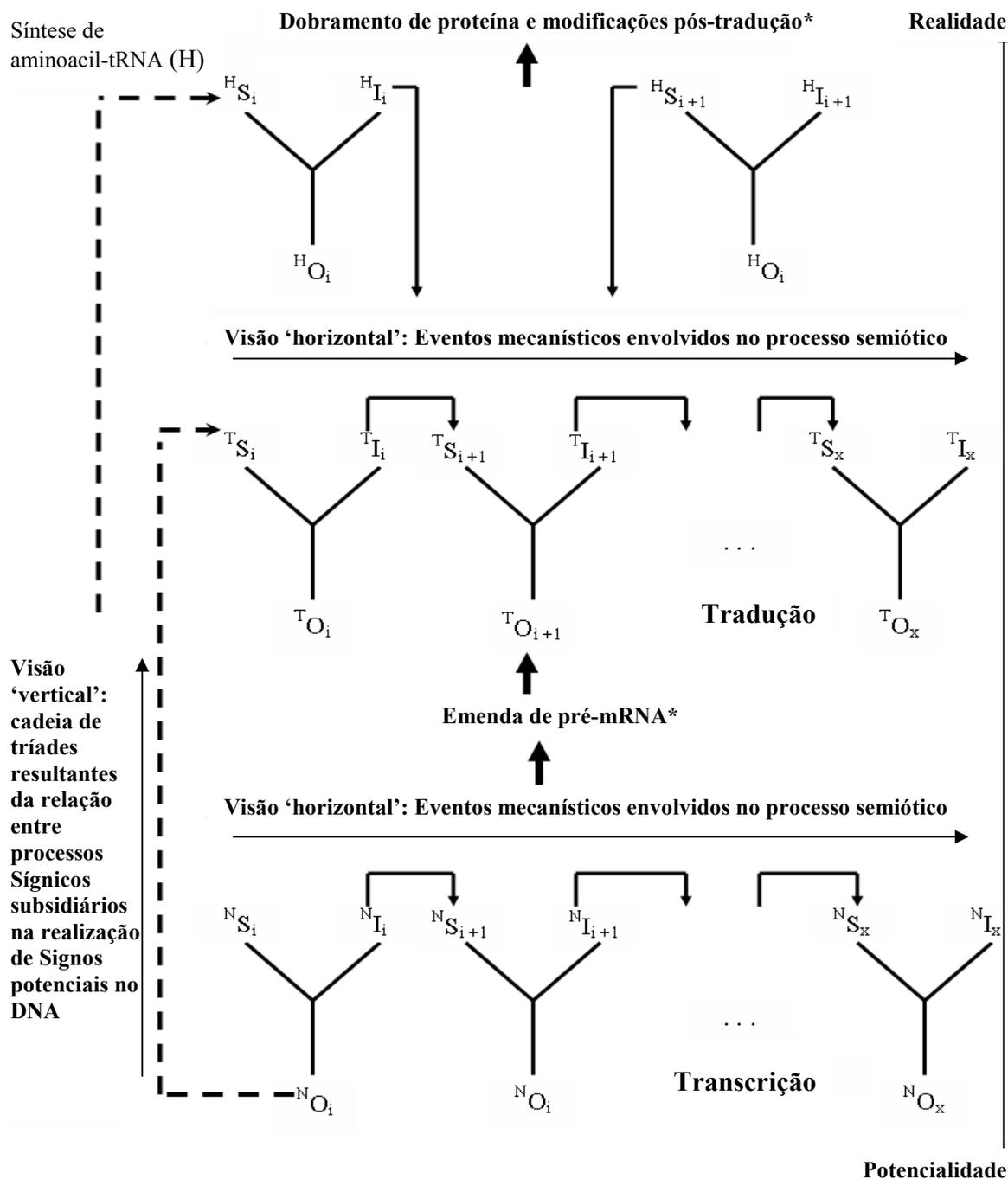


Figura 7. Visão geral dos processos semióticos envolvidos na realização de signos potenciais no DNA. A letra 'N' é usada para identificar elementos da tríade no nível da transcrição. $^N S$: Signos Simples no DNA (trincas de nucleotídeos); $^N O$: Objetos Imediatos Simples no pré-mRNA (códon); $^N I$: Interpretantes Imediatos na transcrição, o espectro de interpretabilidade estabelecido pelas regras de pareamento de bases pelos quais nucleotídeos específicos no DNA determinam nucleotídeos específicos no mRNA. As setas representam o movimento do subsistema interpretativo, RNA polimerase, para o próximo Signo Simples, quando uma tríade é realizada. Uma seta vertical tracejada à esquerda indica que códon no mRNA, Objetos Imediatos Simples da transcrição, se tornam Signos potenciais na tradução. A letra 'T' identifica os elementos da tríade no nível da tradução. $^T S$: Signos Simples no mRNA (códon); $^T O$: Objetos Imediatos Simples (tRNAs particulares com anticódon específicos); $^T I$: Interpretantes Imediatos na tradução, o espectro de interpretabilidade estabelecido pelas regras de pareamento pelas quais nucleotídeos específicos no mRNA são pareados com nucleotídeos no tRNA. As setas horizontais indicam o movimento dos ribossomos para o próximo Signo Simples, quando uma tríade é realizada na tradução. Na parte superior da figura, a letra 'H' identifica os elementos de uma tríade no nível da síntese de aminoacil-tRNA. $^H S$: Signos: estrutura tridimensional de tRNAs; $^H O$: Objetos Imediatos Simples: aminoácidos específicos; $^H I$: Interpretante Imediato, o espectro de interpretabilidade de cada códon como um Signo Simples, estabelecido pelas regras do código genético. Outra seta tracejada vertical à esquerda indica que o mesmo elemento, um tRNA com anticódon que une um códon no mRNA desempenha papéis funcionais diferentes de Objeto Imediatos simples e de Signo Simples em diferentes tríades.

Pela soma das cadeias de tríades individuais, nós obtemos uma visão global da informação genética como um processo semiótico envolvendo: uma unidade de transcrição no DNA, como um Signo Composto; a seqüência linear de aminoácidos numa proteína, como o Objeto Imediato Composto do Signo; e o espectro de interpretabilidade das possibilidades de reconstrução das sequencias de aminoácidos a partir do Signo no DNA, como o Interpretante Imediato do Signo Composto.

Podemos observar que a investigação do presente trabalho visa aplicar o mesmo modelo sobre um dos processos ainda não analisados, a saber, a emenda alternativa de RNA – posicionada, na Figura 7, entre a transcrição e a tradução. A partir da análise esboçada acima, poderíamos chegar à conclusão de que os Signos potenciais no DNA são genes, entendidos como as unidades de transcrição. Uma vez que sua capacidade de representar um dado RNA ou polipeptídeo depende do contexto celular, i.e., das condições de contorno colocadas pelo nível macro-semiótico, estes genes só podem ser considerados como recursos desenvolvimentais fornecedores de moldes moleculares – o gene-D de Moss (2001, 2003). Como as peculiaridades da emenda alternativa podem ser acomodadas nesta abordagem semiótica e quais as conseqüências desta análise para os conceitos de gene e informação? Antes de respondermos a esta questão, será preciso inserir o modelo semiótico do processo básico da emenda, que independe de esta ser constitutiva ou alternativa.

Da discussão da Figura 7 mais acima segue que a introdução de uma análise semiótica da emenda de RNA implicará, primeiramente, num pequeno ajuste do modelo exposto: Objetos Imediatos resultantes da transcrição passam a ser signos conectados à cadeia de tríades da emenda e não da tradução, como sugere as setas interrompidas na Figura 7, que ligam as cadeias da transcrição e da tradução. Vale ressaltar que *o spliceossomo, que é o sistema interpretativo da emenda de RNA*, reconhece Signos com um tamanho intermediário entre os Signos Simples

(códon no pré-mRNA) e o Signo Composto (toda a molécula de pré-mRNA), na medida em que define éxons. Tendo em conta que éxons podem codificar inteiramente, ou em grande parte, domínios funcionais de uma proteína (Downes 2004), poderíamos chamar um éxon, assim, de *Signo Subcomposto* e chamar a forma semioticamente disponível do domínio protéico funcional que ele codifica, de *Objeto Imediato Subcomposto*, que se encontrará no mRNA maduro – Objeto Imediato Composto – resultante da emenda (Figura 8).

Após a definição de éxons, ocorre um rearranjo no spliceossomo em que os sítios de emenda que definem um íntron são conectados em definitivo e esta transferência é necessária porque apenas pares intrônicos de sítios de emenda têm um sentido funcional de fato e podem dar prosseguimento à reação catalítica (Berget 1995; Black 2000, 2003). Assim, a definição de éxon constitui não uma via alternativa de emenda, mas uma rota indireta para a efetivação da reação de emenda, i.e., para a definição do íntron. Em vista disso, em termos de nossa modelagem semiótica da emenda de RNA, o reconhecimento de Signos Subcompostos parece, a nosso ver, mais propriamente um ‘degrau’ para o reconhecimento dos íntrons, também Signos subcompostos, mas que têm um Objeto Dinâmico diferente. Se a seqüência de um íntron não deve fazer parte do mRNA maduro apto à tradução – i.e., o Objeto Imediato Composto no processo de emenda –, segue que não há, neste último, um Objeto Imediato Subcomposto representado pelo Signo Subcomposto intrônico e, sequer, um correspondente na seqüência de aminoácidos que indicará, após a tradução, um domínio funcional no Objeto Dinâmico, i.e., no produto protéico final.

Por conseguinte, um íntron é um Signo Subcomposto no pré-mRNA que não está realmente relacionado com a disponibilidade semiótica da proteína funcional indicada por um gene, ou melhor, por uma unidade de transcrição no DNA. Neste caso, o Interpretante Dinâmico é a realização de uma regularidade de excisão de seqüências contendo códon de terminação

prematureos (CTP). O Objeto Dinâmico, por sua vez, é a instrução de que um código de leitura aberto (*open reading frame*) deve ser produzido, i.e., o processo posterior de tradução não deve ser interrompido; este OD é tornado semioticamente disponível pelo fato de que uma estrutura intrônica em laço – um dos subprodutos da catálise da emenda (ver subseção 3.2) – é o Objeto Imediato do Signo Subcomposto Intrônico em questão. Códon de terminação (UAA, UAG e UGA) estão envolvidos na terminação do processo semiótico de tradução. Nenhum deles é reconhecido por um tRNA mas por proteínas chamadas ‘fatores de liberação’, que atuam no sítio ribossômico ocupado, no caso de outros códon, por um aminoacil-tRNA. O OD da tradução é a instrução de que o processo deve ser interrompido e este OD é tornado semioticamente disponível pelo fato de que uma molécula de água, em vez de um aminoácido, é o Objeto Imediato do Signo simples em questão.

Podemos inferir, portanto, que dois Interpretantes Dinâmicos são aludidos na emenda de RNA: o primeiro, concerne à emenda de éxons no mRNA e consiste na instanciação da regularidade de que Signos Subcompostos que não contêm CTP deverão ser emendados; o segundo concerne à excisão de íntrons e é subordinado, no fim das contas, ao ID que é a realização de uma regra do código genético de que UAA, UAG e UGA interrompem o processo de tradução.

No que tange à emenda alternativa de RNA, um primeiro ponto que salientamos e que também é válido para a emenda constitutiva versa sobre o papel dos fatores regulatórios positivos e negativos que se ligam aos elementos regulatórios intrônicos e exônicos. Este parece ser um processo essencialmente mecanicista, dependente das flutuações do gradiente desses fatores no nucleoplasma, que interagem por livre difusão (Maniatis e Reed 2002).

Sabe-se que o acoplamento da emenda alternativa à transcrição e a outros eventos de processamento de RNA, como a poliadenilação, e também à exportação do RNA para o

citoplasma, aumenta a probabilidade de interação entre os fatores de emenda e contribui efetivamente para acelerar a cinética da interação. Isso representa uma grande vantagem no reconhecimento específico dos sítios de emenda e na competição com outras proteínas RNA-ligantes não-específicas, que só têm acesso ao pré-mRNA por difusão (ibid.). Apesar de ser comum falar de reconhecimento nesses casos não nos parece adequado considerar fatores de regulação de emenda como subsistemas interpretativos da célula. Adicionalmente, os referidos acoplamentos são tão mecanicistas quanto o movimento da RNA polimerase de uma tríade para outra, bem como do spliceossomo funcional sobre cada tríade da emenda. O processo semiótico é estabelecido na direção vertical ilustrada na Figura 7, não na horizontal.

Uma questão colocada pela emenda alternativa concerne ao grau de diferença entre as variantes protéicas (isoformas) que podem ser produzidas a partir de um mesmo transcrito primário. Isoformas podem ter funções biológicas completamente distintas se a emenda de um éxon alternativo implicar a presença ou não de um dado domínio funcional. Este caso sugere que cada isoforma protéica é um Objeto Dinâmico distinto, cujo respectivo Objeto Imediato – seqüência de aminoácidos que resulta do processo de informação genética – é obtido num padrão de emenda alternativa específico. O Signo composto potencial no DNA pode ser considerado uma sobreposição de vários genes potenciais, cada um sendo atualizado ao findar de um dado padrão de emenda. Nesta etapa do processo de informação genética, pode-se dizer que a célula ‘geneia’, i.e., atualiza genes numa seqüência madura de mRNA – em outras palavras, genes seriam os Objetos Imediatos resultantes da informação. Alternativamente, poderíamos considerar que o ‘genear’ corresponde ao próprio processo de informação genética, nos aproximando da proposta de Neumann-Held (2001) do gene molecular processual.

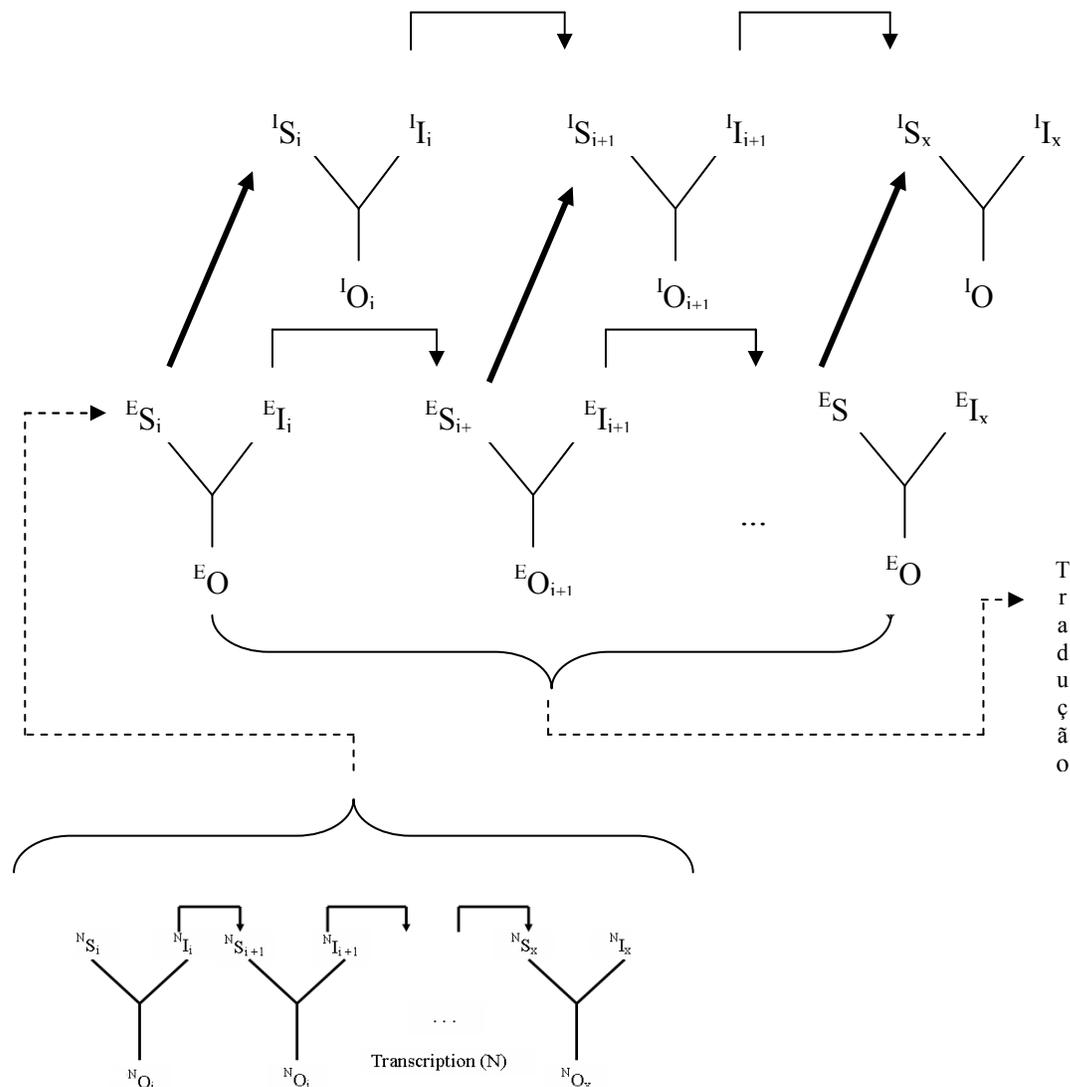


Figura 8. Integração do processo de emenda de RNA na realização de signos potenciais no DNA. O processo semiótico de transcrição é mostrado na parte inferior da figura. O Objeto Imediato Composto, pré-mRNA, resultante deste processo é um Signo potencial para o processo semiótico de emenda de RNA. Neste, Signos Subcompostos Exônicos (${}^E S$) – éxons – são constituídos pelo que, no processo anterior de transcrição, corresponde a uma dada seqüência de Objetos Imediatos Simples (indicada pela seta tracejada à esquerda). Note-se que o sistema interpretativo da emenda, o spliceossomo, não reconhece Signos Simples (códon) no pré-mRNA. O reconhecimento de ${}^E S$ pelo spliceossomo corresponde à definição de éxons, um degrau (setas diagonais espessas) para a definição de íntrons, a qual permite a reação catalítica de excisão e emenda. ${}^E O$: Objeto Imediato Subcomposto Exônico, éxon presente no mRNA emendado. ${}^E I$: Interpretante Imediato da emenda de éxons, o espectro de interpretabilidade estabelecido pela regularidade de que Signos Subcompostos que não contêm códon de terminação prematuros (CTP) deverão ser emendados. O ${}^E O$ desta primeira etapa da emenda é a forma semioticamente disponível de um domínio funcional na proteína, i.e., no Objeto Dinâmico do processo de realização gênica que deverá ser encontrado por experiência colateral da célula. Na transferência dos sítios que definem um éxon para os que definem um íntron, o sistema interpretativo spliceossômico reconhece um Signo Subcomposto Intrônico (${}^I S$). ${}^I I$: Interpretante Imediato Intrônico, o espectro de interpretabilidade estabelecido pela regularidade de que Signos Subcompostos que contêm CTP deverão ser excisados. ${}^I O$: Objeto Imediato Subcomposto Intrônico, uma molécula intrônica em forma de laço como subproduto da catálise da emenda. Apesar de as setas horizontais indicarem a continuidade de emenda de novos éxons e excisão de novos íntrons, vale ressaltar que o spliceossomo não reconhece esses elementos necessariamente em seqüência. O resultado do processo de emenda é um Objeto Imediato Composto, mRNA, que se torna um Signo potencial para a tradução.

Contudo, a descrição de informação como um processo semiótico é suficientemente geral para ser aplicada a outros processos no sistema biológico, e.g., as vias de sinalização que regulam padrões específicos de expressão gênica. Assim, para evitar ambigüidades, parece mais apropriado considerar o gene como o transcrito de mRNA obtido após a emenda alternativa.

Também é possível prescindirmos da manutenção da relação unívoca entre gene e proteína, concebendo o primeiro como a própria unidade de transcrição interrompida por íntrons, i.e., como o Signo composto potencial no DNA que, assim, funcionaria como um recurso desenvolvimental com capacidade de fornecer moldes moleculares – o gene-D *sensu* Moss (2001, 2003). Neste caso, é possível falar de uma relação ‘um gene-várias isoformas protéicas’. Note-se, contudo, que em ambos os casos, a seqüência de DNA é incapaz de especificar por si mesma uma isoforma protéica, i.e., especificar o Objeto Imediato que indica um Objeto Dinâmico específico.

3.4. Posição da emenda alternativa na escala do estocástico ao altamente regulado

A emenda alternativa é chamada de ‘variação de emenda’ por McGuire e colaboradores (2008, p. 2), que reservam a primeira denominação apenas para a variação de emenda que é *regulada e funcionalmente significativa*. Se, por um lado, esses autores querem salientar a distinção entre as variações resultantes de erros e os eventos programados e funcionais de emenda alternativa⁴⁷, por outro, eles concluem que ambos são igualmente informativos no que concerne à maneira como os sítios de emenda são reconhecidos: e.g., a despeito de o evento de emenda ter sido funcional ou não, o seu padrão específico será provavelmente influenciado pelos mecanismos utilizados para identificar e processar os sítios de emenda.

Em vista disso, a descrição biomolecular parece bastar para a compreensão de como os sítios de emenda são reconhecidos – se ao longo do íntron, ou se ao longo do éxon –, visto que esta

⁴⁷ Vale ressaltar que a programação implicada aí inclui várias causas, como mudanças ontogenéticas e estímulos ambientais e, portanto, está subordinada a níveis de organização epigenéticos – no mínimo, celular –, tornando injustificada à alegada programação inerente à molécula de DNA.

‘decisão’ não depende de regulação sistêmica, mas antes de restrições colocadas pelo tamanho dos íntrons e éxons. Outra sugestão de que os mecanismos de emenda estão intimamente ligados à estrutura gênica é o fato de serem os éxons constitutivos – os quais são sempre emendados – e os íntrons constitutivos – os quais são sempre eliminados – facilmente reconhecidos, i.e., mais evidentes do que suas contrapartes alternativas, visto que contêm sítios de emenda mais fortes e maior comprimento em geral (Cáceres & Kornblihtt 2002). Há, neste fato, um ensejo para asserções de que os sítios de emenda adjacentes a éxons e íntrons alternativos constituem *sinais fracos*, com *menor conteúdo de informação* (cf. McGuire et al. 2008, p.11 – itálico nosso), bem como um ensejo para fortalecer uma abordagem reducionista, de descrição de estruturas e mecanismos biomoleculares.

Todavia, podemos clarificar o quão disputável são estas visões ao ponderarmos que o grau de conservação das seqüências consenso⁴⁸, no qual, em grande parte, reside a força dos sítios de emenda, contribui para explicar a emenda constitutiva, porém, é estéril à explicação da emenda alternativa – se éxons e íntrons menos evidentes podem conter igualmente pouca ‘informação’ nas seqüências que os delimitam – as quais são seqüências consenso ditas degeneradas, não constituindo, em geral, informação suficiente para determinar se um sítio irá reunir um spliceossomo e funcionar na emenda (cf. Black 2003) –, o que explica eles serem eventualmente reconhecidos para emenda e excisão e, sobretudo, os seus padrões de emenda diferencial?

A existência de elementos regulatórios no RNA que realcem os sítios de emenda de éxons e íntrons alternativos é necessária, mas vale notar que ela é insuficiente para definir quando e onde esses sítios serão realçados. A discriminação dos pares de sítios de emenda apropriados num dado contexto é alcançada a partir da ligação de múltiplas proteínas aos elementos regulatórios no RNA, juntamente com interações proteína-proteína. É a formação de um complexo específico

⁴⁸ Seqüência consenso é uma seqüência idealizada na qual cada posição representa a base mais freqüentemente encontrada quando várias seqüências reais são comparadas (Lewin 2004).

RNA-multiproteínas que permite o reconhecimento das junções éxon-íntron ou de sítios de emenda crípticos, mais fracos. As abordagens tradicionais que visam o entendimento da regulação da emenda alternativa usam modelos de casos individuais simples⁴⁹, que se baseiam na identificação desses elementos regulatórios *cis*-atuantes e fatores *trans*-atuantes⁵⁰ que em conjunto impõem a regulação de padrões de emenda, e.g., dependentes do genótipo sexual (Matlin et al. 2005).

Matlin e colaboradores (ibid.) destacam, contudo, que a identificação de proteínas cito-específicas – cuja presença desencadeia a regulação de padrões de emenda dependentes do tipo celular, estágio desenvolvimental e via de sinalização – é apenas uma das estratégias de regulação existentes: na maioria dos casos de emenda alternativa em *Drosophila melanogaster*, mamíferos e vírus, são combinações precisas e taxas relativas de fatores atuantes positiva e negativamente, que ditam as decisões regulatórias. Em outras palavras, esse tipo de emenda não é determinado simplesmente pela presença de um conjunto de fatores moleculares regulatórios, mas por ‘códigos celulares’ estabelecidos por *combinações específicas* destes fatores (ibid., p.387; *itálicos nossos*). Revelar detalhes dos códigos que regulam casos mais extremos de emenda alternativa certamente demandará a complementação das abordagens tradicionais por outras mais globais e sistêmicas, que identifique elementos e fatores regulatórios, detecte alvos em RNAs para reguladores da emenda, mas que simultaneamente detecte um grande número de eventos concomitantes de emenda alternativa e de cascatas de sinalização a eles acoplados.

⁴⁹ Entre os casos mais estudados, está a emenda alternativa dos genes *Sex-lethal (Sxl)* e *transformer (tra)*, que participam na determinação do sexo de *D. melanogaster*, e envolve a produção de um único transcrito por gene: apenas nas fêmeas, cada gene resultará numa proteína funcional (ver Lopez 1998).

⁵⁰ ‘*Cis*’ e ‘*trans*’ são termos usados na genética para identificar elementos localizados na mesma molécula e em moléculas diferentes, respectivamente. Sítios no DNA ou RNA (e.g., promotores, terminadores, realçadores, silenciadores) são *cis*-atuantes quando afetam apenas as atividades de seqüências em sua própria molécula, o que geralmente implica que não codificam, mas são alvos de reconhecimento por elementos *trans*-atuantes. Estes, por sua vez, são produtos gênicos (moléculas de RNA ou proteínas) difusíveis que agem sobre qualquer outra molécula.

Um adendo crucial, a nosso ver, consiste em se explicar como a confluência desses processos biológicos ocorre de maneira completamente organizada a ponto de estabelecer códigos celulares. Podemos estimar o quão difícil pode ser esta tarefa, conjecturando sobre como esses códigos celulares de fatores regulatórios mediam entre o reconhecido papel potencial da diversidade de *Dscams* de *Drosophila* na conectividade neuronal e a existência de um código de reconhecimento celular que permite que neurônios se distingam um do outro (Hattori et al. 2007). No cérebro em desenvolvimento, são necessários receptores de membrana neuronais capazes de detectar sinais difusíveis ou ancorados na matriz celular, que direcionam o crescimento de axônios para as células com as quais deverão fazer conexões sinápticas. Neste processo, a mistura de múltiplas células diferentes num milieu ‘tipo *spaghetti*’ torna útil um mecanismo que possibilite a segregação apropriada dos dendritos e axônios provenientes do mesmo neurônio – ditos ‘irmãos’ (Zipursky et al. 2006, p.581; Shapiro 2007). A diversidade de *Dscam* pode suprir esta função porque cada neurônio expressa um conjunto único de isoformas e também porque a ligação homofílica isoforma-específica entre ramos irmãos promove auto-evitação (*self-avoidance*) – i.e., a ligação preferencial entre moléculas idênticas media o reconhecimento e a conseqüente repulsão entre dendritos e axônios do mesmo neurônio, mas não entre os de diferentes neurônios⁵¹ (Hattori et al. 2007). Neste último caso, a ligação é heterofílica, na qual cada célula expressa diferentes moléculas que se ligam umas às outras, gerando enfeixamentos e sinapses. A produção de isoformas com propriedades de reconhecimento distintas é vantajosa para quaisquer organismos com sistema nervoso altamente complexo, caracterizados pelo pequeno número de genes, uma vez que é bastante improvável que cada evento de reconhecimento entre neurônios

⁵¹ Mais precisamente, a identidade real das isoformas de *Dscam* que são expressas é irrelevante, desde que cada neurônio expresse um conjunto diferente de isoformas que os neurônios vizinhos (Olson et al. 2007). Esta é uma condição crítica para satisfazer o modelo de que a diversidade de *Dscam* promove auto-reconhecimento (*self-recognition*) e auto-evitação (*self-avoidance*) de extremidades neuronais irmãs (cf. Hattori et al. 2007, p.225) – em outras palavras, dispensável dentro de um único neurônio, mas essencial dentro de uma população de neurônios, a diversidade de *Dscam* garante a cada um deles uma superfície celular de identidade única (ibid. p. 226).

seja mediado pelo produto de um gene diferente (Zipursky et al. 2006); antes, são gradientes de moléculas de membrana e diferentes combinações de um número limitado de proteínas de reconhecimento que especificam padrões de circuitos neuronais (ibid., p.582).⁵²

A caracterização mecanicista dos sistemas vivos fundamentada em fluxos de matéria e energia é capaz de explicar o que Hulswit (2001) caracteriza como ‘eventos Peirceanos’ (p. 344), os quais são simples, são cegos e brutos – acontecem simplesmente – e não têm qualquer ordem interna ou coerência; em contraposição, o ‘processo Peirceano’ é complexo, uma seqüência contínua de eventos com uma tendência definida, a qual tanto define a unidade, coerência ou ordem interna de um processo – distinguindo-o de outros –, quanto estabelece as chamadas condições de contorno que determinam, no sentido de restringir, a produção de eventos particulares. Para entendermos como a rede organizada de mecanismos, intra- e intercelulares, de sinalização molecular influencia a dinâmica da expressão diferencial do material genético, supomos que cada padrão de expressão genética e cada cascata individual de sinalização não devem ser concebidos meramente como eventos em si mesmos, como uma compulsão determinada pela condição particular das coisas e prosseguindo de maneira perfeitamente determinada (cf. CP 1.212); antes, devem ser considerados como restringidos, determinados pelas condições de contorno colocadas pela referida rede de sinalização dentro de e entre células particulares.

⁵² A emenda alternativa adiciona uma nova dimensão, sobretudo, à complexidade dos vertebrados, em grande medida delegada aos sistemas nervoso e imune, que são baseados em mecanismos intrincadamente conectados e capazes de gerar variação a partir de poucos genes, conforme ressaltam Szathmáry, Jordán e Pál (2001). O aumento na especificidade de reconhecimento celular em vertebrados parece ter evoluído a partir da duplicação e divergência de éxons ou de genes de loci como os de receptores neuronais relacionados à caderina, via expressão graduada de proteínas de reconhecimento (e.g. efrinas e receptores ephs), ou pela associação combinatória de diferentes sistemas de reconhecimento; por sua vez, o aumento da complexidade de circuitos em insetos parece ter emergido, em parte, da diversificação intensa do locus *Dscam* (Zipursky et al. 2006). Este, além de codificar receptores de membrana com importante papel na conexão neuronal, é necessário ao reconhecimento de patógenos no sistema imune de insetos, como mostram estudos recentes. Sabe-se que, pelo menos no grupo de éxon 4, a emenda alternativa de transcritos de *Dscam* é regulada no desenvolvimento e de maneira tecido-específica. Visto que tal regulação é conservada evolutivamente e, sobretudo, porque humanos e moscas compartilham a maioria dos genes envolvidos na regulação da emenda alternativa, princípios aplicáveis à regulação deste processo em humanos podem ser revelados por seu estudo em *Drosophila* – que constitui um sistema excelente para aplicação de tecnologias genômicas, bioquímicas, de cultura de tecidos e de genética de ponta (Graveley 2005).

4. Considerações Finais

Em suma, argumentamos acima que a compreensão dos códigos celulares demanda que os vejamos como intrínsecos às entidades – e.g., *redes* de eventos biomoleculares – e aos processos – que conferem direção às redes de eventos – constituintes de contextos celulares e supra-celulares específicos. Nesta conjuntura, um mesmo código celular pode ser instanciado por diferentes componentes biomoleculares, visto que o primordial na produção de eventos é o caráter geral do resultado, que condiz com uma tendência definida num dado contexto. Assim, segue naturalmente que a produção de inventários de biomoléculas e de reações bioquímicas é insuficiente para capturarmos na devida conta a complexidade dos organismos, pois sempre haverá soluções múltiplas que se ajustam a um caráter geral. É interessante, por exemplo, que a regulação da emenda alternativa no desenvolvimento adquire robustez – pela *redundância nas funções essenciais na emenda* – devido ao fato de ser necessário apenas um membro diferente das famílias de proteínas regulatórias para interagir com um elemento de RNA específico e também devido a esses membros poderem ser intercambiáveis na habilidade de atender a tal requerimento (Black 2003).

Conforme Blencowe (2006), esforços estão sendo agora direcionados para estabelecer o repertório completo tanto das variantes relevantes de transcritos geradas por emenda alternativa, quanto dos papéis específicos de tais variantes na fisiologia normal e patológica, e da maneira como a emenda alternativa é coordenada num nível global para alcançar funções cito- e tecido-específicas. Estes procedimentos levam em conta que, ao invés do simples cômputo do número de genes ou do número de interações entre genes, uma medida mais sensível da complexidade genômica e que se correlacionaria melhor à complexidade biológica em termos de morfologia e comportamento reside na conectividade das redes de regulação (Szathmáry et al. 2001). O inventário dos fatores de transcrição, dos genes que eles regulam e do transcriptoma gerado – i.e.,

do conjunto completo de transcritos de RNA –, bem como o inventário dos elementos e fatores regulatórios envolvidos no processamento de RNA, como os da emenda alternativa, poderiam fornecer um esboço dessa conectividade.

Todavia, a atuação altamente combinatória dos fatores múltiplos positivos e negativos – que variam de forma sutil entre os diferentes tecidos e estágios desenvolvimentais – e dos elementos regulatórios de RNA – bastante similares nos diversos tecidos –, influenciando a quantidade final de um transcrito emendado (Black 2000), suscita algumas suspeitas. Afirmações de que ‘avanços rápidos na biologia de rede indicam que redes celulares são governadas por leis universais’ (Barabási & Oltvai, 2004, p. 101) nos instigam a questionar se o que está em jogo nestas abordagens mais globais consta, de fato, como uma posição não-reducionista e não se trata apenas de uma abordagem reducionista em larga escala. Parece intrigante o argumento pela necessidade de catalogar todas as moléculas e interações, ao invés de se recorrer a abordagens mais sistêmicas que revelariam de forma mais eficiente e barata como as redes celulares funcionam (ibid.).

De acordo com Strohman (2002, p. 701), apesar da complexidade da relação genótipo-fenótipo não ser novidade para os biólogos moleculares, estes são incapazes de explicá-la, principalmente devido um importante elemento está sendo negligenciado: a noção de Polanyi de ‘condições de contorno’, que na biologia de sistemas são conceituadas como níveis de restrições ou restrições de controle. Estes são encontrados em níveis de regulação, muitos dos quais acima do nível do genoma. Conforme argumentamos, uma abordagem semiótica de processos biológicos pode iluminar como a regulação pode ser tratada como uma influência de cima a baixo de um sistema sobre seus componentes, que pode ser, por sua vez, concebida em termos de condições de contorno para a operação de entidades e processos de nível inferior (Salthe 1985). A

dependência que a função gênica tem de um contexto exemplifica de forma notável o papel da regulação nos sistemas vivos.

Para satisfazer o desejo pragmático de que genes sejam objetos componíveis de um banco de dados, possibilitando aos biólogos, por exemplo, agruparem diferentes trabalhos sobre o mesmo gene – e possibilitando, adicionalmente, que os genes sejam objetos patenteáveis para uso econômico (Falk 2001; Griffiths 2002) –, cabe não só falar de um *locus* da *Dscam* – ou de qualquer outro *locus* –, como é também possível investigar, nos cromossomos da drosófila, uma seqüência que assim possa ser denominada. Há regras claras, critérios explícitos para se chamar de gene uma dada região do genoma. Um deles, por exemplo, são os claros limites estabelecidos pelas unidades de transcrição. Embora haja sobreposição entre as mesmas, o que é reconhecidamente um aspecto da complexidade estrutural e dinâmica dos sistemas genéticos, diferentes contextos e propósitos de investigação podem delimitar e enfatizar apenas uma unidade de transcrição dentre as múltiplas que podem estar sobrepostas. O desafio trazido pela emenda alternativa de RNA, que possibilita a produção de inúmeros transcritos a partir de uma mesma região do genoma, pode ser acomodado se levarmos em conta que estes transcritos podem estar relacionados funcionalmente – i.e., ter funções similares. Assim, por que não ‘mapear’ uma população de RNAs a uma região genômica e chamar esta última de gene?

De fato, isto tem sido feito nas próprias pesquisas da emenda alternativa e nas pesquisas em que o critério da unidade de transcrição é empregado como ferramenta para estudo de função e evolução do genoma. O gene identificado como unidade de transcrição mantém as propriedades de estabilidade e permanência que tradicionalmente lhe são atribuídas, bem como a característica de unidade estrutural. Todavia, não apenas será difícil concebê-lo como unidade de função, como também é bastante plausível que qualquer abordagem funcional do gene na biologia molecular do desenvolvimento demande a expansão do foco no DNA para o foco no sistema desenvolvimental

mais amplo no qual ele está imerso (Griffiths 2002). Tal análise funcional e sistêmica também demandará que tenhamos em conta o quanto da configuração flexível e dinâmica do genoma é preciso se abstrair em contextos mais pragmáticos de investigação.

Constitui um bom começo entender que uma unidade de transcrição no DNA não pode, por si só, ser determinante de características e que ela cumpre apenas o importante papel de ser um recurso desenvolvimental fornecedor de moldes moleculares para a produção de vários produtos gênicos, com funções diversas. Este é precisamente o entendimento que fundamenta o gene-D ('D', de desenvolvimental), *sensu* Moss (2001, 2003), i.e., um gene que é definido por sua seqüência molecular, mas indeterminado em relação a qualquer resultado fenotípico, em virtude, em grande parte, da inquestionável relevância de outros fatores causais biomoleculares.

A partir da análise semiótica do processo de emenda e da discussão sobre a regulação do mecanismo da emenda alternativa podemos colher elementos que fundamentam nosso argumento de que uma análise semiótica desse processo revela a plausibilidade de que a célula não contém genes, *a priori*, mas os constrói dentro de contextos específicos nos aproximamos da sugestão de Keller (2005). Quiçá, estimar os possíveis frutos de uma biologia molecular do gene nesta nova conjuntura permanecerá como uma tarefa árdua, porém, o motivo não mais será a descrença num futuro mais brilhante do que a extinção do conceito de gene (ver Portin 1993; Gelbart 1998; Keller 2002) ou a resignação com a plethora de significados no uso corrente deste termo de maneira altamente nebulosa e ambígua (cf. Kitcher 1982). O desafio será sobrepujar a forte tendência da metafísica ocidental, que remonta a época dos filósofos pré-socráticos, a favor de 'entidades', de estabilidades fixadas, ao invés de processos, de modos de mudança e adotar a perspectiva dos críticos da metafísica de substâncias, os quais incluem vários pensadores como Whitehead, Peirce, Weiss, Alexander, Lloyd Morgan, entre outros (ver Rescher, 1996). Esses pensadores defendem uma filosofia de processos, em que processos são considerados mais

fundamentais ontologicamente do que entidades e na qual se enfatiza a contingência, a emergência, a novidade e a criatividade. Como salienta El-Hani (2007), a partir dessa perspectiva, o problema de individualizar genes processuais seria visto como algo natural, uma vez que processos são, peculiarmente, mais difíceis de se individualizar por serem dependentes de contexto. E acrescenta que, ‘se chegarmos à conclusão de que genes são mais propriamente conceituados como processos, nós não devemos nos esquivar da tarefa de individualizá-los, ainda que seja uma tarefa enorme’ (ibid., p. 304).

Quanto à integração da abordagem semiótica à análise funcional, é importante destacar que uma célula, como intérprete global, usa signos no DNA como base para sintetizar um Objeto Dinâmico que lembra suficientemente um Objeto Dinâmico passado que não mais existe, mas que resultou em experiências adaptativas bem-sucedidas. Note-se que, ainda assim, não há um ensejo, aqui, para atribuímos função às seqüências de DNA para explicar por que estas seqüências existem (conforme uma abordagem neo-teleológica). De fato, uma abordagem sistêmica como a de Cummins ([1975]1998) tem como limitação a não-distinção entre função e acidente (Nunes-Neto e El-Hani 2008) necessária a uma abordagem evolutiva como esta implicada na longa história da relação entre certos tipos de Signos e Objetos. Se algum erro neste processo de realização de signos potenciais no DNA favorecesse uma dada capacidade do sistema celular, ele poderia ser considerado funcional, segundo Cummins. A diferença entre se considerar ou não um acidente como um item funcional revela, no fim das contas, que a perspectiva etiológica selecionista preserva o pensamento teleológico, ao passo que a abordagem da análise funcional desenvolvida por Cummins se mantém, de acordo com ele mesmo, distante de compromissos teleológicos. Nunes-Neto e El-Hani (2008) ressaltam que há um uso legítimo do conceito de função nas abordagens etiológicas selecionistas, assim como há um uso legítimo do conceito de função no âmbito da análise funcional; a aceitação desta tese, segundo eles, é tão

mais facilitada quanto mais for compreendida a importância e legitimidade do consenso dualista de Godfrey-Smith (1993), o qual pode ser apoiado por uma distinção entre biologia funcional e evolutiva.

Dessa forma, o uso de ambas as abordagens sobre funcional seria frutífero desde que delimitados os domínios de aplicação. Esta pode ser uma sugestão para trabalhos futuros que podem relacionar a adoção deste consenso dualista com uma diferenciação dos tipos de signos, como definidos por Peirce. Uma abordagem evolutiva iria requerer uma análise da relação entre Signo e Objeto estabelecida como um hábito natural evoluído (o que fundamenta a classificação peirceana de símbolos). Já uma análise funcional sistêmica demandaria uma ênfase maior no aspecto de ‘aqui e agora’ inerente à relação entre Signo e Objeto, i.e., demandaria o foco em como a contigüidade espaço-temporal entre estes dois componentes semióticos propicia a estabilidade atual do sistema em que se inserem, o intérprete.

Enfim, analisamos acima os indícios de que focar apenas em moléculas pode obscurecer um tipo de ação no mundo biológico caracterizado por oferecer restrições e não por realizar ou produzir um efeito – mais precisamente, referimo-nos às restrições colocadas pelos diferentes níveis hierárquicos de organização daquele em que um dado fenômeno é observado. Entender como um poder restritivo favorece padrões de soluções múltiplas pode ser a chave para a compreensão da elevada plasticidade evolutiva dos sistemas vivos. De um lado, o custo desta plasticidade é a redução no número de maneiras nas quais as partes podem ser arranjadas, visto que elas precisam ‘estar a serviço’ do todo (ver Juarrero 2000). De outro lado, qualquer forma concebível de se ‘estar a serviço’ deste todo é bem-vinda e em vez de estruturas particulares e *estabilidades* de interações físico-químicas, o mais crucial é a estabilidade em relação a prováveis perturbações – a qual pode ser descrita apropriadamente pelo termo ‘robustez’ (Keller 2007, p.9). Um sistema auto-regulado e robusto alcança novas maneiras de persistir – novas e múltiplas

soluções – através da combinação dos papéis funcionais de suas partes numa organização modular e maleável, que constrói redes cada vez mais complexas e estáveis (persistentes), mais bem explicadas em termos das *relações dinâmicas entre eventos*. Assim, processos – definidos por Rescher (1996, p. 38) como ‘uma família coordenada de ocorrências que são ligadas sistematicamente umas às outras seja causalmente ou funcionalmente’ – adquirem uma primazia ontológica em relação a entidades e a estabilidades em geral.

A nosso ver, uma abordagem semiótica Peirceana dos sistemas biológicos acomoda mais apropriadamente, num pensamento orientado a processos, idéias correntes que se mostram decisivas para a compreensão da complexidade dos organismos, como robustez, modularidade, redes complexas e emergência. Adicionalmente, a partir deste arcabouço teórico – que vem a propósito para a descrição dos tipos de *entidades inerentemente relacionais* que podem emergir desta dinâmica – poderá ser construído também um *novo léxico* que dê conta desta interatividade, sendo estes os requisitos apontados por Keller para seguirmos um caminho mais promissor do que a tentativa de

‘[...] edificar uma biologia a partir de substantivos, uma ciência construída ao redor de entidades’ (Keller 2005, p.107).

E com ela estamos de acordo de que

‘Talvez seja o momento para uma biologia edificada a partir de verbos, uma ciência construída ao redor de processos’(ibid.)

Referências

- Barabási, Albert-László & Oltvai, Zoltán N. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* 5:101-113.
- Bateson, Gregory. 1972. Steps to an Ecology of Mind. New York: Ballantine.
- Beurton, P. (2000). A unified view of the gene, or how to overcome reductionism. In: P. Beurton, R. Falk, & H.-J. Rheinberger (Eds). *The concept of the gene in development and evolution:*

historical and epistemological perspectives (pp. 286–314). Cambridge: Cambridge University Press.

- Black, D.L. 2000. Protein Diversity from Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology. *Cell*, 103:367-370.
- Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, vol. 72, pp. 291-336, 2003.
- Blencowe, B.J. Alternative splicing: new insights from global analyses. *Cell* 126, 37-47 (2006).
- Bruggeman, Frank J.; Westerhoff, Hans V. & Boogerd, Fred C. (2002). Biocomplexity: A Pluralist Research Strategy is Necessary for a Mechanistic Explanation of the “Live” State”. *Philosophical Psychology* 15 (4), 411-440.
- Bruning, Jackeline (1997) Genuine triads and teridentity. In *Studies in the logic of Charles Sanders Peirce*. (eds.) Nathan Houser, D. Roberts and J. Evra, 252-270. Bloomington and Indianapolis: Indiana University Press.
- Boue, S.; Letunic, I. & Bork, P. 2003. Alternative Splicing and Evolution. *BioEssays*, 25: 1031-1034.
- Cáceres, J.F. & Kornblihtt, A.R. 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends in Genetics*, 18 (4):186-193.
- Carroll SB, Grenier JK and Weatherbee SD. 2005. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell, Oxford, 258 pp.
- Colapietro, Vincent (1989). *Peirce’s Approach to the Self: A Semiotic Perspective on Human Subjectivity*. State University of New York Press.
- Cummins, R. [1975]1998. Functional Analysis. In: Allen, C.; Bekoff, M.; Lauder, G. (orgs). *Nature’s Purposes – Analyses of Function and Design in Biology*. Cambridge, MA: MIT Press, pp.169-196.
- Cummins, R. 2002. Neo-teleology. In: Ariew, A.; Cummins, R.; Robert, P.; Perlman, M. (eds.) *Functions: new essays in philosophy of psychology and biology*. Oxford: Oxford University Press, pp. 157-172.
- Disponível em <https://netfiles.uiuc.edu/rcummins/www/HomePage/Cummins.html>>. Acesso em: 16 mar. 2006.
- Downes, S.M. Alternative splicing, the gene concept, and evolution. 2004. *History and Philosophy of the Life Sciences*, 26: 91-104.
- El-Hani, C. N.; Queiroz, J. & Emmeche, C. 2006. A semiotic analysis of the genetic information system. *Semiotica* 160(1/4): 1-68.
- El-Hani, C. N.; Arnellos, A. & Queiroz J. 2007. Modeling a semiotic process in the immune system: signal transduction in B-cell activation. *TripleC – Cognition, Communication, Cooperation* 5(2): 24-36.
- Emmeche, C. 1999. The Sarkar challenge: is there any information in a cell? *Semiotica*, 127: 273–93.
- Fetzer, James H. 1988. *Signs and Minds: An Introduction to the Theory of Semiotic Systems*. In: Fetzer, James (Ed.) *Aspects of Artificial Intelligence*. The Netherlands. Dordrecht. pp 133–161.

- Graveley, B.R. 2005. Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell*, 123 (1): 65-73.
- Graveley, B., Kaur, A., Gunning, D., Zipursky, L., Rowen, L. & Clemens, J.C. 2004. The organization and evolution of the Dipteran and Hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA*, 10:1499-1506.
- Griffiths, P.E. & Neumann-Held, E. 1999. The many faces of the gene. *BioScience*, 49 (8): 656-662.
- Hattori, D.; Demir, E.; Kim, HW et al. 2007. Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* 449: 223-227.
- Hausman, Carl. 1993. Charles Sanders Peirce's Evolutionary Philosophy. Cambridge University Press.
- Hirschman JE, Balakrishnan R, Christie KR, et al. 2006. Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. *Nucleic Acids Res.* (34 Database):D442–D445.
- Hoffmeyer, J.. 2001 Life and Reference. In: *Biosystems*, 60 (1-3): 123-130.
- Houser, Nathan (1997). Introduction: Peirce as a logician. In *Studies in the Logic of Charles Sanders Peirce*. Nathan Houser, D. Roberts & J. Evra (eds.), 1-22. Bloomington and Indianapolis: Indiana University Press.
- Hulswit, M. (2001). Semeiotic and the Cement of the Universe: a Peircean Process Approach to Causation, *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, Summer, XXXVII (3): 339-363.
- Ideker, Trey; Timothy Galitski and Leroy Hood. 2001. *A new approach to decoding life: systems biology*. In: Annual Review of Genomics and Human Genetics, 2: 343-372.
- Jablonka, Eva. 2002. Information: Its interpretation, its inheritance, and its sharing. *Philosophy of Science* 69: 578-605.
- Juarrero, A. 2000. Dynamics in Action: Intentional Behavior as a Complex System. *Emergence*, 2(2): 24-57.
- Keller, Evelyn Fox. 2005. The century beyond the gene. *Journal of Bioscience*, 30 (1): 101-108.
- Keller, Evelyn Fox. 2007. The disappearance of function from 'self-organizing systems'. In: F.C.Boogerd, Bruggeman, F.J., Hofmeyr, J-H.S. and Westerhoff, H.V (eds.). Amsterdam: Elsevier, 304-327.
- Küppers, B.O. 1990. *Information and the Origin of Life*. Cambridge, Mass: MIT Press.
- Kwan, T. et al. 2007. Heritability of alternative splicing in the human genome. *Genome Res.* 17: 1210-1218.
- Jeffares DC, Mourier T, Penny D. 2006. The biology of intron gain and loss. *Trends Genet.* 22:16–22.
- Johansen, Jorgen Dines (1993). *Dialogic Semiosis*. Bloomington and Indianapolis: Indiana University Press.
- Lander, E.S. et al. (International Human Genome Sequencing Consortium). Initial sequencing and analysis of the human genome. *Nature*, vol. 409 (6822), pp. 860-921, 2001.

- Lewin, B. 2004. *Genes VIII*. Upper Saddle River-NJ: Pearson Education.
- Liszka, J. 1996. *A General Introduction to the Semeiotic of Charles Sanders Peirce*. Bloomington, In: Indiana University Press.
- Lopez, A.J. Alternative Splicing of Pre-mRNA: Developmental Consequences and Mechanisms of Regulation. *Annu. Rev. Genet.*, vol. 32, pp. 279-305, 1998.
- Magen, A. & Ast, G. 2005. The importance of being divisible by three in alternative splicing. *Nucleic Acids Research*, vol. 33 (17), pp. 5574-5582.
- Matlin, A.J.; Clark, F. & Smith, C.W.J. 2005. Understanding Alternative Splicing: Towards a Cellular Code. *Nature*, 6: 386-398.
- Mattick J. 2004. RNA regulation: a new genetics? *Nature Reviews Genetics*, 5: 316–323.
- McGuire, A.M., Pearson, M.D., Neafsey, D.E., Galagan, J.E. 2008. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology*, 9:R50.
Disponível em: <http://genomebiology.com/2008/9/3/R50>. Acesso em 25 de Junho de 2008.
- Modrek, B. & Lee, C.J. 2002. A genomic view of alternative splicing. *Nat Genet*, 30:13–19.
- Murphey, Murray (1993). *The Development of Peirce's Philosophy*. Cambridge: Harvard University Press.
- Neumann-Held, Eva. 2001. Let's talk about genes: The process molecular gene concept and its context. In: *Cycles of Contingency: Developmental Systems and Evolution*, Susan Oyama, Paul E. Griffiths and Russell D. Gray (eds.), pp. 69-84. Cambridge-MA: MIT Press.
- Nunes-Neto, N.F. 2008. Bases epistemológicas para um modelo funcional em Gaia. Dissertação. Universidade Federal da Bahia. Salvador, Bahia.
- Olson, Sara, Blachette, J.P., Savva, Y. Yeo, G.W, Yeakley, J.M, Rio, D.C, Graveley, B.R. 2007. A regulator of Dscam mutually exclusive splicing fidelity. *Nature Structural & Molecular Biology*, 14:1134-1140.
- Parker, Kelly (1998). *The Continuity of Peirce's Thought*. Nashville:Vanderbilt University Press.
- Peirce, Charles Sanders (1931–1935). *The Collected Papers of Charles Sanders Peirce*. Electronic edition reproducing Vols. I–VI [C. Hartshorne & P. Weiss (Eds.), Cambridge-MA: Harvard University Press, 1931–1935], Vols. VII–VIII [A. W. Burks (Ed.), same publisher, 1958]. Charlottesville: Intalex Corporation. (Aqui referido como CP, seguido do volume e número do parágrafo).
- Peirce, Charles Sanders. (1967). *Annotated Catalogue of the papers of Charles S. Peirce*. Amherst-MS: University of Massachusetts. ROBIN, R. (Ed.) [References to manuscripts and letters by Charles S. Peirce — MS and L — are in accordance with this catalogue.]
- Peirce, Charles Sanders. ([1893-1913]1998). *The Essential Peirce: Selected Philosophical Writings*. Vol. II. (Ed.) Peirce Edition Project. Bloomington and Indianapolis: Indiana University Press. (Here referred as EP2, followed by the number of the page.)
- Queiroz, João . 2004. Semiose segundo C.S.Peirce. São Paulo: EDUC - Editora da Puc-SP,. v. 500. 207 p
- Queiroz, J. El-Hani, C.N. 2006. Towards a multi-level approach to the emergence of meaning processes in living systems. *Acta Biotheoretica*, 54(3): 174:-206.

- Queiroz, J.; Emmeche, C. & El-Hani, C. N. 2005. Information and semioses in living systems: A biosemiotic approach. *S.E.E.D. Journal: Semiotics, Evolution, Energy, and Development* 5 (1): 60-90.
- Queiroz, J.; Emmeche, C. & El-Hani, C. N. 2008. A Peircean approach to information and its relationship with Bateson's and Jablonka's Ideas. *American Journal of Semiotics*,.
- Queiroz, J.; Emmeche, C. & El-Hani, C. N.. 2006. Semiosis as an Emergent Process. Transactions of the Charles Sanders Peirce Society, Buffalo, New York, v. 42, n. 1, p. 78-116.
- Rescher, Nicholas (1996). *Process Metaphysics: An Introduction to Process Philosophy*. New York: State University of New York Press.
- Rheinberger, Hans-Jörg & Müller-Wille, Staffan. 2008. Gene Concepts. In: Sahotra Sarkar and Anya Plutynsky (eds). *A Companion to the Philosophy of Biology*, Oxford: Blackwell, pp. 3-21.
- Rocha, L.M. 1997. *Evolutionary Systems and Artificial life*. Lecture Notes. Disponível em http://informatics.indiana.edu/rocha/ss504_02.html. Acesso em 23 de Julho de 2008.
- Rosenberg, A. & McShea, D. 2008. *Philosophy of Biology – a contemporary introduction*. New York: Routledge.
- Roy SW & Gilbert W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet*.7:211–221.
- Salthe, Stanley N. 1985. *Evolving Hierarchical Systems: Their Structure and Representation*. New York: Columbia University Press.
- Santaella-Braga, Lucia (1994). Peirce's broad concept of mind. *European Journal for Semiotic Studies* 6 (3/4), 399-411.
- Savan, David (1986). Response to T.L.Short. *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, Summer, XXII (2), 125-143.
- Schellenberg, MJ; Ritchie, DB; MacMillan, AM. 2008. Pre-mRNA splicing: a complex picture in higher definition. *Trends Biochem Sci.*, 33(6): 243-6.
- Schmucker, D. et al. 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101: 671–684.
- Shapiro, L 2007. Self-Recognition at the Atomic Level: Understanding the Astonishing Molecular Diversity of Homophilic Dscams. *Neuron*, 56 (1):10 - 13
- Simon H. 1996. Disappearance of function from 'self-organizing systems.' 317. In: *The Sciences of the Artificial*, 3rd edition. Cambridge, MA: MIT Press.
- Stamm, S.; Ben-Ari, S.; Rafalska, I. et al. Function of alternative splicing. *Gene*, vol. 344, pp. 1 – 20, 2005.
- Staley, J.P. & Guthrie, C. 1998. Mechanical Devices of the Spliceosome: Motors, Clocks, Springs and Things. *Cell*, 92: 315-326.
- Strohman, R. 2002. Maneuvering in the complex path from genotype to phenotype. *Science*, 296: 701-703.

- Szathmáry, E.; Jordan, F. & Pál, C. Can genes explain biological complexity? *Science*, vol. 292 (5520), pp. 1315-1316, 2001.
- Thanaraj, T.A.; Stamm, S.; Clark, F. et al. ASD: the alternative splicing database. *Nucleic Acids Res.*, vol. 32, pp. D64–D69, 2004.
- Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., Darnell, R.B., 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302: 1212– 1215.
- Venter, J.C. et al. 2001. 1994. The sequence of the human genome. *Science*, 291 (5507): 1304-1351.
- Williams, Nigel. 1997. Biologists cut reductionist approach down to size. *Science*, 277 (5325), 476-477.
- Zipursky SL, Wojtowicz WM, Hattori D. 2006. Got diversity? Wiring the fly brain with Dscam. *Trends Biochem Sci.*, 31(10):581-8.

EXPLICAÇÕES SOBRE GENES EM LIVROS DIDÁTICOS
DE BIOLOGIA DO ENSINO MÉDIO
PUBLICADOS NO BRASIL

1. Introdução

Entre agosto de 2005 e fevereiro de 2006, foi realizada uma avaliação de livros didáticos de biologia do ensino médio publicados no Brasil, dentro do Programa Nacional do Livro para o Ensino Médio (PNLEM), do Ministério da Educação. Nesta avaliação, os livros foram analisados quanto aos seus aspectos conceituais, pedagógico-metodológicos, relativos à construção do conhecimento científico e relativos à construção da cidadania (para detalhes sobre a metodologia empregada e os resultados gerais da análise, ver El-Hani, Roque & Rocha, 2007a,b). Ao longo desta análise, foi possível perceber tanto avanços quanto problemas persistentes na abordagem dos conteúdos em várias áreas da biologia, incluindo a genética e a biologia celular e molecular (El-Hani et al 2007). Neste artigo, abordaremos em maiores detalhes um dos problemas conceituais encontrados pela equipe de biologia do PNLEM, relativo ao tratamento do conceito de gene, com o intuito de trazer contribuições para uma abordagem do assunto no ensino médio que seja mais rica e atualizada, bem como epistemologicamente mais bem fundamentada. É importante destacar desde logo que, embora o problema conceitual abordado neste artigo tenha sido identificado na avaliação do PNLEM, os procedimentos e os dados aqui discutidos não fazem parte daquela avaliação, representando, pois, resultados independentes. Investigamos, contudo, a mesma amostra de 18 livros didáticos analisados pela equipe de biologia do PNLEM.

Com base nos resultados do PNLEM, tomamos como focos de investigação o ‘conceito molecular clássico’ e a ‘concepção informacional’ de gene, visto que aqueles resultados mostraram seu predomínio nos livros didáticos de biologia do ensino médio analisados. Além disso, entre os desafios ao gene molecular clássico, concentramos nossa atenção no processo de *splicing* alternativo, uma vez que este é um conteúdo já abordado em alguns livros de ensino médio, que dá vez a discussões importantes e férteis sobre as possibilidades de explicação do conceito de gene no contexto contemporâneo.

Nas últimas três décadas, uma série de fenômenos colocou desafios importantes para o chamado ‘conceito molecular clássico’ de gene, assim nomeado por Griffiths e Neumann-Held (1999) e Stotz e colaboradores (2004) (ver tb. El-Hani, 2005, 2007). De acordo com este conceito, um gene é um segmento de DNA que codifica um produto funcional (polipeptídeo ou RNA). Entre os desafios, encontramos a existência de genes interrompidos, o *splicing* alternativo de RNA, os transposons, os genes superpostos e nidados, a edição de RNA mensageiro (RNAm) etc. (ver, por exemplo, Fogle 1990, 2000; Falk 1986, 2000; Pardini e Guimarães 1992; Griffiths e Neumann-Held 1999; Keller 2000; Moss 2001, 2003; El-Hani et al. 2006; El-Hani 2005, 2007; Neumann-Held e Sutter 2006). De acordo com El-Hani (2005, 2007), os problemas com este conceito de gene podem ser sumariados com base em três características das relações entre DNA e RNAs/polipeptídeos estabelecidas pela genética e biologia molecular: (1) situações em que um único segmento de DNA corresponde a muitos RNAs/polipeptídeos, como no caso do *splicing* alternativo de RNA; (2) situações em que muitos segmentos de DNA correspondem a um RNA/polipeptídeo, como no caso dos rearranjos genômicos no sistema imune, envolvidos na geração da diversidade de receptores de antígenos de linfócitos B e T; (3) ausência de

correspondência entre segmentos de DNA e RNAs/polipeptídeos, como no caso da edição de RNAm.

Como um exemplo dos desafios ao conceito molecular clássico, examinaremos os genes interrompidos e uma de suas conseqüências, o *splicing* alternativo. Genomas eucarióticos possuem genes interrompidos, nos quais seqüências codificantes – éxons – são intercaladas com seqüências não-codificantes – íntrons –, o que requer um processo de retirada dos íntrons do transcrito primário e a subsequente emenda de éxons num RNA mensageiro maduro, para que a síntese protéica seja possível. Esse processo, por si só, é suficiente para desafiar a relação de um para um entre a unidade de transcrição e um produto funcional, RNA ou polipeptídeo, conforme sustentada no conceito molecular clássico. Mas a situação se agrava ainda mais com a combinação diferencial de éxons, que possibilita que múltiplas proteínas relacionadas (isoformas) sejam formadas a partir de um único gene (Black 2003; Stamm et al 2005), a depender, por exemplo, do genótipo sexual, do estado de desenvolvimento/diferenciação celular, da idade da célula e/ou da ativação de uma via particular de sinalização celular. Pode-se dizer, assim, que fenômenos como a existência de genes interrompidos e o *splicing* alternativo tornam difícil sustentar a idéia geral de genes como unidades (sejam estruturais, funcionais ou informacionais), que é parte essencial do conceito molecular clássico (El-Hani 2005; Fogle 1990; Solha 2005; Solha e Waizbort 2007; Santos 2005).

A emenda alternativa requer que as concepções de gene não se limitem ao simples esquema capturado em fórmulas como ‘um gene-uma proteína ou polipeptídeo’ (Falk 1986; Fogle 1990). E nem mesmo a reformulação para o esquema ‘um gene-várias proteínas ou polipeptídeos’ poderia acomodar suficientemente o desafio da emenda alternativa ao conceito molecular clássico. Visto que padrões de emenda são determinados pelo contexto epigenético da célula, o gene perderia boa parte do seu aspecto de especificidade, em relação aos polipeptídeos que seriam produzidos. Segundo Keller (2000), a situação é tal que não se pode esclarecer dessa maneira onde afinal estaria o gene. O gene estaria naquele segmento de DNA que pode gerar múltiplas isoformas protéicas? Ou estaria em cada RNAm emendado, adotando-se a fórmula ‘um RNA maduro-uma proteína’? Nesse segundo caso, também seria necessário acomodar o fato de que o próprio RNAm pode ser modificado (edição de RNA) e de que o transcrito final pode ser emendado a partir de éxons derivados de diferentes transcritos primários (*trans-splicing*). Se a intenção for manter uma relação de um para um entre gene e produto gênico, o RNAm também não será, assim, o local mais adequado para situar-se o gene. Adicionalmente, caso demarcados em RNAs, os genes existiriam no zigoto recém-formado apenas como possibilidades (Keller 2000), e, desse modo, não teriam a permanência e a estabilidade tipicamente atribuídas aos genes. Tais genes demarcados no RNA talvez não fossem encontrados nos cromossomos e, muitas vezes, sequer no núcleo, visto que a versão final do transcrito pode ser concluída apenas no citoplasma. É claro que poderíamos eliminar do conceito de gene as idéias de estabilidade e permanência, ou de existência nos cromossomos ou no núcleo. A questão, então, seria a de se o conceito assim redefinido poderia cumprir os papéis explicativos que lhe cabem na genética e biologia celular e molecular. Apenas para ilustrar as dificuldades que seguiriam de tal redefinição, como poderíamos pensar nos genes como entidades que mediam a herança de características de uma geração a outra se eles não possuíssem suficiente estabilidade e permanência, e existissem no zigoto apenas como possibilidades? Torna-se aparente que a tentativa de manter o conceito de gene como unidade poderia terminar, assim, tendo um preço alto demais, afastando nossa compreensão dos genes dos requisitos explicativos que nos fizeram construir tal conceito em primeiro lugar (El-Hani, 2005, 2007).

Neste artigo, nosso interesse recai, em particular, sobre a possibilidade de introduzir no ensino médio, mediante recontextualização didática, a explicação de fenômenos que desafiam o

conceito molecular clássico, bem como os próprios debates sobre o conceito de gene. Estes debates têm implicações importantes para o ensino de biologia, tanto no nível superior, em que livros didáticos têm um papel crucial no treinamento de cientistas e de professores de ciências em formação inicial, quanto no ensino médio, em que livros didáticos desempenham papel central nas práticas pedagógicas, influenciando tanto o planejamento quanto a efetiva condução das aulas (Gericke e Hagberg 2007b; Xavier et al 2006). Neste nível de ensino, eles são, frequentemente, a única fonte impressa disponível para professores e estudantes. Desse modo, é fundamental compreender os conhecimentos, os valores e as práticas que tais materiais veiculam e, assim, introduzem nas salas de aula.

Não pretendemos defender, contudo, que livros didáticos do ensino médio devam oferecer visões altamente atualizadas dos desenvolvimentos na genética e na biologia celular e molecular. Apenas propomos que o tratamento dos sistemas genéticos em tais livros deve aproximar-se mais da complexidade destes sistemas, a fim de evitar interpretações equivocadas por parte dos estudantes, que possam levar a equívocos persistentes em seu entendimento de idéias fundamentais para a construção de sua compreensão do conhecimento científico, bem como para o exercício de sua cidadania. Entre estes equívocos, encontramos, por exemplo, as visões deterministas sobre as relações entre genes e características fenotípicas. Estas interpretações equivocadas são fontes potencialmente importantes de simplificações excessivas das relações genótipo-fenótipo e da estrutura e dinâmica dos sistemas genômicos, que encontramos nos discursos sobre genes em diferentes esferas de nossa sociedade. Afinal, o ensino médio constitui, para a maioria das pessoas, o último momento de educação formal acerca de genes e de seu papel nos sistemas vivos.

Como ponto de partida, investigamos a frequência de ocorrência de diferentes conceitos de gene nos livros, sobretudo o ‘conceito molecular clássico’ de gene, bem como a frequência e o modo de abordagem de fenômenos que desafiam esse conceito, em particular, o *splicing* alternativo. Visamos com essa análise verificar se os livros didáticos de ensino médio apresentam o ‘gene molecular clássico’ como um conceito incontroverso e bem estabelecido na biologia, ou se o colocam como um problema a ser resolvido, quiçá apresentando até mesmo alguns desafios a este conceito de gene.

Investigamos também o tratamento dos genes como unidades ou carreadores de informação, ou seja, a concepção ‘informacional’ do gene (Stotz et al., 2004), na qual ‘informação genética’ é, em geral, simplesmente entendida como informação sequencial no DNA (Sarkar 1998). É controverso, contudo, se a informação genética pode ser de fato entendida nesses termos (El-Hani et al. 2006) e encontramos na literatura, inclusive, um reconhecimento de que o discurso da informação na biologia se apóia numa série de metáforas ainda pouco elucidadas, à espera de uma teoria da informação biológica que possa atribuir-lhes sentido mais preciso (Griffiths 2001; Stotz et al. 2004). Entretanto, uma teoria consistente e largamente aceita da informação biológica não foi ainda construída. Sintomaticamente, o uso do conceito de ‘informação genética’ pode não estar sendo acompanhado, nos livros de ensino médio, por qualquer referencial teórico que possa elucidá-lo, dando sentido mais preciso a esta metáfora. Este é, pois, mais um ponto a ser investigado.

Com essa investigação, pretendemos ainda mostrar *como* o discurso sobre genes e *se* a chamada semiotização espontânea da biologia – entendida como a implementação de um vocabulário com conotações semióticas pelos cientistas, sem se darem conta de sua natureza semiótica (ver Emmeche 1999) – se propagam nesse estágio da transposição didática do conhecimento biológico e alcançam o nível médio de escolaridade.

Por fim, avaliamos de que maneira os modelos de função gênica encontrados nos livros analisados estão relacionados aos modelos históricos de função gênica e examinamos se os livros

fazem referências a aspectos concernentes à natureza da ciência e do modo de produção dos conhecimentos científicos, juntamente com a abordagem de elementos de história das ciências, ao tratar de concepções sobre genes.

2. Breve retrospectiva: Bases históricas dos modelos de função gênica

A variação conceitual foi uma característica do entendimento sobre genes ao longo da história da genética e da biologia molecular. O fato de o termo ‘gene’ possuir diversos significados não constitui, em si, um problema, e pode continuar a ser heurísticamente útil, desde que o contexto de aplicação de cada significado esteja bem demarcado, de modo a não resultar em ambigüidade e confusões semânticas. Não é o que ocorre, entretanto, na comunidade científica, na qual se observa grande variação na compreensão sobre genes em diversos contextos e subdisciplinas, aparentemente sem critérios claros para sua demarcação (Stotz et al. 2004). É preciso ter clareza, contudo, de que a reformulação desse conceito central da biologia não requer, necessariamente, a construção de um conceito único de gene, com o propósito dificilmente alcançável de dar conta da ampla diversidade de significados que podem e têm sido atribuídos a esse termo. Para servir a propósitos práticos, o conceito de gene não precisa ser inteiramente geral e uma diversidade de conceitos de gene pode ter, inclusive, maior poder explicativo e heurístico. Para tanto, basta que o domínio de aplicação de cada conceito seja bem delimitado pela comunidade científica (El-Hani 2007).

Quando o foco é a previsão de fenótipos a partir de genótipos, como costuma ocorrer, por exemplo, na genética médica, é possível abstrair a complexidade da relação genótipo-fenótipo. Por sua vez, uma definição molecular complexa de gene é muito mais útil para um biólogo molecular do que para um estudante que não esteja envolvido em pesquisas genéticas (Carlson 1991). De acordo como Rosenberg (1985), algumas explicações encontradas na genética têm grande aplicabilidade prática, o que assegura sua presença tanto em livros didáticos, quanto nas ciências biológicas. Por isso, não somos a favor do uso exclusivo das explicações mais recentes e modernas sobre genes nos livros didáticos de ensino médio. O gene, como qualquer objeto de estudo, possui muitos níveis e aspectos a serem percebidos e investigados; logo, os vários níveis da organização biológica – molecular, celular, organísmico, populacional – e os diferentes aspectos de um gene – transmissão, recombinação, mutação, função, evolução – podem e devem requerer distintos modelos explicativos. Segundo Gericke e Hagberg (2007a,b), livros didáticos tendem a combinar esses aspectos em modelos híbridos de função gênica, os quais podem gerar dificuldades na aprendizagem, visto que encerram problemas de consistência interna e externa entre os modelos misturados. Alternativamente, os livros didáticos poderiam evitar ambigüidades no uso de diferentes explicações sobre genes para diferentes propósitos, se explicitassem: (i) que estão usando modelos; (ii) o modelo que está sendo usado num dado momento; (iii) o porquê de diferentes modelos serem usados em paralelo na explicação de um dado assunto pelo livro; e (iv) o escopo e a limitação de cada modelo (Gericke e Hagberg 2007a).

A função gênica pode ser entendida, basicamente, como a contribuição para um traço observável, para a produção de polipetídeos e RNAs ou para um processo em um organismo. Os diferentes modelos de função gênica podem sustentar ou não uma relação de *um-para-um* entre gene e função. Também podem variar entre estes modelos: (i) a concepção de gene, sobretudo no que se refere ao seu estatuto ontológico e à sua estrutura ou substância material; (ii) o aspecto de passividade ou de atividade gênica; (iii) a natureza dos processos que relacionam os genes e outras entidades que compõem o modelo; (iv) as idéias sobre outras entidades, além do gene, que influenciam suas característica ou sua função; (v) as idéias sobre quais e quão diferentes são os níveis de organização utilizados para explicar a função gênica; (vi) o nível no qual o gene e a

função gênica são definidos; e (vii) idéias sobre quais aspectos do gene – transmissão, recombinação, mutação, função ou evolução – são enfatizados (ver Gericke e Hagberg 2007b).

Gericke e Hagberg (2007b) definiram cinco modelos de função gênica (Mendeliano, Clássico, Bioquímico-Clássico, Neoclássico e Moderno), a partir da investigação sistemática das características citadas acima, bem como da verificação das seguintes características adicionais: (i) o objetivo principal de cada modelo; (ii) a maneira pela qual um novo modelo supera as deficiências explicativas dos antecedentes; (iii) as características do modelo anterior que foram modificadas e incorporadas ao novo modelo; e (iv) as deficiências explicativas do novo modelo. Como cada conceito de gene pode estar associado a um modelo histórico de função gênica, utilizamos a classificação acima como parâmetro para identificar e categorizar os diferentes conceitos de gene que aparecem nos livros que analisamos. Conceitos, modelos e teorias rivais coexistem em qualquer época na comunidade científica, mas foge ao escopo do presente trabalho dar conta de cada uma das visões sobre gene que foram contemporâneas num dado período. Em vez disso, nos debruçamos sobre os conceitos e modelos que foram mais largamente aceitos em cada época, correlacionando-os aos modelos identificados por Gericke e Hagberg (2007b).

A idéia da existência de uma unidade hereditária responsável pela transmissão e determinação de cada característica ou traço fenotípico de um organismo teve sua origem no conceito de ‘fator’, empregado por Mendel como um conceito abstrato de natureza instrumental, uma unidade de cálculo para expressar a regularidade da transmissão de caracteres fenotípicos em cruzamentos, sem uma hipótese sobre possíveis entidades materiais, reais, que corresponderiam a eles (Falk 1986). Segundo Gericke e Hagberg (2007b), descrever a função do gene dentro de um organismo é apenas uma consequência secundária, tendo menor importância dentro deste modelo, cujo propósito principal é explicar a transmissão genética.

No modelo mendeliano, há entidades no nível macroscópico – o traço ou caráter manifesto – e num nível abstrato – o gene (*ibid.*). A função do gene é definida de ‘cima para baixo’, extrapolando-se a existência de um gene abstrato a partir de um traço manifesto. O modelo não correlaciona o gene a qualquer contraparte material, implicando nada além de uma estrutura *hipotética* física ou química na célula germinativa. Portanto, ele não oferece uma explicação fisiológica da função ou transmissão do gene e esta constitui uma deficiência explicativa do modelo mendeliano, segundo Gericke e Hagberg (2007b). O gene, nesse modelo, não é mais que um conceito abstrato, não sendo acompanhado por uma hipótese sobre alguma entidade física que pudesse corresponder a ele – dito de outra maneira, uma visão instrumentalista do gene predomina em tal modelo e uma visão realista sobre o estatuto do gene só viria a ocorrer posteriormente, na história da genética.

Por fim, o gene mendeliano é considerado passivo, na medida em que não se faz distinção entre genótipo e fenótipo, sendo o genótipo visto, no entendimento de Gericke e Hagberg (2007b), como uma espécie de fenótipo em miniatura. Nesse modelo, a mera existência do gene é necessária e suficiente para a manifestação de um caráter definido – o que implica uma relação unitária gene/traço – e, em vista disso, aspectos ambientais e epigenéticos não são considerados. Tudo isto é reflexo do fato de que, nos primeiros anos de construção da genética clássica, não havia ainda uma distinção clara entre o caráter manifesto (o fenótipo) e o potencial herdado de apresentar o caráter (genótipo). Essa distinção foi feita pelo geneticista dinamarquês W. L. Johannsen, em 1908, o que o levou, então, a propor, em 1909, um termo para designar as unidades que constituiriam o genótipo, ‘gene’. Os genes de Johannsen ainda eram entendidos, contudo, como construtos instrumentais.

Com o estabelecimento da teoria cromossômica da herança pelo grupo de T. H. Morgan, em 1911, prevaleceu uma compreensão sobre genes que resultou da combinação de análises de cruzamentos com debates entre embriologistas e estudiosos da citologia e reprodução. Esta

compreensão corresponde ao que Gericke e Hagberg (2007b) denominam o ‘gene clássico’, não mais entendido como um construto instrumental, mas como uma entidade real, uma partícula indivisível no cromossomo, no qual genes estariam organizados como as contas de um colar, analogia que ainda encontramos em livros didáticos de biologia atuais.⁵³ Não se conhecia, contudo, a estrutura molecular dos genes, dos quais se tinha apenas a vaga noção de que fossem ou atuassem como enzimas na determinação de um traço. Assim, a função gênica ainda era definida a partir de um traço, através de sua redução explicativa a um gene físico no nível celular. Em vista disso, surgiram inconsistências internas no modelo, porque, por um lado, os métodos de pesquisa se baseavam na relação de um-para-um entre gene e traço, e, por outro, fenômenos como a poligenia (vários genes influenciando um único caráter) e a pleiotropia (um único gene influenciando vários caracteres) já eram conhecidos. Isso terminou por se refletir numa separação bem marcada entre a genética da transmissão e a genética fisiológica. Não se sabia quais eram exatamente os processos naturais envolvidos na produção de um traço fenotípico, apenas que envolviam o conceito de um gene, mais ativo do que o mendeliano, que *determina* traço, e sem a influência de qualquer fator ambiental.

Na genética clássica, o gene foi entendido inicialmente como uma unidade de função, mutação e recombinação (Mayr, 1982). Mas ficou eventualmente claro que genes não são unidades nem de recombinação nem de mutação. Assim, terminou por prevalecer a idéia de que o gene seria uma unidade de função. Sintomaticamente, a partir da década de 1940, a compreensão dos genes passou a ser fortemente influenciada pelo entendimento das reações bioquímicas, mudando-se a ênfase da transmissão para a função gênica. O foco no aspecto funcional do gene é apontado por Gericke e Hagberg (2007b) como uma das diferenças cruciais entre o modelo Clássico e o modelo Bioquímico-Clássico, que impedem que o segundo seja uma mera extensão do primeiro. De análises de cruzamentos, citologia de animais e plantas e estudos sobre recombinação e mutação, a ênfase na genética passou a recair sobre experimentos genéticos e bioquímicos com microorganismos, além dos estudos da função gênica em geral e de processos desenvolvimentais em particular, como nas pesquisas com drosófilas.

Inicialmente, os resultados da bioquímica apontavam para um gene que poderia ter sua função explicada por sua redução à conexão entre um traço fenotípico e uma enzima específica produzida pelo gene. O gene era identificado como um ativo produtor de enzimas, as substâncias que determinam traços. Entretanto, não havia referência a qualquer ocorrência de processo real desse tipo, de modo que, como no modelo Clássico, as reações bioquímicas causais e mecanicistas ainda são descritas como idealistas, neste novo modelo que, apesar de se apoiar nos resultados da bioquímica e da biologia molecular, ainda utilizava ferramentas conceituais da genética clássica. O ‘gene Bioquímico-Clássico’ continuava sendo uma entidade cuja estrutura molecular era desconhecida e, assim, este modelo não explicava como os processos bioquímicos se desenrolavam quando o gene produzia uma enzima. Os problemas de consistência interna eram ainda maiores do que no modelo Clássico: a conexão do gene à enzima era abordada ‘de baixo para cima’, mas permanecia a abordagem de ‘de cima para baixo’, que conectava o traço à enzima no nível celular a partir de uma redução explicativa.

A compreensão de que nem todos os produtos gênicos são enzimas levou à reformulação deste modelo, passando-se a empregar aos esquemas ‘um gene-uma proteína’ e, depois, ‘um

⁵³ Falk (1986, p. 169) ressalta que o conceito de gene evoluiu através da ‘confrontação dialética’ entre a abordagem instrumentalista e a realista, seguindo um padrão que ele denomina de ‘bonecas russas’ – referindo-se à ‘matryoshka’, boneca que se encontra dentro de outra, que, por sua vez, se encontra dentro de outra), em que descobertas sobre a natureza química do gene levavam à elaboração de novas definições funcionais, que, por sua vez, geravam significados estruturais mais profundos, levando a significados funcionais ainda mais profundos, e assim sucessivamente.

gene-um polipeptídeo’, quando se descobriu proteínas constituídas por várias cadeias polipeptídicas, codificadas por genes distintos. A compreensão de que RNAs também podem ser produtos gênicos levou, enfim, ao esquema ‘um gene-um polipeptídeo ou RNA’. A idéia de unidade foi preservada em todos esses esquemas e foi subsequentemente reforçada pelo modelo de estrutura do DNA proposto por Watson e Crick, em 1953.

Este modelo explicou de uma só tacada a natureza da seqüência linear dos genes, o mecanismo de replicação gênica e de síntese de RNA a partir de DNA, a natureza das mutações e a separação entre mutação, recombinação e função no nível molecular (El-Hani, 2005). Além disso, este modelo estabeleceu com vigor uma visão realista sobre genes, na medida em que havia agora uma contrapartida material clara para o conceito de gene. Estavam colocadas as bases para a definição de genes a partir de estruturas moleculares, e não a partir de traços fenotípicos – como havia sido o caso por toda a história da genética até aquele momento –, bem como para a construção do conceito molecular clássico de gene.

Assim, os dois modelos restantes da classificação de Gericke e Hagberg (2007b) refletem a transição da genética clássica para a genética e biologia molecular. O conceito molecular clássico (assim nomeado por Griffiths e Neumann-Held, 1999; Stotz et al., 2004) sobrepõe a idéia mendeliana de unidade hereditária ao modelo molecular de gene (Fogle, 1990), configurando-se, assim, como um modelo híbrido, no sentido de que, historicamente, já mistura duas concepções distintas sobre genes. De acordo com ele, o gene é um segmento contínuo, cuja seqüência de bases codificantes não sofre interrupções; discreto, por ser uma unidade individual que não se sobrepõe a outros genes; com começo e fim bem definidos; e com localização constante. De acordo com este conceito, genes podem ser considerados unidades de estrutura – pelas características descritas acima –, de função – por produzirem um único polipeptídeo ou um único RNA, que, por sua vez, teria uma função única –, e, a partir da incorporação do discurso da informação à biologia, também de informação. A sobreposição freqüente, no discurso sobre gene, do conceito molecular clássico à concepção informacional pode ser identificada dentro do modelo Neoclássico, no qual, conforme a classificação de Gericke e Hagberg, genes são tipicamente vistos como portadores de informação, em sua seqüência de bases, para produzir um polipeptídeo ou RNA.

Por fim, podem ser extraídos do modelo Moderno, na classificação de Gericke e Hagberg (2007b), conceitos de gene que levam em conta os novos desafios da era genômica: o gene não é uma unidade de estrutura, função ou informação; o gene pode ser visto como uma entidade de existência temporária e descontínua; o gene requer *processos* desenvolvimentais no contexto celular e fatores epigenéticos para sua expressão funcional; diferentes partes do DNA podem cooperar de múltiplas formas para gerar diferentes produtos no tempo e espaço. Face a estes problemas, alguns autores propõem a definição de gene como uma seqüência linear de DNA que não tem um papel único no desenvolvimento, mas corresponde a uma única ‘norma de reação’⁵⁴ de produtos gênicos, numa variedade de condições celulares. Assim, o gene poderia ser relacionado a vários fenótipos, através de diferentes proteínas, produzidas, por exemplo, por *splicing* alternativo. Griffiths e Neumann-Held (1999) incluem estas idéias num ‘conceito molecular contemporâneo de gene’, que, para eles, é ainda uma resposta conservadora, visto que não dá conta das condições epigenéticas necessárias para o desenvolvimento, por exemplo, para o disparo de vias de sinalização celular específicas, que definem padrões específicos de *splicing* alternativo. Em contrapartida, eles propõem o ‘conceito molecular processual de gene’ (*process molecular gene*), segundo o qual o gene é entendido não somente como uma seqüência específica

⁵⁴ De acordo com Lewontin (2000), podemos entender ‘norma de reação’ como uma representação gráfica da diversidade de fenótipos que um determinado genótipo pode condicionar, em função do ambiente. Ver seção 4.4 do presente trabalho.

de DNA, mas como todo o processo molecular subjacente à capacidade de expressar um produto gênico particular (ver tb. Neumann-Held 2001).

Há outros conceitos que enfatizam a relevância dos processos que mediam a relação entre uma seqüência no DNA e um dado produto gênico, sem adotar uma visão processual de gene. É o caso do ‘conceito sistêmico de gene’, que como proposto por Pardini e Guimarães (1992), é uma combinação de uma ou mais seqüências de ácidos nucleicos (DNA ou RNA) que corresponde a um produto (polipeptídeo ou RNA), mas que só é definida num determinado contexto de um sistema, que pode ser a célula, ao interagir com o ambiente, ou o próprio ambiente, no caso de sistemas subcelulares ou pré-celulares. Assim, uma seqüência no DNA pode ter múltiplos significados, cada um respectivo a um contexto específico no qual ele se insere. Uma idéia semelhante, mas que não inclui seqüências de RNA é a proposta de Fogle (1990, 2000) de que genes sejam tratados como conjunto de domínios no DNA, entendendo-se por ‘domínio’ uma seqüência de nucleotídeos que pode ser distinguida de outras seqüências com base em suas propriedades estruturais e/ou atividades, como, por exemplo, éxons, íntrons, promotores, *enhancers*, operadores etc. Os domínios podem ser combinados de várias maneiras para formar, nos termos de Fogle (1990, 2000), um ‘Conjunto de Domínios para Transcrição Ativa’, ou, nos termos mais correntes, um gene. Na visão desse autor, dessa forma seria possível assimilar a complexidade e diversidade da arquitetura genética, abandonando-se a necessidade de encontrar uma unidade única para delimitar uma região de informação genética, ou, dito de outra maneira, superando a sobreposição da idéia mendeliana de unidade e da compreensão molecular das bases da herança que se afirmou no conceito molecular clássico de gene. Ainda dentro da perspectiva de que o significado de um gene para seu intérprete, a célula, ou a significância biológica de um gene, não pode ser encontrada nas seqüências de DNA de um cromossomo, mas no contexto que estas se inserem, El-Hani, Queiroz e Emmeche (El-Hani et al. 2006) empregam a moldura teórica estabelecida pela teoria pragmática do signo de C.S. Peirce, para construir um entendimento de genes como signos e da informação genética como semiose.

A seguir, analisamos duas explicações sobre genes que exemplificam como o reconhecimento de desafios ao conceito de gene, bem como da importância do contexto no qual o gene se insere podem ser utilizados para fomentar, por um lado, a dispersão de visões deterministas genéticas e, por outro, o entendimento das seqüências no DNA como recursos relevantes, mas meramente potenciais para um fenótipo, a serem utilizados no desenvolvimento.

3. Retórica do ‘gene-para’ um traço: ferramenta conceitual ou fonte de determinismo genético?

O conceito de gene sempre esteve em contínua evolução e seu significado está imerso em controvérsias praticamente desde sua criação, como discutimos acima com base na literatura filosófica e histórica sobre a genética. Mas os avanços da biologia molecular e da genômica nos últimos trinta anos tornaram difícil até mesmo definir e localizar o que é um gene. Como sumariza Falk (1986), uma série de descobertas sobre a organização e a dinâmica do genoma tem tornado claro que genes não são discretos – há genes superpostos e nidados –, nem contínuos – há íntrons dentro dos genes; eles não têm necessariamente uma localização constante – há transposons –, nem uma função claramente definida – há pseudogenes; eles não são unidades nem de função – há genes que sofrem *splicing* alternativo, genes que codificam proteínas multifuncionais, e uma forte dependência da ação gênica em relação aos contextos celular e supracelular –, nem de estrutura – há muitos tipos de seqüências *cis*-atuantes que influenciam a transcrição, genes interrompidos etc.

Na conjuntura atual de esforços para refinar o arcabouço teórico da biologia, sobretudo no que concerne ao conceito de gene, se destaca a ‘teoria dos sistemas de desenvolvimento’ (*developmental systems theory*), proposta a partir dos *insights* de pesquisadores de inúmeras áreas (Oyama [1985]2000; Oyama et al. 2001), insatisfeitos com as formulações dicotômicas gene/ambiente e ávidos por uma nova forma de pensar sobre conceitos biológicos centrais como herança, desenvolvimento e evolução. A tese central dessa teoria contra o determinismo genético, é a paridade causal dos recursos desenvolvimentais, que rejeita a distinção entre causas essenciais – tradicionalmente atribuídas aos genes – e causas acessórias e destaca o desenvolvimento como um terceiro elemento que estaria faltando nas explicações deterministas da relação genótipo-fenótipo. De acordo com esta visão, o gene, como uma seqüência no DNA, é apenas um recurso desenvolvimental, situado ontologicamente no mesmo plano que outras biomoléculas, como RNAs, proteínas, oligossacarídeos etc., destituído de qualquer privilégio causal na manifestação de traços fenotípicos num organismo (Oyama, 2000; Moss, 2003). Esta idéia de paridade, contudo, não é uma posição consensual e há, na literatura, propostas que mantêm a atribuição de privilégio causal ao gene, sem deixar de reconhecer a devida importância dos processos desenvolvimentais que permitem conectar um gene a um traço fenotípico, como a explicação de genes como signos no DNA e de informação como um processo semiótico (El-Hani *et al.* 2006).

Subjacente a ambas as reformulações, entretanto, está a visão de que genes no DNA não podem ser determinantes de características, mas apenas recursos desenvolvimentais que fornecem possíveis moldes para a produção de vários produtos gênicos, seja diretamente, na síntese de RNA, ou indiretamente, na síntese de proteínas. Fora da comunidade científica, entretanto, o discurso do tipo ‘um gene para um traço fenotípico’ é continuamente extrapolado para significar que o gene determina, ou ainda, que o gene possui a instrução para a produção de um traço. Este tipo de retórica, tão familiar nos dias atuais, encerra uma forma de *pré-formacionismo constitutivo*, no qual se subentende que a totalidade de um organismo pode ser decomposta em genes. Moss (2001, 2003) argumenta que esse pré-formacionismo emerge da mistura de dois significados diferentes presentes no discurso atual sobre genes, cada um, herdeiro de uma das duas maiores tendências históricas de explicação da ordem biológica: a epigênese e o pré-formacionismo. O gene epigenético é, segundo este autor, uma unidade de transcrição no DNA e, como recurso desenvolvimental, é definido por sua seqüência molecular, porém, é indeterminado em relação a qualquer resultado fenotípico. Esse gene é denominado por Moss (*ibid.*) ‘gene-D’ (‘D’, de desenvolvimental) e é fundamentado na tese de paridade causal entre as biomoléculas.

Enquanto o gene-D, de acordo com a visão realista sobre genes como entidades moleculares, corresponde a uma entidade material, o gene-P (o ‘gene para fenótipo’; ‘P’, de pré-formacionista, *sensu* Moss 2001, 2003), pode ser, em uma de suas versões, concebido como uma entidade abstrata, empregada quando uma diferença entre fenótipos não resulta da *presença* de duas capacidades gênicas qualitativamente diferentes, mas, antes, da *ausência* da capacidade de se produzir uma proteína normal. Em outras palavras, um fenótipo que usualmente é mascarado quando um dado produto gênico está presente – sendo este produto um dos recursos necessários para a expressão do fenótipo considerado normal –, vem à tona quando o organismo é incapaz de produzir aquele recurso molecular normal. Uma vez que este recurso pode estar ausente por muitos desvios possíveis na seqüência do DNA normal e também por outras falhas ao longo de sua via de produção, nenhuma estrutura molecular específica pode ser correlacionada à anormalidade fenotípica. Esta é, na verdade, o resultado altamente complexo do que acontece no organismo na ausência do recurso molecular normal.

Em sua outra versão, o gene-P pode ser validamente empregado como uma ferramenta conceitual que representa toda e qualquer inabilidade de produção do recurso molecular normal, sendo, portanto, indeterminado quanto à seqüência no DNA, e definido a partir do fenótipo

‘anormal’, o que constitui uma espécie de *pré-formacionismo instrumental*. De acordo com Moss (2003, p. 45), neste tipo de pré-formacionismo, “falar de um *gene para um fenótipo* é falar *como se, e somente como se*, ele determinasse diretamente o fenótipo”. Esse caso difere do pré-formacionismo constitutivo – o qual considera a existência de segmentos no DNA que são, ao mesmo tempo, gene-D e gene-P –, já que não se compromete com a hipótese realista de que o conceito de um gene determinante de fenótipos ou de diferenças fenotípicas representa algo realmente existente no genoma ou no corpo. O uso válido do gene-P instrumental leva em conta que nenhum gene no DNA pode determinar fenótipos e que, ainda que qualquer gene-P possa ser materializado numa entidade molecular, cuja presença confirmaria a ausência de uma seqüência normal, esta materialização do gene-P seria reconhecida como apenas um dos possíveis desvios da seqüência normal que podem ser relacionados à anormalidade fenotípica. É o caso, por exemplo, do gene para o câncer de mama – BRCA1 – e do gene para a fibrose cística, ambos, genes-P.

O gene-P instrumental participa de um ‘jogo’ explicativo semelhante ao executado pelo gene mendeliano na análise de cruzamentos e previsão de fenótipos em genealogias, mas não está confinado aos métodos puramente clássicos – como vimos acima, uma entidade molecular pode ser tratada como gene-P –, sobretudo porque é aplicado *no contexto atual*, em que o conceito de gene já tem uma contraparte material e a estrutura complexa do DNA já é conhecida. Isso implica que o gene-P, em relação ao gene mendeliano, só é válido e só pode ser empregado para um conjunto limitado de traços fenotípicos e dentro de certos limites contextuais, por exemplo, quando sua aplicação traz benefícios médicos e/ou econômicos imediatos (Moss 2003). Por exemplo, pode ser prático e útil para um médico ou pesquisador falar de forma preditiva de um ‘gene para fibrose cística’, na ausência de um entendimento molecular-desenvolvimental mais aprofundado dos processos que resultam nessa patologia. Nesse caso, o gene-D, recurso desenvolvimental, que nos indivíduos sem histórico da doença está localizado no locus da fibrose cística, não é um gene para a função normal dos pulmões, não é um gene para fenótipo: é apenas um dentre milhares de genes e outros recursos que estão envolvidos na função e no desenvolvimento pulmonares. Misturado ao gene-D, o uso do gene-P é inválido, ilegítimo, e a retórica do ‘gene-para’ um traço, em vez de ferramenta conceitual, passa a ser fonte de determinismo genético, como argumenta Moss (2003).

A partir dessas considerações, examinamos como os livros de ensino médio abordam: (i) os conceitos de gene formulados nos primeiros anos da genética, como o gene mendeliano, o gene clássico e o bioquímico-clássico; (ii) o entendimento atual de genes e os desafios enfrentados pelo conceito molecular clássico; (iii) as propostas mais recentes de reformulação do conceito de gene; (iv) e os ‘genes para fenótipos’, se como gene-P instrumental ou como gene-P sobreposto ao gene-D, ou mesmo ao gene molecular clássico, numa visão do DNA como programa para o desenvolvimento, de acordo com a qual a informação para a construção do organismo estaria armazenada na seqüência de bases do DNA.

4. Metodologia

A amostra de dezoito livros didáticos de biologia analisada no presente trabalho foi a mesma submetida por editoras brasileiras ao Programa Nacional do Livro Didático do Ensino Médio (PNLEM), do Ministério da Educação. Esta amostra foi escolhida por ser considerada bastante representativa do universo de livros didáticos de biologia do ensino médio publicados no Brasil, na medida em que este é um grande programa governamental, para o qual as editoras

direcionam a maioria dos materiais didáticos que publicam.⁵⁵ Estas obras didáticas, algumas compostas por volumes únicos e outras, por três volumes, cada um voltado para um ano do ensino médio, são listadas no Anexo 1.

4.1. Fase Exploratória

Todos os capítulos dos livros foram analisados com base em técnicas de análise de conteúdo (Bardin 2000). Em particular, utilizamos a *análise categorial*, a qual envolve operações de desmembramento do texto em unidades, as categorias, que são construídas segundo reagrupamentos analógicos – reagrupamentos de elementos do texto que compartilham certos caracteres – cujo critério caracterizador pode ser sintático (ocorrência de signos lingüísticos precisos) ou semântico (ocorrência de um mesmo significado num dado contexto). Num primeiro momento, realizamos uma leitura de sondagem do conteúdo dos livros para o planejamento das operações de recorte do texto, tratamento dos dados e categorização. Essa análise piloto nos forneceu uma primeira impressão a respeito dos conceitos de genes presentes nos livros.

Quanto à forma de recorte do texto, determinamos as unidades de registro e as unidades de contexto, que constituem os elementos a se ter em conta em nossa análise. As unidades de contexto serão discutidas mais abaixo. Em relação às *unidades de registro*⁵⁶, optamos pelo recorte a um nível semântico, ou seja, pela análise de unidades de significação complexas, os temas. Diferente do recorte lingüístico, de palavras ou frases precisas, que visa, em geral, constatar a pura existência e a freqüência de um dado material lingüístico, o recorte de temas se preocupa com o sentido de dado conteúdo do texto e a variação deste sentido em diferentes contextos. Adicionalmente, a categorização temática é de ordem psicológica, sendo geralmente utilizada para estudar motivações de opiniões, de atitudes, valores, crenças e tendências (Bardin 2000). Nossos temas, como unidades de registro, consistem de excertos (frases condensadas, resumos, afirmações ou alusões) nos quais os autores exprimem idéias a respeito de unidades perceptíveis, como as palavras ‘gene’ e ‘DNA’.

4.2. Categorias para a análise

Com base na fase exploratória, as seguintes categorias de conceitos de gene foram propostas para a análise dos livros didáticos: ‘gene molecular clássico’, ‘gene informacional’, ‘gene-P’ e ‘gene-D’ (ambos *sensu* Moss 2001, 2003), ‘gene mendeliano’, ‘gene clássico’, ‘gene bioquímico-clássico’. Estas categorias foram obtidas a partir tanto dos materiais didáticos analisados, quanto da literatura filosófica e educacional. Desse modo, nem são ‘pré-categorias’, obtidas somente a partir dos marcos teóricos assumidos no projeto, nem ‘pós-categorias’, construídas apenas à luz da exploração dos livros didáticos. Tratam-se de um produto intermediário que consideramos bastante adequado para o tipo de pesquisa que empreendemos, que não pode, de um lado, declinar do aprofundamento da compreensão da diversidade de conceitos de gene que somente um estudo do conhecimento acumulado a respeito poderia dar, e, de outro, não pode somente impor categorias construídas previamente à realidade investigada, o que foi evitado pela construção das categorias mediante a fase exploratória de análise aqui descrita. Podemos compreender este procedimento como uma espécie de diálogo entre conhecimento acadêmico acumulado sobre conceitos de gene e o conhecimento escolar veiculado nos livros didáticos, mediado, claro, pela pesquisadora que realizou o estudo exploratório dos

⁵⁵ Para maiores detalhes, ver

<http://portal.mec.gov.br/seb/index.php?option=content&task=view&id=648&Itemid=666>.

⁵⁶ Unidade de registro é “a unidade de significação a codificar” e que “corresponde ao segmento de conteúdo a considerar como unidade de base, visando a categorização e contagem frequencial” (Bardin 2000, p. 104).

livros. Como resultado deste diálogo com o conhecimento escolar presente nos livros, algumas categorias derivadas da literatura não foram incorporadas na análise, como o gene processual molecular (Griffiths e Neumann-Held 1999, Neumann-Held 2001), a interpretação de genes como conjuntos de domínios no DNA (Fogle 1990), o conceito sistêmico de gene (Pardini e Guimarães 1992, Guimarães e Moreira 2000), a interpretação dos genes como signos (El-Hani et al. 2006). Na discussão de nossos resultados, esse diálogo também cumprirá papel importante, na análise das implicações das ausências⁵⁷ de determinados conceitos de genes no conhecimento escolar, conforme expresso nos livros didáticos analisados.

Basicamente, as idéias sobre genes contidas nas unidades de registro correspondem a uma ou mais características epistemológicas de modelos de função gênica identificadas por Gericke e Hagberg (2007a), que foram encontradas no conhecimento escolar veiculados pelos livros. Estas características são mostradas na Tabela 1. Em nossa análise, reconhecemos conceitos de gene através destas características epistemológicas, que nos permitem inferir tanto sobre a natureza do gene quanto sobre o modelo de função gênica implícito num dado contexto, com base na combinação de variantes de cada uma dessas características. Cada unidade de registro pode fazer alusão a mais de um conceito de gene e esse aspecto foi considerado em nosso procedimento de contagem.

A categorização dos conceitos de gene de uma unidade de registro vai depender fortemente da *unidade de contexto*⁵⁸ da qual faz parte. Da análise piloto, inferimos que, em geral, o significado do gene não muda conforme capítulos, mas conforme unidades de contexto maiores, que correspondem a disciplinas da biologia que abordam os genes de maneiras distintas e para diferentes propósitos. Estas unidades, que podem conter vários capítulos de um livro, foram escolhidas como nossas unidades de contexto, tendo sido categorizadas a partir do título e do conteúdo dos capítulos dos livros. Categorizamos as seguintes unidades de contexto: ‘glossário’, ‘caracterização da vida’ – que trata de assuntos como origem, natureza, diversidade dos seres vivos –, ‘biologia celular e molecular’, ‘genética’ e ‘evolução’.

Com base nesses indicadores, elaboramos um protocolo visando padronizar a análise dos livros. No protocolo referente a cada livro, catalogamos as unidades de registro que poderiam conter os seguintes elementos: conceitos de gene e idéias sobre função gênica; referências a diferentes tipos de seqüências de DNA; menções de desafios ao conceito molecular clássico, sobretudo, o *splicing* alternativo; reconhecimento das dificuldades trazidas por esses desafios ao conceito de gene; abordagens do conceito de informação genética; metáforas e analogias sobre genes e DNA; abordagem de história e natureza da ciência e da relação modelo/realidade. As obras foram analisadas tanto qualitativamente, com base na literatura histórica e filosófica sobre o conceito de gene, como quantitativamente, mediante a categorização e enumeração dos extratos de acordo com a frequência de aparição de concepções significativas.

Por fim, para cada unidade de contexto, identificamos um modelo de função gênica implícito. A partir da Tabela 2, foi derivada, para cada contexto, uma tabela em que destacamos as variantes epistemológicas nele encontradas, com base num padrão de sombreamento (ver Tabelas 4 a 8, na subseção 5.2.2). A intensidade do sombreamento indica o número de obras em que cada variante está presente: 1 a 6 (cinza claro), 7 a 12 (tom intermediário de cinza) e 13 a 18 (cinza escuro). As variantes não-ocorrentes no contexto não são destacadas (branco).

Em seguida, observamos se há predominância de variantes de um modelo histórico em particular, o que permitiu categorizarmos os modelos de cada contexto de acordo com o modelo

⁵⁷ As ausências ou omissões destes conceitos constituem o elemento pontual de análise do discurso que aplicamos no presente trabalho.

⁵⁸ A unidade de contexto corresponde a um segmento do texto cujas dimensões (superiores às da unidade de registro) são ótimas para que se possa compreender a significação exata desta última unidade (Bardin 2000).

histórico com o qual possuem o maior número de variantes epistemológicas em comum, no maior número de obras analisadas.

Levando-se em conta que o modelo identificado para um contexto poderia ser híbrido, caso contivesse sobreposições de suas variantes epistemológicas com as de outros modelos históricos de função gênica, estimamos o grau de hibridização desse modelo, através da comparação das características epistemológicas de cada modelo encontrado àquelas dos modelos históricos de função gênica (Tabela 2), e também investigamos a presença de variantes não-históricas – que não são encontradas em qualquer modelo histórico – e de idéias incorretas quanto ao conhecimento científico.

Ponderamos, contudo, que modelos históricos compartilham, naturalmente, algumas variantes epistemológicas, uma vez que características de um modelo anterior podem ser incorporadas a um novo. Por conseguinte, todos os modelos históricos de função gênica aqui apresentados são naturalmente híbridos (ver Tabela 2). Todavia, ao classificarmos, em nossa análise, o *modelo de um dado contexto* como híbrido, não estamos nos referindo a essa hibridização natural (histórica), mas àquela resultante da mistura de variantes epistemológicas *peculiares* de diferentes modelos. Dessa forma, o grau de hibridização do modelo identificado para cada contexto foi calculado como o percentual dos tipos de variantes que não estão associados a este modelo, dentre todos os tipos de variantes ocorrentes no contexto.

De acordo com Bardin (2000), o uso da associação ou equivalência (ou, ainda, sobreposição) de temas como indicador se assenta no postulado de que elementos associados numa manifestação da linguagem estão, ou estarão, igualmente associados no pensamento do locutor (ou do destinatário), o que, em nosso entendimento, constituirá uma importante ferramenta numa análise futura da congruência de conhecimentos, valores e práticas dos autores da recontextualização didática dos sistemas genéticos.

Tabela 1. Características epistemológicas e suas variantes contidas nas unidades de registro (adaptado a partir de Gericke e Hagberg 2007a).

Características epistemológicas		Variantes das características epistemológicas	
1	Relação entre estrutura e função do gene	1a	O gene é uma entidade abstrata e, assim, não tem estrutura
		1b	O gene é uma partícula no cromossomo
		1c	O gene é um segmento de DNA
		1d	O gene consiste de um ou vários segmentos de DNA com vários propósitos
		1e*	O gene é carreador e/ou unidade de informação*
2	Relação entre nível de organização e definição da função gênica	2Ia	O modelo tem entidades no nível fenotípico e conceitos abstratos ⁽ⁱ⁾
		2Ib	O modelo tem entidades nos níveis fenotípico e celular
		2Ibx	<i>O modelo tem entidades nos níveis fenotípico, celular e molecular</i>
		2Ic*	O modelo tem entidades no nível molecular*
		2Icx	<i>O modelo tem entidades nos níveis celular e molecular</i>
		2Icy*	O modelo tem entidades nos níveis fenotípico e molecular*
		2IIa	A correspondência entre gene e sua função é um-para-um
2IIb	A correspondência entre gene e sua função é vários-para-vários		
3	A abordagem ‘real’ para se definir a função do gene: de cima para baixo/de baixo para cima	3a	A função do gene é definida de ‘cima para baixo’
		3b	A função do gene é definida de ‘baixo para cima’
		3c	A função do gene é definida por um processo subjacente à capacidade de expressar um produto gênico particular
4	A relação entre genótipo e fenótipo	4a	Não há separação entre genótipo e fenótipo
		4b	Há separação, sem explicação, entre genótipo e fenótipo
		4c	Há uma separação entre genótipo e fenótipo com enzima como uma explicação causal intermediária ⁽ⁱⁱ⁾
		4d	Há uma separação entre genótipo e fenótipo com uma explicação de processo bioquímico ⁽ⁱⁱⁱ⁾
5	As relações idealistas versus naturalistas nos modelos	5Ia	Há relações idealistas no modelo, sem referência a um processo natural
		5Ib	Há relações naturalistas no modelo, com descrição detalhada do processo bioquímico natural de expressão gênica
		5IIa	As relações no modelo são causais e mecanicistas (interações químicas de genes determinam traços independentemente do contexto)
		5IIb	As relações no modelo são orientadas para processos e holísticas (a função do gene depende do contexto em que ele se insere)
6	O problema da redução explicativa	6a	Há redução explicativa do nível macro a conceitos abstratos
		6b	Há redução explicativa do nível macro ao nível celular
		6bx	<i>Há redução explicativa do nível macro ao nível molecular</i>
		6c	Não há redução explicativa
7	Relação entre fatores genéticos e ambientais na constituição do fenótipo e no desenvolvimento	7a	Fatores ambientais não são considerados
		7ax	<i>Entidades ambientais + entidades genéticas resultam num traço/produto/função^(iv)</i>
		7b	Entidades ambientais incluídas pelo sistema desenvolvimental
		7c	Entidades ambientais mostradas como parte de um processo

⁽ⁱ⁾ Variantes italicizadas seguidas de ‘x’ indicam modelos híbridos de função gênica encontrados por estes autores. ⁽ⁱⁱ⁾ Variantes marcadas com ‘*’ foram introduzidas por nós e não constam no original. ⁽ⁱⁱⁱ⁾ No original, aparece a expressão ‘nível macro’ – de macroscópico – e não ‘nível fenotípico’. Evitamos usar ‘macro’, por ser termo relativo que se aplica a qualquer comparação entre níveis de organização. E, no lugar de ‘nível simbólico’, usamos ‘conceitos abstratos’. ^(iv) O gene é entendido como uma entidade no cromossomo que produz uma enzima. ^(v) Há referência aos processos bioquímicos que mediam a expressão de um dado fenótipo a partir do DNA. ^(vi) Compreende-se a relação como uma interação aditiva, na qual cada fator se relaciona ao produto, mas um fator não influencia o outro significativamente.

Tabela 2. Variantes de características epistemológicas dos modelos históricos de função gênica (adaptado de Gericke e Hagberg 2007a). A variante ‘1e*’ foi introduzida por nós e não consta no trabalho destes autores.

Modelos históricos de função gênica	Características epistemológicas								
	1	2I	2II	3	4	5I	5II	6	7
Modelo Mendeliano	1a	2Ia	2IIa	3a	4a	5Ia	5IIa	6a	7a
Modelo Clássico	1b	2Ib	2IIb	3a	4b	5Ia	5IIa	6b	7a
Modelo Bioquímico-clássico	1b	2Ib	2IIa e 2IIb	3a e 3b	4c	5Ia	5IIa	6b	7a
Modelo Neoclássico	1c e 1e*	2Ic	2IIa	3b	4d	5Ib	5IIa	6c	7b
Modelo Moderno	1d	2Ic	2IIa	3c	4d	5Ib	5IIb	6c	7c

Em nossa investigação, foram tão relevantes quanto à frequência de aparição de concepções significativas sobre gene, algumas variáveis de inferência de natureza não-linguística. Por exemplo, na categorização, são relevantes as direções adotadas pelo autor: favoritismos, expressados no glossário, que é um contexto particularmente importante, porque reflete mais explicitamente a posição do autor ao se deparar com a necessidade de fornecer uma definição para gene; e o tempo do verbo que, no pretérito, indica a abordagem do gene num contexto histórico e, no presente, indica o que o autor identifica como o entendimento atual sobre gene, seja como conceito bem estabelecido, ou como conceito controverso. Também importa o tamanho da unidade de contexto: quanto mais ampla ela for, maior o número de co-ocorrências de diferentes características epistemológicas, o que facilita a identificação do modelo de função gênica implícito.⁵⁹ A ordem, posição relativa e co-ocorrências em cada unidade de contexto também são relevantes. Conceitos de gene podem vir sobrepostos – implicando equivalência – ou apenas associados – abordados simultaneamente –, e algumas associações podem ser mais relevantes do que outras.

4.3. Validação interna

Para aumentar a validade interna dos achados deste estudo, uma análise independente dos mesmos dados brutos foi realizada por outra pesquisadora de nosso grupo, de modo a determinar-se a taxa de concordância das duas análises. Em seguida, foram construídas interpretações com as quais todos os pesquisadores envolvidos no projeto (num total de 4) concordaram (sempre que possível), através de um processo de discussão. A taxa média de concordância entre as pesquisadoras que realizaram as duas análises independentes foi de 89,94%. As categorizações nas quais houve diferença entre as duas pesquisadoras foram submetidas à discussão, envolvendo também os dois pesquisadores seniores do projeto, chegando-se a uma conclusão compartilhada sobre qual categorização deveria ser preferida. Desse modo, os resultados que serão relatados a seguir correspondem a um consenso compartilhado pelos quatro pesquisadores envolvidos no presente projeto.

⁵⁹ Os exercícios propostos, sobretudo os de ‘múltipla escolha’, não foram considerados na contagem de ocorrências de conceitos de gene. Visando simplificar nossa análise, optamos por registrar apenas os exercícios que apresentavam explicitamente, na pergunta ou na resposta sugerida no ‘Manual do Professor’, um conceito de gene.

O presente projeto também apresenta validade externa, na medida em que trabalhamos com uma amostra significativa de livros didáticos de biologia do ensino médio publicados no Brasil. Nossos resultados podem ser, pois, extrapolados para o conjunto de tais livros.

5. Resultados e Discussão

5.1. Frequência de ocorrência dos conceitos de gene

A Figura 1 sumaria os resultados de nossa análise, no que tange à frequência de ocorrência de conceitos de gene em obras de biologia do ensino médio recomendadas (1a) e excluídas (1b) do PNLEM/2007.

Pode-se observar na figura que o conceito de gene-P se mostrou prevalente em 12 das 18 obras analisadas, chegando a uma frequência média nas 18 obras de 25,05 ocorrências/obra, a maior encontrada em nossos dados. A alta frequência desse conceito se deve, em grande parte, ao conteúdo extenso dos capítulos de genética, em que aparecem muitos exemplos de análises de genealogias/heredogramas (*pedigrees*) e de probabilidades de herança de fenótipos, que abordam os métodos da genética clássica, bem como a genética médica. Nesses contextos, há maior probabilidade de ocorrer gene-P, quando se abstrai o entendimento atual da complexidade da relação entre genótipo e fenótipo, para ter como foco a previsão do segundo a partir do primeiro. As ocorrências do conceito de gene-P podem ser exemplificadas pelos seguintes excertos:

‘O gene para olhos castanhos situado num cromossomo é alelo do gene para olhos verdes, situado no cromossomo homólogo’ (Borba e Cançado 2005, vol. 3. p. 15).

‘[...] planta Rr é de fenótipo vermelho, mas o genótipo mostra que, apesar de vermelha, ela apresenta o gene para a cor branca’ (Borba, Crozetta e Lago 2005, p.262).

Ressaltamos que, em sua maioria, os exemplos de genes-P encontrados implicam a idéia de determinação de fenótipos ou diferenças fenotípicas, como no exemplo que segue:

‘[...] em Recife, o caráter olhos claros, proveniente da ocupação holandesa, ainda é difuso entre a população. Como, depois de tanto tempo, o gene que o determina, que é recessivo, ainda está presente na população e não foi completamente substituído pelo alelo dominante, que determina olhos escuros?’ (Gainotti e Modelli 2005, p. 286).

No entanto, a idéia de determinação muitas vezes aparece implícita no uso do termo ‘condicionar’, como nos trechos abaixo:

‘O gene que condiciona a cor da casca da semente de ervilha [...] é um gene pleiotrópico, determinando também a cor da flor e a presença de uma mancha roxa nas estípulas foliares’ (Amabis e Martho 2005, vol. 3, p. 40).

‘Um gene – denominado pleiotrópico – pode determinar o aparecimento simultâneo de várias características. Um exemplo é o alelo humano que condiciona a fibrose cística, doença que altera muitas secreções do corpo’ (Favaretto e Mercadante 2005, p. 128).

‘[...] hoje são conhecidos alguns casos de herança em que um caráter é condicionado por dois ou mais pares de genes. Nesses casos, os vários pares interagem, colaborando para determinar a característica [...] Um exemplo clássico de interação é o das diferentes formas de cristas de galinhas: simples, rosa, ervilha e noz’ (Silva Júnior e Sasson, vol. 3, p.132).

‘O gene P é o responsável pela pelagem aguti e dominante sobre o seu alelo p, que condiciona a pelagem preta. A pigmentação é, então determinada pela ação do gene C, dominante sobre o seu alelo c, que impede a formação dos pigmentos’ (Borba e Cançado 2005, vol. 3, p. 52).

Os exemplos acima mostram claramente como genes-P, à semelhança das explicações sobre genes oferecidas pela genética clássica, são definidos a partir do fenótipo e indeterminados quanto à seqüência molecular, mesmo nos casos em que a relação entre gene e traço não é de um-para-um, como na pleiotropia, interação gênica e herança quantitativa. Apesar de o gene-P soar ‘puramente clássico’, ou seja, mendeliano, em oposição a molecular (Moss 2003, p. 45), em certos contextos, uma entidade molecular pode ser tratada como um gene-P, como argumenta o próprio Moss, em particular, quando se fala de um gene para fenótipo como se, e apenas como se, ele determinasse diretamente o fenótipo. Este será um uso válido do gene-P desde que a seqüência molecular em questão seja considerada apenas um, dentre muitos possíveis, motivos da ausência de uma seqüência molecular normal. Esta sutileza no uso válido do gene-P dificilmente aparece clara no discurso dos autores de livros didáticos, mas podemos citar alguns exemplos que mais se aproximam dessa idéia:

‘A disponibilidade de um teste molecular para uma determinada doença genética tem sido importante para a confirmação do diagnóstico [...] Estes achados têm grande importância para a família, pois podem permitir, por meio da análise de seu DNA, por exemplo, classificar uma irmã do afetado como portadora ou não da alteração. A disponibilidade desses testes representou grande avanço, pois possibilitou a detecção de certeza quanto ao gene da DMD pertencente a estas famílias [...] o fato de poder definir se uma consulente é portadora ou não da alteração específica que causa a doença (mutação) muda bastante o efeito do aconselhamento genético. Antes de dispormos dos testes de DNA, por meio da análise da genealogia e de testes bioquímicos estimávamos o risco de uma mulher [...] ser portadora do gene defeituoso’ (Borba e Cançado 2005, vol. 3, p. 19).

‘A atuação torna-se ainda mais incerta quando mulheres não pertencentes a essas famílias são testadas para BRCA1 e verifica-se que são portadoras de alterações neste gene [...] Qual o risco de elas virem a desenvolver câncer? No momento não podemos responder a essa pergunta pois está se verificando que esse gene pode apresentar inúmeras alterações, e não se sabe ainda quais estão associadas com o desenvolvimento do câncer [...] outros genes, como o BRCA2, também associados com câncer de mama, estão sendo identificados’ (Borba e Cançado 2005, vol. 3, p. 19).

Nos exemplos acima, o ‘gene da DMD’ (Distrofia Muscular de Duchenne)’ e os genes BRCA1 e BRCA2 são formas alternativas da ausência de uma seqüência molecular normal que têm sido correlacionadas, em maior ou menor grau, a uma anormalidade fenotípica e, como genes-P, permitem falar preditivamente sobre fenótipos, com a diferença de que probabilidades baseadas em heredogramas ou genealogias foram suplantadas por probabilidades baseadas em sondagem molecular, como mencionado a seguir:

‘Antes da era da genômica, a genética trabalhava com o estudo das características – os fenótipos – para chegar aos genes, mapeá-los etc. Agora, essa nova genética segue o caminho inverso, pois a partir da decifração dos genes (DNA), chega-se às proteínas e às suas funções nos organismos’ (Silva Júnior e Sasson 2005, vol. 3, p. 173).

Citamos ainda, como uso válido do gene-P, o exemplo do ‘gene para fibrose cística’, freqüente nas abordagens da genética médica nas obras analisadas:

‘Para desenvolvê-la, a pessoa precisa ter as duas cópias do gene para fibrose cística alteradas, tanto a proveniente da mãe quanto a proveniente do pai [...] Graças ao desenvolvimento da Biologia Molecular foi possível isolar o gene responsável pela fibrose cística e compreender seu modo de ação’ (Amabis e Martho 2005, vol. 1, pp. 116-117).

‘Há poucos anos, ao contratar um seguro de saúde, uma norte-americana foi obrigada a realizar um teste de DNA. O exame revelou que ela era portadora de um gene recessivo para a fibrose cística. A doença, que ataca os pulmões e pode ser mortal, só se manifesta na presença de dois genes ligados à anormalidade’ (Paulino 2005, vol. 3, p. 128).

Diante dos exemplos acima, pelo fato de o ‘jogo’ explicativo do gene-P não ser confinado aos métodos puramente clássicos, torna-se mais fácil misturar seu significado com o de gene-D, considerada por Moss (2001, 2003) uma das fontes do determinismo genético, ou mesmo sobrepô-lo ao gene molecular clássico, caso freqüente em nossa análise. Sobreposições do gene molecular clássico e do gene-P e do gene molecular clássico e do gene informacional foram freqüentes em todas as obras analisadas, mas no primeiro caso, o percentual médio (39,06%) foi maior, tendo a sobreposição ultrapassado 50% em algumas obras (Tabela 3). Entre os exemplos de gene molecular clássico sobreposto a gene-P, citamos:

‘Atualmente, sabe-se que o **gene** (ou **gen**) é uma seqüência de nucleotídeos em um DNA. Cada gene é responsável pela síntese de uma proteína, e conseqüentemente, por uma ou mais características do indivíduo, pois as proteínas podem ter funções estruturais e reguladoras do metabolismo. Os genes estão alojados nos cromossomos e são didaticamente representados por letras, números e símbolos. Por exemplo, o gene para cor de pele normal é simbolizado por A e o gene para albinismo, por a’ (Cheida 2005, p. 283).

Nesse tipo de sobreposição, a função gênica é definida ‘de baixo para cima’, fazendo-se referência a um gene que codifica uma proteína específica, que, por redução explicativa, media a determinação genética do fenótipo. Casos como a pleiotropia podem ser acomodados, como no exemplo acima, com o reconhecimento de funções múltiplas para uma mesma proteína. É importante ressaltar que tal sobreposição aponta para um modelo híbrido de função gênica, visto que a explicação molecular clássica de gene implica apenas a colinearidade entre um gene e sua proteína específica, ou melhor, seu polipeptídeo ou RNA específico, mas silencia quanto à conexão deste gene a um fenótipo (Gericke e Hagberg 2007a).

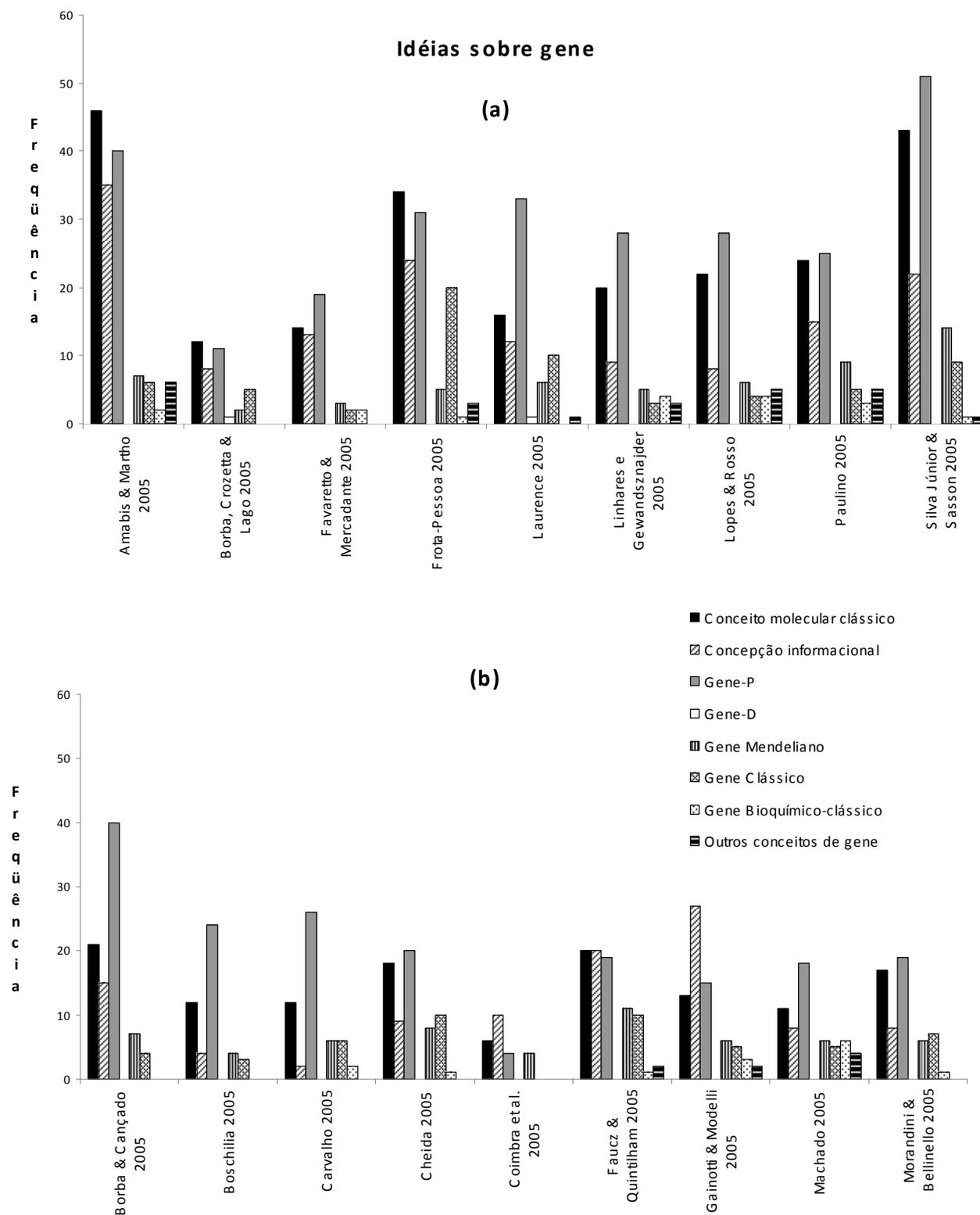


Figura 1. Frequência de ocorrência de conceitos de gene nas obras de biologia de ensino médio analisadas. Em (a), obras recomendadas e, em (b), obras excluídas do PNLEM/2007.

Na Figura 1, podemos ver também que o conceito molecular clássico foi o mais frequente em apenas 4 obras, sendo que, em uma delas, ele se mostrou tão frequente quanto a concepção

informativa de gene. Contudo, o gene molecular clássico foi o segundo conceito mais frequente nas mesmas 12 obras em que o gene-P predominou, sendo que sua frequência média, nas 18 obras, foi de 20,05 ocorrências/obra, a segunda maior encontrada em nossos dados. Esse resultado pode ser considerado até mesmo mais relevante do que a predominância do gene-P, uma vez que o gene molecular clássico permeia o conteúdo de biologia das três séries do ensino médio e seu número de ocorrências é relativamente elevado em todos os contextos em que genes são discutidos, e não apenas em trechos que tratam de procedimentos específicos da pesquisa genética, como é o caso do gene-P. Além disso, na maioria das vezes em que genes são definidos explicitamente pelos livros didáticos investigados, o conceito molecular clássico é empregado. As seguintes passagens oferecem exemplos de ocorrência do gene molecular clássico:

‘Um gene corresponde à porção da molécula de DNA capaz de codificar a síntese de uma proteína. De acordo com a seqüência de bases nitrogenadas que possui nos nucleotídeos que o constituem, o gene determina a seqüência de aminoácidos na molécula protéica produzida’ (Paulino 2005, vol. 1, p. 81).

‘O RNA pode ter função estrutural (nos ribossomos, por exemplo). Então, também é um gene o segmento de DNA que codifica a síntese de RNA estrutural, ainda que não venha a traduzir-se na produção de uma proteína’ (Favaretto e Mercadante 2005, p. 95).

‘Nas células do corpo humano, temos cerca de 30 mil genes diferentes. Portanto, é provável que também tenhamos, aproximadamente, 30 mil proteínas diferentes. O DNA, pelos genes, é que dita a síntese de proteínas nos seres vivos’ (Cheida 2005, vol. 3, p. 58).

‘Os genes são compostos, em média, por mil pares de nucleotídeos que aparecem em uma seqüência específica na molécula de DNA’ (Coimbra 2005, vol. 2, p. 86).

Em nove obras, o gene aparece como portador de informação para fenótipos, sem a menção de sua relação com *apenas* uma determinada proteína – em outras palavras, sem sobreposição também ao gene molecular clássico –, o que resulta num percentual médio de gene informativo sobreposto a gene-P de 8,89% (Tabela 3). Em seis dos livros analisados, por sua vez, ocorre gene molecular clássico sobreposto, simultaneamente, ao gene-P e ao gene informativo, como ilustra a seguinte passagem:

‘Cada gene é um segmento da molécula de DNA e contém a informação para a síntese de uma molécula de RNA, que poderá codificar determinado polipeptídeo ou proteína. Cada alelo de um gene contém diferenças nessas informações, codificando RNAs diferentes, porém relacionados ao caráter determinado por aquele gene’ (Lopes e Rosso 2005, p. 436).

Pela Figura 1, vemos ainda que os genes mendeliano, clássico e bioquímico-clássico foram menos frequentes, com médias de ocorrência para as dezoito obras de 6,39, 6,73 e 1,72 ocorrências/obra, respectivamente. Em geral, estas explicações de gene apareceram no contexto de descrições da história da biologia, como ilustram os seguintes exemplos de gene mendeliano:

‘Quando Mendel propôs o seu modelo para a hereditariedade, os fatores hereditários (genes) eram entidades puramente abstratas: ninguém, na verdade, sabia ao certo onde se encontravam, nem que processo fazia com que existissem duas ‘doses’ de cada gene nas células somáticas e uma ‘dose’ apenas nos gametas’ (Gainotti e Modelli 2005, p.278).

‘Mendel não vê os fatores cuja existência postula, apenas **deduz** sua presença pela análise dos cruzamentos’ (Silva Júnior e Sasson 2005, vol. 3, p.14).

Entre os exemplos de ocorrência do gene clássico, podemos citar:

‘[...] os ‘fatores de Mendel’ (os genes) deveriam estar contidos nos cromossomos. Desta forma, propuseram formalmente que os cromossomos continham os genes. A Teoria cromossômica da herança é um dos fundamentos da genética e explica o local onde se encontra o suporte físico dos princípios mendelianos’ (Faucz e Quintilham 2005, vol. 1, p. 51).

‘No começo do século XX, a Ciência encontrou os fatores hereditários no interior dos cromossomos das células, os genes’ (Cheida 2005, p. 283).

Por sua vez, o gene bioquímico-clássico pode ser exemplificado pela seguinte passagem:

‘Existem [...] espécies que formam hifas, como ocorre com a *Neurospora crassa*. Esta tornou-se clássica nos experimentos de Genética no início dos anos 1940 como modelo experimental em que os pesquisadores George Beadle e Edward Tatum propuseram que cada gene era responsável pela formação de uma enzima (um gene-uma enzima), tendo recebido por isso o prêmio Nobel em 1958’ (Lopes e Rosso 2005, p. 231).

Fora do contexto histórico, os genes mendeliano, clássico e bioquímico-clássico aparecem em versões atualizadas, sobrepostos, sobretudo, ao conceito molecular clássico de gene. Em 13 das obras analisadas, o gene molecular clássico apareceu sobreposto ao gene mendeliano, quando o primeiro era descrito como unidade hereditária (Tabela 3). Por exemplo, Paulino (2005) define, no glossário, gene como um segmento de DNA para a síntese de proteína, mas, ao definir cromossomo, o conceito molecular clássico aparece sobreposto ao gene mendeliano:

‘Cromossomo - É constituído de cromatina e contém os genes, que são as unidades hereditárias do organismo’ (p. 311).

Outro exemplo é encontrado na seguinte passagem:

‘Genes são apenas segmentos de DNA. Porém, são segmentos bem específicos que controlam as estruturas e funções celulares. Eles são considerados a unidade funcional da herança biológica; são a seqüência de bases que usualmente codificam para uma seqüência de aminoácidos’ (Faucz e Quintilham 2005, vol. 3, p. 15).

Similarmente, há exemplos de gene molecular clássico sobreposto ao gene clássico, quando genes são definidos como segmentos de DNA que se organizam em série, como as contas de um colar, ou como partículas que se enfileiram nos cromossomos – o que também traz implícita a idéia mendeliana de unidade. Como exemplos, podem ser citados os seguintes trechos:

‘Por intermédio da hereditariedade, os pais transmitem suas características genéticas para os filhos [...] A semelhança é determinada pelo DNA (ácido desoxirribonucléico), uma molécula linear que contém os genes, as partículas que transmitem esses padrões hereditários’ (Machado 2005, p. 15).

‘Os genes estão enfileirados nos cromossomos como as contas de um colar’ (Favaretto e Mercadante 2005, p. 143).

‘As *partículas hereditárias* que Mendel chamou de fatores hereditários são hoje conhecidas como *genes*’ (Morandini e Bellinello 2005, p. 402).

Tabela 3. Percentual de sobreposições do gene molecular clássico e da concepção informacional a outros conceitos de gene. Estas sobreposições foram relativamente as mais significativas em nossa análise.

Autores	Gene molecular clássico						Gene informacional
	Concepção informacional	Gene-P	Concepção informacional e gene-P	Gene mendeliano	Gene clássico	Gene bioquímico-clássico	Gene-P
Amabis e Martho 2005	36,9%	50,0%	2,2%	-	-	-	5,7%
Borba e Cançado 2005	38,0%	43,0%	-	-	4,8%	-	6,7%
Borba, Crozetta e Lago 2005	33,0%	41,7%	-	8,3%	8,3%	-	-
Boschilia 2005	8,3%	66,7%	-	-	-	-	-
Carvalho 2005	-	33,3%	-	8,3%	25,0%	8,3%	50,0%
Cheida 2005	22,2%	33,3%	-	-	16,7%	-	-
Coimbra et al. 2005	33,3%	10,0%	10,0%	10,0%	-	-	30,0%
Faucz e Quintilham 2005	30,0%	45,0%	10,0%	30,0%	20,0%	5,0%	25,0%
Favaretto e Mercadante 2005	28,6%	28,6%	-	-	14,3%	-	-
Frota-Pessoa 2005	26,5%	32,4%	-	5,9%	26,5%	-	4,2%
Gainotti e Modelli 2005	69,2%	30,8%	23,1%	7,7%	-	23,1%	14,8%
Laurence 2005	18,8%	37,5%	-	6,2%	31,2%	-	-
Linhares e Gewandsznajder 2005	25,0%	50,0%	5,0%	5,0%	10,0%	15,0%	11,1%
Lopes e Rosso 2005	9,09%	40,9%	4,5%	9,1%	13,6%	9,1%	12,5%
Machado 2005	9,09%	45,4%	-	18,2%	36,4%	-	-
Morandini e Bellinello 2005	5,88%	47,6%	-	11,8%	17,6%	-	-
Paulino 2005	25,0%	25,0%	-	4,2%	4,2%	12,5%	-
Silva Júnior e Sasson 2005	20,93%	41,9%	-	4,6%	7,0%	2,3%	-
Média (%) para as 18 obras	24,4%	39,1%	3,0%	7,2%	13,1%	4,2%	8,9%

A visão de que o gene é, ao mesmo tempo, uma partícula e uma sequência de nucleotídeos no DNA é incongruente e pode resultar em ambigüidades, configurando, assim, um exemplo de como o uso de modelos híbridos de função gênica – aqui manifestos na sobreposição de diferentes conceitos de gene – pode prejudicar a compreensão dos estudantes.

Por fim, registramos apenas duas ocorrências de gene-D nas obras analisadas (Figura 1), o que é compreensível, visto que a idéia de gene como recurso desenvolvimental está vinculada a uma abordagem recente na filosofia da biologia, que ainda não tem ampla circulação entre

geneticistas e biólogos moleculares, a saber, a teoria dos sistemas desenvolvimentais (Sterelny e Griffiths 1999; Oyama 2000; Oyama et al 2001). Pitombo et al. (2007) apresentam resultados de uma análise de conceitos de gene em livros textos de biologia celular molecular do ensino superior e também apontam para uma pequena ocorrência do conceito de gene-D. Logo, a ausência desse conceito no ensino médio é naturalmente esperada. Nos exemplos de gene-D encontrados em nossa análise, está implícita a idéia de gene como mera possibilidade herdada para uma característica, em virtude de sua interação com outros fatores – epigenéticos e ambientais – com os quais tem paridade causal, como exemplificado abaixo:

“É o veículo da hereditariedade ou a unidade de potencial hereditário existente no cromossomo e que, ao interagir com outros genes, com o citoplasma e o meio, condiciona o aparecimento de um dado caráter. Os genes distribuem-se linearmente ao longo de cada cromossomo e possuem como constituinte fundamental o DNA que tem função genética” (Borba, Crozetta e Lago 2005, p.259).⁶⁰

É interessante notar neste exemplo que mesmo uma abordagem mais moderna sobre genes aparece sobreposta a idéias mendelianas e clássicas, o que aponta, mais uma vez, a extensão da hibridização de idéias sobre genes no conhecimento escolar veiculado nestes livros didáticos.

Alguns conceitos de gene tiveram ocorrência insignificante ou não foram encontrados nos livros analisados (categoria ‘Outros’, na Figura 1). Um exemplo é o ‘gene evolutivo’, que foi introduzido por Williams (1996) e elaborado por Dawkins (1982). De acordo com este conceito, o gene é um segmento qualquer de DNA, começando e terminando em pontos arbitrários de um cromossomo, que compete com segmentos alelomórficos pela região do cromossomo em questão. Trata-se de um conceito que foi bastante influente na biologia evolutiva, apesar das críticas lançadas contra ele (ver Griffiths e Neumann-Held 1999; Sterelny e Griffiths 1999) e da diminuição de sua influência em anos recentes. Ele enfatiza que servir como unidade de seleção pode ser uma função do gene. A ausência desse conceito nos livros analisados não surpreende, visto que a evolução não cumpre papel organizador do conhecimento biológico na amostra de livros avaliada no PNLEM/2007, como discutem Rocha e colaboradores (Rocha *et al.* 2007). E este conceito está vinculado a uma interpretação de uma tradição de pesquisa bastante específica, a biologia evolutiva.

Ainda quanto à categoria ‘Outros’, entre os conceitos raros ou ausentes, podemos citar algumas idéias mais recentes sobre genes, formuladas na biologia ou na filosofia da biologia e que podem ser inseridas no modelo Moderno de função gênica, de acordo com a classificação de Gericke e Hagberg (2007b): o ‘conceito molecular processual de gene’; proposto por Neumann-Held (2001); o ‘conceito sistêmico de gene’, proposto por Pardini e Guimarães (1992); o ‘Conjunto de Domínios para Transcrição Ativa’, sugerido por Fogle (1990, 2000); e o entendimento de genes como signos como propõem El-Hani e colaboradores (El-Hani et al. 2006). Estes conceitos capturam importantes aspectos da complexidade e diversidade genômica e sua abordagem em livros didáticos de ensino médio, mesmo sem maior aprofundamento, poderia levar a uma melhor compreensão da natureza e da dinâmica dos sistemas genômicos.

5.2. Modelos de função gênica

Cada modelo de função gênica, em geral, encerra uma combinação particular de funções atribuídas ao gene, e algumas destas são compartilhadas por dois ou mais modelos. Portanto, nos livros analisados, menções a alguma função gênica foram consideradas como fortes indícios, mas

⁶⁰ Vale ressaltar que, conforme pudemos inferir do contexto de explicação, o verbo ‘condicionar’, aqui, não significa ‘determinar’ como nos trechos citados mais acima.

não instâncias incontestáveis de um modelo particular de função gênica. Antes, as ocorrências de funções atribuídas ao gene foram interpretadas como mais um elemento a se ter em conta, ao lado das ocorrências de conceitos de gene, para a investigação da combinação de características epistemológicas presentes, que forneceria um quadro mais complexo, a saber, o modelo de função gênica para um dado contexto. Em vista disso, conforme fizemos com os conceitos de gene na seção anterior, nossos resultados quanto às idéias sobre função gênica são primeiramente apresentados independentemente do modelo de função gênica implícito. Em seguida, mostramos a combinação desses primeiros resultados em modelos de função gênica para cada categoria de contexto.

5.2.1. Frequência de ocorrência de funções atribuídas ao gene

A Figura 2 mostra a predominância da função gênica de causar ou determinar fenótipo ou diferença fenotípica, cuja frequência média foi de 29,22 ocorrências/obra, seguida da visão da função de um gene como a de codificar a estrutura primária de polipeptídeos ou RNAs, com frequência média de 22,78. Estes achados são condizentes com a predominância dos conceitos de gene-P e de gene molecular clássico, aos quais essas idéias estão respectivamente relacionadas.

Entre as funções mais mencionadas, estão também as seguintes, por ordem de frequência: ‘transmitir caracteres hereditários’, ‘controlar metabolismo celular’ e constituir um ‘programa ou instrução para a função celular e/ou o desenvolvimento’ (Figura 2), cujas frequências médias foram, respectivamente, 8,28, 4,33 e 3,89 ocorrências/obra. Aqui, não houve uma correspondência exata com a ordem das frequências de conceitos de gene: o gene informacional, que foi o terceiro conceito de gene mais frequente, não está vinculado à terceira função gênica mais frequente – ‘transmitir caracteres hereditários’ – mas sim, às duas frequências seguintes de funções gênicas – ‘controlar metabolismo celular’ e constituir um ‘programa ou instrução para a função celular e/ou o desenvolvimento’. Esta inversão pode ser explicada pelo fato de que, embora a função de ‘transmitir caracteres hereditários’ seja mais diretamente associada aos genes mendeliano e clássico, que foram menos frequentes do que o gene informacional em nossa análise, ela também aparece vinculada aos conceitos mais frequentes de gene-P e de gene molecular clássico, implícita na sobreposição destes à idéia de gene como unidade de herança.

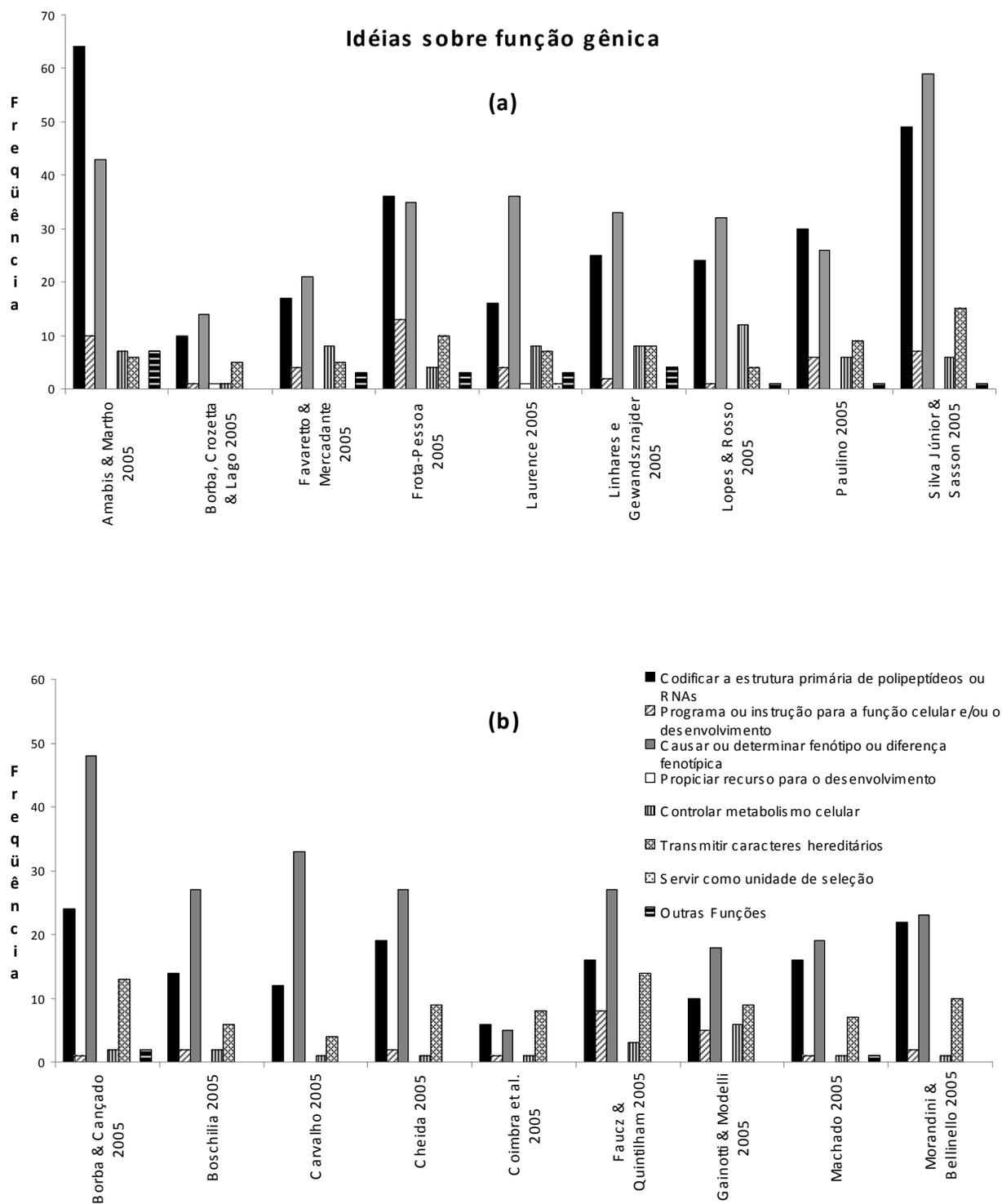


Figura 2. Frequência de ocorrência de idéias sobre função gênica nas obras de ensino médio analisadas. Em (a), obras recomendadas e, em (b), obras excluídas do PNLEM/2007.

Similarmente às frequências de ocorrência de gene-D e de gene evolutivo, as frequências de ocorrências da função de ‘propiciar recurso para o desenvolvimento’ e a de ‘servir como unidade de seleção’ não foram significativas. Funções gênicas associadas às explicações de genes mais recentes não foram encontradas (Figura 2).

5.2.2. Modelos de função gênica por categoria de contexto

Observamos nos livros analisados que, em geral, modelos de função gênica são abordados como se fossem eles próprios os fenômenos. Não se explicita o papel dos modelos, os propósitos específicos e as limitações de cada um, nem as diferenças entre modelo e realidade, perdendo-se de vista aspectos da natureza da ciência e do conhecimento científico que têm papel importante na educação em ciências.

Em alguns casos, contudo, um tratamento mais adequado dos modelos de função gênica, tendo-se na devida conta ao menos parte dos aspectos acima, é encontrado. Os seguintes trechos servem de exemplo:

‘Embora o número de exceções, ao que parece seja maior do que o de regras, as leis de Mendel têm e tiveram o imenso mérito de constituir um modelo inicial, que explicou de forma coerente – pela primeira vez – os mecanismos hereditários. O fato de que esse modelo tem sido reformulado e adaptado, à medida que novos conhecimentos surgem, sem dúvida não torna as leis mendelianas menos importantes’ (Silva Júnior e Sasson 2005, vol. 3, p. 131).

É importante considerar, além disso, que não foi registrada a ocorrência de nenhum modelo histórico genuíno, ou seja, os modelos de função gênica presentes nas obras não correspondiam a modelos efetivamente propostos na história da genética. Em todas as obras, há mistura e transferência de atributos de diferentes modelos de função gênica que podem ser identificados na história da genética. O resultado é a prevalência nos livros de modelos híbridos. Estes, por sua vez, implicam a idéia de que o conhecimento científico cresce linearmente e é independente do contexto de sua produção (Gericke e Hagberg 2007b). Se assim fosse, acreditar-se-ia que o somatório dos modelos históricos – mesmo aqueles que se contradizem entre si e são oriundos de contextos distintos – se aproximaria assintoticamente da ‘melhor descrição’ de um fenômeno. Dessa forma, os diversos significados atribuídos ao gene em diferentes contextos dos livros didáticos poderiam ser interpretados como partes de um todo coerente. Porém, adotando essa perspectiva, os estudantes não adquirem a capacidade de pensar *sobre* modelos, mas apenas de pensar *com* modelos (*ibid.*), cujas fronteiras são fracamente definidas. A aprendizagem torna-se, assim, confusa e prejudicada. A idéia equivocada de que combinar diferentes modelos conduz a um todo coerente impossibilita que qualquer abordagem histórica consistente das ciências seja feita, visto que não se pode ter acesso, dessa maneira, a nenhuma mudança ou progressão entre os diferentes modelos (*ibid.*).

Em todos os livros, nos modelos presentes nos cinco contextos – ‘glossário’, ‘caracterização da vida’, ‘biologia celular e molecular’, ‘genética’ e ‘evolução’ – foram encontrados atributos de dois ou mais modelos históricos (ver Tabelas 4 a 8, derivadas da Tabela 2, e que mostram o modelo encontrado em cada contexto, para o conjunto de obras analisadas). Ademais, nossos resultados apontam que os cinco modelos de função gênica das categorias de contexto possuem variantes epistemológicas não-históricas, indicando que o uso de modelos híbridos nos livros didáticos de ensino médio é a regra, e não a exceção.

As variantes epistemológicas do modelo histórico Neoclássico predominaram nos cinco contextos e, em geral, ocorreram em um número maior de obras analisadas, em relação às

variantes dos outros modelos históricos. Em vista disso, os modelos encontrados nos cinco contextos podem ser categorizados como mais próximos do modelo Neoclássico (borda destacada em negrito nas Tabelas 4 a 8), no qual se incluem, como discutido acima, o conceito molecular clássico e a concepção informacional de gene. A diferença entre os contextos reside, então, no percentual de hibridização do modelo Neoclássico, na frequência e no percentual de sobreposição entre o gene molecular clássico e o informacional, e na abordagem da relação genótipo-fenótipo, traduzida nas idéias não-históricas sobre função gênica. Neste último quesito, importa, sobretudo: se o modelo inclui entidades nos níveis macro e molecular e se há redução explicativa do primeiro ao segundo (2Icy* e 6bx); e se o modelo implica determinismo genético ou interacionismo aditivo entre genes e fatores ambientais (7ax).

Quatro das obras analisadas (Faucz e Quintilham 2005; Frota-Pessoa 2005; Paulino 2005; Silva Júnior e Sasson 2005) apresentam glossário, contexto no qual o modelo Neoclássico possui percentual de hibridização de 27% - apenas três do total de 11 tipos de variantes ocorrentes não pertencem a este modelo (Tabela 4). Pela Tabela 9, vemos que, dentro desse contexto, predomina o gene molecular clássico (34,6%), seguido do gene informacional (26,9%) e da descrição do gene como unidade de herança (23,1%), aspecto inerente ao gene mendeliano. Em contrapartida, o conceito molecular clássico se sobrepõe, mais frequentemente, à idéia de unidade hereditária (56%), seguida da idéia de segmento de DNA como informação para uma proteína (33%). Embora a frequência de sobreposição entre gene molecular clássico e gene-P, ao longo dessas quatro obras, seja a maior dentre suas sobreposições (Tabela 3), ao menos no glossário, a redução explicativa do fenótipo ao nível molecular não é comum. Ainda assim, e apesar do glossário ser o contexto de menor percentual de hibridização do modelo Neoclássico, há exemplos nele contidos, com enorme grau de sobreposições, como o excerto que segue abaixo:

‘Gene: porção do DNA que contém a informação necessária para a realização de uma característica genética específica. Unidade biológica que definem características específicas de cada indivíduo, controlando estruturas e funções celulares. Unidade funcional da hereditariedade’ (Faucz e Quintilham, vol.1, glossário).

Nos capítulos de introdução à biologia e à origem e diversidade da vida – categoria de contexto ‘caracterização da vida’ –, predomina fortemente o gene informacional (64,4%), enquanto as idéias de gene como determinante de fenótipo (11,9%) e gene como segmento de DNA que codifica um produto funcional (13,6%) também estão presentes, mas de maneira secundária (Tabela 9). Nesse contexto, é menos comum a menção de detalhes moleculares dos sistemas genéticos, os quais são apresentados, em particular, como fornecedores das informações para o desenvolvimento dos organismos, permitindo a transmissão de características entre as gerações. O grau de hibridização do modelo Neoclássico, nesse contexto, é de 47% - oito das 17 variantes epistemológicas encontradas não são deste modelo (Tabela 5).

No contexto em que são tratados conteúdos de biologia celular e molecular, a concepção informacional de gene também predomina (45,2%), mas é pequena a diferença relativa ao conceito molecular clássico (38,1%), o que também pode ser ilustrado se compararmos a diferença de sombreamento das variantes 1c e 1e* nas tabelas 4 a 8. Neste contexto, são enfatizados, de maneira causal e mecanicista, os processos que resultam na transcrição de RNAs e tradução de polipeptídeos. Este é o contexto que traz maior sobreposição entre o gene molecular clássico e a concepção informacional (48%), além de já apresentar maior ocorrência de visões deterministas sobre a etiologia dos traços, em relação aos dois contextos anteriores (comparar variantes 6bx e 7ax nas Tabelas 4, 5 e 6). Neste contexto, a sobreposição do gene molecular clássico ao gene-P é de 21%. Adicionalmente, pode-se afirmar que é no contexto da biologia celular e molecular que o modelo Neoclássico é abordado de forma mais consistente,

visto que suas variantes epistemológicas aparecem num maior número de obras e seu grau de hibridização (44%) só é maior do que nos contextos do glossário e de evolução, nos quais as informações sobre gene são mais sucintas e as variantes epistemológicas, abordadas em menor número de obras.

Ao compararmos nossos resultados para as três primeiras categorias com a análise de Pitombo e colaboradores (no prelo) de conceitos de gene em livros-texto de biologia celular e molecular do ensino superior, encontramos algumas semelhanças: na análise destes autores, a concepção informacional de gene e o conceito molecular clássico, em seguida, predominam, sobretudo nos contextos de biologia e genética molecular. Além disso, estes dois conceitos estão presentes numa maior diversidade de contextos do que outras idéias sobre gene e, nos glossários, predomina o gene molecular clássico. Por fim, nos resultados relatados por estes autores, o modelo de função gênica predominante atribuído ao gene a função de codificar a estrutura primária de RNAs e proteínas. Esses resultados corroboram a prevalência do modelo Neoclássico de função gênica, tanto no ensino médio, quanto no superior, e, no que se refere ao gene informacional, apontam quão difundida é esta concepção, a despeito da ausência de uma sólida fundamentação teórica para a atribuição de significado ao conceito de informação na biologia contemporânea.

Nossos resultados mostram, ainda, que o contexto da genética apresenta algumas peculiaridades. Nele, a frequência de ocorrência do gene-P é mais expressiva, o que é devido, em grande parte, ao conteúdo extenso dos capítulos de genética nos livros analisados contendo grande quantidade de exercícios que tratam de heredogramas, como discutido em 5.1. No conjunto das 18 obras, há cerca de 400 ocorrências de gene-P no contexto da genética, enquanto o número de ocorrências nos demais contextos não ultrapassam 25. Em consequência, neste contexto, o gene-P predomina, e com grande vantagem, sobre os outros conceitos de gene (ver Tabela 9): o percentual de ocorrência do gene-P (45%) é quase o dobro da segunda maior frequência neste contexto, a saber, a do gene molecular clássico (24%). Este último, por sua vez, encontra-se, em 64% dos casos, sobreposto a gene-P, o que mostra a prevalência de idéias relacionadas ao determinismo genético neste contexto. Esta interpretação também é apoiada, como se pode ver nas tabelas 4 a 8 (*cf.* a variante *6bx*), pela redução explicativa do nível fenotípico ao nível molecular e pela relação entre as ocorrências das variantes 3a e 3b, relativas à definição do gene ‘de cima para baixo’ e ‘de baixo para cima’, respectivamente. No contexto de ‘genética’, a variante *6bx* aparece no maior número de obras analisadas em relação aos outros contextos, o que também vale para as variantes 3a e 3b, sendo que neste contexto, 3a aparece em um número de obras similar ao de 3b (Tabela 7). É também no contexto de genética, contudo, que fatores não-genéticos passam a ser considerados, aplicando-se a fórmula ‘fatores ambientais + fatores genéticos resultam num traço/produto/função’ (variante *7ax*). Isso mostra como este modo de reconhecer o papel do ambiente, por privilegiar a compreensão da relação entre fatores genéticos e ambientais de maneira aditiva, não é suficiente para que se deixe de lado o determinismo genético (*cf.* El-Hani, 1995, 1997). Por conseguinte, no contexto da genética, sobressaem diferentes nuances de pré-formacionismo, seja no uso do gene-P ou nas abordagens de gene mendeliano (11,4%) e gene clássico (10,9%), que, embora escassos, têm, em geral, maior frequência do que nos outros contextos (Tabela 9). Ademais, o elevado percentual de hibridização do modelo Neoclássico (58%) neste contexto (ver Tabela 7) indica que, em prol de uma linguagem que favoreça a herança e a previsão de fenótipos a partir de genes, livros didáticos de ensino médio podem não contemplar adequadamente a relevância de fatores epigenéticos na determinação de traços.

Tabela 4. Modelo de função gênica encontrado no contexto ‘glossário’.

Modelos históricos de função gênica	Características epistemológicas									
	1		2I	2II	3	4	5I	5II	6	7
Modelo Mendeliano	1a		2Ia	2IIa	3a	4a	5Ia	5IIa	6a	7a
Modelo Clássico	1b		2Ib	2IIb	3a	4b	5Ia	5IIa	6b	7a
Modelo Bioquímico-clássico	1b		2Ib	2IIa	3a	4c	5Ia	5IIa	6b	7a
				2IIb						
Modelo Neoclássico	1c	1e*	2Ic	2IIa	3b	4d	5Ib	5IIa	6c	7b
Modelo Moderno	1d		2Ic	2IIa	3c	4d	5Ib	5IIb	6c	7c
<i>Idéias não-históricas sobre função gênica</i>				2Icy*					6bx	

⁽ⁱ⁾ Modelo identificado: Neoclássico (borda em negrito). ⁽ⁱⁱ⁾ Sombreamento indica variantes ocorrentes no contexto. Cinza claro: ocorrências em 1-6 obras. ⁽ⁱⁱⁱ⁾ Percentual de hibridização: 27% (percentual dos tipos de variantes que não estão associados ao modelo Neoclássico, dentre os tipos de variantes ocorrentes no contexto). ^(iv) Variantes com asteriscos e as italicizadas têm o mesmo significado da Tabela 1.

Tabela 5. Modelo de função gênica encontrado no contexto ‘caracterização da vida’.

Modelos históricos de função gênica	Características epistemológicas									
	1		2I	2II	3	4	5I	5II	6	7
Modelo Mendeliano	1a		2Ia	2IIa	3a	4a	5Ia	5IIa	6a	7a
Modelo Clássico	1b		2Ib	2IIb	3a	4b	5Ia	5IIa	6b	7a
Modelo Bioquímico-clássico	1b		2Ib	2IIa	3a	4c	5Ia	5IIa	6b	7a
				2IIb						
Modelo Neoclássico	1c	1e*	2Ic	2IIa	3b	4d	5Ib	5IIa	6c	7b
Modelo Moderno	1d		2Ic	2IIa	3c	4d	5Ib	5IIb	6c	7c
<i>Idéias não-históricas sobre função gênica</i>			2Ibx	2Icy*					6bx	7ax

⁽ⁱ⁾ Modelo identificado: Neoclássico (borda em negrito). ⁽ⁱⁱ⁾ Sombreamento indica variantes ocorrentes no contexto. Cinza claro: ocorrências em 1-6 obras. Tom intermediário de cinza: ocorrências em 7-12 obras. Cinza escuro: ocorrências em 13-18 obras. ⁽ⁱⁱⁱ⁾ Percentual de hibridização: 47% (percentual dos tipos de variantes que não estão associados ao modelo Neoclássico, dentre os tipos de variantes ocorrentes no contexto). ^(iv) Variantes com asteriscos e as italicizadas têm o mesmo significado da Tabela 1.

Tabela 6. Modelo de função gênica encontrado no contexto ‘biologia celular e molecular’.

Modelos históricos de função gênica	Características epistemológicas									
	1	2I	2II	3	4	5I	5II	6	7	
Modelo Mendeliano	1a	2Ia	2IIa	3a	4a	5Ia	5IIa	6a	7a	
Modelo Clássico	1b	2Ib	2IIb	3a	4b	5Ia	5IIa	6b	7a	
Modelo Bioquímico-clássico	1b	2Ib	2IIa	3a	4c	5Ia	5IIa	6b	7a	
			2IIb	3b						
Modelo Neoclássico	1c	1e*	2Ic	2IIa	3b	4d	5Ib	5IIa	6c	7b
Modelo Moderno	1d	2Ic	2IIa	3c	4d	5Ib	5IIb	6c	7c	
<i>Idéias não-históricas sobre função gênica</i>			<i>2Icy*</i>					<i>6bx</i>	<i>7ax</i>	

⁽ⁱ⁾ Modelo identificado: Neoclássico (borda em negrito). ⁽ⁱⁱ⁾ Sombreamento indica variantes ocorrentes no contexto. Cinza claro: ocorrências em 1-6 obras. Tom intermediário de cinza: ocorrências em 7-12 obras. Cinza escuro: ocorrências em 13-18 obras. ⁽ⁱⁱⁱ⁾ Percentual de hibridização: 44% (percentual dos tipos de variantes que não estão associados ao modelo Neoclássico, dentre os tipos de variantes ocorrentes no contexto). ^(iv) Variantes com asteriscos e as italicizadas têm o mesmo significado da Tabela 1.

Tabela 7. Modelo de função gênica encontrado no contexto de ‘genética’.

Modelos históricos de função gênica	Características epistemológicas									
	1	2I	2II	3	4	5I	5II	6	7	
Modelo Mendeliano	1a	2Ia	2IIa	3a	4a	5Ia	5IIa	6a	7a	
Modelo Clássico	1b	2Ib	2IIb	3a	4b	5Ia	5IIa	6b	7a	
Modelo Bioquímico-clássico	1b	2Ib	2IIa	3a	4c	5Ia	5IIa	6b	7a	
			2IIb	3b						
Modelo Neoclássico	1c	1e*	2Ic	2IIa	3b	4d	5Ib	5IIa	6c	7b
Modelo Moderno	1d	2Ic	2IIa	3c	4d	5Ib	5IIb	6c	7c	
<i>Idéias não-históricas sobre função gênica</i>		<i>2Ibx</i>	<i>2Icy*</i>					<i>6bx</i>	<i>7ax</i>	

⁽ⁱ⁾ Modelo identificado: Neoclássico (borda em negrito). ⁽ⁱⁱ⁾ Sombreamento indica variantes ocorrentes no contexto. Cinza claro: ocorrências em 1-6 obras. Tom intermediário de cinza: ocorrências em 7-12 obras. Cinza escuro: ocorrências em 13-18 obras. ⁽ⁱⁱⁱ⁾ Percentual de hibridização: 58% (percentual dos tipos de variantes que não estão associados ao modelo Neoclássico, dentre os tipos de variantes ocorrentes no contexto). ^(iv) Variantes com asteriscos e as italicizadas têm o mesmo significado da Tabela 1.

Tabela 8. Modelo de função gênica encontrado no contexto ‘evolução’.

Modelos históricos de função gênica	Características epistemológicas									
	1		2I	2II	3	4	5I	5II	6	7
Modelo Mendeliano	1a		2Ia	2IIa	3a	4a	5Ia	5IIa	6a	7a
Modelo Clássico	1b		2Ib	2IIb	3a	4b	5Ia	5IIa	6b	7a
Modelo Bioquímico-clássico	1b		2Ib	2IIa	3a	4c	5Ia	5IIa	6b	7a
				2IIb	3b					
Modelo Neoclássico	1c	1e*	2Ic	2IIa	3b	4d	5Ib	5IIa	6c	7b
Modelo Moderno	1d		2Ic	2IIa	3c	4d	5Ib	5IIb	6c	7c
<i>Idéias não-históricas sobre função gênica</i>				2Icy*					<i>6bx</i>	<i>7ax</i>

⁽ⁱ⁾ Modelo identificado: Neoclássico (borda em negrito). ⁽ⁱⁱ⁾ Sombreamento indica variantes ocorrentes no contexto. Cinza claro: ocorrências em 1-6 obras. Tom intermediário de cinza: ocorrências em 7-12 obras. Cinza escuro: ocorrências em 13-18 obras. ⁽ⁱⁱⁱ⁾ Percentual de hibridização: 33% (percentual dos tipos de variantes que não estão associados ao modelo Neoclássico, dentre os tipos de variantes ocorrentes no contexto). ^(iv) Variantes com asteriscos e as italicizadas têm o mesmo significado da Tabela 1.

Tabela 9. Percentual de ocorrência de conceitos de gene por categoria de contexto.

Categoria de Contexto	Conceitos de Gene						
	Conceito molecular clássico	Concepção informacional	Gene-P	Gene-D	Gene mendeliano	Gene clássico	Gene bioquímico-clássico
Glossário	34,6%	26,9%	7,7%	-	23,1%	7,7%	-
Caracterização da vida	13,6%	64,4%	11,9%	-	5,1%	3,4%	1,7%
Biologia celular e molecular	38,1%	45,2%	7,4%	0,3%	1,3%	4,2%	3,5%
Genética	24,0%	6,5%	45,0%	0,1%	11,4%	10,9%	2,1%
Evolução	32,4%	14,7%	50,0%	2,9%	-	-	-

Quando conteúdos de evolução são abordados, predomina o gene-P (50%), seguido do gene molecular clássico (32,4%), enfatizando-se que um dado fenótipo, mais particularmente uma anormalidade fenotípica, é produzido quando ‘a’ seqüência molecular normal está ausente, com foco nas alterações das estruturas moleculares do DNA (mutações). Em outras palavras, não há referência a formas alternativas da ausência do recurso molecular considerado normal e o gene no DNA é abordado como determinante de fenótipo. A discussão sobre genes e fenótipos é conduzida empregando-se invalidamente o gene-P, que é sobreposto à grande parte das ocorrências de gene molecular clássico. Portanto, há redução explicativa do nível fenotípico ao nível molecular, ainda que esta não seja tão acentuada quanto nos contextos de ‘genética’ e de ‘biologia celular e molecular’ (*cf.* variante *6bx* nas Tabelas 6, 7 e 8).

Gericke e Hagberg (2007a) usam os modelos históricos que identificaram em trabalho anterior (Gericke e Hagberg 2007b) para investigar como a função gênica é abordada em livros didáticos de biologia e química de ensino médio da Suécia. Nós utilizamos a mesma metodologia

básica destes autores, partindo da investigação da presença das características epistemológicas de modelos históricos de função gênica, em um conteúdo pré-definido que abordava os genes para, em seguida, identificarmos o modelo de função gênica implícito neste conteúdo. E da mesma forma, estimamos o grau de hibridização de tal modelo, com base na presença de variantes típicas de outros modelos históricos. Em sua análise, eles descrevem modelos de função gênica para cada capítulo no qual o conteúdo sobre gene é abordado e seus resultados mostram que, em sua maioria, os modelos apresentados são híbridos e podem ser categorizados, com base na predominância de variantes epistemológicas, como Neoclássicos. Considerando que quanto menor o conteúdo, menor a probabilidade de variação do significado do gene, é compreensível que na análise sueca tenham sido classificados, para cada um dos capítulos examinados, outros modelos além do Neoclássico. Afinal, diferentemente do que ocorre para um contexto mais amplo que envolve vários capítulos, é mais provável que em cada capítulo se enfatize um conjunto mais restrito de aspectos, que podem ser associados a um dado modelo histórico de função gênica.

Por exemplo, Gericke e Hagberg identificaram, para alguns capítulos, modelos mendelianos pouco híbridos ou mesmo totalmente coerentes com o modelo histórico. Em nossa análise, os atributos desse modelo, quando presentes, contribuíram para aumentar o percentual de hibridização do modelo predominante, o Neoclássico, num contexto maior que incluía vários capítulos. Em ambas as análises, o modelo Neoclássico aparece de forma mais consistente nos contextos de citologia e biologia molecular. Isso significa que entre os menores percentuais de hibridização está o do modelo Neoclássico, tanto o identificado em nossa categoria de contexto ‘biologia celular e molecular’ quanto os identificados em alguns dos capítulos que compõem as seções de citologia e biologia molecular dos livros suecos. Particularmente em nossa análise, este modelo predomina também nas outras categorias de contexto, mesmo nas de genética e evolução, nas quais predomina, na análise dos livros suecos, o modelo Mendeliano. Isso se deve à maior relevância que atribuímos, em comparação com aqueles autores, ao entendimento atual dos genes – que, em nossa análise aparece em todas as categorias de contexto e se vincula ao modelo Neoclássico. Nos livros que analisamos, nenhum traz a idéia do gene como uma entidade abstrata, sem estrutura conhecida, a não ser em contexto histórico, e mesmo as idéias dos genes como ‘unidades de herança’ e ‘partículas no cromossomo’, quando aparecem, são sobrepostas ao gene molecular clássico.

Assim como Gericke e Hagberg (2007a), que observaram uma baixa frequência de ocorrência do modelo Moderno nos livros suecos, e apenas em capítulos de citologia e biologia molecular, em nossa análise também registramos uma baixa frequência de variantes epistemológicas exclusivas deste modelo. Basicamente, registramos apenas a variante 1d – ‘o gene consiste de um ou vários segmentos de DNA com vários propósitos’ –, que apareceu em cerca de 7-12 obras no contexto de biologia celular e molecular (*cf.* 1d na Tabela 6), quando os autores consideravam desafios ao conceito molecular clássico de gene, como seqüências não-codificantes de DNA com função conhecida, como os promotores, genes interrompidos, *splicing* e *splicing* alternativo promotores, íntrons etc. Em outros contextos, sua ocorrência foi pequena ou nula. As variantes do modelo Moderno 3c, 5Ib e 7c, que focam os *processos* desenvolvimentais dos quais depende a expressão gênica, não foram observadas em qualquer contexto de nossa análise. Este resultado indica que, nas obras que analisamos, as conseqüências da complexidade dinâmica dos genomas não são abordadas em seu devido alcance e sequer são relacionadas à dependência dos genes ao contexto celular. Prevalece a abstração dessa complexidade, bem como da complexidade estrutural dos genomas, em prol do foco numa relação de um para um entre gene, proteína e fenótipo que não se aplica na maioria dos processos desenvolvimentais necessários à produção das características dos organismos. Assim, as obras que analisamos dão

lugar a visões deterministas genéticas, que também foram relatadas com frequência na análise de Gericke e Hagberg (2007a).

A diferença metodológica em relação à nossa análise, na qual foram identificados modelos de função gênica para categorias de contexto, e não para capítulos, não impediu a comparação de nossos resultados com aqueles obtidos por estes autores, que mostrou claramente uma similaridade substancial na abordagem da função gênica em livros didáticos brasileiros e suecos.

5.3. O papel do ambiente: visões deterministas e interacionistas sobre a origem das características

O entendimento da relação genótipo-fenótipo que se pode discernir das sobreposições entre gene molecular clássico, gene informacional e gene-P encontradas em nossos resultados (ver acima) está intimamente relacionado com a compreensão do DNA como um ‘programa para o desenvolvimento’ ou ‘para a função celular’, de acordo com a qual toda a instrução para a construção e o funcionamento de um ser vivo estaria depositada nas seqüências de nucleotídeos daquela molécula. Conforme El-Hani, Queiroz e Emmeche (2006) argumentam, quando a informação é concebida dessa forma, o “DNA se torna um tipo de reservatório do qual toda ‘informação’ numa célula flui e ao qual ela deve ser, em última análise, reduzida” (p. 4). Assim, a maquinaria celular seria apenas uma ferramenta necessária para ‘rodar’ o programa genético, afetando-o, grosso modo, somente por sua presença ou ausência. Este modo de compreender o desenvolvimento e a função celular foi objeto de muitas críticas (e.g., Oyama 2000; Nijhout 1990; Sarkar 1996; El-Hani 1997; Keller 2000) e tem sido uma das fontes mais importantes das visões deterministas genéticas, tanto na comunidade científica, quanto no domínio público, em termos gerais. Nossos resultados indicam que os livros didáticos de biologia do ensino médio têm contribuído para a construção e perpetuação na sociedade de um discurso determinista sobre genes e seu papel nos sistemas biológicos. Afinal, este discurso está presente em quase todas as obras analisadas (17 das 18 obras), como os seguintes trechos ilustram:

‘Apenas recentemente as novas descobertas sobre o código mostraram como ele é mesmo um programa sofisticado [...] Quando falamos em código e decodificação estamos sendo quase literais. As instruções gênicas estão armazenadas no DNA e no RNA, ambos ácidos nucléicos’ (Amabis e Martho 2005, vol. 1, p. 258).

‘A estrutura conhecida como **gene** corresponde a um segmento ou pedaço da molécula de DNA [...] Nos genes estão as informações responsáveis pelas características do indivíduo, como a cor dos olhos, a cor dos cabelos, a forma do nariz e, no caso de uma aranha, até mesmo o tipo de teia que ela tece para capturar suas presas. Desse modo, o DNA funciona à semelhança de um programa de computador e o organismo corresponderia a um computador que trabalhasse segundo às ordens do DNA’ (Linhares e Gewandszajder 2005, p. 14).

‘De que jeito as ‘informações próprias de cada espécie’ são conservadas de geração em geração? [...] O DNA existe em praticamente todos os seres vivos e contém a programação genética da espécie, ou seja, todas as informações relativas às características daquele tipo de ser vivo’ (Silva Júnior e Sasson 2005, vol.1, p. 18).

Em uma das obras (Paulino 2005), a metáfora do DNA como programa chega a ser descrita como ‘poética e realística’ numa questão de vestibular, da Universidade Federal do Rio de Janeiro, reproduzida pelo livro, que cita, sem datar, Dawkins:

‘Está chovendo DNA lá fora. Estão chovendo instruções lá fora; estão chovendo programas para a formação de novas árvores’ (Dawkins, citado por Paulino 2005, vol. 1, p. 87).

A idéia de instrução ou de programa genético também pode aparecer, de forma mais atenuada, quando se aborda o contexto celular, como nos trechos abaixo:

‘Uma bactéria, uma samambaia, um cachorro ou qualquer outro ser vivo desenvolvem as características típicas de suas espécies a partir das instruções genéticas e do arcabouço celular que receberam pela reprodução de seus pais’ (Amabis e Martho 2005, vol. 1, p.6).

‘Célula, unidade mínima de um organismo capaz de atuar de maneira autônoma, sendo a responsável pela aquisição de seu sustento e realizando funções para as quais foi geneticamente programada’ (Faucz e Quintilham 2005, vol. 1, p. 29).

Na verdade, como apontam Sterelny & Griffiths (1999), não há deterministas num sentido extremo, no qual adotariam uma versão caricatural do determinismo genético, que consistiria na visão de que há fatores biológicos (usualmente genes) cuja presença em um organismo significa que, não importando quais outros fatores estejam presentes, certo resultado seguirá. Assim, a expressão ‘determinismo genético’ é aplicável, de fato, a visões mais moderadas, e frequentemente mais vagas, como aquelas encontradas nos livros didáticos que analisamos, nos quais se reconhece a participação do ambiente na determinação dos fenótipos, mas se mantém a supremacia causal dos genes no DNA (*cf.* tb. El-Hani 1995, 1997). Achados como a predominância do gene-P e sua freqüente sobreposição ao gene molecular clássico sugerem que, nos livros analisados, admite-se um interacionismo fraco, como na visão de que ambientes extremos (que tornam o organismo inviável), e somente eles, podem impedir que os genes presentes num organismo sejam expressos. Nesta visão, cada gene está correlacionado a apenas um fenótipo, uma vez que se reconhece como único papel do ambiente o fornecimento de recursos necessários para rodar o suposto programa genético.⁶¹ Em nossa análise, essa idéia aparece na menção do ambiente como um fator secundário, um mero disparador da expressão gênica, como exemplificam os trechos abaixo:

‘Outro exemplo da interação entre genótipo e ambiente na manifestação do fenótipo refere-se à produção de clorofila nas plantas. Os genes envolvidos na síntese desse pigmento são ativos somente na presença de luz. Plantas germinadas no escuro não produzem clorofila, apresentando fenótipo albino’ (Amabis e Martho, vol. 3, p. 34)

‘A influência ambiental sobre a expressão de um gene recebe o nome de peristase. Há vários exemplos [...] Um deles é o da drosófila [...] alguns representantes dessa espécie apresentam asas enroladas – também chamadas asas curly – uma característica

⁶¹ No extremo oposto dessa vertente de determinismo genético, estão os deterministas ambientais, que concordam, por exemplo, que genes são necessários para garantir que uma prole seja da mesma espécie de seus pais, ainda que todas as suas características restantes sejam determinadas por fatores sociais (ver Figura 3B). Em outras palavras: na versão fraca de interacionismo, genes e ambientes representam, reciprocamente, apenas restrições de natureza ampla sobre o espectro de resultados que cada um pode produzir (Sterelny e Griffiths 1999). Nos livros analisados, quando aparece essa versão fraca de interacionismo, fatores ambientais aparecem como restrições de natureza ampla para genes e não o contrário, o que sugere que genes são os fatores essenciais para o desenvolvimento e, portanto, que permanece um determinismo genético, ‘disfarçado’.

determinada geneticamente. No entanto, essas asas enrolam-se somente a temperatura em torno dos 25°C. Aos 16°C, por exemplo, elas parecem asas normais' (Carvalho 2005, vol. 3, p. 21).

Nota-se nos trechos acima que o ambiente é descrito meramente como algo que influencia a expressão gênica, resultando na presença ou na ausência de determinado fenótipo, e não como um fator capaz de alterar qualitativamente o resultado fenotípico.

Contudo, algumas das obras analisadas introduzem o interacionismo de forma mais consistente do que nesta versão fraca, que admite apenas restrições extremas e se mostra mais um determinismo mal disfarçado do que uma visão verdadeiramente interacionista. Em geral, aquela visão interacionista mais consistente é introduzida pelos livros por meio do conceito de 'norma de reação', seja implícita, ou explicitamente. Como exemplo, podemos citar os seguintes excertos:

'Chamamos de **norma de reação** o conjunto de fenótipos possíveis produzidos pelo mesmo genótipo em condições ambientais diferentes' (Linhares e Gewandsznajder 2005, p. 379).

'O mais comum é que um mesmo genótipo produza uma gama variada de fenótipos dentro de certos limites. A gama de variação fenotípica que um genótipo pode expressar é denominada **norma de reação**' (Amabis e Martho 2005, vol. 3, p. 43).

'Falamos dos genes como se eles fossem determinantes exclusivos dos caracteres, mas isso é uma simplificação. Cada gene age sob a influência dos demais genes do indivíduo e dos fatores ambientais que atuam sobre ele. O ambiente atua mais no efeito de certos genes do que no de outros. Os grupos sanguíneos não são alteráveis pelo ambiente. A cor dos olhos só é modificada em casos de certas doenças raras. Por outro lado, as conseqüências de um acidente físico dependem pouco do genótipo da vítima. Entre esses extremos situa-se a maioria dos caracteres' (Frota-Pessoa 2005, vol. 3, p. 94).

'O genótipo determina, portanto, uma escala de variação fenotípica para o indivíduo, sendo que o meio ambiente determina em que ponto dessa escala o indivíduo está' (Lopes e Rosso 2005, p. 438).

Outro ponto que merece destaque são os casos em que um mesmo livro traz tanto visões deterministas ou as versões de interacionismo fraco, quanto a versão mais consistente do interacionismo, como ocorre em Amabis e Martho (2005), Linhares e Gewandsznajder (2005) e Frota-Pessoa (2005), por exemplo. Não relatamos, contudo, qualquer distinção na ocorrência de determinismos, visões mais fracas, e visões mais fortes de interacionismo, de acordo com as categorias de contexto, o que indica em geral, pouca coerência na abordagem das relações genótipo-fenótipo nestas obras.

Vale destacar que, ao discutir a determinação das chamadas características multifatoriais, a maior parte das obras analisadas incorpora a idéia de um interacionismo aditivo, o qual implica, como o nome sugere, a existência de parcelas de influência gênica e ambiental na determinação de uma característica, como ilustra o trecho abaixo, ao dizer que os efeitos dos genes,

'[...] se combinam e confundem com influências ambientais [...] A inteligência é, portanto, multifatorial e sua variação é causada, em partes praticamente iguais, pelos genes e pelo ambiente' (Frota-Pessoa 2005, vol. 3, p. 96-97).

Um dos problemas suscitados pelas visões interacionistas aditivas é que continuam fomentando o debate, ainda que mais moderado, entre deterministas genéticos e ambientais: elas tornam provável que a discussão prossiga em termos de qual fator, gene ou ambiente, tem a maior ‘cota’ de contribuição para uma característica, o que sempre dará ensejo a visões que privilegiam o papel dos genes, como mostra o seguinte exemplo:

‘Os genes determinam e a cultura modela os padrões de comportamento sexual’ (Frota-Pessoa 2005, vol. 2, p.182).

Adicionalmente, como ressalta Lewontin (2002), o mais provável é que, ao longo do desenvolvimento, fatores genéticos e não-genéticos interajam de forma complexa, não-aditiva, na gênese de traços, tornando-se impossível determinar tais parcelas de contribuição, sobretudo sobre caracteres complexos, como aqueles relativos ao comportamento, à personalidade, à inteligência etc. A Figura 3 ilustra diferentes padrões de normas de reação para dois alelos hipotéticos – g_1 e g_2 . Em ‘A’, qualquer que seja o ambiente (exceto nos casos extremos), para cada genótipo, haverá um único fenótipo. Em ‘B’, ambos os alelos produzem os mesmos fenótipos, que variam apenas com o ambiente. Essas duas primeiras normas de reação representam os determinismos genético e ambiental, nos quais mudanças (com exceção dos extremos) no ambiente e no genótipo, respectivamente, não alteram o produto.

As duas normas seguintes (C e D) representam duas formas de interacionismo. Na interação aditiva, uma mudança em uma das variáveis altera o produto, mas sempre da mesma forma, não importando o valor da outra variável. Assim, se por um lado, a diferença entre os fenótipos de g_1 e g_2 é sempre a mesma nos diversos ambientes, por outro, para uma dada mudança ambiental, os fenótipos respectivos de g_1 e g_2 têm o mesmo tipo de alteração (ambas as normas de reação serão ascendentes ou ambas, descendentes). Dessa forma, é possível que um genótipo seja considerado sempre ‘superior’ a outro em relação a um dado traço. A título de exemplo, seria possível dizer que homens serão sempre mais agressivos que mulheres, já que seu nível de testosterona seria mais alto que o delas em qualquer ambiente, ou que um dado grupo étnico seria indiscutivelmente mais inteligente do que outro. Os prejuízos na formação da cidadania dos estudantes causados por esse tipo de abordagem dispensam maiores comentários.

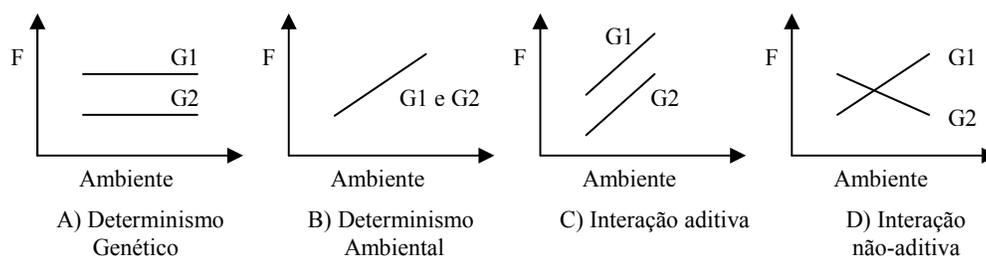


Figura 3. Normas de Reação e diferentes modelos da relação genótipo-ambiente (adaptada de Sterelny e Griffiths 1999).

No interacionismo não-aditivo, os fenótipos de genótipos distintos variam de forma diferente com o ambiente. No caso mostrado na figura 3D, enquanto a norma de reação de um é ascendente, a do outro é descendente. Assim, é possível que uma característica se expresse mais num dado ambiente, e menos em outro, bem como que qualquer relação de ‘superioridade’ entre dois genótipos se inverta a depender do ambiente. Retomando os exemplos citados acima, mulheres poderiam ser mais agressivas que homens e um grupo étnico alcançar o ‘patamar de

inteligência’, em determinados ambientes. Oportunamente, algumas obras analisadas abordam o interacionismo de forma mais complexa e realista do que na formulação básica ‘gene mais ambiente é igual a fenótipo’:

‘No outro extremo da escala herança-ambiente encontramos caracteres muito complexos como comportamentais (inteligência, vontade, talento em alguma atividade específica) que, embora tendo uma base genética, são enormemente influenciados pelas condições sociais e culturais. O estudo desses caracteres é muito complexo, seja porque é difícil dar a eles uma definição precisa, seja porque é praticamente impossível separar, nesses casos, o componente genético daquele ambiental’ (Gainotti e Modelli 2005, p. 284).

‘Enquanto em características físicas (como altura) o grau de influência dos genes pode chegar a 90%, em relação à personalidade a interferência genética é bem mais fraca [...] Nesse caso, o ambiente, sob a forma de estímulos externos, ou o próprio esforço da pessoa podem mudar bastante o resultado final’ (Linhares e Gewandszajder 2005, p. 383).

As questões mais interessantes na compreensão sobre o desenvolvimento biológico requer que avancemos além das dicotomias subjacentes aos tradicionais debates nas ciências biológicas e sociais, como ‘genes e ambiente’, ‘natureza e criação’ (*nature* e *nurture*), ‘biologia e cultura’. Consideramos notável encontrar, em livros didáticos de ensino médio, o reconhecimento da inadequação de se adotar visões deterministas para a previsão de fenótipos, em geral, e o reconhecimento das questões éticas envolvidas, como em Lopes e Rosso (2005):

‘[...] nunca será possível determinar a priori e de maneira absoluta o futuro de um indivíduo a partir de seus genes [...] Do mesmo modo que um fenótipo é determinado por uma interação entre genes e ambiente, as decisões éticas necessariamente terão de ser determinadas por uma interação entre o conhecimento científico e o ambiente no qual a ciência existe. Para tanto, precisamos garantir a todas as pessoas uma compreensão mínima dos avanços da genética’ (p. 438-439).

Entre esses avanços, inclui-se a compreensão dos desafios ao gene molecular clássico, os quais ressaltam a importância do contexto que cerca o DNA, desde o ambiente mais imediato constituído pelo contexto nuclear e celular até contextos mais amplos, em níveis de organização biológica superiores, até chegar ao organismo como um todo.

5.4. Desafios ao conceito molecular clássico

De acordo com Xavier e colaboradores (2006, p. 3), a nova biologia, que estes autores entendem como a “integração entre as novas tecnologias do DNA e novas aplicações em Genética, que inclui a Biotecnologia e a Biologia Molecular”, é um dos temas recorrentes na mídia desde o final dos anos 1960, devido ao seu forte apelo social e à sua influência direta na vida das pessoas. Os autores destacam que este tema não se encontra suficientemente representado nos livros didáticos de biologia do ensino médio, os quais dedicam apenas cerca de 1-2% de seu conteúdo para a abordagem da nova biologia. De acordo com estes autores, estes livros estão pouco atualizados quanto a temas que consideram necessários ao bom entendimento da nova tecnologia, como organismos transgênicos, projeto genoma, clonagem de mamíferos, células-tronco, teste de paternidade, variabilidade genética, melhoramento genético, DNA recombinante, hibridação, seqüenciamento, enzimas de restrição, plasmídios, eletroforese, PCR e, em particular, no que concerne a um dos interesses do presente trabalho, a abordagem de íntrons e éxons. Estes autores (*ibid.*) apontam, ainda, que os livros de biologia do ensino médio refletem

uma preocupação de inculir no aluno apenas o referencial genético clássico, sem ampliar os temas, atualizar as informações sobre genética e contextualizá-la em relação às novas tecnologias. Em sua visão, a genética tradicional tem um espaço muito grande dedicado ao seu estudo, chegando a 20% de um livro didático. Esses achados estão em consonância com os resultados de nosso estudo, que mostram que o extenso conteúdo de genética clássica dos livros de biologia do ensino médio contribuiu para o predomínio do gene-P e para a forte hibridização das idéias sobre genes, como discutido acima.

Em nossa análise, os temas referentes à nova biologia são, em geral, abordados nas obras analisadas, embora não de maneira tão significativa quanto os conteúdos de genética clássica. A análise mais detalhada quanto à presença e à atualização de temas como engenharia genética e temas afins é encontrada nos trabalhos resultantes da avaliação do PNLEM/2007 (ver, por exemplo, El-Hani, Roque & Rocha, 2007a,b). No presente artigo, dos temas da nova biologia, enfatizamos a discussão sobre a abordagem dos desafios atuais da biologia molecular e da genômica ao conceito molecular clássico de gene. Quanto à atualização frente a esses desafios, observamos, nas 18 obras analisadas, os seguintes números de ocorrências de menções a diferentes fenômenos que trazem dificuldades para aquele conceito: genes interrompidos e *splicing* (28 ocorrências); DNA não-codificante (14); promotores (9); ‘DNA lixo’ (9); seqüências regulatórias (5); DNA repetitivo (4); pseudogenes (3); *splicing* alternativo (3); operon *lac* (2); genes dentro de genes (*nested genes*) e superpostos (1), edição de RNA (1) e transposons (1).

Um ponto que merece destaque é o uso da expressão ‘DNA lixo’ para descrever seqüências de DNA não-codificante. Embora tenha sido largamente usada no passado, inclusive pela comunidade científica (de onde o termo certamente foi transposto para o conhecimento escolar), esta expressão pode ser hoje considerada inapropriada, por tratar-se de uma metáfora que negligencia e desvaloriza qualquer outro tipo de função que as seqüências de DNA possam ter, além da codificação de seqüências de aminoácidos de proteínas e de nucleotídeos de RNAt e RNAr. Entre estas outras funções, temos muitas de grande importância, sobretudo regulatórias, muitas vezes associadas a micro-RNAs, que têm sido detectadas com freqüência cada vez maior em regiões não-codificantes (Fire 1999; Grosshans e Slack 2002; Hannon 2002; Lenz 2005). Um aspecto positivo, entretanto, é que, das sete obras que utilizam esta expressão, quatro vão além da afirmação de que se trata de seqüências de DNA que “não correspondem a genes” e “sem função”, e reconhecem a inadequação da expressão ‘DNA-lixo’, como as seguintes passagens exemplificam:

‘[...] esse termo pode passar a impressão de que a maior parte do DNA humano não tem função; o mais correto seria afirmar que os cientistas não compreendem completamente sua função’ (Laurence 2005, p. 654).

‘[...] Esta última denominação é pouco adequada, por dar a impressão de que se trata de DNA inútil, o que não é verdade. Já se descobriu que muitos tipos de DNA não-codificante desempenham funções importantíssimas na estrutura e no funcionamento dos cromossomos’ (Amabis e Martho 2005, vol. 1, p. 249).

‘O DNA não-codificante tem sido chamado de DNA-lixo, pois aparentemente não tem função. Embora ainda não se saiba exatamente a função de alguns trechos desse DNA, recentemente descobriu-se que o DNA não-codificante: participa da estruturação dos cromossomos, pode apresentar alguns trechos correspondentes a genes que, ao longo da evolução de uma espécie, deixaram de ter função; assim, desempenham papel importante na evolução’ (Lopes e Rosso 2005, p. 429).

[...] Esse “deserto” genético está sendo chamado de “DNA-lixo”, ou, como alguns preferem, “suposto DNA-lixo”, já que seu papel biológico ainda não é bem conhecido. Suspeita-se que esse DNA possa atuar, por exemplo, como uma espécie de “material de reserva” no reparo de falhas de genes’ (Paulino 2005, vol. 1, p. 194).

Enquanto a maioria dos desafios ao gene molecular clássico é mencionada de forma sucinta pelos livros, o processo de *splicing* é o que apresenta, relativamente, a melhor descrição, incluindo desde a explicação do que são éxons e íntrons, até a discussão das implicações da presença de íntrons para a síntese protéica, até a descrição da excisão de íntrons e emenda de éxons no processamento de RNA. Metade das obras analisadas (9) trata dos genes interrompidos e, entre estas, oito apresentam esquemas ilustrativos. Contudo, apenas três abordam o *splicing* alternativo e somente duas destas discutem suas implicações para a compreensão do que são genes.

Para facilitar a transposição didática do tema, alguns autores recorrem a metáforas informacionais, como o seguinte trecho ilustra:

‘Portanto, o que comanda a síntese da proteína é a mensagem codificada no RNA, depois da exclusão dos íntrons, e não a do DNA. É como se mandássemos um artigo para um jornal, traduzido do inglês para o português, com vários pedaços de letras embaralhadas, que não fazem sentido, entremeadas ao texto correto, obrigando o redator a cortar os trechos incompreensíveis (íntrons) e colar, em ordem, os outros trechos (exons), notando que a continuidade não foi prejudicada’ (Frota-Pessoa 2005, vol. 3, p. 36).

Uma das abordagens mais completas é encontrada em Amabis e Martho (2005). Os autores se referem à inexistência de “[...] colinearidade entre as cadeias polipeptídicas e os segmentos de DNA que as codificam” e comentam que “a instrução para a síntese de proteínas nos genes eucarióticos é geralmente interrompida por trechos da molécula que não codificam aminoácidos” (vol.3, p. 144). Citam a origem dos termos ‘éxon’ e ‘íntron’, utilizam o termo ‘*splicing*’, mas apresentam tradução como ‘emenda e corte’, citam que a molécula de RNA recém-sintetizada sofre uma série de modificações até ser transformada no RNAm, mas que a remoção dos íntrons é a mais importante. Citam até mesmo o complexo enzimático que catalisa o *splicing* (*spliceossomo*) e descrevem sucintamente as etapas por meio das quais esse complexo se liga ao RNA e remove os íntrons. Por fim, introduzem o *splicing* alternativo, colocando a questão sobre a vantagem evolutiva da presença de íntrons (vol.3, pp. 144-147). Além disso, comentam as implicações desse fenômeno para nossa compreensão do que são genes, como voltaremos a discutir mais abaixo.

Consideramos satisfatória a abordagem dos genes interrompidos e do *splicing* em seis das nove obras recomendadas na avaliação do PNLEM: Amabis e Martho (2005), Frota-Pessoa (2005), Laurence (2005), Lopes e Rosso (2005), Paulino (2005) e Silva Júnior e Sasson (2005). Contudo, a abordagem do *splicing* alternativo e, sobretudo, o reconhecimento desse processo como um desafio crucial (ver Fogle 1990, Santos 2005; El-Hani 2005, 2007; Solha 2005; Solha e Waizbort 2007) à idéia geral de unidade – seja estrutural, funcional ou informacional - implícita no conceito molecular clássico de gene parece ainda incipiente (ver Tabela 10), uma vez que aparece em apenas dois livros.

Similarmente, livros-texto de biologia celular e molecular do ensino superior abordam muitos dos resultados experimentais que puseram em crise o conceito molecular clássico de gene,

porém, sem considerar apropriadamente as conseqüências desses achados para o conceito de gene propriamente dito (Pitombo et al., no prelo).

Tabela 10. Obras analisadas que abordam genes interrompidos, *splicing* alternativo e as implicações do *splicing* alternativo para o conceito molecular clássico de gene. Em negrito, obras recomendadas na avaliação do PNLEM.

Genes interrompidos	<i>Splicing</i> alternativo	Reconhecimento do <i>splicing</i> alternativo como desafio ao gene molecular clássico
Amabis e Martho (2005) Faucz e Quintilham (2005)	Amabis e Martho (2005)	Amabis e Martho (2005)
Frota-Pessoa (2005) Gainotti e Modelli (2005)		
Laurence (2005)	Laurence (2005)	Laurence (2005)
Lopes e Rosso (2005) Machado (2005)		
Paulino (2005)		
Silva Júnior e Sasson (2005)	Silva Júnior e Sasson (2005)	
9 obras	3 obras	2 obras

Ressaltamos ainda que éxons, íntrons e o processo de *splicing* muitas vezes são descritos nessas obras com o uso de termos inadequados, por exemplo, mediante o emprego de metáforas vulgarizadas, que, como afirma Alice Lopes (1997), podem impedir que os estudantes compreendam as diferenças entre os conceitos usados nos contextos científico e cotidiano.

Outras expressões podem, por sua vez, superestimar o papel de seqüências codificantes no contexto celular, como aquelas que atribuem uma capacidade de ação ao próprio éxon, que nada tem de dinâmico, pois apenas fornece os sítios para a ligação do complexo enzimático que conduz o processo de *splicing*:

“[...] Terminada a transcrição, o núcleo da célula contém uma fita de RNA e a dupla hélice do DNA restaurada. No entanto, após essa meticulosa operação ocorre algo fantástico: a fita de RNA se parte em pedaços! Vários fragmentos da fita, os **íntrons**, são descartados e destruídos. Os outros fragmentos, os **éxons**, se juntam pelas pontas, na ordem em que estavam antes da fragmentação, reconstituindo a fita, que fica mais curta. São os éxons da nova fita que levam a mensagem para o citoplasma, onde a proteína correspondente será sintetizada” (Frota-Pessoa 2005, vol. 3, p. 36).

Na Tabela 11, são mostradas idéias e termos utilizados, nas obras analisadas, para descrição dos genes interrompidos e do *splicing*. Referências à ‘edição’ – como em “[...] o RNA [...] com a instrução genética devidamente ‘editada’, passa para o citoplasma” (Amabis e Martho 2005, vol. 3, p. 146) – para descrever o *splicing* são ambíguas, gerando confusão com outro processamento pós-transcricional, a ‘edição de RNA’. Um problema similar ao uso indiscriminado do termo ‘informação’ se encontra no uso do termo ‘significado’ como propriedade dos éxons. Ambos os casos carecem de um referencial teórico que lhes forneça uma definição mais precisa, como discutimos acima. No excerto abaixo, aparece o que poderíamos

considerar como uma idéia pré-semiótica, implícita no emprego dos termos ‘significação’, ‘mensagem’ etc., que exemplifica, contudo, como a ausência de um arcabouço teórico apropriado para o uso desses termos de forma mais precisa, pode endossar o entendimento de genes como independente do contexto no qual ele se insere:

‘Sempre se imaginou que alguma coisa passa dos pais para os filhos e os torna parecidos. De fato, deve haver, por exemplo, no ovo da galinha, algum tipo de mensagem que obrigue o pintinho a se desenvolver dentro dos padrões da espécie. Essa mensagem e sua significação estão contidas no código genético [...] o código é o conjunto de sinais usados para exprimir a mensagem’ (Frota-Pessoa 2005, vol. 3, p. 24).

Quanto à abordagem do *splicing* alternativo, uma das obras o faz da seguinte forma:

‘[...] hoje há vários indícios de que, em alguns casos, um mesmo gene, trabalhando em células diferentes, é capaz de produzir mais de um tipo de proteína [...] pelo fato de existir mais de uma forma de cortar o mesmo pré-RNA, o que resultaria, em cada caso, em segmentos diferentes utilizados como éxons’ (Silva Júnior e Sasson 2005, vol. 1, p. 262).

Considerando-se que esses autores caracterizam gene como um ‘Segmento da molécula de DNA, no qual está codificada uma característica hereditária’ (Glossário, vol. 1, p. 393) e não mencionam a existência de controvérsias sobre o estatuto dos genes ao fazerem alusão ao *splicing* alternativo, pode-se afirmar que eles conciliam o desafio trazido por esse fenômeno com a idéia de unidade no DNA, em última análise, assimilando o *splicing* alternativo, de tal maneira que ele termina por não suscitar qualquer questionamento quanto ao modo como eles explicam os genes e sua relação com o fenótipo.

Outro livro que trata do *splicing* alternativo, por sua vez, coloca em questão a compreensão dos genes a partir do exame desse fenômeno numa seção intitulada ‘Reformulações no conceito de gene’:

‘Reformulações no conceito de gene [...] Com o seqüenciamento do genoma humano, os pesquisadores encontraram, no entanto, evidências de que o DNA funciona de maneira mais complexa do que se imagina. Um único gene pode, por exemplo, produzir três tipos diferentes de RNAm, cada um codificando a produção de uma proteína diferente. É como se existissem genes dentro de genes! Além disso, uma mesma molécula de RNAm pode sofrer alterações após a transcrição, também resultando em proteínas diferentes’ (Laurence 2005, p. 656).⁶²

⁶² A expressão ‘gene dentro de genes’, utilizada nesse trecho para abordar o *splicing* alternativo, pode ser confundida com outro aspecto da complexidade estrutural dos genomas, os genes nidados, entendidos como genes contidos numa seqüência de nucleotídeos mais ampla, com sítio de terminação de transcrição posterior ao seu.

Tabela 11. Idéias e termos utilizados para descrever éxons, íntrons e *splicing*.

Éxons	Íntrons	<i>Splicing</i>
Seqüências com significado (Gainotti e Modelli 2005)	Seqüências sem significado (Gainotti e Modelli 2005) (Silva Júnior e Sasson 2005)	Amputação (Gainotti e Modelli 2005)
Com informação (Faucz e Quintilham 2005) (Frota-Pessoa 2005) (Paulino 2005)	Sem informação (Faucz e Quintilham 2005) (Frota-Pessoa 2005) (Paulino 2005)	Remoção de íntrons e união de éxons (Gainotti e Modelli 2005)
Codificantes (Lopes e Rosso 2005) (Amabis e Martho 2005)	Não-codificantes (Lopes e Rosso 2005) (Amabis e Martho 2005)	Processamento (Faucz e Quintilham 2005) (Frota-Pessoa 2005)
Fragmentos ou porções ativas (Machado 2005) (Paulino 2005)	Fragmentos inibidores (Machado 2005)	Descarte (Frota-Pessoa 2005)
	Sem função (Gainotti e Modelli 2005)	Corte e emenda (Amabis e Martho 2005)
	Espaçadores de éxons (Faucz e Quintilham 2005)	Processo de emendas ou <i>splicing</i> (Machado 2005)
		Edição (Amabis e Martho 2005)
		Maturação (Lopes e Rosso 2005) (Paulino 2005)
		Remoção e solda (Paulino 2005)

Por fim, Amabis e Martho (2005, vol. 3, p. 147) afirmam que mais de 60% dos genes humanos apresentam *splicing* alternativo e que isso explicaria por que o número de tipos de proteínas humanas é tão superior ao número de genes. Eles descrevem o *splicing* alternativo da seguinte maneira:

‘Os cientistas descobriram que uma mesma molécula de pré-RNA mensageiro pode sofrer tipos diferentes de *splicing* em diferentes tipos celulares. Em outras palavras, nos diferentes tipos de células pode haver diferentes tipos de segmentos eliminados, de modo que o mesmo pré-RNA mensageiro é cortado e montado de diferentes maneiras, dependendo do tipo de célula. Esse fenômeno é chamado de *splicing* alternativo’ (Amabis e Martho 2005, vol. 3, p. 147).

O gene DSCAM de drosófila é citado por eles como um dos casos mais notáveis de *splicing* alternativo, no qual a combinação diferencial de 115 éxons possibilita a produção de milhares de proteínas. Por fim, introduzem ‘a problemática de se definir gene’ (p. 147), apresentando um texto extraído de ‘O século do gene’, de Evelyn Fox-Keller (2002), que coloca explicitamente o problema do conceito de gene entendido como unidade de estrutura e função e discute genes interrompidos, seqüências repetitivas de DNA, genes superpostos, DNA críptico,

transcrição reversa, genes nidados e promotores múltiplos como desafios ao conceito molecular clássico. Além disso, trata-se de um texto que chega a propor que talvez o termo ‘gene’ deva ser abandonado, cunhando-se novas palavras para desempenhar papéis dos quais aquele termo tem sido incumbido⁶³.

Salientamos, como ponto positivo, o modo como uma série de dificuldades, incluindo o *splicing* alternativo, é discutida em alguns livros, – Amabis e Martho (2005), Laurence (2005) e Lopes e Rosso (2005) – sendo considerados como desafios ao conceito molecular clássico de gene. Estas abordagens indicam perspectivas de como introduzir no conhecimento científico escolar, no nível médio de escolaridade, não um novo conceito de gene, ou mesmo um conceito de informação, mas sim os próprios debates sobre genes e as dificuldades empíricas que o conceito molecular clássico e a concepção informacional enfrentam. Por exemplo, Amabis e Martho (2005) enfatizam, no manual do professor (vol.1, p.66) que acompanha o livro, que a definição ‘exata’ de gene ainda é tema de debate entre os cientistas, fornecendo estímulos para esse tipo de discussão entre alunos e professores na sala de aula. Também no manual do professor (p.77), Lopes e Rosso (2005) citam os genes interrompidos, mas declaram que o conceito de gene trabalhado no livro é o mais simplificado e que, numa atividade sugerida, considerarão apenas os éxons na explicação da síntese de RNA. Vale dizer que o simples reconhecimento da adoção de uma simplificação já facilita a introdução da contenda a respeito do conceito de gene na sala de aula.

Destacamos que, se tais controvérsias forem inseridas, uma visão mais crítica sobre a relação genótipo-fenótipo, o determinismo genético, a tecnologia do DNA recombinante etc. poderá ser construída. Esta possibilidade é exemplificada numa passagem de Laurence (2005), que traz uma das raras ocorrências do gene-D, ou seja, da interpretação dos genes como recursos desenvolvimentais, lado a lado com outros recursos que desempenham papéis centrais na ontogenia, e que explicita a condição do DNA como mero *potencial* herdado para uma característica:

“Todo indivíduo traz, portanto, a possibilidade de desenvolvimento de uma série de características, cuja manifestação é influenciada ou condicionada pelo ambiente” (p. 678).

6. Considerações Finais

Em nossa análise de 18 livros didáticos de biologia do ensino médio, foi possível verificar que, ao privilegiarem determinados aspectos dos fenômenos de transmissão do material genético e da função gênica, os autores estão, ao mesmo tempo, negando ou silenciando muitos outros. Como ressaltam Nascimento e Martins (2005), o que ‘não é dito’ é altamente relevante, como elemento de análise do discurso, ou seja, trata-se de um silêncio com significado, não se limitando às palavras enunciadas o papel de condicionar a atribuição de sentidos pelos sujeitos que interagem discursivamente. Dito de outra maneira, quando não dizemos certas coisas, também estamos influenciando o processo de atribuição de significado de nossos interlocutores. Na maioria dos livros analisados, o reconhecimento de que alguns fenômenos que são discutidos por eles correspondem a desafios ao gene molecular clássico constitui uma *ausência* notável. O mesmo pode ser dito do próprio reconhecimento da existência de debates sobre o conceito de gene, ou da ausência de referência à importância do contexto intracelular e extracelular, e dos

⁶³ Em artigo posterior, Fox Keller (2005) não se compromete com a idéia de que o termo ‘gene’ deveria ser abandonado, como o fez em *O Século do Gene*.

processos de sinalização celular, para a compreensão da função gênica. Não nos parece que esses aspectos não possam ser recontextualizados como parte do conhecimento escolar do ensino médio, como sugere, inclusive, a existência de livros que o fazem na amostra analisada.

Nossos resultados apontam que, em livros didáticos de ensino médio, prevalecem as explicações de genes no DNA: (i) com informação para um fenótipo, (ii) com funções gênicas de determinar as características dos organismos e, (iii) de produzir um polipeptídeo ou RNA. Por um lado, o foco na previsão de fenótipo a partir do genótipo se concentra na descrição de procedimentos da genética clássica e da genética médica, cujo conteúdo é extenso nos livros de ensino médio. Por outro, a concepção informacional e o conceito molecular clássico de gene estão difundidos numa ampla diversidade de contextos, ainda que não haja um significado claro para o conceito de ‘informação genética’ na biologia atual e ainda que uma série de recentes descobertas desafie o gene molecular clássico. Os livros, com algumas notáveis exceções, não utilizam esses achados para desafiar essas idéias. Assim, as concepções de gene informacional, gene molecular clássico e gene-P estão fortemente presentes e intimamente relacionadas no ensino médio, denotando modelos híbridos de função gênica, nos quais há forte predominância de aspectos do modelo neoclássico, misturados a visões deterministas da relação genótipo-fenótipo.

O argumento que aqui empregamos contra o uso de modelos híbridos em livros didáticos do ensino médio não implica, entretanto, a necessidade de adoção de um único e abrangente modelo ou conceito de gene, que inclua toda a diversidade de significados e funções epistêmicas conectada a esse termo, mas sim a proposta de co-existência de uma diversidade de conceitos e modelos de ‘gene’, mas com domínios bem delimitados de aplicação (El-Hani 2007). Contudo, assim como ocorre atualmente na comunidade científica, observamos que os livros analisados não delimitam os domínios de aplicação para os diferentes conceitos de gene, levando a uma sobreposição entre as visões clássicas e atuais sobre gene, que conduzem a uma série de dificuldades na compreensão dos estudantes. Ressaltamos ainda que, mesmo na comunidade científica, sobreposições de significados e hibridização dos modelos de função gênica ocorrem numa extensão mais limitada do que a que vemos nos livros didáticos.

Pitombo e colaboradores (no prelo, 2007), numa análise de idéias sobre genes em livros-texto de biologia celular e molecular do ensino superior, encontraram resultados semelhantes: (i) a despeito dos desafios ao conceito molecular clássico de gene e da ausência de significado da concepção informacional de gene, ambos são amplamente empregados; (ii) quando esses referidos desafios são abordados, os livros não os utilizam como base para propor qualquer discussão explícita sobre o conceito de gene e suas dificuldades, bem como sobre os debates atuais sobre esse conceito; (iii) e há, nos livros analisados, uma proliferação de significados do termo ‘gene’ que faz o conceito parecer vago e confuso, e levando até mesmo a concepções equivocadas. Esses resultados indicam, de um lado, uma possível fonte das idéias dominantes sobre gene nos livros do ensino médio, e, de outro, que os problemas encontrados em nossa análise também aparecem no ensino superior, havendo, portanto, necessidade de recontextualização didática dos debates sobre genes em ambos os níveis de ensino.

Outro achado importante reside na prevalência, nos livros analisados, da visão do DNA como o reservatório de toda a informação para construir um organismo, conforme expressa, por exemplo, nas metáforas do programa genético ou do DNA como ‘livro da vida’. Detectamos até mesmo uma negação de que esta última expressão é uma metáfora:

‘Comparar o genoma a um livro não é uma metáfora. Um livro é uma porção de informações digitais, organizadas de forma linear [...] e definida por um código que utiliza um pequeno alfabeto de sinais. Esses sinais produzem um grande léxico de significados

dependendo da ordem de seus agrupamentos. Assim também é o genoma' (Laurence 2005, p. 641).

Como discute Oyama (2000), não se deve diminuir a importância da conexão entre esse modo usual de explicar a noção de 'informação genética' e o determinismo. Esta autora argumenta que o determinismo genético é inerente à maneira como genes e sua função em sistemas biológicos são tipicamente representados na biologia. Na visão dessa filósofa da biologia, enquanto genes forem representados como se carregassem informação sobre como um organismo se desenvolverá, eles continuarão a ser vistos como causas determinantes dos próprios organismos e de suas características, não importa quanta evidência exista contra essa idéia. Salientamos que uma mudança no modo como a informação é entendida poderá conter esse tipo de conexão com o determinismo (ver a proposta do entendimento de informação como semiose em El-Hani et al. 2006). Em oposição ao determinismo genético, Oyama propôs sua 'teoria dos sistemas de desenvolvimento', na qual a paridade causal entre genes e outros recursos desenvolvimentais é um dos postulados básicos (Oyama 2000; Oyama et al 2001). Essa perspectiva busca destacar um elemento que estaria ausente nas explicações deterministas da relação genótipo-fenótipo, a saber, o desenvolvimento, ao longo do qual genes, organismos e ambientes interagem uns com os outros de tal maneira que cada um deles é tanto causa quanto efeito, de uma maneira complexa (Lewontin 1983, 2000).

Tomando essa tríade como base – genes, desenvolvimento e ambiente –, nosso desafio é, então, caracterizar genes, sua relação com DNA, seu modo de ação, sua regulação e sua relação com o fenótipo de maneira a aproximar o conhecimento escolar sobre genética e biologia celular e molecular de um modo de compreensão que vem se tornando dominante na comunidade científica que se ocupa dessas áreas, no qual o reconhecimento da complexidade da estrutura e dinâmica dos genomas tem sido a tônica, bem como da complexidade dos processos de desenvolvimento e de constituição de características fenotípicas.

Referências

- Bardin L (2000) Análise de conteúdo. Edições 70, Lisboa
- Black D (2003) Mechanisms of alternative pre-messenger RNA splicing. *A Rev Biochem* 72:291-336
- Carlson EA (1991) Defining the gene: an evolving concept. *J hum Genet* 49(2):475-487
- El-Hani CN (1995) O Insustentável Peso dos Genes: A Persistência do Determinismo Genético na Mídia e na Literatura Científica. Dissertação, Universidade Federal da Bahia
- El-Hani CN (2005) Controvérsias sobre o Conceito de Gene e suas Implicações para o Ensino de Genética. Atas do V Encontro de Pesquisa em Educação em Ciências, Bauru, São Paulo, 28 Nov-03 Dez 2005
- El-Hani C (2007) Between the cross and the sword: the crisis of the gene concept. *Genet Molec Biol* 30(2):297-307
- El-Hani C, Queiroz J, Emmeche C (2006) A semiotic analysis of the genetic information system. *Semiotica* 60:1-68
- El-Hani CN, Roque N, Rocha PLB (2007a) Brazilian High School Biology Textbooks: Results from a National Program. In: Proceedings of the IOSTE International Meeting on Critical Analysis of School Science Textbook, University of Tunis, Hammamet, 7-10 February 2007

- El-Hani CN, Roque N, Rocha PLB (2007b) Livros Didáticos de Biología do Ensino Médio: Resultados do PNLEM/2007. Atas do VI Encontro de Pesquisa em Educação em Ciências (ENPEC), Universidade Federal de Santa Catarina, Florianópolis,
- EL-Hani CN, Roque N, Vanzela, AL et al (Equipe de avaliação de Biologia, Programa Nacional do Livro para o Ensino Médio, Brazil) (2007) Brazilian High School Biology Textbooks: Main Conceptual Problems in Genetics and Cell & Molecular Biology. In: Proceedings of the IOSTE International Meeting on Critical Analysis of School Science Textbooks, University of Tunis, Hammamet, 494-504, 7-10 Feb 2007
- El-Hani CN (1997) Explicações Causais do Desenvolvimento: São os Genes Suficientes?. *Cad Hist Filos Ciênc (UNICAMP)*, Campinas-SP 7(1):123-168
- Emmeche C (1999) The Sarkar challenge to biosemiotics: is there any information in a cell? *Semiotica* 127(1/4): 273-293
- Falk R (1986) What is a gene? *Stud Hist Philos Sci* 17:133-173
- Falk R (2000) The gene – A concept in tension. In: Beurton P, Falk R, Rheinberger H.-J (eds) *The concept of the gene in development and evolution*. Cambridge University Press, Cambridge-UK, pp. 317-348
- Fogle T (1990) Are genes units of inheritance? *Biology and Philosophy* 5:349-371
- Fogle T (2000) The dissolution of protein coding genes. In: Beurton P, Falk R, Rheinberger H.-J (eds) *The concept of the gene in development and evolution*. Cambridge University Press, Cambridge-UK, pp. 3-25
- Gericke N, Hagberg M (2007a) The phenomenon of gene function in textbooks for upper secondary school in Sweden – A comparative analysis with historical models of gene function. In: Proceedings of the IOSTE International Meeting on Critical Analysis of School Science Textbooks, University of Tunis, Hammamet, 554-563, 7-10 Feb 2007
- Gericke N, Hagberg M (2007b) Definition of Historical Models of Gene Function and their Relation to Students' Understandings of Genetics. *Sci Educ* 16:849-881
- Griffiths P (2001) Genetic information: A metaphor in search of a theory. *Philos Sci* 68:394-403
- Griffiths P, Neumann-Held E (1999) The many faces of the gene. *BioScience* 49:656-662
- Keller E (2000) *The century of the gene*. Harvard University Press, Cambridge-MA
- Keller E (2005) *The century beyond the gene*. *J Biosci* 30:3-10
- Lewontin R (2000) *The Triple Helix: Gene, Organism, and Environment*. Harvard University Press, Cambridge-MA
- Lopes AC (1997) Conhecimento escolar: inter-relações com conhecimentos científicos e cotidianos. *Contexto & Educação*, Ijuí, 11(45): 40-59
- Mayr E (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Harvard University Press, Cambridge-MA
- Moss L (2001) Deconstructing the gene and reconstructing molecular developmental systems. In: Oyama S, Griffiths PE, Gray RD (eds) *Cycles of contingency: Developmental systems and evolution*. MIT Press, Cambridge-MA, pp. 85-97
- Moss L (2003) *What genes can't do*. MIT Press, Cambridge-MA
- Nascimento T, Martins I (2005) O texto de genética no livro didático de ciências: uma análise retórica crítica. In: *Investigações em Ensino de Ciências*. 10:1-21. Disponível em <http://www.if.ufrgs.br/public/ensino/revista.htm>. Citado em 23 Set 2007
- Neumann-Held E, Rehmann-Sutter C (2006) *Genes in development*. Duke University Press, Durham
- Pardini M, Guimarães R (1992) A systemic concept of the gene. *Rev Brasil Genet* 15:713-721
- Pitombo M, Almeida, A, El-Hani, C (no prelo) Conceitos de gene e idéias sobre função gênica em livros didáticos de biologia celular e molecular do ensino superior. *Cont Educ*

- Pitombo M., Almeida, AMR, El-Hani CN (2007) Gene Concepts in Higher Education Cell & Molecular Biology Textbooks. In: Proceedings of the IOSTE International Meeting on Critical Analysis of School Science Textbooks, University of Tunis, Hammamet, 855-864, 7-10 Feb 2007
- Oyama S (2000) The ontogeny of information: Developmental systems and evolution (2nd Ed.). Cambridge University Press, Cambridge-UK
- Oyama S, Griffiths P E, Gray R D (eds) (2001) Cycles of contingency: Developmental systems and evolution. MIT Press, Cambridge-MA
- Rocha PLB, Roque N, Silva SAH et al. (2007) Brazilian High School Biology Textbooks: Main Conceptual Problems in Evolution and Biological Diversity. In: Proceedings of the IOSTE International Meeting on Critical Analysis of School Science Textbooks, University of Tunis, Hammamet, 893-907, 7-10 Feb 2007
- Rosenberg A (1985) The structure of biological science. Cambridge University Press, Cambridge
- Santos VC (2005) Splicing alternativo: implicações para os conceitos de gene e informação. Monografia, Universidade Federal da Bahia
- Sarkar S (1998) Forty years under the central dogma. *T Biochem Sci* 23: 312-316.
- Solha GCF, Waizbort RF (2007) Os genes interrompidos: o impacto da descoberta dos íntrons sobre a definição de gene molecular clássico. *Rev Sociol. Brasil Hist Ciênc* 5:63-82
- Solha GC (2005) Os genes interrompidos: introdução histórica ao impacto da descoberta dos íntrons (1977) na controvérsia sobre a definição de gene molecular clássico (1960). Dissertação, Casa de Oswaldo Cruz – Fiocruz
- Stamm S, Ben-Ari S, Rafalska, I et al (2005) Function of alternative splicing. *Gene*. 344:1–20
- Sterelny K, Griffiths, P E (1999) Sex and death: An introduction to the philosophy of biology. The University of Chicago Press, Chicago
- Stotz K, Griffiths, P, Knight R (2004) How biologists conceptualize genes: An empirical study. *Stud Hist Philos Biol Biomed Sci* 35:647-673
- Xavier M, Freire A, Moraes M (2006) A nova (moderna) biologia e a genética nos livros didáticos de ensino médio. *Ciênc Educ* 12(3): 275-289
- Fire A (1999) RNA-triggered gene silencing. *Trends Genet* 15(9):358-363
- Grosshans H and Slack FJ (2002) Micro-RNAs: small is plentiful. *J Cell Biol* 156(1):17-21
- Hannon GJ (2002) RNA interference. *Nature* 418:244-251
- Lenz G (2005) The RNA interference revolution. *Braz J Med Biol Res* 38(12):1749-1757

ANEXO**Lista dos livros didáticos analisados**

- Amabis JM, Martho GR (2005) *Biologia*. Moderna, São Paulo
- Borba AA, Cançado OFL (2005) *Biologia*. Positivo, Curitiba
- Borba AA, Crozetta MA, Lago SR (2005). *Biologia*. IBEP, São Paulo
- Boschilia C (2005). *Biologia sem segredos*. RIDEEL, São Paulo
- Carvalho W (2005) *Biologia em foco*. FTD, São Paulo
- Cheida LE (2005) *Biologia integrada*. FTD, São Paulo
- Coimbra MA, Rubio PC, Corazzini R et al (2005) *Biologia – Projeto escola e cidadania para todos*. Editora do Brasil, São Paulo
- Faucz FR, Quintilham CT (2005) *Biologia: Caminho da vida*. Base, Curitiba
- Favaretto JA, Mercadante C (2005) *Biologia*. Moderna, São Paulo
- Frota-Pessoa O (2005) *Biologia*. Scipione, São Paulo
- Gainotti A, Modelli A (2005) *Biologia*. Scipione, São Paulo
- Laurence J (2005) *Biologia*. Nova Geração, São Paulo
- Linhares S, Gewandsznajder F (2005) *Biologia*. Ática, São Paulo
- Lopes S, Rosso S (2005) *Biologia*. Saraiva, São Paulo
- Machado SW (2005) *Biologia*. Scipione, São Paulo
- Morandini C, Bellinello LC (2005) *Biologia*. Atual, São Paulo
- Paulino WR (2005) *Biologia*. Ática, São Paulo
- Silva Júnior C, Sasson S (2005). *Biologia*. Saraiva, São Paulo

Considerações Finais

O caminho percorrido nesta dissertação parte da constatação de que foi dado início a uma nova era interessada nas propriedades não-lineares e adaptativas de sistemas dinâmicos complexos, em que as visões de causalidade linear devem ser substituídas pelas análises de redes interagindo com o ambiente e operando através de diferentes níveis de regulação: genético, epigenético, morfogenético e organísmico (Kay 2000). O novo alvo da biologia pós-genômica é ir além de determinismos monogênicos e poligênicos, e mesmo além da genômica funcional: é preciso ir além do gene como salienta a filósofa da biologia E.F. Keller em seu artigo ‘*The century beyond the gene*’ (2005), no qual destaca que ‘trazer o genoma à vida’ é a meta de novos programas de pesquisa, que demarcam um ponto de virada na biologia molecular ao passarem da fase inicial de seqüenciamento genômico à abordagem sistêmica da complexidade biológica (p. 102).

A ênfase em processos é um aspecto notável nos debates atuais sobre a evolução e o desenvolvimento biológico e instila diálogos ainda mais originais no que tange à abordagem processual dos referentes ontológicos do conceito de gene e de informação genética. Nesse caso, ela implica mais incisivamente a ousadia de desafiar o pensamento ocidental em sua tendência de reificação, de favorecimento de entidades, estabilidades e permanências, ao invés de *processos*, *atividades* e *mudanças*. Se os passos nessa direção ainda são incipientes, o mesmo não se pode dizer dos olhares na biologia teórica e na filosofia da biologia, cada vez mais voltados para uma abordagem orientada a processos. Num dos exemplos mais explícitos, Keller salienta não apenas a necessidade de se construir um arcabouço teórico apropriado, mas vai além, ao sugerir que ainda mais fundamental é a necessidade de se construir um arcabouço lingüístico apropriado à biologia, como expressa no que chama de um ‘mero palpite’:

‘Por muito tempo nós tentamos edificar uma biologia a partir de substantivos, uma ciência construída ao redor de entidades. Talvez seja o momento para uma biologia edificada a partir de verbos, uma ciência construída ao redor de processos’ (Keller 2005, p.107).

Idéias relacionadas foram propostas em inúmeras ocasiões com o intuito de sanar lacunas na biologia molecular e na genética desenvolvimental (ver Neumann-Held 2001, Oyama et al. 2001). Parece sensato pensar que, na querela sobre a origem das características dos organismos, a demanda mais basal seja por uma mudança de hábito, por novas formas de pensar e falar que, nos termos de Keller (2005), ‘nos leve além do paradigma de construir o todo a partir das partes, e comece a acomodar a co-construção histórica de partes e todos’ (p. 107). De acordo com a visão desta autora, novas técnicas de análise e, sobretudo, um novo arcabouço conceitual, baseado numa *epistemologia dinâmica e relacional*, e também uma nova linguagem são inexoráveis para a biologia molecular pós-genômica. Que ferramentas teóricas e metodológicas poderão avançar nessa direção é uma questão em aberto, algo ainda em desenvolvimento na comunidade de pesquisadores – porém, ao atentar para os palpites acima, um biossemioticista não pode evitar a sensação de um quê de *déjà vu*.

Numa visão biossemiótica, processos de interpretação de signos têm lugar precisamente onde, em cada ciclo de desenvolvimento, a reconstrução simultânea de partes e de todos demanda a concepção dos sistemas vivos como representações, em seu ecossistema, da história evolutiva de sua espécie. Nessas situações, supõe-se que a visão processual precisa ir além, até mesmo, da ênfase em processos físicos e incorporar processos de memória, representação, categorização e outros normalmente percebidos como pertencentes ao nível cognitivo. Este ponto abre espaço para uma dialética interessante.

A introdução do pensamento informacional na biologia provocou uma mudança de *gestalt* que ressuscitou referências a concepções associadas, de alguma maneira, à teleologia, um tema que acabou virando tabu nas ciências da vida. Atualmente, biólogos moleculares podem, por exemplo, objetar quanto à existência real de semiose em sistemas vivos não humanos, apontando-a como uma atribuição antropomórfica e até mesmo desnecessária: assim ela deve ter sido, eles alegariam, em muitos avanços no conhecimento de fatos sobre a vida nos níveis molecular e celular no século passado. Um biossemiotista, por sua vez, pode replicar que tal progresso na biologia foi influenciado pela introdução de um vocabulário semiótico e postular que este movimento espontâneo foi gerado pelo pensamento (intuitivo) dos biólogos de que um processo interpretativo estaria envolvido e de que a característica essencial dos sistemas vivos seria a semiose e não sua natureza química. A tréplica a esta suposição sustenta que conceitos semióticos se propagaram nas ciências biológicas devido meramente às suas qualidades essencialmente metafóricas, que os tornam dispositivos convenientes para propor hipóteses e para expor, mais simplesmente, um conhecimento físico-químico mais preciso (cf. Emmeche 1999).

Porém, é saliente o quanto se vêm dando conta na biologia, ainda que implicitamente, do potencial heurístico da concepção de signo e de outros conceitos semióticos para a compreensão dos sistemas vivos. A investigação biológica direciona paulatinamente seus holofotes para a compreensão da ação do signo, um elemento ainda ausente nas contendas sobre genes e informação genética, e cuja relevância geral Peirce descreve da seguinte maneira:

‘A mim parece que a função essencial de um signo é tornar eficientes relações ineficientes, – não colocá-las em ação, mas estabelecer um hábito ou regra geral por meio de que elas irão agir oportunamente. [...] O conhecimento, de alguma forma, as torna eficientes; e um signo é algo que por conhecê-lo nós conhecemos algo mais [...] todo nosso pensamento e conhecimento está em signos’ (CP 8.332).

Diante deste trecho, é plausível supor que muitos enigmas na compreensão da complexidade dos organismos podem se dissolver com a aplicação de um arcabouço semiótico que coloque a ação e a interpretação de signos em primeiro plano, em contraposição às reações bioquímicas salientadas nas explicações reducionistas de vias celulares metabólicas e desenvolvimentais. Como ressaltou o biólogo celular Paul Nurse (citado por Williams 1997, p. 476-477), num simpósio sobre reducionismo em 1997, mesmo um sistema biológico tão pequeno quanto uma única célula exibe uma profusão de características que estão ausentes no nível molecular, como informação posicional, compartimentalização, cinética de sinalização, oscilações, ritmos etc. O enfoque substancial em interações moleculares pode eclipsar princípios – não evidentes nas leis que governam constituintes básicos – indispensáveis para explicar a emergência de propriedades em níveis mais elevados de organização. E se é produtivo conceber os organismos como sistemas complexos processadores de informação, é possível que apenas análises moleculares não sejam suficientes e que, em alguns casos, sejam desnecessárias (Williams 1997, p.476-477). Neste contexto, a abordagem da ação de signos como um processo emergente sistêmico pode iluminar a relevância de interações mais sofisticadas, que acoplam redes de processos biológicos situados em contextos específicos.

À primeira vista, uma abordagem semiótica soa como um movimento na contracorrente da genética e da biologia molecular mais em voga. Todavia, argumentamos que a modelagem semiótica é compatível e, sobretudo, uma contraparte necessária aos modelos funcionais e mecanicistas de sistemas genéticos e de sinalização celular. Da perspectiva de um biossemioticista, cabe, até mesmo, aventar a hipótese de que muitos avanços da biologia molecular no século passado só foram possíveis em decorrência de uma ‘semiotização espontânea’, em que biólogos passaram a empregar conceitos semióticos como ferramentas

essenciais às explicações dessa ciência, ainda que de maneira ‘superficial, irrefletida, metafórica’ e ‘num nível intuitivo ou subconsciente’ (Emmeche 1999).

Uma abordagem biossemiótica pode contribuir para a necessidade de se recompor as pressuposições ontológicas hegemônicas na biologia. Em contraposição aos modelos diádicos de signo, em que a significação aparece como uma relação direta entre signo e objeto, a significação na semiótica Peirceana é determinada pelo interpretante que o signo elicia. Neste modelo triádico – da relação entre Signo (S), Objeto (O) e Interpretante (I) –, a significância é intrinsecamente relacionada ao potencial para um processo infinito de sucessivas interpretações, entendido como o processo semiótico.

Conforme Merrell (1995, p.78), Peirce enfatiza relações dinâmicas, ao invés de conteúdo, essência, substância. A ênfase em processos é um aspecto notável de sua teoria dos signos. Como autêntico adepto do pensamento processual, Peirce manifesta mais categoricamente a relevância que atribui a processos, em sua ontologia, na qual eventos não pressupõem substâncias, nem são mudanças perceptíveis nas mesmas, mas as precedem ontologicamente. De acordo com ele, a ‘coisa “existente” é apenas individual no sentido de ser uma lei contínua regulando e unificando eventos de uma série de instantes’ (MS 478:47-48). Há fortes indícios de que a concepção de realidade de Peirce envolve uma ontologia de processos no sentido estrito da palavra (Hulswit 2001).

O pensamento processual, numa versão mais fraca, emerge dentro de uma perspectiva epistêmica, como instrumento conceitual mais apropriado e eficaz para o entendimento do mundo. Em sua versão mais forte – e ao que parece, é a adotada por Peirce – é proeminente o aspecto metafísico, ontológico, e qualquer valor epistêmico que essas idéias venham a ter é atribuído ao entendimento de que as características mais universais e essenciais da realidade, alegadamente, envolvem processos – itens mais bem indicados por verbos do que por nomes,

como ressalta Rescher (1996). Nesta última perspectiva, um sistema semiótico pode ser entendido como um aglomerado relativamente estável, tanto temporal quanto espacialmente, de processos semióticos. A influência do pensamento evolutivo sobre Peirce o levou a enxergar o funcionamento de um dinamismo seletivo em todo lugar – não somente no campo biológico, mas também no campo físico, em suas leis e regularidades, e inclusive no desenvolvimento de nosso conhecimento sobre ele (ibid.).

Acreditamos que uma abordagem processual dos sistemas genéticos de informação nos conduzirá dos determinismos para uma perspectiva mais sistêmica, de organismos como sistemas integrados de processos coordenados. É importante ressaltar que um pensamento orientado a processos está interessado, primordialmente, em modos de compreensão, e não em modos de discurso. Portanto, ele não propõe dispensar termos e conceitos que se referem a ‘substâncias’ – como substantivos – a favor de um vocabulário exclusivamente processual – apoiado no uso de verbos –, ainda que esta seja uma solução plausível, como tem sido discutido por alguns filósofos da biologia, como Keller (2005). Antes, a abordagem processual apenas insiste que termos substantivos são, em princípio, redutíveis a uma fala processual (versão mais forte) ou, ao menos, representam uma maneira mais informativa de compreender as ‘coisas’ (versão mais fraca), conforme Rescher (1996). Em vista disso, nossa preocupação mais imediata não é tanto propor um novo vocabulário para o sistema genético de informação, mas investigar em que medida uma abordagem semiótica processual pode beneficiar a compreensão desse sistema no âmbito científico e, sobretudo, vir a facilitar a transposição didática no ensino de ciências.

Em prol disto, essa empreitada deve envolver, primeiramente, a análise do discurso sobre gene e informação ao longo da história da genética e biologia molecular, incluindo a contenda atual sobre esse tema entre biólogos e filósofos da biologia. A essa etapa nos dedicamos no primeiro capítulo da presente dissertação. Em seguida, é imprescindível investigar como é feita

esta abordagem nos demais níveis da transposição didática, como nos meios de divulgação científica, entre professores do ensino superior e médio, em livros didáticos de biologia e entre estudantes. Nosso grupo de pesquisa tem se dedicado a estes três últimos atores da transposição e o presente trabalho, em particular, enveredou na investigação dos livros didáticos de ensino médio.

Nossos resultados sugerem que há, nos livros didáticos de ensino médio, um equivalente à chamada ‘semiotização espontânea’ da biologia (Emmeche 2000), que se encontra, porém, intrinsecamente associada à controversa concepção seqüencial da informação biológica. Será interessante averiguar, em trabalhos futuros, se a transposição do modelo alternativo semiótico que aqui aplicamos para compreensão dos sistemas genéticos de informação pode ser conduzida em termos de uma ‘semiotização estimulada’ no ensino médio, dentro das possibilidades de transposição didática da própria semiótica. Em outras palavras, será bastante profícuo discorrer sobre as possíveis contribuições da semiótica para o ensino de ciências, em particular, para o ensino sobre os sistemas genéticos de informação, analisando também a viabilidade da transposição didática das explicações biossemióticas para o ensino médio de biologia. Esta tarefa, contudo, requer a discussão prévia das possibilidades da semiótica Peirceana como uma moldura teórica para construir o entendimento de genes como signos e de informação genética – e biológica – como semiose.

Com este intuito, adentramos o caminho da aplicação da moldura teórica Peirceana, especialmente aquela desenvolvida no domínio da semiótica, para construir um entendimento de genes como signos e da informação genética como semiose, algo com importantes conseqüências para o ensino de ciências. Isso decorre, sobretudo, do fato de que tal tratamento se contrapõe ao DNA- ou ‘gene-centrismo’ que caracterizou grande parte da biologia da segunda metade do século XX, bem como visões deterministas genéticas associadas. Elas não somente impedem uma

compreensão adequada do funcionamento do sistema genético e do próprio sistema celular, bem como do desenvolvimento do organismo como um todo, mas também têm importantes implicações políticas, sociais, econômicas e éticas.

Ao aplicarmos o modelo semiótico à regulação da emenda alternativa de RNA por vias de sinalização celular, visamos ressaltar a importância desse processo como desafio ao conceito molecular clássico de gene. Com este desenvolvimento, esperamos ter introduzido elementos que poderão futuramente também contribuir para o avanço em relação ao poder preditivo de tais modelos semióticos, que incluirá a proposição de previsões testáveis sobre vias de sinalização e regulação da emenda alternativa.

Referências

- Atlan, H. & Koppell, M. The Cellular Computer DNA: Program or Data? *Bulletin of Mathematical Biology*, vol. 52 (3), pp. 335-348, 1990.
- Beadle, G.W. & Tatum, E.L. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA*, 27: 499-506.
- Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, vol. 72, pp. 291-336, 2003.
- Carlson, E.A.C. 1991. Defining the Gene: An Evolving Concept. *American Journal of Human Genetics*, 49: 475-487.
- Crick, Francis. H. (1958). On protein synthesis. *Symposium of the Society of Experimental Biology* 12:138-163.
- Downes, S.M. Alternative splicing, the gene concept, and evolution. *History and Philosophy of the Life Sciences*, vol. 26, pp. 91-104, 2004.
- Eco, Umberto, ed. 1984. *Semiotica medievale*. Milano: Bompiani.
- Eco, U. e Sebeok, T. A., eds. 1983. *The Sign of Three: Dulpin, Holmes, Peirce*. Bloomington: Indiana University Press.
- El-Hani, C. N. Controvérsias sobre o Conceito de Gene e suas Implicações para o Ensino de Genética. Atas do V Encontro de Pesquisa em Educação em Ciências. 2005.
- El-Hani, C. 2007. Between the cross and the sword: the crisis of the gene concept. *Genetics and Molecular Biology*, 30(2):297-307.
- El-Hani, C. N.; Queiroz, J. & Emmeche, C. 2006. A semiotic analysis of the genetic information system. *Semiotica* 160(1/4): 1-68.
- El-Hani, C. N.; Arnellos, A. & Queiroz J. 2007. Modeling a semiotic process in the immune system: signal transduction in B-cell activation. *TripleC – Cognition, Communication, Cooperation* 5(2): 24-36.
- Emmeche, C. 1999. The Sarkar challenge: is there any information in a cell? *Semiotica*, 127: 273–93.
- Falk, R. 1986. What is a gene? *Studies in the History and Philosophy of Science*, vol. 17, pp. 133-173.
- Fogle, T. Are genes units of inheritance? *Biology and Philosophy*, vol. 5, pp. 349-371. 1990.
- _____. The Dissolution of Protein Coding Genes, in: Beurton, P.; Falk, R. & Rheinberger, H.-J. *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press., pp. 3-25, 2000.
- Gelbart, William (1998). Databases in genomic research. *Science* 282, 659-661.
- Graveley, B.R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics*, vol. 17 (2), pp. 100-107.
- Griffiths, Paul E. 2001. Genetic information: A metaphor in search of a theory. *Philosophy of Science* 68 (3): 394-403.

- Griffiths, P.E. & Neumann-Held, E. 1999. The many faces of the gene. *BioScience*, 49 (8): 656-662.
- Hall, B. K. 2001. The gene is not dead, merely orphaned and seeking a home. *Evolution and Development* vol. 3, pp. 225-228.
- Hanson, M.R. Protein products of incompletely edited transcripts are detected in plant mitochondria. *The Plant Cell*, vol. 8, pp. 1-3. 1996.
- Hoffmeyer, J. & Emmeche, C. 1991. Code-Duality and the Semiotics of Nature. In: Anderson, Myrdene and Merrell, Floyd (eds.), *On semiotics Modeling*. Berlin: Mouton de Gruyter, 1991, p. 117-166. Reprinted with annotations in *Journal of Biosemiotics*, 2005, 1: 35-85.
- Horowitz, N.H. & Leupold, 1951. U. Some recent studies bearing on the one gene-one enzyme hypothesis. *Cold Spring Harbor Symp. Quant. Biol.* 51: 65-74.
- Jablonka, Eva. 2002. Information: Its interpretation, its inheritance, and its sharing. *Philosophy of Science* 69: 578-605.
- Kampa, D.; Cheng, J.; Kapranov, P.; Yamanaka, M.; et al.. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Research*, vol. 14, pp. 331-342. 2004.
- Kay, Lily. 2000. *Who wrote the book of life? – A history of the genetic code*. California: Stanford University Press.
- Keller, Evelyn Fox. 2002. O Século do Gene. Belo Horizonte: Editora Crisálida.
- Keller, Evelyn Fox. 2005. The century beyond the gene. *Journal of Bioscience*, 30 (1): 101-108.
- Kitcher, P. Genes. *British Journal for the Philosophy of Science* vol. 33, pp. 337-359. 1982.
- Koch, Walter 1971. *Varia Semiotica*. Hildesheim: Olms.
- Köller, Wilhelm. 1977. Der sprachtheoretische Wert des semiotischen Zeichenmodells. In: Spinner, Kaspar H., ed. *Zeichen, Text, Sinn – Zur Semiotik des literarischen Verstehens*, pp. 7-77. Göttingen: Vandenhoeck.
- Lewin, B. 2004. *Genes VIII*. Upper Saddle River-NJ: Pearson Education.
- Lewontin, R.C. The organism as the subject and object of evolution. *Scientia*, vol. 118, pp. 63-83, 1983.
- _____. (2000). *The Triple Helix: Gene, Organism, and Environment*. Cambridge-MA: Harvard University Press.
- Liszka, J. 1996. *A General Introduction to the Semeiotic of Charles Sanders Peirce*. Bloomington, In: Indiana University Press.
- Magen, A. & Ast, G. The importance of being divisible by three in alternative splicing. *Nucleic Acids Research*, vol. 33 (17), pp. 5574-5582, 2005.
- Maynard Smith, J. The concept of information in biology. *Philosophy of Science*, vol. 67, pp. 177-194, 2000.
- Mayr, E. O Desenvolvimento do Pensamento Biológico. Brasília: Editora Universidade de Brasília, 1998.
- Merrell, Floyd (1995). *Peirce's Semiotics Now*. Toronto: Canadian Scholar's Press.

- Morris, Charles. 1964. *Signs, language and behavior*. New York: Prentice-Hall.
- Moss, L. (2001). Deconstructing the gene and reconstructing molecular developmental systems. In *Cycles of Contingency: Developmental Systems and Evolution*, Susan Oyama, Paul E. Griffiths and Russell D. Gray (eds.), 85-97. Cambridge-MA: MIT Press.
- Moss, L. 2003. *What Genes Can't Do*. Cambridge-MA: MIT Press.
- Neumann-Held, Eva. 2001. Let's talk about genes: The process molecular gene concept and its context. In: *Cycles of Contingency: Developmental Systems and Evolution*, Susan Oyama, Paul E. Griffiths and Russell D. Gray (eds.), pp. 69-84. Cambridge-MA: MIT Press.
- Nöth, W. 1995. *Handbook of Semiotics*. Bloomington/In, Indiana University Press.
- Nijhout, H. F. 1990. Metaphors and the role of genes in development. *Bioessays* 12(9):441-446.
- Oyama, S. 1985. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd Ed. Cambridge: Cambridge University Press. 1st Ed.
- Oyama, S.; Griffiths, P.E. & Gray, R.D. (eds.) 2001. *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge-MA: MIT Press.
- Pardini, M.I.M.C. & Guimarães, R.C. A systemic concept of the gene. *Rev. Brasil. Genet.*, vol. 15 (3), pp. 713-721, 1992.
- Peirce, Charles Sanders (1931–1935). *The Collected Papers of Charles Sanders Peirce*. Electronic edition reproducing Vols. I–VI [C. Hartshorne & P. Weiss (Eds.), Cambridge-MA: Harvard University Press, 1931–1935], Vols. VII–VIII [A. W. Burks (Ed.), same publisher, 1958]. Charlottesville: Intelix Corporation. (Aqui referido como CP, seguido do volume e número do parágrafo).
- Pitombo, M. A., Almeida, A. M. R. & El-Hani, C. N. (2008). Conceitos de gene e idéias sobre função gênica em livros didáticos de biologia celular e molecular do ensino superior. *Contexto & Educação (Brasil)* 77: 81-110.
- Pitombo, M. A., Almeida, A. M. R. & El-Hani, C. N. (2008). Gene concepts in higher education cell and molecular biology textbooks. *Science Education International* 19(2): 219-234.
- Portin, P. 1993. The concept of the gene: short history and present status. *Quarterly Review of Biology*, vol. 56, pp. 173-223.
- Queiroz, J. & El-Hani C. 2007. On Peirce's notion of information: remarks on De Tienne's position. *Cognitio* 8(2): 289-298.
- Queiroz, J.; Emmeche, C. & El-Hani, C. N. 2005. Information and semioses in living systems: A biosemiotic approach. *S.E.E.D. Journal: Semiotics, Evolution, Energy, and Development* 5 (1): 60-90.
- Rescher, N. 1996. *Process Metaphysics: An Introduction to Process Philosophy*. New York: State University of New York Press.
- Rheinberger, Hans-Jörg. 2000. Gene concepts: Fragments from the perspective of molecular biology." In Peter Beurton, Raphael Falk, and Hans-Jörg Rheinberger (eds.), *The Concept of the Gene in Development and Evolution. Historical and Epistemological Perspectives*. Cambridge University Press, Cambridge: 219-239.
- Salthe, Stanley N. 1985. *Evolving Hierarchical Systems: Their Structure and Representation*. New York: Columbia University Press.

- Sarkar, Sahotra 1996. Biological information: A skeptical look at some central dogmas of molecular biology. In: *The Philosophy and History of Molecular Biology: New Perspectives*, Sahotra Sarkar (ed.), pp. 187-231. Dordrecht: Kluwer.
- Sarkar, S. 1998. *Genetics and Reductionism*. New York: Cambridge University Press.
- Shannon, Claude E. & Weaver, Warren. 1949. *The Mathematical Theory of Communication*. Urbana-IL: University of Illinois Press.
- Solha G. e Waizbort R. (2007) Os genes interrompidos: o impacto da descoberta dos íntrons sobre a definição de gene molecular clássico. *Rev Socied. Brasil Hist Ciênc*, 5:63-82
- Solha G. 2005. Os genes interrompidos: introdução histórica ao impacto da descoberta dos íntrons (1977) na controvérsia sobre a definição de gene molecular clássico (1960). Dissertação, Casa de Oswaldo Cruz – Fiocruz
- Stamm, S.; Ben-Ari, S.; Rafalska, I. et al. 2005. Function of alternative splicing. *Gene*, 344:1-20.
- Sterelny, K. & Griffiths, P. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press, 1999.
- Stotz, Karola, Griffiths, Paul E. and Knight, Rob 2004. How biologists conceptualize genes: An empirical study. *Studies in the History and Philosophy of Biological & Biomedical Sciences* 35: 647-673.
- Stuart, C. I. J. M. 1985. Bio-informational equivalence. *Journal of Theoretical Biology* 113: 611-636.
- Thanaraj, T.A.; Stamm, S.; Clark, F. et al. ASD: the alternative splicing database. *Nucleic Acids Res.*, vol. 32, pp. D64–D69, 2004.
- Venter, J.C. et al. 2001. 1994. The sequence of the human genome. *Science*, 291 (5507): 1304-1351.
- Waters, C. K. Genes made molecular. *Philosophy of Science*, 61: 163-185.

