



**João Pedro
Oliveira de Faria**

**Genética da Evolução: tradução comentada de
artigo científico**



**João Pedro
Oliveira de Faria**

Genética da Evolução: tradução comentada de artigo científico

Projecto apresentado à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Tradução Especializada, realizada sob a orientação científica da Professora Doutora Maria Teresa Murcho Alegre, Professora Auxiliar do Departamento Línguas e Culturas da Universidade de Aveiro

Dedico este trabalho à minha família e amigos, pelo apoio e motivação constante.

o júri

presidente

Prof. Doutora Maria Eugénia Tavares Pereira
Professora Auxiliar da Universidade de Aveiro

Prof. Doutora Margarida Sâncio da Cruz Fardilha
Professora Auxiliar Convidada da Universidade de Aveiro (arguente)

Prof. Doutora Maria Teresa Murcho Alegre
Professora Auxiliar da Universidade de Aveiro (orientadora).

agradecimentos

Agradeço à Professora Doutora Maria Teresa Murcho Alegre, o seu constante acompanhamento, as suas revisões, comentários e contributos foram de facto vitais para a execução deste projecto.

Agradeço à Doutora Anabela Cachada e ao Doutor Pedro Pato Martins, os seus incontáveis contributos facilitaram imenso a execução deste projecto. São dois bons amigos.

Agradeço aos meus pais pelo esforço que fizeram para me darem a oportunidade de seguir a via académica na área que realmente gosto. Os constantes apoios, ensinamentos e, acima de tudo, amizade, trouxeram-me até aqui.

Agradeço aos colegas de trabalho pela motivação para este trabalho, e por criarem um ambiente de trabalho fantástico.

Agradeço aos amigos por não colaborarem directamente neste projecto, caso contrário não teria qualquer momento de diversão ou convívio.

Todos temos alguém especial que desperta, no verdadeira sentido da palavra, o melhor que há dentro de nós. Não sou excepção. Por isso, a ela dou o meu mais sincero agradecimento.

palavras-chave

Tradução Especializada, Terminologia, Ferramentas CAT, Glossário, Genética, Evolução.

resumo

No contexto da Tradução Especializada, este projecto aborda a metodologia e as problemáticas inerentes, utilizando como texto de partida um artigo científico relativo à Genética da Evolução. O projecto é composto pela análise textual, análise do tema, abordagem do Tradutor ao tema, metodologia do Tradutor, tradução comentada e glossário.

Trata-se de um tema que necessita de uma abordagem expositiva, de modo a fornecer uma compreensão mais detalhada e exacta do processo de Tradução Especializada.

keywords

Specialized Translation, Terminology, CAT tools, Glossary, Genetics, Evolution.

abstract

This project will approach the methodology and problematic of Specialized Translation, using a scientific paper on the Genetics of Evolution as source text. The project comprises textual analysis, specialized field analysis and its approach by the Translator, the Translator's methodology, commented translation and a glossary.

It is a subject that requires an expositive approach, in order to provide a detailed and exact understanding of the Specialized Translation process.

Índice

1. Introdução.....	1
2. Contextualização e enquadramento.....	5
3. Objectivos.....	7
4. A evolução do estudo da Evolução.....	9
5. Descrição e Análise do Texto de Partida.....	15
6. Características do Texto de Partida.....	17
6.1 Macro-estrutura.....	17
6.2 Micro-estrutura.....	21
6.2.1 Sintaxe.....	21
6.2.2 Voz passiva.....	23
6.2.3 Grupos nominais.....	24
6.2.4 Siglas.....	25
6.2.5 Frases interrogativas.....	27
6.2.6 Argumentação.....	30
7. Metodologia da Tradução.....	35
7.1 Ferramentas.....	37
7.1.1 Critério de seleção.....	37
7.1.2 Preparação do texto.....	39
7.1.3 Análise de ferramentas.....	40
7.1.4 Extração terminológica, tradução e definição.....	44
7.2 Problemas de tradução.....	49
8. Conclusão.....	61
9. Glossário.....	63
10. Bibliografia.....	89
11. Anexos.....	93

1. Introdução

Enquadra-se o presente projeto no Mestrado em Tradução Especializada, compreendendo técnicas e conhecimentos na área da Tradução Especializada adquiridos na Licenciatura em Tradução e Mestrado em Tradução Especializada, existindo também um forte contributo fora do âmbito académico, relativo à experiência e domínio de ferramentas e conceitos, obtidos através da experiência profissional como tradutor e gestor de projetos numa empresa de tradução, ao longo de cerca de dois anos. É de destacar o Mestrado em Tradução Especializada, pois forneceu-me os conhecimentos e capacidades necessárias para elaborar um projeto de Tradução Especializada, especificamente na área da Saúde e Ciências da Vida.

A motivação deste projeto partiu do interesse que surgiu após um trabalho académico que me aproximou de um tema que compreende a área de estudo do objeto de tradução utilizado no presente trabalho: os Estudos sobre a Evolução. O referido trabalho focava a vertente de tradução audiovisual e, apesar de o resultado prático desse projeto não trazer qualquer benefício direto para o presente trabalho, os conhecimentos adquiridos no momento de estudo da área da evolução foram úteis, nomeadamente após a sua conjugação com conhecimentos adquiridos nas disciplinas de Nomenclatura do Conhecimento Científico e Fundamentos em Ciência, especificamente o tema da Genética. A conjugação destes conhecimentos fez com que um tema tão distante de problemáticas da tradução, como o é a genética da evolução, não permanecesse tão inacessível ou incompreensível. Quando aliada a um forte e sincero interesse, a compreensão de um determinado tema, por distante que seja, torna-se de facto acessível. Este é, afinal, o ponto de partida para um projeto de Tradução Especializada: estudo e compreensão genérica do tema de especialização.

O presente projeto aborda não só o tema da especialização no âmbito da tradução como também modernos fatores que impactam diretamente no próprio *modus operandi* do tradutor especializado, os quais necessitam obrigatoriamente de ser abordados. Como é sabido, a tradução especializada não consiste apenas num processo de procura de equivalências para terminologia específica; o tradutor especializado tem que ler e escrever tal como os membros dos grupos sociais dessas áreas leem e escrevem. Para compreender um texto médico na língua de partida, há que o ler como um

médico, e para o traduzir corretamente há que escrever como um médico, na língua de chegada. Trata-se de um processo que não requer, necessariamente, reformulação (Robinson, 1997). Talvez seja discutível ou até radical, esta constatação de Robinson, mas o conceito base (o tradutor especializado necessita de garantir uma elevada compreensão e uma fiel reprodução escrita da linguagem utilizada) integra a própria definição de Tradução Especializada.

Relativamente aos referidos fatores, estes não alteraram o processo, simplesmente aceleraram-no. No entanto, a veloz sociedade na qual vivemos influencia diretamente o modo como trabalhamos. A globalização e as novas tecnologias de informação alteraram o nosso mundo para sempre, e os seus efeitos repercutem-se em processos que à partida não necessitam de renovação, como é o caso da Tradução Especializada. A facilidade de acesso a meios de informação, particularmente a Internet, e o desenvolvimento de ferramentas específicas para uma gama praticamente ilimitada de áreas de trabalho tornou o trabalho de Tradução Especializada incrivelmente mais acessível, rápido e fácil. No entanto, a premissa original da Tradução Especializada, a emulação da leitura e escrita do grupo social de uma determinada área de especialidade, está a ficar esquecida devido à acessibilidade, rapidez e apoio de que o tradutor atualmente beneficia. Trata-se de um paradoxo bem conhecido: a facilidade de acesso à informação “engana” o tradutor, pois este parte do princípio que dispõe de tudo o que necessita para a Tradução Especializada e acaba por esquecer o pilar mais importante desta vertente da Tradução. O resultado é visível nos Curriculum Vitae de inúmeros tradutores que afirmam ser especializados em várias áreas do conhecimento, por vezes dezenas, que muitas vezes são tão díspares como Engenharia Mecânica e Oftalmologia, a título de exemplo (Stolze 2001; Gouadec 2007). Seria de facto um indivíduo extraordinário, o tradutor especializado em várias dezenas de áreas de conhecimento complexas e totalmente díspares.

Não obstante este, chamemos-lhe, paradoxo da facilidade do acesso à informação, os fatores modernos serão para sempre indissociáveis à Tradução Especializada, e será necessária uma maturação da Tradução Especializada moderna para que o profissional da tradução recue um pouco no tempo e reconsidere se deverá designar-se *Especializado em*, caso não possua a capacidade de “ler e escrever como um especialista da área” (Robinson 1997).

Por estes motivos, um principal objetivo deste projeto é a defesa da Tradução Especializada, auxiliada pelo comentário de um processo de tradução que recorreu às mais modernas ferramentas e meios de apoio à Tradução Especializada. O sucesso de parte deste objetivo, designadamente o texto traduzido, irá determinar o cumprimento de um outro objetivo não menos importante: a difusão de informação relevante da área aos especialistas da biologia evolutiva e do

desenvolvimento. De facto, este segundo objetivo requer uma tradução eficaz, sustentada pelo conhecimento geral da matéria e pelo conhecimento específico da terminologia.

A concretização destes objetivos utiliza um modelo prático e genérico da Tradução Especializada, composto pelo processo de pré-tradução, tradução e revisão: as três fases elementares do processo de tradução da norma EN15038. Não é consensual afirmar que o objetivo da teoria da tradução é melhorar a qualidade da tradução, simplesmente porque os teóricos da tradução têm objetivos e responsabilidades diferentes dos praticantes da tradução (Toury, 1995). Por este motivo, a opção recaiu na utilização de um modelo funcional e não numa teoria da tradução. Este modelo, devidamente implantado, contribui para o cumprimento de ambos os objetivos previamente mencionados: o processo de pré-tradução abrange o estudo, compreensão e investigação do tema, necessários para a efetiva especialização do tradutor; a qualidade da tradução está dependente, em última análise, do processo de tradução (auxiliado pelo resultado da pré-tradução) e revisão.

As fases deste modelo funcional serão pois descritas e comentadas, fornecendo assim a fundamentação necessária para validar as técnicas utilizadas, mas também para validar o próprio projeto, pois sem descrição e comentário não seria possível a defesa e justificação de escolhas tradutológicas ou critérios de seleção ou utilização de recursos e métodos.

2. Contextualização e enquadramento

Este trabalho é o projeto final do Mestrado em Tradução Especializada, na vertente de Saúde e Ciências da Vida.

A estrutura do projeto foi construída em prol dos temas que julguei mais pertinentes abordar no âmbito deste Mestrado, nomeadamente a área das Ciências da Vida e o processo de Tradução Especializada. Deste modo, o projeto encontra-se dividido em partes distintas, sendo que o objeto de estudo é o processo de tradução do artigo. No entanto, todos os restantes processos estão diretamente relacionados com a tradução e merecem atenção e descrição detalhada.

A ideia de elaborar um projeto relacionado com a Biologia Evolutiva do Desenvolvimento surgiu no seguimento de um projeto de legendagem na disciplina de Tradução Audiovisual, no qual foi utilizado como objeto de tradução um documentário sobre Charles Darwin, narrado em Inglês. Trata-se de um tema que despertou a atenção e, chegado o momento de escolha de um objeto de tradução para este projeto, foi efetuada a pesquisa por um artigo científico relacionado com esta área. Um importante pré-requisito na escolha do artigo científico foi o conteúdo rico em elementos que possibilitassem uma relevante exploração na área da tradução especializada. A pesquisa por textos académicos demonstrou que seria relativamente fácil encontrar um artigo que cumprisse com estes requisitos. Esta pesquisa trouxe benefícios, nomeadamente a familiarização com o conceito de *evo-devo*, ou biologia evolutiva do desenvolvimento, sendo esta a área que mereceu destaque e sob a qual recaiu a escolha, pois é reconhecida como um dos campos mais vanguardistas e atuais na área da biologia que trouxe uma dramática compreensão dos processos de desenvolvimento e de como estes podem explicar a diversificação evolutiva que origina as inúmeras formas de vida existentes (Brakerfield, 2011). A terminologia do estudo da biologia evolutiva do desenvolvimento é vanguardista, recente e altamente específica, o que a torna ideal para este trabalho.

A escolha recaiu sobre o artigo “The Loci of Evolution: How Predictable is Genetic Evolution?” por diversos motivos, nomeadamente o facto de conter a referida terminologia, e conter vários elementos na macro e micro-estrutura que suscitam questões de tradução de interesse.

Este texto revelou-se de uma complexidade imensa para o leitor não-especializado, como se pode constatar apenas com a leitura do resumo. Na literatura académica, o propósito do resumo é comunicar sucintamente toda a complexa pesquisa efetuada. Neste caso, a complexidade presente

no resumo está também presente em todo o texto – uma frase selecionada ao acaso terá grande probabilidade de necessitar da compreensão de conceitos avançados para ser corretamente interpretada. Trata-se, portanto, do objeto ideal para ser analisado num projeto de Tradução Especializada, porque permite a integração de todos os conceitos e técnicas de tradução aprendidos no meio acadêmico: estudo e análise do tema com recurso a obras da especialidade (artigos científicos, livros técnicos, documentários, manuais escolares, entre outros); processo de pré-tradução com recurso a ferramentas CAT, produção de corpus especializado, base terminológica e consequente glossário.

3. Objetivos

O objetivo deste trabalho académico é, acima de tudo, enriquecer competências de tradução através da produção de um projeto de tradução especializada de raiz. O resultado será a compilação de todos os conhecimentos adquiridos no Mestrado, juntamente com valiosíssimas competências adquiridas na profissão de tradutor.

Um dos objetivos deste projeto é formular uma aplicação prática dos conteúdos do mesmo. Neste sentido, este projeto foi conceptualizado tendo em conta uma possível utilização futura por parte do público-alvo. O primeiro público-alvo do projeto é constituído pelos docentes que irão avaliar o projeto. Este passo é essencial, pois a validação e uma avaliação positiva por parte de especialistas na área da tradução irá determinar a sua adequação a um público-alvo mais abrangente, nomeadamente estudantes de tradução, tradutores, estudantes de biologia e biólogos.

Começando pelo estudante de tradução, justifica-se esta opção por se tratar de um relatório de uma tradução especializada e como tal, um dos objetivos é demonstrar como se processa um projeto de tradução especializada na íntegra e também como é que um tradutor se “especializa” numa área concreta. A experiência como estudante numa licenciatura em Tradução e Mestrado em Tradução Especializada deu as competências necessárias para fazer Tradução Especializada, e este projeto demonstra como é que um tradutor se especializa em determinada área, que métodos e ferramentas utiliza e que tipos de problemas surgem no processo. O tradutor profissional poderá também beneficiar do conteúdo deste projeto, nomeadamente os produtos finais do processo de tradução, ou seja, a própria tradução do artigo – que poderá ser utilizada no contexto de criação de um corpus com textos paralelos na área em questão – e o glossário, que fornece terminologia de elevada especificidade que, apesar de se encontrar num par linguístico relativamente comum, necessita de pesquisa aprofundada e conhecimento da área.

Relativamente ao estudante de biologia, a utilidade deste projeto justifica-se por vários motivos; um glossário bilingue com definição em português que poderá ajudar este estudante a melhor compreender determinados conceitos; a tradução de um texto científico relevante e atual, que possibilita o acesso a um estudo nem sempre disponível ao estudante – apesar do contacto regular que este tem com a língua inglesa, o inglês técnico tem uma dificuldade acrescida, e é necessário um contacto prolongado com este tipo de texto para uma compreensão integral; a tradução deste

documento, que pode despertar o interesse a estudantes que pretendam aprofundar conhecimentos, independentemente do seu domínio da língua inglesa. Também os especialistas nesta área, biólogos e/ou docentes ou investigadores da área da biologia, nomeadamente da biologia evolutiva e do desenvolvimento têm, quase obrigatoriamente, que lidar com a língua inglesa regularmente, porque de facto os estudos mais relevantes são acedidos a partir de publicações científicas internacionais, sendo que a maioria, e com maior fator de impacto, encontram-se na língua inglesa (“Nature”, “The Journal of Cell Biology”, “PNAS”, “Molecular Biology and Evolution”, “Proceedings of the Royal Society: Biological Sciences”, etc.). Regra geral, os especialistas deste campo estudam e escrevem na língua inglesa, o que descarta a necessidade de recorrer a serviços de tradução. No entanto, este fenómeno faz com que a mensagem nem sempre seja transmitida como desejado, porque estes autores não são, por norma, especialista na língua, mas antes indivíduos com sólidos conhecimentos técnicos que dominam os conceitos científicos. Assim, para se obter um resultado final desejável – um artigo disponibilizado em inglês com conteúdo e forma perfeitos – há a necessidade de recorrer a um linguista. Neste caso, defende-se a colaboração do autor do artigo com um tradutor especializado neste campo, visto a Tradução Especializada implicar a fusão necessária entre a compreensão de conceitos e o domínio da língua.

Serve assim este projeto para demonstrar a validade deste tipo de tradução mesmo em áreas da ciência tão inovadoras e vanguardistas como a aqui representada.

4. A evolução do estudo da Evolução

Existem mais de 9000 espécies de pássaros, mais de 300 000 espécies de escaravelhos e apenas uma pequena fração das espécies de peixes são conhecidas. O número de espécies conhecidas que habitam o planeta é superior a dois milhões. Charles Darwin não sabia porquê. Sabemos atualmente que uma espécie poderá dar origem a várias, e é a Darwin que o devemos agradecer (Feijó, 2009).

Charles Darwin era apenas um rapaz que gostava de colecionar insetos e plantas, não gostava muito de ir à escola. Tanto que abandonou a Universidade de Edimburgh, uma das melhores universidades europeias naquele tempo, mas o seu pai decidiu que ele devia ir para um mosteiro, já que a vida académica não lhe interessava. Lá, teve a oportunidade de estudar ciências naturais, aprofundando o seu interesse no colecionismo de insetos e plantas, e começou a tornar-se num estudante mais sério e empenhado (“What Darwin Never Knew”, 2009). Até que surgiu a oportunidade de embarcar numa viagem de barco, o HMS Beagle, que o levou a conhecer o mundo. Encontrou os primeiros fósseis de enormes animais extintos na Patagónia que o deixaram perplexo e a questionar-se sobre as suas semelhanças extraordinárias com os pequenos armadillos, mas foi nas ilhas Galápagos que se iniciou a grande revelação de Darwin (Darwin, 1859).

Este arquipélago é composto por onze ilhas e cada uma tinha os mesmos animais, mas com diferenças físicas acentuadas. As tartarugas gigantes, por exemplo, eram diferentes de ilha para ilha, assim como os tentilhões. Darwin questionou-se por que Deus se daria ao trabalho de arquitetar tentilhões diferentes para as várias ilhas. Porquê tanta variedade? Talvez fossem todos da mesma espécie, mas com características diferentes. Consoante estudava os tentilhões, percebia que as formas e dimensões dos seus bicos estavam adaptadas à flora local de cada ilha: talvez fosse uma única espécie que se foi adaptando às características de cada uma das ilhas. Talvez.

Darwin regressou da viagem ansioso por continuar este estudo, e percebeu que a resposta estava na plataforma desta variedade, os embriões. Começou a estudar embriões de serpentes e reparou que, nas primeiras fases embrionárias, estes tinham pequeninos ossos no local da perna que não se desenvolviam mais. Talvez isso significasse que fossem descendentes de outras espécies com pernas. Mas mais surpreendente foi a sua descoberta destes mesmos ossos em embriões de peixes. Será que éramos todos descendentes de peixes? A resposta para esta dúvida estava presente numa

característica da cultura britânica: a criação de raças de cães. Os britânicos adoram os seus cães, e cruzam-nos para criar novas raças ou selecionam apenas os “melhores” exemplares de cada espécie para obter os resultados pretendidos. E assim surgiu a teoria da Seleção Natural. A sobrevivência do mais forte e melhor adaptado ao ambiente. Darwin não conseguia comprovar esta teoria, o estudo a nível molecular era inexistente.

Hoje em dia, há vários estudos que comprovam esta teoria: Michael Nachman, investigador da Universidade do Arizona, tem um interesse pouco habitual por ratos do deserto. Numa determinada área do deserto do Arizona havia uma enorme área cujo solo era totalmente negro, resultado de um vulcão que tinha eclodido há muito tempo naquela área. Nessa mesma área, os ratos do deserto eram de cor preta e não de cor amarelada, como os restantes ratos do deserto. Após o seu estudo genético compreendeu que a cor preta era uma mutação, e os ratos que tinham esta mutação sobreviviam muito melhor, devido à sua camuflagem, na área das rochas negras do que nas areias amarelas do deserto (“What Darwin Never Knew”, 2009). Hoje em dia compreendemos o porquê destas mutações. Estas ocorrem a nível genético. Por este motivo, o estudo da Evolução implica o estudo da Genética.

Tudo começa com uma molécula à qual chamamos ADN. A molécula de ADN, que transporta toda a informação necessária ao crescimento e desenvolvimento de um organismo, é diferente de todas as outras. O seu aspeto, por exemplo, é estranho: uma hélice dupla de longa dimensão composta por grandes quantidades de quatro moléculas mais pequenas: adenina, timina, citosina, guanina. Estas moléculas são representadas por letras (A,T,C,G), que estão sempre representadas em pares, designados de par de base, e com essas quatro letras, o ADN consegue fazer “palavras” praticamente infinitas e assim dar origem a tanta variedade. Estas moléculas são sequências muito especiais dos nossos genes, que determinam a nossa aparência física, desde a cor dos olhos e cabelo até às sardas que temos na cara. Estes genes são traduzidos em proteínas, que criam os músculos e cabelos, por exemplo. Há uma outra particularidade nestes genes: podem mudar, ou aliás podem mutar (Feijó, 2009).

No caso dos ratos do deserto, os investigadores percorreram extensas cadeias de ADN do rato negro e do rato amarelo até que encontraram uma pequena diferença num único par de bases: quando a sequência CG CA CC alterava a primeira molécula para uma G (guanina), surgia uma mutação nos genes de amarelo para preto, e o ADN do rato tinha agora a sequência GG CA CC (“What Darwin Never Knew”, 2009). Mutação genética é, no fundo, Evolução.

Darwin compreendia que as espécies evoluíam, e todos os estudos da Evolução comprovam isso mesmo. Mas o que nunca compreendeu foi como é que evoluíam. A resposta a esta dúvida inicia-se

num pequeno inseto, a mosca da fruta, mais conhecida pelo seu nome binominal, *Drosophila melanogaster*. Este inseto é muito utilizado em experiências de Genética, e é um dos mais importantes organismos modelo na área da Biologia. Este pequeno organismo é responsável pela criação do primeiro centro mundial de biologia experimental, que era no fundo uma modesta sala repleta de espécimes de *Drosophila melanogaster*, conhecida como “sala de moscas”, na Universidade de Columbia. Thomas Hunt Morgan era o investigador responsável e mentor do projeto (Feijó, 2009).

De entre os inúmeros estudos da *Drosophila*, destaca-se o estudo das pintas negras existentes nas asas de alguns espécimes. Esta característica, que no fundo é uma mutação, torna o espécime mais atraente para o sexo oposto, o que significa que o sortudo que tiver esta mutação irá encontrar uma parceira mais facilmente. O estudo genético dos dois tipos de mosca (com e sem pintas) não revelou qualquer gene diferente nas espécies. As moscas tinham os mesmos genes, apenas utilizavam-nos de modos diferentes. Para compreender isto, há que ter em conta que nem todo o ADN se traduz em proteínas, aliás, apenas 2% do ADN se traduz em proteínas. Ninguém faz a mínima ideia em quê que se traduz os restantes 98% do ADN. Chamam-lhe a matéria negra do genoma.

Para determinar que cadeia de ADN provocava esta mutação, Sean B. Carroll, investigador na área da evolução, que estuda de perto a regulação *cis*, no contexto da biologia de desenvolvimento, percebeu que a resposta deveria estar em moléculas de ADN que não se traduzem em proteínas. A sua investigação concluiu que o gene que provocava esta pinta era exatamente igual em ambas as moscas, mas havia uma sequência de ADN diferente (Carroll, 2005). Esta diferença deve-se à existência de reguladores nas cadeias de ADN. Estes reguladores funcionam como um interruptor da luz do quarto: a lâmpada está lá, mas tanto pode estar ligada como desligada. Os reguladores *ligam* e *desligam* determinados pares de bases. No caso da mosca, apesar de o ADN ser igual, uma tinha o regulador ativo, na região de ADN responsável pela pinta, e a outra não. Os reguladores não são genes, são apenas uma função dos genes. Exemplificando, a sequência CC AG TA CC com um regulador inativo no terceiro par de bases resultaria numa sequência CC AG ta CC, ou seja, o par de bases TA está fisicamente presente, mas inativo. É essencial compreender a existência e funcionamento destes reguladores para responder à questão de Darwin.

Darwin foi publicamente condenado por teorizar que o ser humano poderia ser descendente de primatas. Hoje em dia esta descendência direta é um facto assumido e mais que comprovado. Tanto que o estudo da descendência humana está agora a recuar muito mais no tempo e a dedicar-se a um organismo que ainda hoje vive no meio de nós, o peixe.

O peixe esgana-gata tem sido alvo de muitos estudos recentes, principalmente porque existem espécies deste peixe que habitam água salgada e água doce, com pequenas diferenças, como é óbvio. Neil Shubin, biólogo evolutivo e autor da obra *Your Inner Fish*, reparou numa diferença vital: o peixe de água salgada podia-se proteger dos outros peixes por causa dos seus espiões externos, que o tornavam difícil de engolir, mas o peixe de água doce não tinha estes espiões, o que o tornava numa presa fácil. Ambos os peixes tinham os mesmos pequenos ossos durante o estado embrionário, mas o peixe de água salgada desenvolvia esses ossos em espiões, enquanto o peixe de água doce tinha uma extensão óssea que mais se assemelhava a uma perna, muito semelhante ao portentoso manatim.

O manatim é um mamífero aquático de água doce que habita geralmente águas costeiras e rasas. O estudo desta espécie conclui ser descendente de um animal muito semelhante (*Tiktaalik roseae*) que poderia apenas defender-se das outras espécies fugindo, por vezes para fora de água. Para tal, desenvolveu a perna e adaptou-se às condições atmosféricas. Esta conclusão foi obtida a partir do estudo destes reguladores, auxiliada pela existência de fósseis que comprovam a existência do referido precursor do manatim, que ficam ativos ou desativos durante o seu desenvolvimento no estado embrionário. O estudo destes reguladores revelou que a sua ativação e desativação são também altamente variáveis, dependendo do tempo e intensidade, ou seja, voltando à comparação com o interruptor da luz, são interruptores que podem estar ligados ou desligados, e a intensidade da luz pode ser elevada ou fraca, e poderá ser ligada durante muito ou pouco tempo (Shubin, 2008).

O estudo da regulação do ADN é, por estes motivos, alvo de elevado interesse. De facto, alteraram completamente o modo como a Biologia estudava as mutações e a evolução, dando assim origem a esta nova área de estudo tão popular hoje em dia que até deu origem a um nome informal muito em voga: *evo-devo*. A *evo-devo*, ou biologia evolutiva do desenvolvimento, estuda como evoluem os processos de desenvolvimento e compara o desenvolvimento entre diferentes organismos de modo a determinar a relação ancestral entre estes (Laublicher, M. 2007).

Trata-se de uma área que está em franca expansão, que é para mim de acarinhado interesse. No entanto, a produção do presente texto não é proveniente desse interesse; todo o conteúdo do presente texto é o resultado do estudo e investigação do tema no âmbito de um projeto de Tradução Especializada.

É minha convicção que uma especialização deverá implicar a compreensão geral do tema e de conceitos-chave. Considero, no mínimo, modestos os meus conhecimentos na fascinante área da biologia evolutiva do desenvolvimento, no entanto estou seguro que são o ponto de partida para

uma compreensão geral do tema e de determinados conceitos, de modo a assegurar o objetivo final da Tradução Especializada: ler e escrever tal como um especialista da área.

5. Descrição e Análise do Texto de Partida

O texto de partida eleito como objeto de estudo intitula-se “The Loci of Evolution: How Predictable is Genetic Evolution?” e foi publicado primeiramente no volume 62-9 da revista *Evolution* (Society for the Study of Evolution) (pp. 2155-2177). Trata-se de uma publicação científica com fator de impacto de 5.6 e é bastante citada no campo da biologia evolutiva (ISI Journal Citation Reports © Ranking: 2010: 6/45 [Evolutionary Biology]).

O artigo em questão é da autoria de David L. Stern e Virginie Orgogozo. David L. Stern, B.S., PhD, é Professor de Ecologia e Biologia Evolutiva na Universidade de Princeton, onde é também diretor do Stern Laboratory, e investigador no Howard Hughes Medical Institute. Virginie Orgogozo, PhD, é investigadora na Universidade Paris Diderot e trabalhou no seu primeiro pós-doutoramento com David L. Stern.

O artigo que desenvolveram em conjunto é um artigo científico de refutação (*rebuttal*) e com revisões por pares, e o objeto de análise é a discussão no meio académico sobre os fatores que determinam a previsibilidade da evolução genética. Segundo os autores do artigo, ainda não estão devidamente esclarecidas, na área da biologia evolutiva, questões essenciais, necessárias para se poder considerar a questão da previsibilidade da evolução genética. Deste modo, o texto analisa as várias hipóteses já existentes através da descrição e crítica de cada uma, acompanhadas de análises e provas empíricas, e compara as hipóteses que são consideradas mais prováveis, para a resposta à questão inicial. Tecnicamente, trata-se de uma comparação entre duas regiões do ADN (codificação e regulação *cis*), sendo que cada uma é analisada em grande pormenor. Os autores recorrem a outros artigos científicos publicados sobre as várias hipóteses, que são criticados ou utilizados para exemplificação. É historicamente importante referir o destaque dado aos artigos dos autores Hopi E. Hoekstra e Jerry A. Coyne, dois biólogos de competências reconhecidas, popularizados por um artigo publicado em 2007, que põe em causa um dos pilares da biologia evolutiva (a evolução morfológica resulta de alterações na regulação *cis*). Foram vários os autores que literalmente atacaram os argumentos de Hoekstra e Coyne (Craig, L.; *Defending Evo-Devo: A Response to Hoekstra and Coyne*, por exemplo), e o próprio título do artigo de David L. Stern e Virginie Orgogozo é nada mais que uma

referência ao artigo de Hoekstra e Coyne que despoletou esta polémica: *The Locus of Evolution: Evo Devo and the Genetics of Adaptation*, (também publicado na revista científica *Evolution*, volume 65, pp. 995–1016). Este artigo despertou a atenção por se tratar de uma crítica que, no fundo, afirma existirem sérias dúvidas sobre princípios básicos da biologia evolutiva – e é particularmente difícil publicar um artigo crítico numa publicação científica. Polémicas aparte, os autores do presente artigo deste projeto aceitaram estas críticas construtivas e decidiram analisar pressupostos antigos tendo em conta as variáveis postas em causa por Hoekstra e Coyne, nomeadamente os efeitos da regulação *cis* na mutação evolutiva. Os autores do artigo apoiam as suas afirmações com argumentos que tinham sido postos de lado pela comunidade da biologia evolutiva relativamente a este tema (genética populacional), além da clara argumentação com recurso a factos, dados estatísticos, citações e gráficos de análise. Os autores prosseguem com uma descrição dos dados utilizados para a sua argumentação e apresentam três distintas previsões relativas à questão da origem da mutação. A conclusão do artigo não é definitiva, pois o argumento da genética populacional não fornece todos os dados necessários, além de admitirem não existir uma certeza sobre a origem das mutações evolutivas. No entanto, terminam concluindo que as mutações subjacentes à evolução fenotípica poderão vir a ser previsíveis, assim que existirem estudos suficientes sobre os dados necessários. Após a conclusão do artigo, existem ainda dois apêndices. O primeiro indica a existência de uma base de dados de artigos (na página online da revista *Evolution*) que fornecem evidências das alterações genéticas que contribuíram para a evolução em raças domésticas, além de referirem o tipo de estudos que incluíram e respetiva justificação. O segundo apêndice dirige-se especificamente ao previamente mencionado artigo de Hoekstra e Coyne. Trata-se de uma análise aos outros argumentos mais importantes do texto de Hoekstra e Coyne que não foram referidos no artigo.

6. Características do Texto de Partida

A caracterização do texto de partida será dividida entre a análise à sua macro-estrutura e micro-estrutura, ou seja, a análise à estrutura global de significado de um texto e a análise aos códigos estilísticos responsáveis pela coerência linear, respetivamente.

6.1 Macro-estrutura

O texto de partida está formatado de acordo com as normas de publicação delineadas pela revista na qual foi editado, contudo existem também vários aspetos que remetem para a macro-estrutura típica do artigo científico. Estes aspetos saltam imediatamente à vista devido à disposição gráfica e registo linguístico. No entanto existem vários outros elementos que serão de seguida analisados.

O artigo encontra-se dividido em duas colunas, com exceção para os elementos textuais que servem de legenda a figuras ou tabelas, visto estarem formatados numa só linha. Esta disposição de texto é habitual em artigos científicos publicados em revistas da especialidade (ver, por exemplo, Wray, 2007; Rockman, 2002). Os vários textos que compõe o artigo seguem uma formatação habitual neste género de artigo científico, no entanto contêm características únicas na organização e disposição dos textos.

A organização dos textos corresponde à fórmula base da escrita argumentativa (Weston, 1996), sendo a sua estrutura composta por uma introdução, desenvolvimento e conclusão. No entanto, estas secções não estão, de todo, explícitas neste artigo, pois não existe qualquer numeração e os referidos termos não são utilizados, com exceção para o capítulo “Conclusion”. É possível, e até provável que esta opção seja intencional por parte do autor, pois trata-se no fundo de uma tentativa de eliminação do preconceito de artigo científico pré-formatado e normalizado com vista a fomentar a argumentação do autor. A grande divisão das três secções do texto é feita apenas com títulos graficamente destacados (tipo de letra único no texto, a itálico, com o dobro do espaçamento e tamanho de letra superior), a designar: “Arguments for the Cis-Regulatory Hypothesis”; “The Data:

Evidence for a Predictive Theory of Genetic Evolution”; “Conclusions”. Estas secções são subdivididas em tópicos, sendo que cada tópico contém um título também ele graficamente distinto, devido à utilização de um tipo de letra mais destacado.

Começando pelo título e nomes dos autores, a formatação é adaptada à restante mancha gráfica e norma de edição da revista: título bastante destacado, seguido dos nomes dos autores, respetivas moradas profissionais e contactos. Excluindo o título, a primeira informação relevante do artigo surge ainda antes do resumo: data de receção e aceitação. Significa isto que o artigo foi sujeito a uma revisão por pares. Neste processo de revisão é comum a identidade dos revisores não ser revelada, no entanto, este artigo contém uma menção de agradecimento a ambos os revisores, que concordaram em revelar a sua identidade (ver *Agradecimentos*, pág.127). Segue-se a referência às palavras-chave do artigo que revelam, no fundo, os temas, ou conceitos, mais relevantes do estudo.

Como é norma do artigo científico, o primeiro texto é um resumo de artigos científicos sobre o assunto abordado, referenciando os conceitos estudados. O objetivo deste resumo é informar o leitor sobre a natureza e o objeto do estudo, as suas teorias e qual o seu objetivo. O primeiro impacto é sem dúvida a falta de título; percebe-se de imediato que este não possui o habitual título “Abstract”, um elemento frequente neste género de literatura científica. Após análise de outros artigos desta mesma publicação (Hoekstra, 2007), é possível constatar que tal se deve a uma opção editorial e não dos autores. Esta secção é, tal como o texto que legenda as figuras e tabelas, formatada de modo diferente do resto do artigo, visto não estar dividida em colunas e ter um tipo de letra destacado.

Este artigo tem também uma característica menos frequente: a editora que aloja gratuitamente a versão digital do projeto está referenciada, no entanto sob a forma de ligação para um website (ver Figura 1) no qual poderá ser descarregado o artigo sem qualquer custo, ou seja, este patrocínio torna o artigo gratuito.

OnlineOpen: This article is available free online at www.blackwell-synergy.com

Figura 1 – Objeto incluído no artigo, com ligação para site da editora da revista *Evolution – International Journal of Organic Evolution*, na qual o artigo foi publicado; a editora disponibiliza gratuitamente o artigo em formato digital.

Os autores optaram por inserir, imediatamente após esta informação, duas citações, com indicação de autor, obra e data de publicação, que no fundo explicam a motivação do autor.

O texto principal do artigo inicia-se com uma extensa introdução que contém a crítica e discussão das hipóteses em estudo. A introdução não é explícita no texto, devido à falta de título. Esta particularidade poderá inicialmente induzir o leitor em erro, caso este considere que a introdução é constituída apenas pelo texto sem título iniciado na pág. 2155 que termina na pág. 2157. Na verdade,

a introdução termina apenas no final da página 2163 e a sua constituição inclui duas tabelas, uma figura e oito, geralmente extensos, tópicos. Esta secção termina no início do desenvolvimento do artigo.

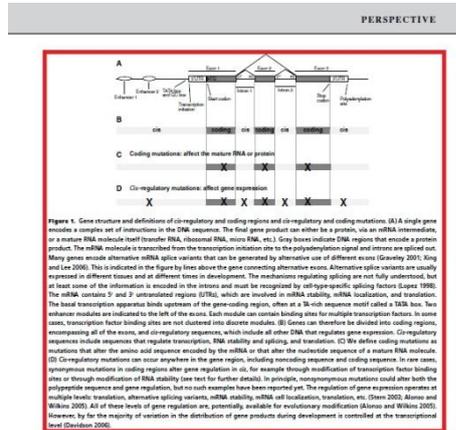
“The Data: Evidence for a Predictive Theory of Genetic Evolution” é o título da secção do texto que compõe o seu desenvolvimento. Nesta secção, os autores fornecem os dados e três distintas previsões, com análise aprofundada. Cada previsão é constituída por um só texto, sem tópicos adicionais, ao contrário de grande parte do texto, sendo que cada texto é acompanhado de figuras ou tabelas. Estas três previsões são, no fundo, o objeto fundamental de estudo deste artigo. Esta secção contém, portanto, quatro tópicos analisados, mas mais importante, quatro figuras e quatro tabelas, dados reveladores do carácter argumentativo que constitui o desenvolvimento do artigo. Todas as figuras e tabelas contêm uma legenda esclarecedora do conteúdo. O exemplo da terceira tabela demonstra como os autores fizeram questão de esclarecer a informação no texto, apesar de ser meramente textual e da legenda fornecer informações acerca do significado dos conceitos utilizados na tabela, algo que, à partida, seria do conhecimento dos especialistas

O texto termina com a conclusão, designada por esse mesmo nome, com apenas um curto texto constituído por quatro parágrafos, seguindo a brevidade típica da conclusão de um artigo científico, sendo bastante objetiva e, por esse motivo, sucinta. A esta secção segue-se uma breve nota de agradecimentos. Justifica-se a colocação de agradecimentos no final do texto, pois a norma do artigo científico indica que esta secção é apenas feita no final e deve apenas referir os contributos significativos para a concretização do estudo.

A parte pós-textual é composta pela Bibliografia – três páginas de bibliografia – e Apêndices – dois Apêndices que complementam o estudo principal do artigo. A bibliografia segue o formato adotado pela revista, que por sua vez aparenta ser uma derivação do formato de bibliografia APA (American Psychological Association), popularizado nos Estados Unidos da América, pois a diferença é mínima: perda de itálico no título da obra citada e de parênteses na data de publicação. O artigo contém ainda dois apêndices. Estes apêndices estão numerados e mantêm a mesma formatação de título e texto utilizados no corpo do artigo, sendo que o primeiro é relativamente curto, ao contrário do segundo. Este é composto por três tópicos, iniciando-se na página 2174 e terminando na página final do documento, 2177.

Importa ainda destacar outros elementos do artigo, nomeadamente imagens, tabelas e legendas. As legendas são utilizadas para melhor compreensão de conceitos ou apresentação de resultados. Certas legendas de imagens ou tabelas são bastante extensas, pois interpretam os relevantes dados apresentados (Figura 2). No exemplo da Figura 2 podemos constatar que a sua legenda explica, passo

a passo, toda a informação exposta no gráfico. Este gráfico contém pouca informação textual, no entanto é o suporte mais adequado para demonstrar o conceito de estrutura de gene e definição das regiões de regulação cis e codificação, assim como as suas mutações. Toda a informação presente no gráfico encontra-se detalhadamente descrita na legenda.



at the genetic level. Therefore, we focus primarily on the two hypotheses that make predictions about the genetic basis of phenotypic evolution, the narrow and broad cis-regulatory hypotheses, together referred to as the cis-regulatory hypothesis.

Arguments for the Cis-Regulatory Hypothesis

Over the past 50 years, many different arguments have been advanced to support the predominant role of cis-regulatory changes in phenotypic or morphological evolution. We believe that these can be parsed into seven discrete arguments. We discuss and critique each one below.

IMPORTANCE OF GENE REGULATION IN LIFE

The origins of the cis-regulatory hypothesis can be traced back, ultimately, to classic experiments on gene regulation in the bacterium *Escherichia coli* (Jacob and Monod 1961). These experiments revealed that levels of enzyme activity are determined primarily by transcriptional regulation. Certain gene products, what we now call transcription factors, bind to specific nucleotides adjacent to the target coding sequence and either recruit or block recruitment of the basal transcription apparatus to the promoter. This is, in principle, a genetic mechanism to control gene expression in response to external signals.

Over the past 50 years, research in developmental biology has shown that this basic mechanism in bacteria also applies to

Figura 2 - Exemplo de figura e respetiva legenda presente no objeto de tradução (destaque a vermelho). A complexidade do esquema representado justifica a extensa legenda. Verifica-se também que este texto não segue a formatação em duas colunas, presente nas restantes secções.

6.2 Micro-estrutura

Relativamente à semântica do texto, este apresenta as habituais características inerentes a este género textual, sendo um texto de teor informativo, de carácter objetivo e conciso, sem espaço para subjetividade por parte do autor. As já mencionadas frases longas são de facto uma característica deste texto, mas há que referir que este contém também frases curtas e objetivas:

This argument has two core assumptions.

Este argumento tem duas suposições base.

Do ponto de vista formal, trata-se de um texto de difícil compreensão para leitores sem especialização nesta área. O vocabulário é específico, e a terminologia específica é abordada pelo autor com total naturalidade, pois este parte do princípio que o público-alvo do texto tem um domínio natural desta terminologia.

Na análise à micro-estrutura do texto, serão mencionados exemplos de várias características, e, caso se justifique comparação, esta terá como base os textos presentes na bibliografia.

6.2.1 Sintaxe

O estudo da sintaxe no âmbito da tradução do par linguístico inglês-português é um dos grandes desafios deste exercício. A organização das palavras numa frase em inglês necessita de análise cuidada no momento de tradução para português, particularmente no caso específico do texto de partida aqui utilizado, devido à utilização frequente da voz passiva e conectores:

For example, the increase in scute expression associated with the production of extra bristles in a Moroccan population of *D. melanogaster* **has been shown to** result from a coding change in a transcription factor gene regulating scute expression (Gibert et al. 2005).

Por exemplo, **demonstrou-se que** o aumento da expressão no gene *scute*, associada à produção de mais pelos numa população Marroquina de *D. melanogaster*, resultava de uma alteração na codificação do fator de transcrição que regula a expressão do gene *scute* (Gibert et al. 2005).

However, in rare cases, *cis*-regulatory mutations may arise in coding regions.

Existem, **no entanto**, casos raros nos quais as mutações na regulação *cis* poderão surgir nas regiões de codificação.

O frequente recurso a expressões numéricas constitui também um desafio neste processo de tradução. A enumeração de argumentos é frequente no texto de partida, no entanto a sua tradução direta não constitui uma boa opção para a língua portuguesa. Assim, optei por traduzir estas expressões de acordo com o contexto:

First, Hoekstra and Coyne (2007) excluded a few traits from consideration not because they have been shown to be neutral or deleterious, but because they have not been proven to be adaptations.

Para começar, Hoekstra e Coyne (2007) deixaram de parte algumas características, não por serem neutras ou deletérias, mas porque não se provou serem adaptações.

Second, they argued that gene duplication allows one gene copy to retain an ancestral function and the other to evolve a new function.

O seu **segundo** argumento indica que a duplicação de genes permite que um gene copiado mantenha uma função antiga e o outro evolua até obter uma nova função.

As mencionadas enumerações são bastante representativas do carácter argumentativo do artigo. No entanto, também a apresentação de dados ou factos recorre muito frequentemente a esta figura de estilo, apenas sob a forma de enumeração simples (não existem enumerações recolectivas ou caóticas ao longo do texto):

The regulation of gene expression operates at multiple levels: translation, alternative splicing variants, mRNA stability, mRNA cell localization, translation, etc.

A regulação da expressão do gene funciona a vários níveis: tradução, variantes alternativas de união, estabilidade do ARNm, localização do ARNm nas células, tradução, etc.

6.2.2 Voz passiva

A voz passiva nem sempre apresenta problemas na tradução de inglês para português. São as frases com voz passiva que não têm tradução direta em português que habitualmente geram dúvidas ao tradutor. A voz passiva enfatiza o processo e não o agente, e no âmbito da tradução para português, a solução passa por alterar a voz passiva e manter a ideia original, ou seja, o processo em si é mais importante que o agente. No caso seguinte, é possível observar como a tradução manteve a ordem dos processos, apesar da eliminação da voz passiva:

To our knowledge, **only two evolutionary changes in phenotype have been shown to derive from synonymous mutations** (Stam and Laurie 1996; Nackley et al. 2006), whereas hundreds of **evolutionary changes in phenotype have been shown to involve nonsynonymous mutations** (see below).

Tanto quanto sabemos, **apenas foi possível demonstrar a existência de duas alterações evolucionárias no fenótipo com origem em mutações sinónimas** (Stam e Laurie 1996; Nackley et al. 1996), enquanto já é **possível demonstrar que centenas de alterações evolucionárias no fenótipo eram afetas às mutações não-sinónimas** (ver abaixo).

6.2.3 Grupos nominais

A tradução de grupos nominais no contexto de linguagem científica complexa, é por vezes um verdadeiro desafio ao tradutor. A dificuldade em detetar estas unidades lexicais extensas que têm, por norma, um nome acompanhado por uma ou mais palavras que o determinam, complementam ou modificam, é por vezes ultrapassada apenas com o auxílio do conhecimento terminológico da área por parte do tradutor. No caso deste par linguístico, a dificuldade é acrescida, porque a língua inglesa produz inúmeros grupos nominais sem determinação do género masculino ou feminino, o que cria ambiguidade na compreensão. No caso do artigo analisado é, de facto, necessária uma compreensão dos conceitos mais aprofundada de modo a obter uma tradução correta de alguns grupos nominais. Nos exemplos seguintes apresenta-se o caso de um grupo nominal (1) composto pelo nome e uma palavra composta por justaposição, cuja tradução original (1a) poderia ser considerada correta, no entanto o Tradutor Especializado, assim como um especialista da área, poderá detetar que o significado desta tradução não estaria correto: o termo “proteína codificada do ADN” será compreendido como uma proteína, existente no ADN, que foi codificada, no entanto o ADN não contém proteínas; o ADN sintetiza proteínas nas células de acordo com instruções contidas no código genético, sendo que esta sintetização compreende o processo de transcrição e tradução. A codificação é a tradução de informação do ADN em, por exemplo, proteínas (Owen et al. 2004). Ou seja, o ADN está a codificar proteínas (1b).

(1) Due to the genetic code, 24% of nucleotide substitutions in **protein-coding DNA** are expected to cause synonymous substitutions if mutations occur randomly (Wilke 2004).

(1a) Devido ao código genético, prevê-se que 24% das substituições nucleotídicas em **proteínas codificada do ADN** causem substituições sinónimas, caso as mutações ocorram aleatoriamente (Wilke 2004).

(1b) Devido ao código genético, prevê-se que 24% das substituições nucleotídicas em **ADN que codifica proteínas** causem substituições sinónimas, caso as mutações ocorram aleatoriamente (Wilke 2004).

Este caso apresenta similares dificuldades encontradas em grupos nominais extensos com palavras que podem ser compreendidas como verbo ou substantivo. Sem contexto e, principalmente, sem conhecimento da área, o tradutor poder-se-ia inclinar para uma tradução incorreta. Os dois casos ([2], [3]) apresentados, contêm palavras que poderiam ser interpretadas como verbos (ver termos destacados), no entanto a tradução deste grupo nominal obriga à compreensão não só do termo mas também do conceito. Senão vejamos: um regulador contém sequências de ADN (2a), não sequencia ADN (Zielenski et al. 1991, Azakami et al. 1992); o fator de transcrição é uma proteína que se liga a sequências de ADN específicas através dos chamados locais de ligação ou, mais especificamente, locais de ligação do fator de transcrição (3a), logo trata-se de um só grupo nominal e não uma oração simples (Cunha e Cintra 1998).

(2) These features cause **cis-regulatory DNA sequences** to evolve faster than coding DNA sequences.

(2a) Estas características fazem com que as **sequências de ADN de regulação cis** evoluam mais rapidamente do que as sequências de ADN da codificação.

(3) In rare cases, synonymous mutations in coding regions alter gene regulation in *cis*, for example through modification **of transcription factor binding sites** or through modification of RNA stability (see text for further details).

(3a) Em casos raros, as mutações sinónimas em regiões de codificação alteram a regulação do gene em *cis* através, por exemplo, da modificação dos **locais de ligação do fator de transcrição** ou através da modificação da estabilidade do ARN (ver texto para mais detalhes).

6.2.4 Siglas

A tradução de siglas foi alvo de pesquisa não apenas de tradução mas também de contexto, ou seja, verificação de frequência de utilização de tradução de determinada sigla no âmbito da área do texto de partida. Um exemplo familiar no campo da tradução é a tradução ou *não-tradução* da sigla do **ácido desoxirribonucleico**, ADN. Objetivamente, não se coloca qualquer problema na tradução deste termo ou respetiva sigla; no entanto é tão frequente surgir este termo como o equivalente inglês, na literatura científica portuguesa. A título de exemplo, um de muitos projetos de investigação da área

da biologia, financiado pela Fundação para a Ciência e Tecnologia (FCT), contém a referida sigla em inglês no título do projeto: “Manipulação de DNA em soluções e interfaces” (Maria Miguel et al, 2004). Trata-se apenas de um indicador objetivo da frequência de utilização, e aceitação geral, da sigla original. É possível verificar a frequência similar de utilização de ambas as siglas até mesmo em motores de busca da internet. O motor de busca *Google*, com parâmetros de resultados exclusivamente em PT-pt, revela que a sigla portuguesa tem 944 000 resultados, enquanto a sigla inglesa difere em menos 211 000, com 733 000 resultados em PT-pt.

O caso mais estudado neste projeto foi indubitavelmente a tradução do termo “*cis-regulatory*”, traduzido como “regulação *cis*”. A dúvida inicial surgiu devido à questão do termo “*cis*”, que desperta impacto imediato, pois surge sempre em itálico e a literatura científica mantém sérias dificuldades na adaptação deste termo para a língua portuguesa. Esta questão será discutida em maior detalhe na página 52. De facto, o termo “*cis*” não se trata de qualquer tipo de sigla, trata-se sim de um termo proveniente do latim *cis*, que significa “no mesmo lado que”. Trata-se da única dúvida relativamente a uma sigla que surgiu no processo de pré-tradução. As restantes siglas sofriam da referida questão da sigla ADN/DNA. A tradução das siglas que se seguem é frequentemente descartada, mesmo em artigos científicos. No entanto, é minha convicção que a tradução destes termos não só é possível como aconselhável, bastando como justificação a compreensão do seu significado por leitores de língua portuguesa. Além do mais, a opinião do autor relativamente à questão do purismo da língua, ou crescente influência de anglicismos não necessita sequer de ser pronunciada: trata-se de uma questão de coerência gramatical e não de estilo. O processo mental de reconhecimento da sigla “ARNm” (Ácido Ribonucleico mensageiro) e conseqüente associação ao seu significado é indubitavelmente mais rápido do que a sigla “mRNA” (Messenger Ribonucleic Acid), por exemplo. Não obstante este facto, existe uma certa intraduzibilidade em determinadas siglas neste artigo, embora sejam raras as ocorrências – a sigla UTR (Untranslated region) não foi traduzida porque seria necessário proceder à criação de uma sigla (a sigla UTR é reconhecida pela comunidade científica portuguesa – a FCT aprovou projetos cuja designação em português mantém esta sigla; como exemplo, destaca-se o projeto de biologia celular e molecular intitulado “Função do 3' UTR do polo de *Drosophila melanogaster* no processamento alternativo de mRNA e na expressão genética”, que curiosamente opta por manter a sigla “mRNA” anteriormente mencionada), processo que deverá ser executado pelos especialistas da área, não da linguística. Para melhor compreensão, seguem-se exemplos de siglas presentes no texto de partida e conseqüente tradução.

BMP4 Bone morphogenetic protein 4	PMO4 Proteína morfogenética do osso 4
RNA Ribonucleic Acid	ARN Ácido Ribonucleico
mRNA messenger Ribonucleic Acid	ARNm Ácido Ribonucleico mensageiro

É também importante mencionar o caso da sigla *DGB*. Neste caso, a sigla manteve-se devido ao mesmo motivo da referida sigla *UTR*, no entanto, o texto continha a composição da sigla, o que permitiu fornecer a tradução do conteúdo da sigla:

Gene is a known or presumptive member of a **differentiation gene battery (DGB)**.

O gene é um membro conhecido ou presumível membro de um **conjunto de genes de diferenciação (DGB)**.

6.2.5 Frases interrogativas

No contexto da análise à sintaxe do texto de partida, é de salientar a questão das frases interrogativas. No contexto do artigo científico aqui utilizado, importa de facto mencionar a ocorrência de frases interrogativas, visto tratar-se de um texto argumentativo. O texto argumentativo é composto essencialmente por frases declarativas, no entanto existem claras vantagens na utilização de frases interrogativas. Os autores deste artigo compreenderam claramente que quando se está a defender uma hipótese, não basta mostrar que irá resolver um determinado problema, há que considerar alternativas (Weston, 1996), e esta técnica apura o carácter argumentativo do texto. Isto refletiu-se na adição de frases interrogativas que visam despertar o interesse do leitor, fomentando deste modo o debate. Além de fomentar o debate, estas frases surgem pontualmente para criar uma relação dialógica com o leitor. Segue-se um exemplo desta tentativa de aproximação com o leitor, através de uma interrogação colocada no final de um tópico, com referência ao título do tópico seguinte:

How common is pleiotropy? Is it universal?

Gene pleiotropy (...)

Quão comum é a pliotropia? É universal?

Pliotropia do gene(...)

As frases interrogativas exprimem muito bem a problemática da organização das palavras, no âmbito da tradução entre a língua inglesa e portuguesa. Estas frases sofrem geralmente muitas alterações na sua tradução de inglês para português, nomeadamente a alteração do sujeito e de modos verbais. A solução passa, invariavelmente, pela alteração da sintaxe da própria frase, caso contrário a frase traduzida não manteria a coerência do texto de partida. Segue-se o caso de uma frase interrogativa com interessante resolução da problemática de tradução.

Is genetic evolution predictable?

- (1) É possível prever a evolução genética?
- (2) Será possível prever a evolução genética?
- (3) Poder-se-á prever a evolução genética?

A escolha final recaiu sobre a segunda hipótese. Qualquer uma das hipóteses poderia ser considerada uma frase declarativa, no entanto a primeira hipótese foi descartada por se encontrar no presente, o que a aproxima mais ainda de uma declarativa. As duas hipóteses seguintes poderão também ser declarativas, no entanto o facto de estarem no futuro já suscita alguma dúvida imediata no momento de leitura, porque de resto o único elemento interrogativo característico é o ponto de interrogação. Entre estas duas hipóteses, a escolha final recaiu sobre a segunda por uma questão de estilo textual.

Existem também casos de frases interrogativas nos quais foi necessária uma reorganização da sintaxe, o que suscitou também dúvida entre a escolha das várias alternativas:

Do coding or *cis*-regulatory mutations cause more phenotypic evolution?

(4) O que provoca maior evolução fenotípica são as mutações na codificação ou na região da regulação *cis*?

(5) O que provoca maior evolução fenotípica: mutações na codificação ou na região da regulação *cis*?

A escolha recaiu na primeira hipótese (4) porque, apesar de apresentar uma estrutura relativamente ambígua, no sentido em que tanto poderá ser uma frase declarativa como interrogativa com a simples alteração da pontuação final, não está dividida como a segunda hipótese (5). Esta divisão entre as duas orações principais acaba por alterar em demasia a frase do texto de partida.

O seguinte caso exemplifica como as frases interrogativas com os mesmos advérbios de interrogação são traduzidas de modo diferente. Neste caso, a tradução da primeira frase efetuou a substituição do advérbio por um pronome, mas na segunda frase manteve-se o advérbio. A contextualização contribuiu também para a escolha desta opção tradutológica.

For example, **how do we expect** particular kinds of mutations to generate particular kinds of phenotypic variation?

Por exemplo, **o que se pode esperar** de certos tipos de mutações que geram tipos específicos de variação fenotípica?

How do we expect population genetic parameters to influence the spread and fixation of different kinds of mutations?

Como se pode esperar que a genética populacional influencie a propagação e fixação de diferentes tipos de mutações?

O próprio título do artigo é uma referência clara à intenção de interação e debate dos autores com o público, neste caso com os autores do artigo que suscitou grande parte da polémica debatida no presente artigo, “The Locus of Evolution: Evo Devo and the Genetics of Adaptation”, Hopi E. Hoekstra e Jerry A. Coyne:

The Loci of Evolution: How Predictable is Genetic Evolution?

Os Loci da Evolução: Quão Previsível é a Evolução Genética?

6.2.6 Argumentação

A argumentação do texto está patente na totalidade da estrutura do texto: introdução, desenvolvimento e conclusão. As frases declarativas, as mais comuns no texto científico, são o exemplo claro desta característica do texto, mas mesmo as frases interrogativas têm como objetivo despertar o diálogo com o leitor, além de darem mais vivacidade ao texto.

A secção inicial do texto é um exemplo bastante demonstrativo da estrutura argumentativa do artigo. A secção analisada inicia-se na primeira página, após o resumo, e precede o capítulo cuja própria designação revela a óbvia componente argumentativa do artigo (“Arguments for the *Cis*-Regulatory Hypothesis”). Numa análise mais detalhada, detetam-se vários elementos que revelam o carácter argumentativo, como o seguinte exemplo, no qual se detetam advérbios numerais ordinais, que implicam a exposição de dados argumentativos:

Natural selection causes predictable changes in phenotypic variation. This predictability exists at two levels. **First**, quantitative genetics provides predictions for the short-term response to selection, given estimates of heritability and the selection differential (Falconer and Mackay 1996). **Second**, selection theory often provides reasonable predictions of how populations will adapt over the long term following a change in the selective regime.

A seleção natural provoca alterações previsíveis na variação fenotípica. Esta previsibilidade existe a dois níveis. **Em primeiro lugar**, a genética quantitativa fornece previsões da resposta a curto-prazo à seleção, tendo em conta as estimativas de hereditariedade e do diferencial de seleção (Falconer e Mackay, 1996). **Em segundo lugar**, a teoria da seleção fornece frequentemente previsões

aceitáveis sobre como as populações se irão adaptar a longo-prazo após uma alteração no processo de seleção.

O parágrafo seguinte continua a linha de raciocínio argumentativa, com uma locução adverbial de modo, seguindo-se de uma frase iniciada com uma locução adverbial invariável. Aqui podemos constatar que a argumentação recorre também à exemplificação, bastante frequente neste artigo:

In contrast, the genetic changes underlying these phenotypic changes have historically not been expected to show predictable patterns. **For example**, it has long been recognized that different genetic causes can generate similar patterns of phenotypic variation (Robertson 1959; Wilkens 1971).

As alterações genéticas subjacentes a estas alterações fenotípicas, **pelo contrário**, não têm demonstrado, historicamente, padrões previsíveis. **Por exemplo**, há muito se reconhece que diferentes causas genéticas podem gerar padrões similares da variação fenotípica (Robertson 1959; Wilkens 1971).

No parágrafo seguinte, os autores são ainda mais explícitos e diretos, na argumentação. Uma análise do segmento seguinte revela argumentação explicitada pelos autores. É revelador o facto de este segmento ser apenas o terceiro parágrafo da introdução. Além do destaque inicial, encontra-se também um articulador argumentativo a meio da frase:

To convince the reader that genetic evolution is predictable in at least some general sense, **we point out that** there is already an uncontroversial general theory of genetic evolution.

Para convencer o leitor de que a evolução genética é previsível, pelo menos num sentido geral, **salientamos** a existência de uma incontestável teoria geral da evolução genética.

O modo argumentativo mantém-se ao longo do texto, como se pode verificar no tópico “Correlation between phenotypic change and change in gene expression”. Este inicia-se com dois exemplos e subsequente explicação:

For example, Abzhanov et al. (2004) have discovered that higher levels of Bone morphogenetic protein 4 (BMP4) expression are correlated with deeper beak shapes among Darwin’s finches. (...) **Therefore**, differential expression of BMP4 provides a reasonable explanation for changes in finch beak shape.

Exemplificando, Abzhanov et al. (2004) descobriram que os níveis elevados de expressão da proteína morfogenética do osso 4 (PMO4) estão correlacionados com os diferentes tamanhos de bicos dos tentilhões de Darwin. (...) **Por conseguinte**, o diferencial de expressão da PMO4 dá uma explicação aceitável para as alterações na forma do bico do tentilhão.

Importa também referir que, apesar de grande parte da argumentação consistir em exemplos de outros casos com respetiva explicação, existem também citações presentes no texto (ver segundo parágrafo do tópico “Conservation of coding sequences across taxa”), assim como dados estatísticos:

In *D. melanogaster*, **about 80%** of the genome is noncoding and in humans and mice **more than 98%** of the genome is noncoding. However, in humans **more than 90%** of this noncoding DNA is occasionally transcribed into RNA although the functional roles, if any, of most of this RNA are currently poorly understood (Mattick 2003; Birney et al. 2007).

Na *D. melanogaster*, **cerca de 80%** do genoma é não-codificante, e em humanos e ratos **mais de 98%** do genoma é não-codificante. No entanto, nos humanos, **mais de 90%** deste ADN não-codificante é por vezes transcrito para o ARN apesar das funções funcionais, se aplicáveis, de maior parte deste ARN terem, atualmente, uma compreensão muito limitada (Mattick 2003; Birney et al. 2007).

O modo argumentativo está, portanto, bastante explícito neste texto. No entanto, importa também mencionar outros elementos – conectores – que surgem ao longo do texto, não tão explícitos como os exemplos anteriormente mencionados, mas que revelam o teor argumentativo presente ao longo do texto. Existem vários conectores adversativos no texto de partida, entre outros que serão de seguida referidos, reveladores do carácter argumentativo:

However, in rare cases, *cis*-regulatory mutations may arise in coding regions.

Existem, **no entanto**, casos raros nos quais as mutações da regulação *cis* poderão surgir nas regiões de codificação.

Apesar da sequência argumentativa não ter conectores privados, existem conectores que estabelecem relação de causa (1), relação de semelhança (2), exemplificação (3), consequência (4), expansão do tópico (5) e de conclusão (6).

(1) Three genes could not be assigned to the DGB or non-DGB category **because** their function is unknown.

(1a) Três genes não podem ser atribuídos à categoria DGB ou não-DGB **porque** a sua função é desconhecida.

(2) **Similarly**, for non-DGB genes, phenotypic differences between species involve significantly more *cis*-regulatory mutations than coding mutations (Fig. 3C; Table 5).

(2a) Do mesmo modo, as diferenças fenotípicas para genes não-DGB envolvem muitas mais mutações na regulação *cis* do que na codificação (Fig. 3C; Tabela 5).

(3) **For example**, evolution of the bicoid gene through duplication in higher dipterans may have promoted rapid early embryonic patterning (...)

(3a) **Por exemplo**, a evolução do gene bicoid através da duplicação em dípteros maiores, pode ter promovido uma rápida padronização embrionária prematura (...)

(4) Genes can **therefore** be divided into coding regions (...)

(4a) Os genes podem, **portanto**, ser divididos em regiões (...)

(5) That is, the conception of pleiotropy that has historically gripped evolutionary biologists and served as the central assumption for recent theoretical treatments is **effectively** based on the effects of null mutations (Stern 2000).

(5a) Isto é, a conceção de pliotropia que tem sido historicamente utilizada por biólogos evolutivos e que serviram de pressuposto central para recentes tratamentos teóricos, é **efetivamente** baseada nos efeitos de mutações nulas (Stern 2000).

(6) **Finally**, the abundance of studies of trait loss observed by Hoekstra and Coyne (2007) may reflect the more mundane fact that when comparing closely related species, where analysis is more straightforward, phenotypic loss is more common than gain (Adamowicz and Purvis 2006).

(6a) **Por fim**, a abundância de estudos da perda de características observada por Hoekstra e Coyne (2007) pode demonstrar factos mais banais do que a comparação de espécies de relação próxima, na qual a análise é mais direta e a perda fenotípica é mais comum do que o ganho (Adamowicz e Purvis 2006).

7. Metodologia da tradução

Translation quality assessment proceeds according to the lordly, but completely unexplained, whimsy of 'It doesn't sound right.'

Peter D. Fawcett (citado em Baker, 1992)

Our profession is based on knowledge and experience. It has the longest apprenticeship of any profession. Not until thirty do you start to be useful as a translator, not until fifty do you start to be in your prime.

Llana Castellano (citada em Baker, 1992)

Unlike medicine and engineering, translation is a very young discipline in academic terms. It is only just starting to feature as a subject of study in its own right, not yet in all but in increasing number of universities and colleges around the world. Like any young discipline, it needs to draw on the findings and theories of other related disciplines in order to develop and formalize its own methods; but which disciplines it can naturally and fruitfully be related to is still a matter of some controversy.

Mona Baker (1992)

A tradução é geralmente considerada uma disciplina mais prática que teórica, e o estudo académico desta área, embora em considerável expansão, necessita de maturação. Não há uma fórmula exata para avaliar uma tradução, e a qualidade de uma determinada tradução não irá necessariamente aumentar com o estudo de teorias de tradução por parte do tradutor. As teorias da tradução e o seu debate académico são claramente importantes, mas mais importante para o tradutor é a sua capacidade em seguir uma metodologia eficaz, pois é a eficácia da metodologia que irá determinar grande parte da qualidade da tradução. A norma europeia de qualidade dos LSP (Language Service

Providers) EN 15038 é no fundo o delinear de uma estratégia prática e eficaz do processo de tradução com vista a assegurar uma tradução com qualidade (algo difícil de quantificar na área da tradução), e por este motivo a metodologia adotada baseou-se em três fases elementares da referida norma, no Capítulo 5, referente à normas dos serviços de tradução (Procedures in translation Services): a fase de pré-tradução (5.3, ao qual se segue a análise dos aspetos linguísticos, análise do texto de partida e terminologia), a fase de tradução (Capítulo 5.4, o qual menciona a importância da terminologia, gramática, léxico, estilo, *locale*, formatação, grupo alvo e propósito da tradução), sendo que este capítulo integra a verificação (5.4.2 *Checking*), e por fim a fase da revisão, também integrada no capítulo do processo de tradução (5.4.3 *Revision*; 5.4.4 *Review*; 5.4.5 *Proofreading*). A metodologia adaptada para este trabalho sumariza, portanto, o Capítulo 5 da norma. Não foram considerados os restantes Capítulos da Norma, visto não terem uma relação direta com o propósito académico deste projeto (como, por exemplo, 4 *Client-TSP relationship*; 6 *Added Value Services*).

7.1 Ferramentas

Actualmente, os tradutores utilizam várias aplicações informáticas, ou ferramentas, feitas especificamente para acelerar ou melhorar o processo de tradução, tais como motores de pesquisa, programas de gestão de terminologia, programas de memórias de tradução, entre outros (Gouadec, D. 2007). O processo de tradução especializada poderá ser bastante beneficiado com o apoio destas ferramentas, nomeadamente na questão da gestão terminológica, pois a Tradução Especializada é uma área cujos projetos de tradução têm, geralmente, uma forte (e complexa) componente terminológica. Neste sentido, foram utilizadas no presente projeto ferramentas de apoio à tradução que contribuíram para a gestão da terminologia mas também com recursos que de outro modo não seriam utilizados de forma dinâmica, como é o caso do corpus, composto por textos paralelos. As ferramentas foram criteriosamente escolhidas e serão de seguida descritas.

7.1.1 Critério de seleção

A eficácia de ferramentas CAT com recurso a memórias de tradução é, segundo T. Altanero, uma mais-valia na tradução especializada:

“...there are tools specifically for translators that may be worth considering if you have a large volume of translations that tend to be repetitive. (...) Translation memory tools such as Trados, Star's Transit, SDLX, and Deja Vu, among others, assist with terminology management, glossary creation, and translation memory database development. They can greatly increase translation speed by matching similar strings from legacy work for the translator to edit. For those with an engineering background, localization development environment tools such as ForeignDesk, among others, are popular because they provide a WYSIWYG (what-you-see-is-whatyou-get) environment.”(Altanero, 2000)

Atualmente, o número de *environment tools* existentes é mais elevado. A seleção de uma ferramenta CAT foi determinada com base na experiência nas várias ferramentas, o custo financeiro da aplicação a utilizar, a qualidade e estabilidade da aplicação e integração das várias ferramentas necessárias para realizar um projeto de tradução especializada seguindo os parâmetros determinados.

O MemoQ, da empresa húngara Kilgray foi de facto a primeira escolha. Além de ser um exemplo singular no campo das *environment tools*, a sua característica WYSIWYG facilita o primeiro contacto com um utilizador regular de outras ferramentas. O custo desta aplicação para estudantes é relativamente baixo, considerando o custo normal da aplicação, ou até de soluções alternativas. A Kilgray disponibiliza a licença de utilização *translator pro* para estudantes, que pode ser instalada em dois terminais, a um preço inferior a 100€. Esta aplicação não tem qualquer limite de utilização, e possui todas as ferramentas necessárias para a execução deste projeto. A comparação direta com a ferramenta líder do mercado, o SDL Trados Studio (SDLTS), é inevitável. No entanto, o MemoQ não só possui grande parte das funcionalidades oferecidas pela SDL, como também tem mais funcionalidades não existentes neste: reconhecimento automático de frases repetidas em tempo-real, número ilimitado de línguas utilizadas, bases terminológicas editáveis (esta funcionalidade necessita da aplicação SDL Multiterm, no caso do SDLTS), pré-visualização em tempo real, e integração total de todas as aplicações na mesma tradução, o que permite resultados imediatos da base terminológica, memória de tradução, corpus e dicionários online enquanto o utilizador traduz o texto. Também é relevante o facto de possuir uma ferramenta de alinhamento, que permite introduzir um texto com a respetiva tradução e aproveitar estes resultados para integrar o corpus, sendo que o resultado é depois segmentado e utilizado tal como uma memória de tradução.

Funcionalidade	MemoQ	SDL Trados Studio
Ambiente de trabalho com funcionalidades integradas	Sim	Não
Gestão de terminologia	Sim	Necessita de MultiTerm
Limite de línguas	Não	5
Pré-visualização em tempo real	Sim	Apenas em documentos Word
Segmentação automática da Memória de Tradução	Sim	Não
Reconhecimento automático de expressões linguísticas	Sim	Processo não automatizado
Edição de Bases Terminológicas	Sim	Necessita de MultiTerm
Pesquisa e substituição de termos no documento integral	Sim	Não
Corpus monolingue e bilingue	Sim	Não

A experiência académica com o MemoQ foi breve, no entanto a opção SDLTS foi utilizada durante a licenciatura, e a opção MemoQ foi utilizada por um período de dois anos, no âmbito académico e profissional. A experiência com ambas as opções foi sem dúvida um importante critério de seleção: a experiência com o SDLTS não foi tão agradável como a alternativa, nomeadamente devido à instabilidade do programa e a sua dificuldade em lidar com documentos de formatação complexa. O facto de a opção MemoQ possuir um ambiente de trabalho no qual todas as funcionalidades podem apoiar a tradução sem que o tradutor tenha que aceder externamente às mesmas foi também preponderante.

7.1.2 Preparação do texto

Após a decisão inicial de utilizar uma ferramenta CAT para traduzir o texto, surgiu de imediato o problema da formatação do mesmo. Como utilizador frequente do MemoQ, percebi de imediato que iria ter que proceder à conversão do texto para um formato o mais simples possível.

O documento original possui vários elementos extratexto: vários tipos de letra, cabeçalhos, rodapés, caixas de texto, legendas, tabelas e diversos tipos de gráficos com texto. O formato mais adequado é, de facto, o formato PDF, porque o documento mantém sempre a forma pretendida, independentemente do tipo de software utilizado para a sua leitura, e mesmo que seja impresso em papel. No entanto, a edição deste formato é bastante complexa, sendo necessário o recurso a software próprio para edição. Infelizmente, não existem ainda ferramentas de CAT suficientemente avançadas para proceder à edição em tempo-real de PDF em simultâneo com a utilização dos vários recursos inerentes a estas aplicações (corpus, memória de tradução, base terminológica, dicionários). Há que ter em conta que a ferramenta CAT processa o texto de uma forma singular, pois divide o texto em vários segmentos. Sendo que cada segmento corresponde geralmente a uma frase (a segmentação é feita automaticamente, com a deteção do ponto final e letra maiúscula), o processamento de texto em formato PDF é infinitamente variável, ou seja, a aplicação não consegue “aprender” a segmentar texto num PDF pelo simples motivo de que este varia de acordo com a preferência do autor. Por estes motivos, o melhor método será transformar o texto num formato simples (formato *txt*, norma UTF-8), e depois proceder à sua edição.

O primeiro passo foi copiar o texto para um documento *txt*. Todos os elementos textuais, com exceção dos cabeçalhos e rodapés, foram incluídos. No caso do texto que acompanha os elementos gráficos, o texto foi colocado de forma ordenada, após ou antes um capítulo ou parágrafo. Deste modo, a tradução de um texto não foi interrompida ou intercalada pela inclusão de outros elementos, como por exemplo legendas de imagens ou texto integrado numa imagem.



Figura 3 Resultado da formatação manual do texto: cada frase corresponde a uma linha.

Verificou-se que a formatação do texto não seria aceite pelo MemoQ, porque este identifica cada linha como uma só frase. Para conseguir que o MemoQ aceitasse uma frase de várias linhas como um segmento e não como várias, foi necessário editar todas as linhas de texto de modo a que uma frase correspondesse a uma só linha (Fig.3). Concluída esta edição, o documento *txt* foi importado para o projeto de tradução no programa MemoQ.

7.1.3 Análise de ferramentas

O MemoQ possui um ambiente de trabalho próprio que permite criar de raiz um projeto de tradução com uma base terminológica, corpus e memória de tradução.

Tecnicamente, trata-se de uma aplicação que necessita de poucos recursos de hardware e utiliza pouca memória virtual. Teoricamente, é necessária apenas 1 GB de memória virtual, 200 mb de memória física e um processador com frequência superior a 2 GHz (requisitos recomendados), no entanto a funcionalidade desta aplicação está limitada a um sistema operativo proprietário

(Microsoft Windows). O emulador *WINE* para Linux – sistema operativo livre e não comercial - e outros sistemas operativos baseados neste *kernel* consegue correr esta ferramenta, assim como o software de virtualização *Parallels Desktop* para hardware Apple Macintosh, desde que possua um processador Intel; a emulação em sistemas operativos *MacOS* não é suficientemente estável para correr esta aplicação, além de representar um custo acrescido na aquisição do referido software de virtualização. Apesar de existirem soluções para outros sistemas operativos, a Kilgray não garante a sua funcionalidade ou estabilidade normal.

A criação de um projeto no MemoQ é assistida pela aplicação MemoQ Project Wizard. O primeiro passo foi dar um nome ao projeto. Foi selecionada a língua de partida, língua de chegada, e incluída a informação adicional ao projeto: Criador do projeto; Descrição; Domínio entre outros (Figura 4). Trata-se de informação que ficará anexada ao projeto e ao utilizador. Caso o utilizador crie um novo projeto com o mesmo domínio ou para o mesmo cliente, por exemplo, a aplicação irá sugerir a utilização de recursos que poderão ser utilizados, como é o caso de Bases Terminológicas ou Corpora.

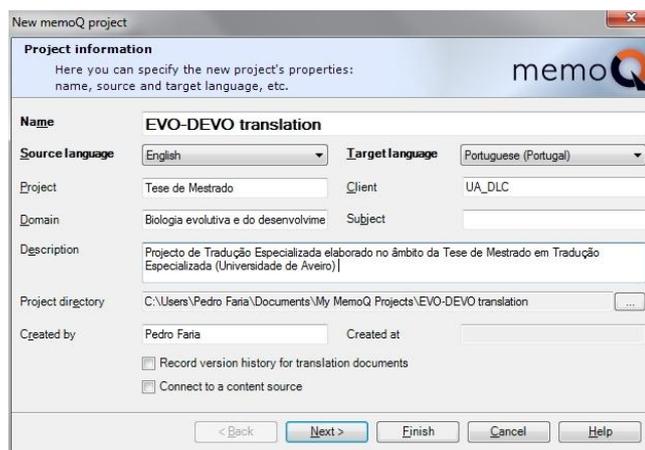


Figura 4 Definições de um novo projeto de tradução

O passo seguinte consiste em adicionar o documento a traduzir. Como referido no Cap. 7.1.2, a aplicação não suporta edição PDF em tempo-real, pelo que o texto foi convertido manualmente para texto simples (formato *txt*). Estando o ficheiro criado, foi importado neste passo para o projeto. A aplicação importou e pós-processou (segmentação) o texto em poucos segundos (Fig.5), e cada segmento correspondia, como previsto, a uma frase.



Figura 5 Importação de documento para projeto

Seguiu-se a criação de uma memória de tradução (MT) (Fig.6). A vantagem desta aplicação no caso das MT é a capacidade de clonar uma MT em segundos, possibilitando assim a utilização da MT principal enquanto o clone é utilizado para observar o conteúdo da MT. Esta característica permite, por exemplo, pesquisar um determinado termo na MT enquanto se efetua a tradução, visto não ser possível utilizar a MT para ambas as funções em simultâneo. A MT poderá ser também utilizada noutros projetos do próprio ou de outro computador, pois pode ser exportada como um único ficheiro, além de possibilitar a sua utilização noutras ferramentas CAT, através da exportação no formato TMX.

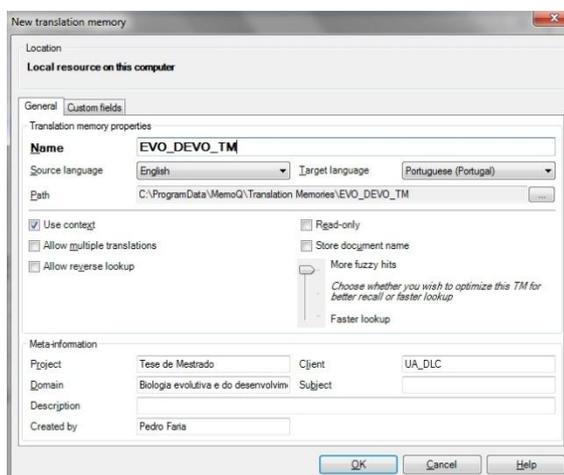


Figura 6 Definições da Base Terminológica

Após a MT, foi criada a Base Terminológica (BT) (Figura 7). Neste caso, existia já um recurso importante a adicionar ao projeto: o resultado da extração terminológica e respetiva tradução. Trata-se no fundo, do glossário produzido no processo de pré-tradução. Este glossário encontrava-se armazenado no formato de folha de cálculo XLS. No entanto, como o MemoQ não importa este

formato como Base Terminológica (embora o faça como documento para tradução), foi necessário proceder à sua conversão para formato CSV, separado por vírgulas (Figura 8). Para efetuar esta conversão, foram eliminadas todas as informações desnecessárias à Base Terminológica, como contextos ou definições. Todo este processo resultou num glossário com termos relativos ao texto e respetiva definição, que foi utilizado *in loco* e em tempo-real durante a tradução.

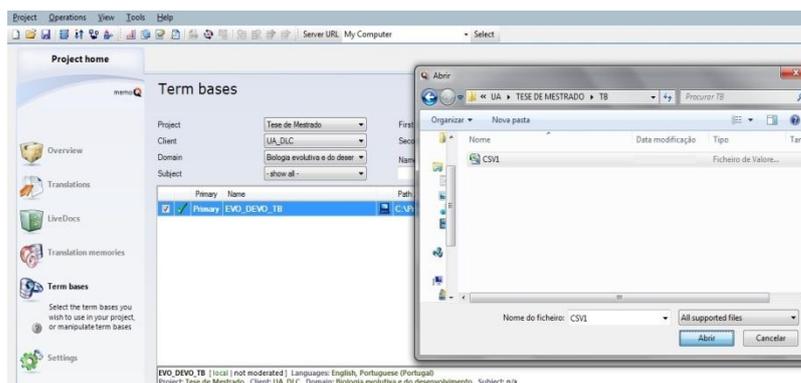


Figura 7 Importação de glossário em folha de cálculo com extração terminológica e tradução.

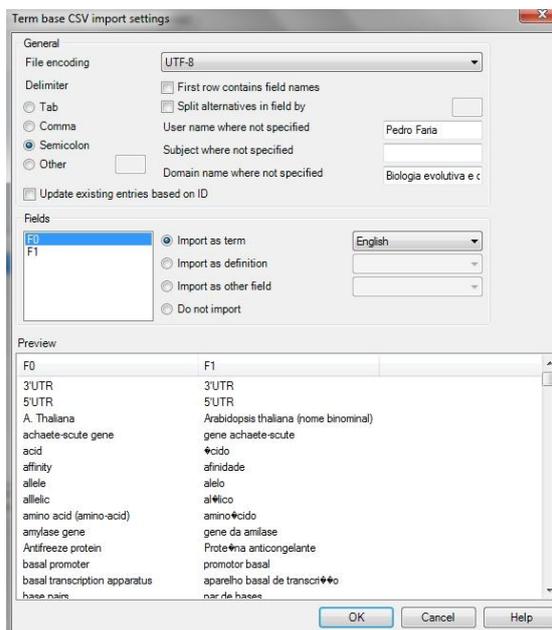


Figura 8 Conversão da folha de cálculo para formato CSV, separado por vírgulas, com codificação UTF-8.

Finda a criação do projeto, Base Terminológica e Memória de Tradução, procedeu-se à criação de um Estando o projeto criado, decidi aproveitar todos os recursos compilados enquanto produzi a BT, corpus linguístico, composto pelos artigos científicos encontrados para obter definições da

terminologia, no momento da criação glossário, mas também outras obras de referência e artigos científicos semelhantes, recolhidos no estudo do tema e na pesquisa terminológica. Todos estes documentos foram introduzidos num corpus, o que na prática funciona como uma pequena biblioteca, que indica em tempo-real quando determinado termo surge nos artigos do corpus; esta funcionalidade foi extremamente útil para compreender o contexto no qual estava inserido determinado termo ou expressão. A título de exemplo, caso surgisse o termo “DNA coding” num segmento de tradução e este se encontrasse destacado numa obra introduzida no Corpus, a aplicação indicaria em que obra e localização exata no texto este surgia, fornecendo assim o contexto. O corpus é composto maioritariamente por textos em inglês, como se pode verificar na Figura 9, mas existem também textos em português e um caso de um documento bilingue inglês-português.



Figura 9 - Relatório de segmentos detetados pela ferramenta de Corpora integrada no MemoQ

7.1.4 Extração terminológica, tradução e definição

O processo de extração terminológica utilizou dois critérios de seleção distintos, porque a finalidade deste era a construção de um glossário e de uma base terminológica. A frequência de ocorrência de termos e a especificidade de termos foram os critérios utilizados.

A primeira necessitou do apoio de uma ferramenta, o Simple Concordance Program (SCP), que fornece dados estatísticos do texto, como por exemplo o número de ocorrências de um determinado termo, ou uma listagem das palavras mais utilizadas. O facto de o texto estar em formato *txt*

também ajudou no processo de extração terminológica e verificação estatística, porque o SCP é compatível com este tipo de ficheiro.

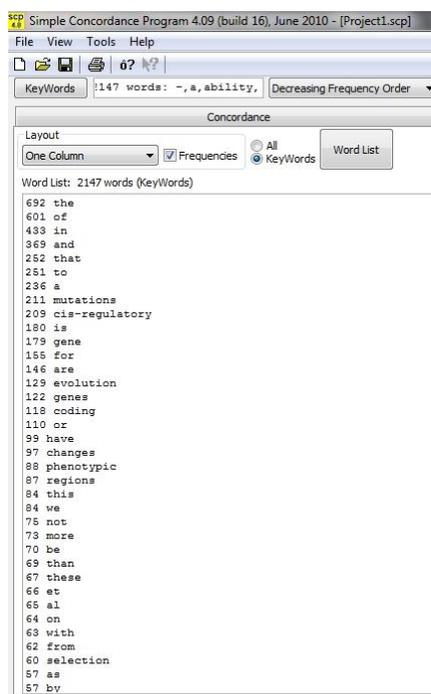


Figura 10 Pormenor do SCP - Word List revela número de ocorrências de palavras

Após a criação de um projeto no SCP, seguiu-se a importação do documento *txt* com o texto do artigo, que depois gerou uma lista de ocorrências (Figura 10). Foram filtradas as palavras mais utilizadas, e as palavras mais relevantes ou específicas, no âmbito da terminologia do texto, foram exportadas para o documento de folha de cálculo XLS. Justifica-se esta opção pela produção de uma base terminológica, integrada na ferramenta CAT, pois esta sugere em tempo-real a tradução de um termo que surge no segmento a traduzir que esteja disponível na Base Terminológica, o que levou a considerar a hipótese de adicionar termos de elevada ocorrência, independentemente da dificuldade prevista para a sua tradução, de modo a acelerar o processo de tradução e manter a consistência terminológica.

Estando concluída a fase de análise do SCP, a terminologia foi guardada e procedeu-se à extração terminológica seguindo o critério de especificidade do termo. Este é o primeiro passo para a criação do glossário.

Este processo requer do tradutor um conhecimento geral da matéria do texto, pois a seleção de termos é feita tendo em conta somente o reconhecimento de termos por parte do tradutor. Tendo em conta que o tradutor é meramente conhecedor do tema (ver Cap.1; 4) e não um especialista neste campo, a seleção de terminologia foi bastante vasta. No entanto, o facto de considerar que a

extração terminológica seria útil para um glossário e uma base terminológica, levou o tradutor a optar por incluir termos não exclusivos deste campo. A terminologia extraída manualmente foi depois integrada no referido documento Excel.

Findo o processo de extração terminológica, procedeu-se à sua tradução.

O processo de tradução da terminologia é geralmente bastante trabalhoso e demorado; no fundo, trata-se da tradução das palavras e expressões cuja tradução é desconhecida e cujo significado pode variar entre diferentes áreas. A metodologia do processo de tradução da terminologia consiste numa primeira pesquisa do termo em dicionários bilingues e bases terminológicas, seguindo-se de um processo de comprovação da utilização adequada com recurso a fontes de elevada fiabilidade e preferencialmente ligadas ao campo da biologia evolutiva e do desenvolvimento. Segue-se uma breve listagem resumida das fontes utilizadas:

Fundação para a Ciência e a Tecnologia – Projetos de I&D

www.fct.pt/apoios/projetos/

Repositório da Universidade de Lisboa – Faculdade de Ciências

repositorio.ul.pt/handle/10451/14

UTL Repository – Teses de Doutoramento

www.repository.utl.pt

Ata Médica Portuguesa

www.actamedicaportuguesa.com

Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA) – Departamento de Genética

www.insa.pt/sites/INSA/Portugues/AreasCientificas/Genetica/

Minerva – Revista de Medicina Interna

www.hsm.min-saude.pt

Ordem dos Médicos – Secção Regional do Norte (SRNOM)

www.nortemedico.pt

Diário da República – Ministério da Saúde

www.dre.pt

Esta metodologia nem sempre trouxe os resultados desejados, e os casos excepcionais merecedores de destaque encontram-se analisados no Capítulo 7.3.

A secção de definição da terminologia específica foi produzida na íntegra pelo tradutor, não existe qualquer definição retirada de uma só fonte, ou seja, cada termo necessita de ser compreendido pelo tradutor, que recorreu a outras definições ou contextualizações do termo. Foram utilizadas fontes em ambas as línguas para obter uma definição. Esta opção facilitou a definição, devido à existência de mais textos e definições na língua inglesa. Por se tratar de um glossário focado na tradução e na definição do conceito, ficaram de fora as definições de classes gramaticais e outras características linguísticas. A definição do conceito é a mais concisa possível, como se pode verificar na Figura 11, e foi elaborada com recurso a literatura compreendida na bibliografia do presente trabalho e aos seguintes sítios da internet:

Biology Online

www.biology-online.org

National Center for Biotechnology Information

www.ncbi.nlm.nih.gov

News Medical

www.news-medical.net

Nature – Scitable

www.nature.com/scitable

	A	B	C	D
10	amino acid (amino-acid)	aminoácido	Os aminoácidos são compostos que apresentam na sua estrutura o grupo carboxílico (-COOH) e o grupo amina (-NH ₂), ligados ou não ao mesmo átomo de carbono da cadeia carbonada.	Nonsynonymous mutations alter the amino-acid sequence and are thus likely to affect protein structure, stability, activity, or binding properties.
11	<i>amylase</i> gene	gene da amilase	enzima catalisador da hidrólise	<i>amylase</i> gene amplification in humans [is] associated with increased starch consumption.
12	Antifreeze protein	Proteína anticongelante	Proteína localizada no fluido extracelular de alguns peixes que vivem em águas muito frias; esta proteína inibe a formação de cristais de gelo	Antifreeze proteins—Fish that live in subzero waters have evolved proteins that capture ice in the bloodstream to prevent ice crystals from puncturing cells.
13	basal promoter	promotor basal	Também conhecido como "core promoter"; trata-se do promotor essencial existente em todos os genes que produzem proteínas	The modules apparently have only weak constraints on position relative to each other and on position relative to the basal promoter.

Figura 11 Tabela com termo original, tradução, definição e contexto. A definição foi elaborada com base nos resultados encontrados na pesquisa da tradução.

7.2 Problemas de tradução

Esta secção destaca uma parte relevante do próprio processo de tradução do texto, ou seja, as dificuldades de tradução da terminologia. Importa desde já referir que, além dos dicionários bilingues, as fontes utilizadas para comprovar a utilização de terminologia são principalmente o site da Fundação para a Ciência e Tecnologia, doravante designada de “FCT”, pois este permite a pesquisa de todos os artigos de investigação científica elaborados em Portugal que concorreram a bolsas de apoio desta fundação; o Repositório Científico de Acesso Aberto de Portugal, doravante designado de “RCAAP”, que agrega conteúdos científicos em acesso aberto/livre existente em repositórios institucionais de entidades nacionais de ensino superior e a Base Terminológica IATE (Inter-Active Terminology for Europe), a base de dados interinstitucional da União Europeia que agrega documentação fornecida por instituições e agências da U.E. desde 2004. Além destes recursos, a especificidade do texto levou a uma pesquisa extralinguística no momento da compreensão etimológica da designação do termo, como foi o caso do nome de certos genes. Para determinar se determinado gene tinha ou não um equivalente em português, ou até se faria sentido existir um equivalente em português, o gene em questão foi pesquisado em bases de dados de sequências genéticas e de proteínas, que informavam a localização exata do gene e forneciam toda a informação existente relativa ao mesmo, como por exemplo, outros nomes alternativos ou até a descrição do gene, o que ajudou em certos casos à compreensão da etimologia do termo. As bases de dados utilizadas foram, para o caso das proteínas, a Universal Protein Resource (UniProt) e, no caso dos genes, o GenBank, do Centro Nacional para Informação de Biotecnologia dos EUA.

Cada problema de tradução detetado contém um curto texto com a sua contextualização, seguindo-se da estratégia utilizada para resolução do problema e tradução do termo, sendo mencionadas as fontes e meios utilizados.

Review

“In this **review**, we first describe and critique the arguments that have been proposed in support of the *cis*-regulatory hypothesis.”

Este termo foi inicialmente traduzido como “estudo”, pois no meu entender tratava-se precisamente de um estudo científico. No entanto, trata-se de uma revisão, ou seja, uma síntese de investigações e estudos sobre um determinado tema, publicados em artigos anteriores. Justifica-se deste modo a escolha do termo “revisão”.

Translation

“In contrast, synonymous mutations do not alter the amino-acid sequence, although they can modify gene function through other mechanisms, such as changes in **translation** efficiency or mRNA stability.”

Comecei por traduzir este termo como “transformação”, devido ao contexto em que estava inserido. Por estar associado ao efeito de mutação sinónima e modificação de funções de um gene, “transformação” parecia mais adequado. No entanto, após pesquisa determinei que a tradução correta seria precisamente o termo “tradução”: trata-se de um processo que conduz à decodificação da mensagem em código contida no ARNm, ou seja, não é necessariamente o processo que transforma as quatro letras do código genético (A,T,C,G) numa proteína, mas sim a tradução destas mesmas letras no código que faz a proteína.

RNA

“Cis-regulatory sequences include sequences that regulate transcription, **RNA** stability and splicing, and translation.”

O tradução do RNA é tão discutível como o termo DNA. Trata-se de uma sigla em inglês cujo significado traduzido é ARN, ou Ácido Ribonucleico. No entanto, por ser cada vez mais habitual a utilização do termo original, este poderá estar a ser cada vez menos utilizado. Optei por utilizar a tradução, por ser largamente conhecida, mesmo no meio académico (23 projetos identificados na Fundação para a Ciência e Tecnologia que utilizam o termo em português; 209 documentos – maioritariamente Dissertações de Mestrado e Teses de Doutoramento – identificadas no Repositório Científico de Acesso Aberto de Portugal que utilizam o termo em português).

mRNA (messenger RNA)

“The final gene product can either be a protein, via an mRNA intermediate, or a mature RNA molecule itself (...)”

A decisão de traduzir o termo RNA para ARN, levou-me a traduzir o termo Messenger RNA utilizando o mesmo critério. Deste modo, a tradução correta será ARN mensageiro. Depois de chegar a esta conclusão, alterei o termo para ARNm.

Codon

“Second, most insertions or deletions in protein-coding regions (those that are not multiples of 3 bp) will disrupt the reading frame and cause a premature stop **codon** or alteration of amino acids in the protein.”

Este termo foi inicialmente traduzido por “códon”, sendo que a tradução foi obtida a partir do dicionário bilingue Português - Inglês Porto Editora. No entanto, verificou-se que este dicionário tem também as suas inconsistências: após uma pesquisa pelo termo “codão” nesse mesmo dicionário, foi comprovado que, de facto, esta é a tradução correta e designa precisamente o significado do original. A descrição do termo continha também uma breve descrição, nomeadamente do codão de iniciação e codão de finalização, ambos referidos no texto de partida. Existem também vários projetos no site da Fundação para a Ciência e Tecnologia que utilizam este termo, assim como no Repositório Científico de Acesso Aberto.

Intron

“The modules can reside 5’ or 3’ of the coding region, in **introns**, or, in rare instances, in exons (Wray et al. 2003).”

A tradução deste termo trouxe o mesmo problema do termo “Codon”, ou seja, utilizei um dicionário online (Infopédia) que demonstrou novamente não ser 100% fiável, pois indicava que a tradução correta era “íntron”(“ intron [‘Intràn]: nome / BIOLOGIA (genética) íntron). No entanto, após pesquisa (inclusive neste mesmo dicionário), concluí que a tradução necessita do sufixo “ão”, ou seja, “intrão”.

mRNA cell localization

“The regulation of gene expression operates at multiple levels: translation, alternative splicing variants, mRNA stability, **mRNA cell localization** (...)”

A tradução desta expressão causou alguma dificuldade, começando logo pelo seu significado. A dificuldade em interpretar se o termo “cell” se referia ao mRNA levou-me a traduzir como “localização de células ARNm”, partindo do princípio que a esta expressão significaria literalmente a localização desta determinada célula do ARNm. No entanto, graças ao estudo de conceitos básicos como o ARN mensageiro, aquando o processo de pré-tradução, percebi que esta tradução, apesar de linguisticamente correta, não faz sentido: o ARNm não é, nem contém, qualquer tipo de célula; o ARNm é uma parte integrante das células. Assim, a pesquisa continuou até se concluir tratar-se da localização do ARNm numa determinada célula (*Propostas de atividades experimentais para extração de DNA*, Cristina Oliveira e Maria Serrano). Deste modo, a tradução desta expressão foi “localização do ARNm nas células”.

Cis-regulatory

“Some of the mutations occurring in coding or in cis-regulatory regions of genes are likely to have specific effects, potentially without any pleiotropic effects.”

O termo “*cis-regulatory*” é um conceito chave do artigo, e a sua tradução sofre várias modificações ao longo do processo de pré-tradução, tradução e revisão. Durante a pré-tradução, o termo foi traduzido como “*cis-regulador*”. Após pesquisa no motor de busca Google, ficou claro que o termo surgia apenas em português do Brasil e em castelhano. O termo foi então dividido, visto tratar-se de uma palavra composta, e foi extraído o significado de “*cis*” e “*regulatory*”. Segundo o dicionário Oxford, “*regulatory*” é um adjetivo, classifica algo que serve para ou pretende regular algo. Um dos exemplos de utilização é até semelhante ao caso estudado: “*regulatory enzymes*”. Relativamente ao termo “*cis*”, o mesmo dicionário revela tratar-se de uma expressão utilizada em Química, e refere-se ou está relacionado com uma estrutura molecular na qual dois átomos ou grupos se encontram do mesmo lado de um determinado plano na molécula. O dicionário informa ainda que poderá ter o mesmo significado que o termo “*trans*”, e é utilizado como prefixo, sendo proveniente do latim “*cis*”

cujo significado é, literalmente, “neste lado de”. Os exemplos fornecidos pelo dicionário são demonstrativos do seu uso na língua inglesa:

“historical on the side nearer to Rome: cisalpine

(of time) closer to the present: cis-Elizabethan

Chemistry (usually cis-) denoting molecules with cis arrangements of substituents:

cis- 1,2-dichloroethene”

Tendo em conta estes resultados, e embora se tratasse de uma palavra proveniente do latim, seria necessária uma correspondência de significado na língua portuguesa. Deste modo, foi efetuada uma busca no Dicionário da Língua Portuguesa (Porto Editora) que revelou vários termos com o mesmo significado, incluindo o termos *cis*. Segundo este dicionário, o termo *cis* é um “prefixo de origem latina, que exprime a ideia de aquém de, do lado de, deste lado de”. Além deste resultado, surgem também palavras compostas por aglutinação com o termo *cis*, como cisalpino (“que está aquém dos Alpes”), cisatlântico (“que fica aquém do Atlântico”), cisdanubiano (“que fica aquém do Danúbio”), entre outros exemplos similares, todos com o mesmo significado: algo que fica aquém de uma determinada região geográfica ou, mais frequentemente, de um determinado rio. Visto tratar-se de um termo latim, e tendo em conta o cariz dos exemplos no qual continua a ser utilizado, é provável que fosse bastante utilizado para indicar que algo se encontra numa posição próxima de um local, sítio ou objeto suscetível de ter dois locais: um próximo e um distante. Esta definição explica integralmente o uso do termo “*cis*” na expressão “*cis-regulatory*”. Conclui-se portanto, que a tradução poderia manter o termo “*cis*”. No entanto, persistia o problema do termo “*regulatory*”. Não deveria ser “regulador” visto não existir qualquer ocorrência em português. A palavra foi então alterada do adjetivo para o substantivo “regulação”, o que resultou em algumas dezenas de ocorrências em português europeu no referido motor de busca, no entanto todas se referiam a uma única obra que utilizava a expressão “*cis-regulação*” no seu título: “Análise Evolutiva à Lógica da Cis-regulação (Diogo Calheiros)”. A tradução manteve-se como *cis-regulação*, durante o processo de tradução. O processo de revisão continha validação terminológica por um especialista da área, e o termo foi alterado para “regulação *cis*”.

Nonsynonymous

“In principle, nonsynonymous mutations could alter both the polypeptide sequence and gene regulation, but no such examples have been reported yet.”

A tradução do termo “nonsynonymous” para “não sinónimo/a” foi imediata. No entanto, este termo adjetivava, neste artigo, não apenas termo presente no referido contexto mas também o termos “substitutions” (substituições). Relativamente ao termo “mutation” adjetivado com esta expressão, foram encontradas várias ocorrências em artigos científicos portugueses, no entanto a grafia nem sempre era igual, pois existiam ocorrências sem hifenização (1) e com hifenização (2) :

(1)“Estas mutações influenciam a virulência e evolução do vírus, particularmente as que originam substituições de aminoácidos (AA) na HA e NA (**mutações não sinónimas**).” Miguel Fevereiro, Darwinismo e os vírus Influenza A. In *Minerva, Revista de Medicina Interna, nº4, março de 2006*).

(2)“Apesar disso, dois haplótipos separados por uma **mutação não-sinónima** foram encontrados nos grupos de *C. kahawae*, revelando a inesperada presença de duas proteínas diferentes ao nível da sequência de aminoácidos numa espécie tão geneticamente uniforme.” Diogo Silva (2010), *Analysis of the neutral and adaptive genetic variation of Colletotrichum kahawae and its relationship with the C. gloeosporioides complex*, (sem publicação; disponível em <http://repositorio.ul.pt/handle/10451/3316>)

A opção final recaiu sobre a grafia sem hifenização, devido à sua ocorrência mais frequente em artigos publicados em revistas especializadas (1) e congressos (ver artigo [3]), enquanto a grafia com hifenização surge mais frequentemente em teses de mestrado (2).

A pesquisa deste termo revelou também a sua utilização adjetivante do termo “substituições”. Curiosamente, estes casos não apresentavam grafias diferentes, ao contrário da adjetivação para o substantivo “mutação”, o que reforçou o pressuposto de que a hifenização não seria a opção correta:

(3)“A diminuição progressiva do número de **substituições não sinónimas** no Env associada á progressão clínica da infeção é consistente com a diminuição da pressão arterial imposta pelos anticorpos neutralizantes sobre o Env em consequência do colapso do sistema imunitário observado nos doentes.” Nuno Taveira, Livro do VII Congresso Virtual HIV/AIDS, 2006

Splicing

“The mechanisms regulating splicing are not fully understood, but at least some of the information is encoded in the introns and must be recognized by cell-type-specific splicing factors (Lopez 1998).”

A tradução genérica deste termo é “união”, de acordo com o Dicionário Inglês-Português Porto Editora, o dicionário online WordReference e a base terminológica europeia IATE. Este termo foi identificado no processo de pré-tradução, e inserido na base terminológica que viria depois a ser utilizada na tradução. No entanto, houve uma falha na identificação deste termo como tendo uma utilização própria no âmbito da biologia. A utilização desta tradução ocorreu sempre de forma errónea no processo de tradução. A validação terminológica por um especialista da área, no processo de revisão, permitiu verificar a suspeita de que este termo não era traduzido na área da biologia. Após a validação, procedeu-se à verificação da sua utilização e, de facto, o termo mantém-se não só no infinitivo mas também quando conjugado noutras formas verbais, como se pode verificar em inúmeros títulos de projetos aprovados pela FCT:

*O papel das proteínas U2AF na regulação de **splicing** (Fonseca, M.; 2006); **Splicing** Alternativo de uma RING E3 ligase de ubiquitina envolvida em stress e desenvolvimento em plantas (Santos, P.; 2006); Alternative **spliced** isoforms of CD6: on the cross-roads of T cell activation and the regulation of lymphocyte development (Carmo, A.; 2006)*

nonA

“For example, a few genes are known to contain transcription factor binding sites in exons (keratin18 in humans and nonA in *Drosophila melanogaster* [Wray et al. 2003]).”

Este termo manteve a sua designação original na tradução para a língua portuguesa. Trata-se de um gene da referida mosca da fruta. No entanto, como a frase na qual estava inserido continha um gene com tradução para português, era necessária a identificação etimológica para dissipar dúvidas sobre o seu significado, assegurando assim que o termo não teria tradução, pois poderia, à primeira vista, tratar-se de uma palavra composta por aglutinação dos termos “non” e “A”, e essa possibilidade suscitava a dúvida de uma possível tradução como “não-A”. Para identificar a sua etimologia, foi utilizada a base de dados de genes e genoma da mosca da fruta *FlyBase*. Este gene estava

devidamente identificado, no entanto não estava presente a informação pretendida, mas existia informação relativa às sequências de ADN e proteínas consistentes com o modelo deste gene. Esta informação foi essencial para compreender a etimologia do termo, pois a determinação da sequência do gene permitiu a sua pesquisa numa base de dados de sequências de proteínas (UniProt) e num banco de genes (GenBank). O GenBank revelou uma descrição detalhada do gene, o que permitiu uma compreensão da sua composição, enquanto o UniProt demonstrou o seu nome proteico (2), o que comprovou tratar-se de um substantivo sem tradução pois o seu significado provinha da junção dos termos “no-on transient A”.

(1) *Drosophila kanekoi* no-on transient A (nonA) gene, partial cds.

(2) No-on transient A

bias/ ascertainment bias/ investigator bias

“Weak selection is expected to **bias** the spectrum of selected mutations toward those with few or no pleiotropic effects (Otto 2004).”

“However, as discussed below (in section “The Data: Evidence for a Predictive Theory of Genetic Evolution”), many sources of **ascertainment bias** strongly inflate the reported contribution of coding changes to evolution.”

“But the most important consequence of **investigator bias** is that the relative number of coding versus cisregulatory mutations causing phenotypic evolution is almost certainly inflated.”

A colocação do termo “bias” no primeiro contexto mencionado implica a sua utilização como verbo, e a tradução direta recaiu sobre o verbo “influenciar”, de acordo com o Dicionário Inglês-Português Porto Editora, sendo depois conjugado consoante o contexto. No entanto, esta tradução gerou alguma dificuldade na compreensão quando o termo era utilizado no grupo nominal do segundo contexto acima indicado. A tradução não foi, de todo, satisfatória. Foi necessária a utilização de um glossário presente no Manual de Normas de Orientação Clínica do CEMBE, Faculdade de Medicina da Universidade de Lisboa, para obter uma tradução correta deste termo. Este glossário demonstra como este termo tem uma utilização e significado específicos no contexto científico, algo que não foi possível compreender a partir de um dicionário genérico como o previamente mencionado. Segue a definição deste glossário:

“um viés define-se como um desvio sistemático do verdadeiro valor de uma variável, fator ou característica. Um viés existe quando as conclusões de um estudo estão sistematicamente longe da verdade, por problemas com a recolha, análise, interpretação, publicação ou revisão dos seus dados. (...) 5) viés de averiguação (ascertainment bias): é a não inclusão sistemática de todas as potenciais classes ou subgrupos de pacientes supostamente representativos na constituição de uma amostra”

A tradução destes termos com recurso ao mencionado glossário, possibilitou uma nova e correta tradução de “investigator bias”. Após pesquisa, determinou-se que o termo “viés do investigador” seria a tradução correta.

fitness

“The probability of fixation of a new mutation depends on the **fitness** effect associated with the mutation and the effective and demographic population sizes (Kimura 1962; Bürger and Ewens 1995).”

O termo “fitness” foi inicialmente traduzido como “aptidão”. A tradução foi obtida através do dicionário bilingue Inglês-Português Porto Editora e da base de dados terminológica IATE. Os resultados sugeriam vagamente, principalmente na base de dados terminológica, que este termo poderia ser utilizado no contexto do artigo analisado, ou seja, na biologia evolutiva do desenvolvimento. No entanto, nenhuma das fontes apresentava resultados relativos à área da biologia. A base de dados terminológica revela que o termo “fitness” poderia ser traduzido como “aptidão” mas nunca apresentou resultados relativos apenas ao termo, pois geralmente o termo “fitness” representava apenas parte de um termo com mais que uma palavra (“physical fitness: aptidão física”; “medical fitness: aptidão física”). Estas duas fontes não forneciam, de facto, fundamentação objetiva para a utilização da tradução escolhida. Assim, foi efetuada pesquisa pelo termo original em artigos científicos desta área escritos em português, através do motor de pesquisa do site da FCT e do RCAAP. Os resultados obtidos permitiram concluir que o termo “fitness” não é traduzido no âmbito da biologia, como demonstrado em vários artigos de investigação e teses. O seguinte exemplo, obtido a partir do RCAAP, demonstra a utilização do termo, e revela que o termo é feminino.

“Uma possível abordagem consiste no seu estudo em tempo real, através de experiências de evolução experimental com a bactéria *Escherichia coli*, de modo a investigar a evolução da **fitness** e os fatores importantes neste processo.”

Estudo do processo adaptativo em populações de Escherichia coli, Batista, J. (2009)

Este termo surgiu também várias vezes como palavra-chave de vários artigos, referenciados no site da FCT, em português:

“Palavra-chave 4 : **‘Fitness’**”

Ecologia evolutiva de populações selvagens de Saccharomyces que colonizam a casca de árvores da família Fagaceae, Sampaio, J. et al (2006)

“Palavra-chave 3: **Fitness**”

Valor adaptativo e base genética de um polimorfismo balanceado de coloração, Seabra, S. et al (2008)

Por fim, uma utilização bem demonstrativa da utilização deste termo é o excerto que se segue do resumo de um projeto de investigação na área da biologia, referenciado no site da FCT, pois não só demonstra a utilização do termo “fitness”, como também espelha uma certa dificuldade de aceitação, utilização e compreensão do termo em português, o que se comprova com a utilização de aspas na primeira ocorrência deste termo.

“Um dos conceitos chave na Teoria da Seleção Natural é o de sucesso reprodutivo ou “**fitness**” relativo de um organismo: é tanto maior quanto maior for o número de descendentes relativamente aos outros membros da população. Indivíduos com maior **fitness** estarão mais representados nas futuras gerações. Portanto, a existência de indivíduos altruistas parece ser contraditória com a visão darwinista da seleção natural. O paradoxo vem do facto do comportamento altruista comportar um custo no seu **fitness**, ao mesmo tempo que beneficia os outros organismos da população.”

O papel da interação social em bactérias na diversidade e evolução bacteriana, Dionísio, F. (2004)

Como se pode verificar através da comparação das referências, não existe consenso sobre a utilização do género relativamente a este termo. Se a referência do RCCP utiliza o género feminino, já a referência da FCT utiliza o género masculino. No entanto, os resultados com a utilização no masculino surgiram em maior número, o que foi preponderante na escolha de género.

8. Conclusão

A tradução comentada e análise do processo de tradução resultou num conhecimento muito mais vasto da área de especialização escolhida, por parte do tradutor. Também o processo de Tradução Especializada aqui exposto revelou detalhes que, apesar de reconhecidos, são raramente estudados durante o processo de tradução, como é o caso da análise da micro-estrutura do texto e a exposição de problemas de tradução que necessitou de ampliar conhecimentos, como o estudo da etimologia de um determinado termo, no sentido de melhor compreender o seu significado. O resultado final deste trabalho académico foi, sem dúvida, extremamente gratificante e enriquecedor para o autor.

Este trabalho académico foi diferente de todos os outros, porque resultou em algo mais do que um enriquecimento curricular do aluno. O objetivo principal é, claramente, apresentar os conhecimentos adquiridos ao longo do percurso académico, mas também demonstrar que este Mestrado tem dois objetivos bem claros: formar um tradutor especializado na área da Saúde e Ciências da Vida, e ensinar ao aluno o próprio processo de especialização, para que possa ser ele próprio a especializar-se na área que desejar, o que será de grande importância na sua carreira profissional como tradutor especializado. O presente trabalho é no fundo um exemplo de como poderá ser posto em prática este segundo objetivo, muito embora já existissem conhecimentos adquiridos durante o Mestrado e a área escolhida esteja relacionada com a área de especialização do Mestrado. Além deste enriquecimento do aluno, o trabalho possui secções que poderão ter de facto uma utilização prática, como é o caso da própria tradução ou do glossário.

Este trabalho tem como objetivo principal demonstrar os conhecimentos do aluno, no entanto, é minha convicção que será também importante se for considerado um real contributo para o público-alvo previamente mencionado (Cap.3), ou seja, se realmente vier a ter *utilidade* no futuro. Do ponto de vista do autor, esta *utilidade* está já a ser posta em prática diariamente: um trabalho académico que, à partida, parecia ser uma “distração” da carreira profissional, acabou por aperfeiçoar dramaticamente o trabalho que efetuo diariamente.

O processo da Tradução Especializada deverá ser observado atentamente e o tradutor especializado deverá exigir de si próprio um aprofundado conhecimento da área de especialização. O mundo em

que vivemos comunica a uma velocidade nunca antes observada, e o tradutor assume cada vez mais o papel de intermediário, mas um bom tradutor só se destacará pela qualidade do seu serviço.

9. Glossário

A informação presente no glossário diz respeito à tradução e definição do termo. Assim, a colocação da informação respeita a importância de cada um, estando pois baseado na formulação utilizada em dicionários bilíngues. A informação presente no glossário segue então os seguintes parâmetros gráficos e de disposição:

termo

tradução

Definição do termo

3'UTR

3'UTR

Sigla original mantém-se. Trata-se de uma região não traduzida (UTR - untranslated region) próxima da extremidade 3' de mRNA, seguindo-se pela cauda poli-A.

5'UTR

5'UTR

Sigla original mantém-se. Trata-se de uma região não traduzida (UTR - untranslated region) localizada na extremidade 5' de mRNA, precedida pela estrutura Cap 5'.

A. Thaliana

Arabidopsis thaliana

Planta da família *Brassicaceae* (nome binominal); utilização modelo no estudo da genética; primeira planta cujo genoma foi completamente sequenciado.

achaete-scute

achaete-scute

A designação completa é “achaete-scute complex”, e trata-se de um grupo de quatro genes da *Drosophila melanogaster* responsáveis pela codificação dos fatores de transcrição hélice-volta-hélice.

acid

ácido

Composto que contém um ou mais átomos de hidrogénio; adjetivo: que tem pH menor que 7 a 25°C.

affinity

afinidade

Tendência que os átomos ou moléculas manifestam para se combinarem entre si; atração química.

allele

alelo

Sequência de uma molécula de ADN (gene) situada no mesmo locus e que corresponde a diferentes versões do mesmo gene.

amino acid (amino-acid)

aminoácido

Compostos que apresentam na sua estrutura o grupo carboxílico (-COOH) e o grupo amina (-NH₂), ligados ou não ao mesmo átomo de carbono da cadeia carbonada.

amylase gene

gene da amilase

Enzima catalisadora da hidrólise.

Antifreeze protein

Proteína anticongelante

Proteína localizada no fluido extracelular de alguns peixes que vivem em águas muito frias; esta proteína inibe a formação de cristais de gelo.

basal promoter

promotor basal

Também conhecido como "core promoter"; trata-se do promotor essencial existente em todos os genes que produzem proteínas.

basal transcription apparatus

aparelho basal de transcrição

Conjunto de fatores de transcrição instalados no promotor, antes da polimerase do ARN.

base pairs

par de bases

Dois nucleotídeos da cadeia ARN ligados um à outro por ligações de hidrogénio.

behavior

comportamento

Reação de um ser perante uma situação ou conjunto de estímulos.

bicoid gene

gene bicoid

Grupo de genes de elevada importância para o desenvolvimento da cabeça e torax no embrião da mosca da fruta, ou *Drosophila melanogaster*.

binding site

local de ligação

Local do ADN ao qual se liga uma proteína.

Bone morphogenetic protein 4 (BMP4)

Proteína Morfogenética do osso 4 (PMO4)

Sigla: BMP4; sinónimo: Bone Morphogenetic Protein 2B (BMP2); Proteína que fomenta a formação de cartilagem e ossos.

bristle

pelo

Cabelo de dimensão mais fina, mais rijo e duro.

broad cis-regulatory hypothesis

hipótese cis-reguladora geral

Hipótese teorizada por Stern, D. e Orgogozo, V. a qual indica que a predominância de mutações no cis-regulador se deve a vários tipos de mutações.

carbamate

carbamato

Composto orgânico derivado do ácido carbâmico.

cell-signaling process

processo de sinalização celular

Este processo é parte do sistema de comunicação celular responsável pela atividade e ação celular. O processo de sinalização celular está dividido em três passos: receção, transdução e resposta. Os erros deste processo na comunicação entre células provocam doenças.

chemoreceptor

quimiorreceptor

Célula nervosa sensorial que deteta e responde a estímulos químicos num determinado ambiente e transmite essa informação ao sistema nervoso central.

chromosom

cromossoma

Estrutura filamentosa do núcleo da célula eucariota que contém a informação genética. Cada cromossoma consiste numa molécula de ADN. O local dum cromossoma onde se situa um determinado gene é o locus desse gene.

cis-regulatory

regulação cis

Sequência de ADN não-codificante localizada num gene, necessária para a expressão desse mesmo gene.

cis-regulatory hypothesis

hipótese da regulação cis

Teoria da biologia evolutiva que indica que grande parte das mutações que causam alterações morfológicas ocorrem na regulação *cis*.

coding

codificação

Também conhecido como Coding Sequence (Sequência de Codificação), CDS, ou coding region (região de codificação). Trata-se de uma região do ADN ou ARN de um gene composta por exões, responsável pela codificação das proteínas.

cuticle pigmentation

pigmentação da cutícula

Coloração anormal da epiderme.

dataset

conjunto de dados

Compilação de dados relevantes ao domínio do artigo.

deleterious effects

efeitos deletérios

Efeitos prejudiciais ou nocivos.

deletion mutant

mutante de deleção

Tipo de mutação de genes no qual a deleção de um ou mais nucleótidos provoca uma alteração na grelha de leitura dos codões do ARNm, o que poderá levar à alteração da sequência de amino-ácidos na tradução de proteínas.

developmental biology

biologia do desenvolvimento

Estudo do processo de desenvolvimento de organismos multicelulares desde a sua forma mais imatura até à forma adulta.

developmental patterning gene

gene padronizante do desenvolvimento

Gene responsável pela criação de um determinado padrão de desenvolvimento.

developmental regulatory architecture

arquitetura reguladora do desenvolvimento

Termo utilizado por D.Stern e V.Orgogozo no artigo utilizado no presente trabalho para designar a composição e construção de um regulador envolvido no processo de desenvolvimento.

DGB member

membro do DGB

Gene integrante de um determinado conjunto de genes de diferenciação.

differentiation gene battery (DGB)

conjunto de genes de diferenciação (DGB)

A sigla DGB não é traduzida para Português. Um DGB é um conjunto de genes responsáveis pelo processo de desenvolvimento ou maturação de uma célula pouco especializada, obtendo estas formas e funções mais distintas.

dipteran

dípteros

Ordem de insetos com um par de asas desenvolvido e outro par cujo desenvolvimento transformou em órgãos rudimentares.

DNA

ADN

Sigla de Ácido desoxirribonucleico; molécula que contém a informação genética necessária a um organismo vivo. Trata-se de um ácido nucleico de cadeia dupla que constitui a base da hereditariedade genética dos organismos vivos.

DNA strand

cadeia de ADN

O ADN é normalmente composto por duas cadeias: senso e antissenso. O termo "cadeia" refere-se ao agrupamento dos átomos na estrutura de ADN.

domesticated species

espécies domésticas

Espécies de seres vivos que passaram pelo processo de domesticação, isto é, o processo de seleção e adaptação de seres a um determinado ambiente.

domesticated races

raças domésticas

Ver DOMESTICATED SPECIES; Raça de seres vivos que passaram pelo processo de domesticação.

Drosophila melanogaster

Drosophila melanogaster

Nomenclatura binominal de uma espécie de mosca conhecida como mosca da fruta. A sua utilização na biologia é bastante comum por ser de fácil utilização, rápida reprodução e por pôr muitos ovos. Foi também um dos primeiros organismos utilizados na análise genética, e o seu genoma foi já totalmente sequenciado.

electron micrograph

micrografia eletrónica

Tipo de fotografia obtida através da utilização de um microscópio, frequentemente utilizada para obter imagens a nível microscópico.

embryonic expression

expressão embrionária

Processo no qual a informação codificada de um gene é convertida nas estruturas de uma célula embrionária.

embryonic patterning

padronização embrionária

Mecanismo que fornece a informação necessária às células para desenvolverem a forma e estrutura anatômica adequada.

endocrine glands

glândulas endócrina

Órgãos que segregam os produtos do sistema endócrino diretamente para a corrente sanguínea.

enhancer

intensificador

Sequência do ADN que intensifica a transcrição de um gene associado quando se liga a proteínas específicas designadas de fatores de transcrição.

enzyme

enzima

Molécula proteica que atua como catalisador das reações metabólicas.

epidermal cells

células epidérmicas

Células que compõe a epiderme, ou seja, a camada externa de um organismo. No caso dos seres humanos, por exemplo, as células da pele são células epidérmicas.

epidermal glands

glândulas epidérmicas

Glândula da epiderme responsável pela segregação de hormonas.

esterase

esterase

Enzima catalisadora da hidrólise de éster em álcool e ácido.

eukaryote

eucariota

Organismo cujas células possuem um núcleo envolvido por uma membrana, onde está contido o ADN repartido nos vários cromossomas. As células eucarióticas geralmente contêm mitocôndrias, cloroplastos e outras estruturas envolvidas por membranas (organitos) que não se observam nas células dos restantes organismos (procariotas).

even-skipped expression

expressão do even-skipped

Processo de conversão da informação codificada no gene even-skipped, neste caso, na banda 2 do embrião da *Drosophila*.

evolution

evolução

Processo resultante de alterações no material genético de populações ao longo do tempo. As adaptações dos organismos ao meio ambiente levam à alteração de genes, novas características e até novas espécies. Estas adaptações são a evolução de um ser.

evolutionary biology

biologia evolutiva

Ramo da biologia que estuda a origem de organismos através das suas mutações.

evolutionary developmental biologist

biólogo especialista em evolução e desenvolvimento

Biólogo especialista no estudo desta área da biologia que investiga os processos de desenvolvimento dos organismos para determinar a relação existente entre eles.

evolutionary genetics

genética da evolução

Estudo da evolução genética, mais especificamente das alterações nos genes de determinadas populações.

exon

exão

Segmento de ADN de um gene que é transcrito e traduzido. Os exões codificam cadeias polipeptídicas específicas codificadas nos ARN mensageiros correspondentes.

expression

expressão

Processo no qual a informação codificada de um gene é convertida num produto do gene.

expression pattern

padrão de expressão

Tipo de processo estandardizado que converte a informação codificada de um gene nas estruturas de uma célula.

file drawer effect

efeito gaveta

Prática de arquivar estudos com resultados negativos, ou seja, sem interesse, por parte de investigadores.

fitness

fitness

Capacidade de um organismo em manter, reproduzir e propagar genes num determinado ambiente.

foregut

intestino dianteiro

Intestino responsável pela fermentação dos alimentos antes de se dirigir para o estômago.

functional constraint

restrição funcional

Processo que ocorre quando uma região de ADN é intolerante à mutação, devido a uma redução na sua capacidade de desempenhar a função codificada.

GC Box

Caixa GC

Local do gene promotor de células de mamíferos que contém a sequência de nucleótidos GGGCGG e no qual se unem os fatores de transcrição.

gene

gene

Unidade do ADN, passível de mutação, responsável pela hereditariedade dos organismos. O significado popular de gene ("bons genes", "genes da cor dos olhos") pode nem sempre significar um gene, no sentido real do termo, mas sim um alelo.

gene duplication*duplicação de genes*

Trata-se, no fundo, de uma cópia adicional de um gene. A duplicação de genes é um mecanismo vital para a evolução, porque quando um gene é duplicado, o gene idêntico pode sofrer alterações e criar dois genes diferentes. A duplicação é o oposto da deleção.

gene expression*expressão do gene*

Transformação, por efeito da transcrição, da informação do gene em ARN mensageiro. Este ARN mensageiro é depois traduzido numa proteína. Este processo é a manifestação fenotípica do gene.

gene loss*perda de gene(s)*

Eliminação de genes nas células de diferenciação.

gene products*produtos do gene*

Material resultante do processo de expressão do gene. Os produtos do gene podem ser, por exemplo, ARN ou proteínas. A atividade do gene é medida de acordo com a quantidade de produtos do gene.

gene rearrangement*reconfiguração dos genes*

Reestruturação das regiões do gene causada pela recombinação do ADN. A reestruturação dos genes ocorre durante o desenvolvimento embrionário, por exemplo, porque esta fase implica a recombinação do ADN. Pode também ser definido como uma alteração estrutural de um cromossoma que provoca mudanças nos loci, ou seja, nos locais dos genes nos cromossomas.

gene-coding region*região codificante do gene*

Região do gene capaz de codificar informação como, por exemplo, traduzir ADN em proteínas.

genetic change*alteração genética*

Alterações nos genes que ocorrem devido a mutação (durante a replicação de ADN) ou evolução por seleção natural. Este processo ocorre devido a erros no momento de replicação de ADN, no caso das mutações (estes são bastante raros - 1 erro por cada 100 milhões de pares-base).

genetic drift*deriva genética*

Processo de alterações na composição genética de uma população devido a eventos aleatórios e não devido a seleção natural.

genomic organization*organização genómica*

Organização da sequência do genoma. A organização genómica não compreende a organização estrutural do genoma.

geometric model of adaptation*modelo geométrico de adaptação*

Modelo de Ronald A. Fisher que possibilita a investigação da dimensão fenotípica de uma determinada mutação. Este modelo consegue localizar os efeitos dos fenótipos na capacidade de adaptação.

hair follicles*folículo capilar*

Conjunto de células que contém uma cavidade na qual nascem outras estruturas. No caso do folículo capilar, a estrutura é uma glândula produtora de sebo que produz pelo. Quantas mais glândulas sebáceas existirem na cavidade do folículo, maior será o tamanho do pelo.

haploid genome*genoma haploide*

O termo haploide refere-se ao número de cromossomas num gâmeta (termo proveniente de Ploidia - quantidade de conjuntos de cromossomas numa determinada célula). As células constituintes do genoma humano, por exemplo, contêm um conjunto de 23 cromossomas.

heritability

hereditariedade

Proporção de variação fenotípica devido a causas genéticas. A hereditariedade determina a importância relativa da variação genética na determinação da variação fenotípica.

heterologous promoter

promotor heterólogo

Região de uma molécula de ADN responsável pelo processo de iniciação da transcrição do ARN mensageiro (após ligação entre a polímerase do ARN com fatores de transcrição), cuja estrutura ou origem evolutiva é diferente do restante ADN, pois deriva de um organismo diferente.

hormone

hormona

Substância química segregada numa determinada parte do corpo, normalmente em pequenas quantidades, e transportada para outro lugar, onde produz uma resposta.

hypothesis

hipótese

Suposição ou explicação preliminar de um determinado fenómeno ou facto, que necessita de ser testada e verificada através de investigação mais aprofundada e experimentação metodológica.

intergeneric

intergenérico

Algo que existe em diferentes géneros ou obtido a partir de diferentes géneros. No caso específico das diferenças intergenéricas, o termo refere-se às diferenças que existem em ambos.

intraspecific variation

variação intraespecífica

Variação entre indivíduos da mesma espécie.

introns

intrões

Sequência não codificadora intercalada entre dois exões de um gene.

keratin8

queratina-8

Proteína membro da família queratina, localizada no cromossoma 12. A função mais importante desta proteína é a manutenção da integridade estrutural de uma célula.

larval trichomes

tricoma larval

O termo tricoma é proveniente do latim trikhoma, que significa "crescimento capilar". Os tricomas larvais são apêndices localizados na epiderme de um organismo em estado larval que irá mais tarde dar origem a pelos.

ligands

ligandos

Molécula que se liga a outra molécula; geralmente é uma molécula pequena que se liga a uma de maior dimensão.

linkage

sistema de ligação

Conjunto de ligações; também se pode referir a grupos de genes no mesmo cromossoma que são herdados em conjunto.

loci

loci

Plural de locus. Região na qual está localizado um determinado gene num cromossoma.

micro RNA

micro ARN

Abreviatura: miRNA; Pequena molécula de ARN localizada em células eucarióticas, que contém menos nucleótidos (somente 22) que o ARN.

morphological

morfológico

Relativo à morfologia (forma e estrutura de organismos).

morphological evolution

evolução morfológica

Desenvolvimento de um organismo provocado por alterações na sua forma ou estrutura (interna ou externa)

mRNA

mARN

ARN que atua como molde para a ligação de aminoácidos ao nível dos ribossomas durante a síntese das proteínas. É transcrito a partir do cistrão original do ADN de que é complementar.

mRNA intermediate

intermediário mARN

ARN mensageiro localizado entre dois determinados locais.

mRNA molecule

molécula mARN

Molécula composta por ARN mensageiro.

mRNA splicing

ligação mARN

Alteração da informação genética após a transcrição, na qual se eliminam os intrões do ARN mensageiro, e são adicionados exões. Esta alteração é necessária para que o ARN pré-mensageiro produza ARN mensageiro maduro que, após tradução, produz proteínas.

mRNA stability

estabilidade de mARN

Retenção da integridade estrutural e resistência à degradação por parte do ARN mensageiro.

multicellular

multicelular

Que consiste em várias células necessárias para o seu funcionamento. O ser humano, qualquer animal e as plantas são organismos multicelulares, por exemplo.

muscle attachment cell

célula de ligação muscular

Células responsáveis pela ligação entre músculos.

mutation

mutação

Alteração na sequência genética. Pode alterar desde uma base nucleótida até vários genes dum cromossoma.

mutation rate

taxa de mutação

Frequência de uma mutação numa determinada população.

narrow cis-regulatory hypothesis

hipótese limitada do regulador cis

Hipótese teorizada por D.Stern, e V.Orgogozo, a qual indica que a predominância de mutações no cis-regulador pode-se dever a mutações na codificação.

nematode

nematode

Organismo de corpo cilíndrico considerado o mais abundante na biosfera.

neurons

neurónios

Célula do sistema nervoso especializado na transmissão de sinais elétricos a longas distâncias.

noncoding DNA

ADN não codificante

ADN que não se codifica em sequências de proteínas.

Non-DBG member

não-membro de DGB

Gene que não integra um determinado conjunto de genes de diferenciação.

nonpleiotropic effects

efeitos não-pleiotrópicos

Efeitos causados num gene que não são pleiotrópicos, isto é, que não estão envolvidos no fenómeno no qual um gene influencia uma ou mais características independentes.

Notch

Notch

Gene neurogénico da *Drosophila melanogaster*.

nucleotide

nucleótido

Nucleosídeo no qual a molécula de açúcar está ligada a um ou mais grupos fosfato. O ADN e o ARN compõem-se de longas cadeias de nucleótidos (polinucleótidos).

oenocytes

oenócitos

Célula secretória localizada na epiderme da região abdominal de larvas.

olfactory recetor

recetor olfativo

Neste contexto, trata-se de um conjunto de células responsável pela deteção de moléculas de odor.

operon

operão

Conjunto de genes ou um segmento de ADN que opera como unidade de transcrição individual. Um operão é composto por um operador, um promotor e um ou mais genes estruturais.

organophosphate

organofosfato

Inseticida que afeta o sistema nervoso de insetos.

Overexpression

superexpressão

Processo no qual um gene tem uma produção extraordinária de produtos de gene.

phenotypic

fenotípico

Com propriedades observáveis, propriedades estas que são produzidas através da interação entre o genótipo e do ambiente.

phenotypic evolution

evolução fenotípica

Processo de desenvolvimento das características observáveis de um organismo.

physiological evolution

evolução fisiológica

Desenvolvimento de características físicas de um organismo.

pleiotropy

pliotropia

Processo que ocorre quando um gene controla ou influencia variadas características fenotípicas.

pluricellular sensory organs

órgãos sensoriais pluricelulares

Órgãos pluricelulares relacionados com um dos cinco sentidos especiais (visão, audição, tacto, gosto e olfacto).

point mutation

mutação pontual

Mutações em moléculas de ADN ou ARN que alteram apenas uma base nucleótida.

polyadenylation

poliadenilação

Processo de sintetização da cauda poli A numa molécula de ARN.

polymorphism

polimorfismo

Coexistência de vários alelos diferentes dum dado gene ou de outra sequência de ADN num grupo de plantas ou de animais da mesma espécie. Os exemplos de polimorfismo incluem os alelos que determinam os grupos sanguíneos na espécie humana e as diferentes cores que se observam em algumas borboletas.

population genetics

genética de populações

Campo da biologia que estuda a composição genética de populações biológicas e as alterações na composição genética resultantes de vários fatores, como por exemplo a seleção natural.

predictability

previsibilidade

Algo que é visto com antecipação; no âmbito da genética evolutiva, esta antecipação requer, geralmente, previsões probabilísticas.

probabilistic prediction

previsão probabilística

Previsão realizada com base num determinado grau de probabilidade.

promoter

promotor

Sequência de ADN que indica à enzima polimerase do ARN o local de início da transcrição dum gene. Podem ligar-se-lhe moléculas reguladoras que afetam a expressão do gene.

protein

proteína

Substância complexa, biologicamente importante, composta de aminoácidos unidos por ligações peptídicas. As proteínas são essenciais para todos os organismos vivos. Como enzimas regulam todos os aspectos do metabolismo.

protein domain

domínio proteico

Unidade estrutural que integra a sequência e estrutura de uma proteína, que por sua vez tem mais que um domínio proteico. Esta unidade é independente da sua proteína e dos outros domínios proteicos.

protein-coding DNA

ADN codificador de proteínas

ADN que se codifica em sequências de proteínas.

pseudogene

pseudogene

Sequência de ADN cuja capacidade de codificação em proteínas foi eliminada. São semelhantes a genes funcionais mas, tanto quanto se sabe, não têm qualquer função.

quantitative genetics

genética quantitativa

Estudo de características genéticas mensuráveis.

receptor protein

proteína receptora

Proteína intracelular que liga um determinado estímulo à actividade da célula. Estes estímulos, ou sinais, emitidos indicam à célula qual será a sua actividade.

regulatory hypothesis

hipótese do regulador

Hipótese que indica que a evolução fenotípica ou morfológica deve-se maioritariamente a alterações do reguladores da expressão dos genes.

regulatory network

rede do regulador

Rede que regula o nível no qual os genes que a integram são transcritos para ARNm. Esta rede é, no fundo, um conjunto de sequências de ADN numa determinada célula que interagem entre si.

repressor binding site

local de ligação de um repressor

Região de proteína reguladora que se liga a um operador, bloqueando assim a transcrição de genes do operão.

ribosomal RNA

ARN ribossomal

ARN presente no ribossoma.

RNA

ARN

Ácido ribonucleico. Ácido nucleico envolvido na síntese de proteínas. Possui normalmente uma cadeia simples, ao contrário do ADN, e consiste num grande número de nucleótidos unidos, cada um dos quais compreende o açúcar ribose, um grupo fosfato e uma de quatro bases (uracilo, citosina, adenina ou guanina). O ARN é copiado do ADN por emparelhamento de nucleótidos livres com as bases complementares da cadeia simples (codificadora) do ADN molde.

RNA transcripts

transcritos de ARN

ARN que passou pelo processo de transcrição de uma molécula de ADN.

Rnase

Ribonuclease (abrev. RNase)

Grupo de enzimas que segmentam o ARN por hidrólise. As ribonucleases reconhecem um ARN específico pela sua estrutura, e desempenham depois importantes papéis no processamento do ARN, em vários processos reguladores e na degradação das moléculas de ARN.

runt gene

gene runt

Um gene de controlo par da *Drosophila melanogaster* que codifica um factor de transcrição com funções de segmentação.

scute

scute

Gene da *Drosophila melanogaster* codificante de proteínas relacionada com o seu mesonoto.

shavenbaby/ovo (svb)

shavenbaby / ovo (svb)

As mutações que eliminam a função deste gene eliminam quase todos os "pelos" da larva, daí o nome shaven + baby.

splice

splice

Processo de remoção de intrões do ARN, que permite a produção de ARNm que contém apenas exões.

splice sites

locais de splicing

Regiões fronteiriças entre exões e intrões.

stable DNA methylation

metilação do ADN estável

Processo de adição de nucleótidos a grupos de metilo, no ADN genómico.

steroid hormones

hormona esteróide

Hormona pertencente à classe do composto químico esteróide. Estas hormonas são segregadas por três glândulas esteróides durante a gravidez através da placenta.

stickleback

esgana-gata

Espécie de peixe de pequenas dimensões, com corpo alongado, fusiforme e sem escamas; a ampla distribuição mundial desta espécie - resultado do grande degelo - e a grande variedade de ecossistemas que habita, fazem com que seja uma espécie altamente variável, morfológica e geneticamente.

superfamily

superfamília

Grupo intermédio entre uma família e uma subordem.

svb expression

expressão do svb

Processo de codificação de informação do gene shavenbaby.

symbiotic bacteria

bactéria simbiote

Bactéria que vive em simbiose com outra bactéria ou com um determinado organismo.

synonymous mutation

mutação sinónima

Mutação que ocorre quando existe uma alteração na sequência genética do ADN, sendo que esta alteração resulta no mesmo produto genético inicial.

TATA Box

Caixa TATA

Sequência de ADN que indica onde é que uma sequência genética pode ser lida e decodificada. Trata-se de uma sequência promotora que indica às outras moléculas onde se inicia a transcrição

taxonomic level

nível taxonómico

Posição relativa na hierarquia taxonómica; cada nível possui subcategorias mais específicas.

transcription factor binding site

local de ligação do factor de transcrição

Local de ligação da proteína que se liga a sequências específicas de ADN, controlando deste modo a transcrição da informação genética de ADN para ARNm.

transfer RNA

ARN de transferência

Molécula adaptadora composta por ARNm, que transporta aminoácidos para a síntese da proteína.

translation

tradução

Processo no qual o código genético do ARNm é decodificado de modo a produzir aminoácidos.

trans-regulatory

trans-regulador

Genes com a capacidade de regular a expressão de outros genes que se encontram em posições distantes.

trichome

tricoma

Elemento capilar epidermal das plantas

trypsinogen

tripsinogénio

Substância inativa libertada pelo pâncreas para o duodeno, que forma a tripsina

10. Bibliografia

Sobre tradução

Altanero, T. (2000), Localisation, Internationalisation, Globalization, and Translation. In S. Thomson (Ed.) *Getting Started: A Newcomer's Guide to Translation and Interpretation*, 47 – 62, American Translators Association.

Cunha, C.; Cintra, L. (1984), *Nova Gramática do Português Contemporâneo*, Companhia Editora Nacional.

Gouadec, D. (2007), *Translation as a Profession*, Amesterdão, Benjamins.

Robinson, D. (1997). *Becoming a Translator*, Londres, Routledge.

Stolze, R. (2001), Creating “presence” in translation. In G.Hansen, J. Malmkjaer, D. Gile (Eds.) *Claims, Changes and Challenges in Translation Studies*, 39-51, Amesterdão, Benjamins.

Toury, G. (1995), *Descriptive Translation Studies and Beyond*, Amesterdão, Benjamins.

Weston, A. (1996), *A Arte de Argumentar* (traduzido por Desidério Murcho), Lisboa, Gradiva.

Sobre biologia e evolução

Azakami, H.; Sugino, H.; Murooka, Y. (2002), Cloning and nucleotide sequence of a negative regulator gene for *Klebsiella aerogenes* arylsulfatase synthesis and identification of the gene as folA. In *J Bacteriol.* 174(7), 2344–2351.

Brakerfield, P. (2011), Evo-Devo and accounting for Darwin's endless forms. In D. Patridge (Ed.) *Philosophical Transactions Of The Royal Society – Biological Sciences*, 2069-2075.

Carroll, S. (2005), *Endless Forms Most Beautiful*, Nova Iorque, W.W.Norton & Company.

Feijó, J.; Gago, M; Vala, F. (2009), *A evolução de Darwin*; Lisboa, Fundação Calouste Gulbenkian.

Darwin, C. (1859). *A evolução das espécies* (tradução de Ana Afonso), 2009, Leça da Palmeira, Planeta Vivo.

Laublicher, M. (2007). Does History Recapitulate Itself? Epistemological Reflections on the Origins of Evolutionary Developmental Biology. In *From Embryology to Evo-Devo: A History of Developmental Evolution*, 2007, Londres, The MIT Press.

Owen, E; Daintith, E. (2004). *The Facts on File Dictionary of Evolutionary Biology*, Market House Books.

Rockman, M.; Wray, G.,(2002) Abundant Raw Material for Cis-Regulatory Evolution in Humans. In *Molecular Biology and Evolution* 19 (11), 1991-2004.

Shubin, N. (2008), *Your Inner Fish*, Londres, Penguin Publishing.

Wray, G., (2007), The evolutionary significance of *cis*-regulatory mutations. In *Nature*, March 2007, Vol.8.

Zielenski, J.; Rozmahel, R.; Bozon, D.; Kerem, B.; Grzelczak, Z.; Riordan, J.; Rommens, J.; Tsui, L.; (2010) Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. In *Genomics*, 1991 May; 10(1), 214-28.

Audiovisual

"What Darwin Never Knew" (2009), Sean B Carrol, John Rubin, Rushmore Denoyeer, Serena Davies, Sara Holt

"What Darwin Didn't Know" (2009) Armand Marie Leroi

"Charles Darwin and the Tree of Life" (2009), David Attenborough

11. Anexos

THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION?

David L. Stern^{1,2} and Virginie Orgogozo^{3,4}

¹Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544

²E-mail: dstern@Princeton.edu

³CNRS, Université Pierre et Marie Curie, Bâtiment A, 5ème étage, case 29, 7 Quai Saint Bernard, 75005 Paris, France

⁴E-mail: Virginie.Orgogozo@snv.jussieu.fr

Received March 12, 2008

Accepted May 22, 2008

Is genetic evolution predictable? Evolutionary developmental biologists have argued that, at least for morphological traits, the answer is a resounding yes. Most mutations causing morphological variation are expected to reside in the *cis*-regulatory, rather than the coding, regions of developmental genes. This “*cis*-regulatory hypothesis” has recently come under attack. In this review, we first describe and critique the arguments that have been proposed in support of the *cis*-regulatory hypothesis. We then test the empirical support for the *cis*-regulatory hypothesis with a comprehensive survey of mutations responsible for phenotypic evolution in multicellular organisms. *Cis*-regulatory mutations currently represent approximately 22% of 331 identified genetic changes although the number of *cis*-regulatory changes published annually is rapidly increasing. Above the species level, *cis*-regulatory mutations altering morphology are more common than coding changes. Also, above the species level *cis*-regulatory mutations predominate for genes not involved in terminal differentiation. These patterns imply that the simple question “Do coding or *cis*-regulatory mutations cause more phenotypic evolution?” hides more interesting phenomena. Evolution in different kinds of populations and over different durations may result in selection of different kinds of mutations. Predicting the genetic basis of evolution requires a comprehensive synthesis of molecular developmental biology and population genetics.

KEY WORDS: *cis*-regulation, development, genetic variation, genome, phenotypic variation.

OnlineOpen: This article is available free online at www.blackwell-synergy.com

“Curiously, the improved understanding of the nature of gene and mutation has not added, so far, to the understanding of evolutionary phenomena.”

E. Mayr *Animal Species and Evolution* (1963, p. 172)

“There are many generalizations in biology, but precious few theories.”

F. Jacob *The Logic of Life* (1973, p. 13)

Natural selection causes predictable changes in phenotypic variation. This predictability exists at two levels. First, quantitative genetics provides predictions for the short-term response to selection, given estimates of heritability and the selection differential (Falconer and Mackay 1996). Second, selection theory often provides reasonable predictions of how populations will adapt over

the long term following a change in the selective regime. These are probabilistic predictions and, due to historical contingency, populations may not evolve as predicted in every case. Nonetheless, in some cases, precise predictions at the phenotypic level have been fulfilled by observation (Herre 1985, 1987). Natural selection thus provides a compelling explanation for phenotypic evolution of life on the earth.

In contrast, the genetic changes underlying these phenotypic changes have historically not been expected to show predictable patterns. For example, it has long been recognized that different genetic causes can generate similar patterns of phenotypic variation (Robertson 1959; Wilkens 1971). Discoveries in molecular and developmental biology over the past 40 years, however, have led some biologists to suggest that mutations that alter the

OS LOCI DA EVOLUÇÃO: QUÃO PREVISÍVEL É A EVOLUÇÃO GENÉTICA?

David L. Stern^{1,2} e Virginie Orgogozo^{3,4}

¹ Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544

² E-mail: dstern@Princeton.edu

³ CNRS, Université Pierre et Marie Curie, Bâtiment A, 5ème étage, case 29, 7 Quai Saint Bernard, 75005 Paris, France

⁴ E-mail: Virginie.Orgogozo@snv.jussieu.fr

Recebido a 12 de março de 2008

Aceite a 22 de maio de 2008

É possível prever a evolução genética? Os biólogos da área da biologia evolutiva do desenvolvimento afirmam ter uma resposta inequivocamente positiva, pelo menos no que diz respeito às características morfológicas. Pensa-se que grande parte das mutações que causam alterações morfológicas se encontra na região de regulação *cis*, e não de codificação, dos genes do desenvolvimento. A “hipótese da regulação *cis*” tem vindo a ser questionada. Esta revisão começa por descrever e criticar os argumentos propostos na defesa da hipótese da regulação *cis*. De seguida são testados os dados empíricos da hipótese da regulação *cis*, com recurso a um estudo abrangente de mutações responsáveis pela evolução fenotípica em organismos multicelulares. As mutações da regulação *cis* representam atualmente cerca de 22% das 331 alterações genéticas identificadas, apesar do número de alterações na região da regulação *cis* publicadas anualmente estar a aumentar rapidamente. Em níveis taxonómicos superiores ao da espécie, verifica-se que as mutações na regulação *cis* capazes de alterar a morfologia são mais frequentes que as alterações na codificação. Verifica-se também, em níveis superiores ao da espécie, que as mutações da regulação *cis* são predominantes nos genes não envolvidos na diferenciação terminal. Estes dados indicam que a simples questão “O que provoca maior evolução fenotípica são as mutações na codificação ou na regulação *cis*?” poderá revelar mais fenómenos de interesse. A evolução poderá resultar na seleção de diferentes tipos de mutação, tendo em conta as variáveis tipo de população e duração. Prever a base genética da evolução requer uma síntese abrangente da biologia molecular e do desenvolvimento e da genética populacional.

Palavras-chave: regulação *cis*, desenvolvimento, variação genética, genoma, variação fenotípica.

“Curiosamente, uma maior compreensão da natureza do gene e da mutação não tem contribuído, até agora, para a compreensão dos fenómenos evolutivos.”

E. Mayr *Animal Species and Evolution* (1963, p. 172)

“Existem muitas generalizações na biologia, mas poucas teorias importantes.”

F. Jacob *The Logic of Life* (1973, p.13)

A seleção natural provoca alterações previsíveis na variação fenotípica. Esta previsibilidade existe a dois níveis. Em primeiro lugar, a genética quantitativa fornece previsões da resposta a curto-prazo à seleção, tendo em conta as estimativas de hereditariedade e do diferencial de seleção (Falconer e Mackay, 1996). Em segundo lugar, a teoria da seleção fornece frequentemente previsões

aceitáveis sobre como as populações se adaptarão a longo-prazo, após uma alteração no processo de seleção. Trata-se de previsões probabilísticas e, devido a contingências históricas, as populações podem não evoluir como previsto em todos os casos. Contudo, em alguns casos, verificaram-se previsões exatas a nível fenotípico através da observação (Herre 1985, 1987). A seleção natural oferece, portanto, uma explicação convincente da evolução fenotípica da vida na Terra.

Já as alterações genéticas subjacentes a estas alterações fenotípicas não têm histórico que permita prever estes padrões. Por exemplo, há muito se reconhece que os padrões similares da variação fenotípica podem ser gerados por diferentes causas genéticas (Robertson 1959; Wilkens 1971). No entanto, as descobertas feitas nos últimos 40 anos na biologia molecular e do desenvolvimento levaram alguns biólogos a sugerir que as mutações que alteram a

regulation of gene expression are more likely to contribute to phenotypic evolution, particularly changes in morphological pattern, than mutations that alter the amino acid sequence of a protein. That is, this hypothesis claims that the genetic causes of evolution are predictable, at least at some scales of genomic organization. We use the term predictability in the sense normally implied by evolutionary genetics as probabilistic predictions.

To convince the reader that genetic evolution is predictable in at least some general sense, we point out that there is already an uncontroversial general theory of genetic evolution. Nonsynonymous mutations are predicted to contribute more to phenotypic evolution than synonymous mutations. There is, of course, a good reason for this prediction. Nonsynonymous mutations alter the amino-acid sequence and are thus likely to affect protein structure, stability, activity, or binding properties. In contrast, synonymous mutations do not alter the amino-acid sequence, although they can modify gene function through other mechanisms, such as changes in translation efficiency or mRNA stability. In addition, there is empirical evidence that nonsynonymous mutations have contributed more to phenotypic evolution than synonymous mutations. Due to the genetic code, 24% of nucleotide substitutions in protein-coding DNA are expected to cause synonymous substitutions if mutations occur randomly (Wilke 2004). To our knowledge, only two evolutionary changes in phenotype have been shown to derive from synonymous mutations (Stam and Laurie 1996; Nackley et al. 2006), whereas hundreds of evolutionary changes in phenotype have been shown to involve nonsynonymous mutations (see below). Therefore, evolutionary biologists are already familiar with the kinds of arguments and evidence that support the contention that some types of mutations contribute more to phenotypic evolution than others.

In this review we focus on whether evolutionarily relevant mutations occur preferentially in *cis*-regulatory regions. Since the 1960s various authors have wielded diverse arguments and data to predict that changes in *cis*-regulatory regions are more likely to underlie phenotypic evolution than other types of genetic changes (see for example Jacob and Monod 1961; Wallace 1963; Zuckerkandl 1963; Britten and Davidson 1969, 1971; King and Wilson 1975; Wilson 1975; Jacob 1977; Raff and Kaufman 1983; Carroll 1995; Gerhart and Kirschner 1997; Akam 1998; Stern 2000; Davidson 2001; Wray et al. 2003; Davidson 2006; Wray 2007). This idea has come under attack recently (Alonso and Wilkins 2005; Hoekstra and Coyne 2007). Hoekstra and Coyne have argued that there is no reason to expect a preponderance of evolutionarily relevant mutations in any particular gene regions (Coyne and Hoekstra 2007; Hoekstra and Coyne 2007).

We must first define how we are using the terms “*cis*-regulatory” and “coding.” The *cis*-regulatory region of a gene encompasses all of the DNA elements (enhancer, promoter, 5'UTR, 3'UTR, introns, etc.) that regulate its expression in *cis*, in other

words that act directly on the gene-coding region located on the same DNA strand, without encoding intermediary factors (Fig. 1). The coding region is the part of a gene that encodes the final gene product, either a protein or a mature RNA (Fig. 1). One can distinguish three main types of mutations: (1) coding changes, which alter the amino-acid sequence or the mature RNA nucleotide sequence; (2) *cis*-regulatory changes, which alter gene expression; and (3) genetic changes that alter both the coding and the *cis*-regulatory regions of one or several gene(s) (gene loss, gene duplication, gene rearrangement, etc.). Coding mutations always occur in coding regions and most *cis*-regulatory mutations occur in *cis*-regulatory regions. However, in rare cases, *cis*-regulatory mutations may arise in coding regions. For example, a few genes are known to contain transcription factor binding sites in exons (*keratin18* in humans and *nonA* in *Drosophila melanogaster* [Wray et al. 2003]). In this review, when we refer to *cis*-regulatory regions, we explicitly mean nucleotides that may be altered to change gene expression irrespective of their precise physical location in a gene region.

The prediction that *cis*-regulatory mutations have played a predominant role in evolution has been stated in many forms. All forms include components of two separate issues. First, most authors have generated predictions specifically for morphological variation, whereas others have considered all kinds of phenotypic changes (morphology, behavior, physiology, etc.). Second, the predominance of *cis*-regulatory mutations has been invoked either relative to coding mutations, what we call the “narrow *cis*-regulatory hypothesis,” or relative to any other type of mutation, the “broad *cis*-regulatory hypothesis.” For example, the broad *cis*-regulatory hypothesis for all phenotypes predicts that *cis*-regulatory mutations should be the predominant cause of phenotypic evolution, in contrast to coding changes, changes in alternative splice sites, gene duplication events, whole gene deletions, gene rearrangements, gene fusions, etc.

Discussions of developmental evolution have not always distinguished clearly between the effects of *cis*-regulatory and coding mutations. Indeed, another oft-mentioned hypothesis, named hereafter the “regulatory hypothesis,” is that phenotypic or morphological evolution is caused mostly by regulatory changes, that is changes in the regulation of gene expression. Regulatory changes, however, can result from mutations in *cis*-regulatory or coding regions, for example in the coding region of a transcription factor that regulates the target gene. The regulatory hypothesis focuses on mutations that alter gene regulation through any means whereas the narrow and broad *cis*-regulatory hypotheses focus on *cis*-regulatory mutations. The regulatory hypothesis predicts simply that phenotypic evolution will be, in most cases, associated with changes in gene expression. It makes no clear prediction about the molecular nature of the mutations underlying evolution. In this review, we will address whether evolution is predictable

regulação da expressão do gene têm maiores probabilidades de contribuir para a evolução fenotípica, em particular as alterações no padrão morfológico do que as mutações que alteram a sequência de aminoácidos de uma proteína. Ou seja, esta hipótese afirma que as causas genéticas da evolução são previsíveis, pelo menos a certos níveis da organização genómica. O termo previsibilidade é aqui utilizado no sentido de previsões probabilísticas, tal como aplicado na genética da evolução.

Para convencer o leitor de que a evolução genética é previsível, pelo menos num sentido geral, salientamos a existência de uma incontestável teoria geral da evolução genética. Prevê-se que as mutações não-sinónimas contribuam mais que as mutações sinónimas para a evolução fenotípica. Existe, de facto, um bom motivo para justificar esta previsão. As mutações não-sinónimas alteram a sequência de aminoácidos e, por este motivo, têm maior tendência para afetar a estrutura, estabilidade, atividade ou propriedades ligantes das proteínas. As mutações sinónimas, pelo contrário, não alteram a sequência de aminoácidos, no entanto, podem modificar a função dos genes através de outros mecanismos como, por exemplo, através de alterações na eficiência da tradução ou da estabilidade do ARNm. Além do mais, existem evidências empíricas de que as mutações não-sinónimas têm contribuído mais para a evolução fenotípica do que as mutações sinónimas. Graças ao código genético, prevê-se que 24% das substituições nucleotídicas no ADN que codifica proteínas causem substituições sinónimas, caso as mutações ocorram aleatoriamente (Wilke 2004). Tanto quanto sabemos, foi apenas possível demonstrar a existência de duas alterações evolucionárias do fenótipo com origem em mutações sinónimas (Stam e Laurie 1996; Nackley et al. 1996), mas foi possível demonstrar que centenas de alterações evolucionárias do fenótipo envolviam mutações não-sinónimas (ver abaixo). Por conseguinte, os biólogos evolutivos estão familiarizados com os tipos de argumentos e provas que apoiam a opinião de que alguns tipos de mutações contribuem mais para a evolução fenotípica que outros.

Esta revisão trata, essencialmente, de avaliar se as mutações evolutivamente relevantes ocorrem preferencialmente nas regiões de regulação *cis*. Desde a década de 1960 que vários autores apresentam argumentos e dados no sentido de prever se as alterações nas regiões da regulação *cis* têm uma maior probabilidade de estarem na base da evolução fenotípica que outros tipos de alterações genéticas (ver por exemplo Jacob e Monod 1961; Wallace 1963; Zuckerkandl 1963; Britten e Davidson 1969, 1971; King e Wilson 1975; Wilson 1975; Jacob 1977; Raff and Kaufman 1983; Carroll 1995; Gerhart e Kirschner 1997; Akam 1998; Stern 2000; Davidson 2001; Wray et al. 2003; Davidson 2006; Wray 2007). Recentemente, esta ideia tem vindo a ser contestada (Alonso e Wilkins 2005; Hoekstra e Coyne 2007). Segundo Hoekstra e Coyne, não há motivos para se esperar que haja domínio de mutações relevantes para a evolução numa região dos genes em particular (Coyne e Hoekstra 2007; Hoekstra e Coyne 2007).

Devemos começar por definir como utilizamos os termos “regulação *cis*” e “codificação”. A região de regulação *cis* de um gene abrange todos os elementos do ADN (intensificador, promotor, 5'UTR, 3'UTR, intrões, etc.) que regulam a sua expressão em *cis*, ou seja, atuam diretamente na região codificante de genes localizada na mesma cadeia de ADN, mas não codificam fatores intermediários (Fig. 1). A região de codificação é a parte do gene que codifica o produto final do gene, quer seja uma proteína ou ARN maduro (Fig.1). Existem três principais tipos de mutações: (1) alterações na codificação,

que modificam a sequência de aminoácidos ou a sequência de nucleótidos do ARN maduro; (2) alterações na regulação *cis*, que modificam a expressão do gene; e (3) alterações genéticas que modificam tanto a região de codificação como a de regulação *cis* de um ou mais genes (perda de genes, duplicação de genes, rearranjo de genes, etc.). As mutações da codificação ocorrem sempre nas regiões de codificação e a maioria das mutações da regulação *cis* ocorrem nas regiões de regulação *cis*. Existem, no entanto, casos raros nos quais as mutações da regulação *cis* poderão surgir nas regiões de codificação. Por exemplo, sabe-se que alguns genes contêm locais de ligação do fator de transcrição em exões (queratina18 nos humanos e nonA na *Drosophila melanogaster* [Wray et al. 2003]). Nesta revisão, quando nos referimos às regiões da regulação *cis*, referimo-nos especificamente aos nucleótidos que podem ser alterados para modificar a expressão do gene, independentemente da sua localização física exata na região do gene.

A previsão de que as mutações da regulação *cis* têm um papel predominante na evolução tem sido confirmada de várias formas. Todas estas formas integram componentes de duas questões individuais. Em primeiro lugar, a maioria dos autores produziu previsões específicas para a variação morfológica, enquanto outros consideraram todos os tipos de alterações fenotípicas (morfologia, comportamento, fisiologia, etc.). Em segundo lugar, o domínio das mutações de regulação *cis* tem sido evocado tanto em relação às mutações da codificação, o que designamos por “hipótese limitada da regulação *cis*,” ou em relação a qualquer outro tipo de mutação, que designamos de “hipótese geral da regulação *cis*”. Por exemplo, a hipótese geral da regulação *cis* prevê, para todos os fenótipos, que as mutações de regulação *cis* sejam a causa predominante da evolução fenotípica, ao contrário das alterações na codificação, alterações em zonas alternativas de união, duplicação de genes, supressão integral dos genes, rearranjo de genes, fusões de genes, etc.

O debate sobre a evolução do desenvolvimento nem sempre distingue de modo claro os efeitos da regulação *cis* das mutações da codificação. Na verdade, outra das hipóteses frequentemente referida, doravante designada de “hipótese da regulação”, aponta para que a evolução fenotípica ou morfológica se deva, maioritariamente, a alterações da regulação, ou seja, alterações na regulação da expressão do gene. No entanto, as alterações na regulação podem resultar de mutações na região de regulação *cis* ou de codificação, como por exemplo a região de codificação de um fator de transcrição que regule o gene alvo. A hipótese da regulação concentra-se nas mutações que alteram a regulação do gene através de qualquer meio, enquanto as hipóteses limitada e geral da regulação *cis* se focam nas mutações de regulação *cis*. A hipótese da regulação prevê a associação da maioria dos casos de evolução fenotípica às alterações na expressão do gene, mas não fornece uma previsão clara sobre a natureza molecular das mutações subjacentes à evolução. Nesta revisão, iremos abordar a questão da previsibilidade

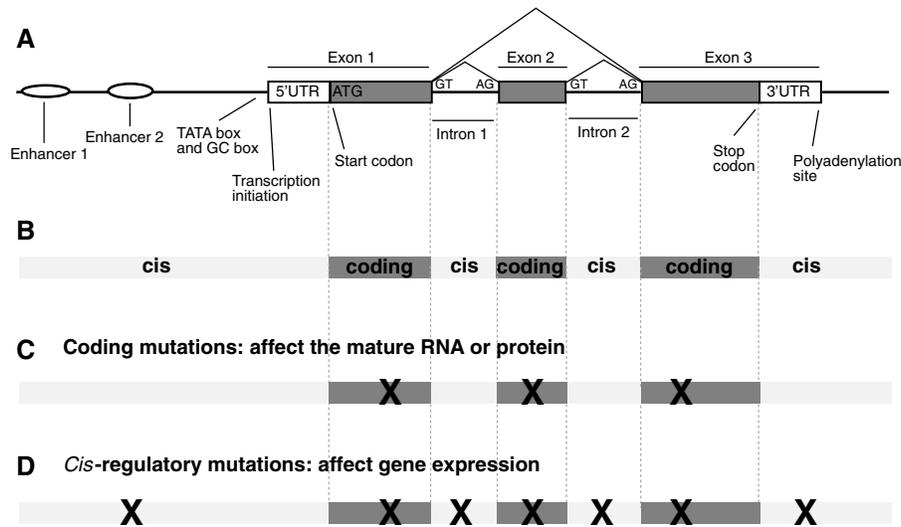


Figure 1. Gene structure and definitions of *cis*-regulatory and coding regions and *cis*-regulatory and coding mutations. (A) A single gene encodes a complex set of instructions in the DNA sequence. The final gene product can either be a protein, via an mRNA intermediate, or a mature RNA molecule itself (transfer RNA, ribosomal RNA, micro RNA, etc.). Gray boxes indicate DNA regions that encode a protein product. The mRNA molecule is transcribed from the transcription initiation site to the polyadenylation signal and introns are spliced out. Many genes encode alternative mRNA splice variants that can be generated by alternative use of different exons (Graveley 2001; Xing and Lee 2006). This is indicated in the figure by lines above the gene connecting alternative exons. Alternative splice variants are usually expressed in different tissues and at different times in development. The mechanisms regulating splicing are not fully understood, but at least some of the information is encoded in the introns and must be recognized by cell-type-specific splicing factors (Lopez 1998). The mRNA contains 5' and 3' untranslated regions (UTRs), which are involved in mRNA stability, mRNA localization, and translation. The basal transcription apparatus binds upstream of the gene-coding region, often at a TA-rich sequence motif called a TATA box. Two enhancer modules are indicated to the left of the exons. Each module can contain binding sites for multiple transcription factors. In some cases, transcription factor binding sites are not clustered into discrete modules. (B) Genes can therefore be divided into coding regions, encompassing all of the exons, and *cis*-regulatory sequences, which include all other DNA that regulates gene expression. *Cis*-regulatory sequences include sequences that regulate transcription, RNA stability and splicing, and translation. (C) We define coding mutations as mutations that alter the amino acid sequence encoded by the mRNA or that alter the nucleotide sequence of a mature RNA molecule. (D) *Cis*-regulatory mutations can occur anywhere in the gene region, including noncoding sequence and coding sequence. In rare cases, synonymous mutations in coding regions alter gene regulation in *cis*, for example through modification of transcription factor binding sites or through modification of RNA stability (see text for further details). In principle, nonsynonymous mutations could alter both the polypeptide sequence and gene regulation, but no such examples have been reported yet. The regulation of gene expression operates at multiple levels: translation, alternative splicing variants, mRNA stability, mRNA cell localization, translation, etc. (Stern 2003; Alonso and Wilkins 2005). All of these levels of gene regulation are, potentially, available for evolutionary modification (Alonso and Wilkins 2005). However, by far the majority of variation in the distribution of gene products during development is controlled at the transcriptional level (Davidson 2006).

at the genetic level. Therefore, we focus primarily on the two hypotheses that make predictions about the genetic basis of phenotypic evolution, the narrow and broad *cis*-regulatory hypotheses, together referred to as the *cis*-regulatory hypothesis.

Arguments for the *Cis*-Regulatory Hypothesis

Over the past 50 years, many different arguments have been advanced to support the predominant role of *cis*-regulatory changes in phenotypic or morphological evolution. We believe that these can be parsed into seven discrete arguments. We discuss and critique each one below.

IMPORTANCE OF GENE REGULATION IN LIFE

The origins of the *cis*-regulatory hypothesis can be traced back, ultimately, to classic experiments on gene regulation in the bacterium *Escherichia coli* (Jacob and Monod 1961). These experiments revealed that levels of enzyme activity are determined primarily by transcriptional regulation. Certain gene products, what we now call transcription factors, bind to specific nucleotides adjacent to the target coding sequence and either recruit or block recruitment of the basal transcription apparatus to the promoter. This is, in principle, a generic mechanism to control gene expression in response to external signals.

Over the past 30 years, research in developmental biology has shown that this basic mechanism in bacteria also applies to

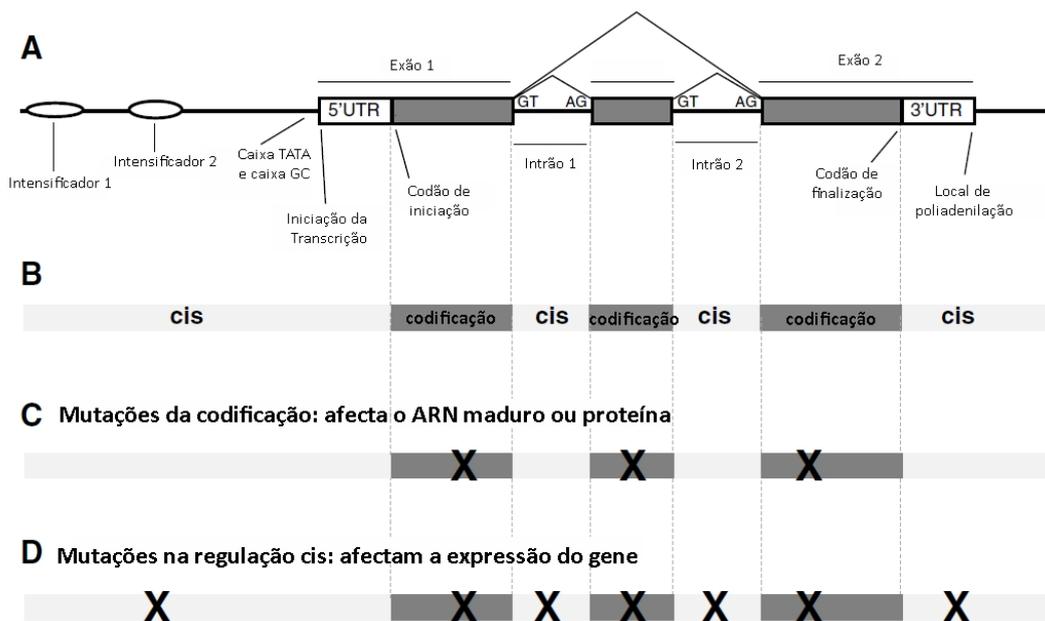


Figura 1. A estrutura do gene e as definições das regiões de regulação *cis* e de codificação e as mutações da regulação *cis* e codificação. (A) Um único gene codifica um conjunto complexo de instruções na sequência de ADN. O produto final do gene tanto pode ser uma proteína, através de um intermediário ARNm, ou a própria molécula de ARN madura (ARN de transferência, ARN ribossômico, micro ARN, etc.). As caixas a cinzento indicam as regiões de ADN que codificam um produto proteico. A molécula ARNm é transcrita do local de início de transcrição para o sinal de poliadenilação, e os intrões são separados. Vários genes codificam variantes alternativas de união do ARNm, que podem ser criadas através da utilização alternativa de diferentes exões (Graveley 2001; Xing e Lee 2006). Tal é indicado na figura sob forma de linhas acima do gene que liga exões alternativos. Os mecanismos que regulam a união não são totalmente compreendidos, mas pelo menos alguma da informação é codificada nos intrões e terá de ser reconhecida por fatores de união específicos do tipo de células (Lopez 1998). O ARNm contém as regiões 5' e 3' não traduzidas (UTR), e estas estão envolvidas na estabilidade e localização do ARNm e tradução. O aparelho basal de transcrição efetua a ligação acima da região codificante do gene, geralmente numa sequência rica em TA designada de caixa TATA. Os dois módulos intensificadores estão sinalizados à esquerda dos exões. Cada módulo pode conter locais de ligação para múltiplos fatores de transcrição. Em certos casos, os locais de ligação dos fatores de transcrição não se encontram reunidos em módulos discretos. (B) Os genes podem, portanto, ser divididos em regiões de codificação, abrangendo todos os exões e sequências da regulação *cis*, o que inclui todo o restante ADN que regula a expressão do gene. As sequências da regulação *cis* incluem sequências que regulam a transcrição, a estabilidade e união do ARN, e a tradução. (C) Podemos definir mutações da codificação como mutações que alteram a sequência de aminoácidos codificada pelo ARNm ou como mutações que alteram a sequência do nucleótido de uma molécula ARNm madura. (D) As mutações na regulação *cis* podem ocorrer em qualquer local da região dos genes, inclusive na sequência não-codificante e na sequência de codificação. Em casos raros, as mutações sinónimas em regiões de codificação alteram a regulação do gene no *cis* através, por exemplo, da modificação dos locais de ligação do fator de transcrição ou através da modificação da estabilidade do ARN (ver texto para mais detalhes). Em princípio, as mutações não-sinónimas podem alterar tanto a sequência de polipeptídeos como a regulação do gene, mas ainda não foram registados quaisquer exemplos. A regulação da expressão do gene funciona a vários níveis: tradução, variantes alternativas de união, estabilidade do ARNm, localização do ARNm nas células, tradução, etc. (Stern 2003; Alonso e Wilkins 2005). Todos estes níveis da regulação de genes têm uma potencial utilização numa modificação evolutiva (Alonso e Wilkins 2005). No entanto, a maioria da variação na distribuição dos produtos genéticos durante o desenvolvimento é controlada ao nível da transcrição (Davidson 2006).

da evolução a nível genético. Iremos assim focar essencialmente as duas hipóteses que elaboram previsões sobre a base genética da evolução fenotípica, as hipóteses limitada e geral da regulação *cis*, designadas em conjunto como hipótese da regulação *cis*.

Argumentos para a Hipótese da Regulação Cis

Nos últimos 50 anos, foram apresentados diversos argumentos que apoiam o papel predominante das alterações de regulação *cis* na evolução morfológica ou fenotípica. Consideramos que estes podem ser divididos em sete argumentos distintos, que seguidamente serão discutidos e criticados.

IMPORTÂNCIA DA REGULAÇÃO DOS GENES NA VIDA

As origens da hipótese da regulação *cis* remetem, em última análise, às clássicas experiências em regulação de genes na bactéria *Escherichia coli* (Jacob e Monod 1961). Estas experiências demonstraram que os níveis de atividade enzimática são determinados essencialmente pela regulação da transcrição. Certos produtos genéticos, que atualmente designamos de fatores de transcrição, ligam-se a nucleótidos específicos adjacentes à sequência de codificação alvo e recrutam, ou bloqueiam a recruta, do aparelho basal de transcrição pelo promotor. Este constitui, em princípio, um mecanismo genérico para controlar a expressão do gene em resposta a sinais externos.

Nos últimos 30 anos, a investigação na biologia do desenvolvimento demonstrou que este mecanismo básico em bactérias também se aplica

development of multicellular organisms (Ptashne and Gann 2002). With a few exceptions, all the cells of a multicellular organism are genetically identical, and the phenotypic differences between cells (heart muscle, neurons, hair follicles, etc.) are determined by gene regulation. Different sets of genes are turned on and off in different parts of the body at different times in development. From these facts, it became reasonable to extrapolate to the hypothesis that much of phenotypic diversity between species is caused by changes in gene expression. This was the essential argument supporting the regulatory hypothesis. However, this argument does not explicitly provide support for the *cis*-regulatory hypothesis, because regulatory changes might arise through either *cis*-regulatory or coding mutations.

CORRELATION BETWEEN PHENOTYPIC CHANGE AND CHANGE IN GENE EXPRESSION

A vast body of comparative data has revealed that changes in expression patterns of developmental patterning genes are often correlated with evolved phenotypes. The singular fact that makes these correlations compelling, beyond their sheer number, is that we understand how transcriptional changes of these genes could, based on their molecular properties, alter the phenotypes being studied. For example, Abzhanov et al. (2004) have discovered that higher levels of *Bone morphogenetic protein 4 (BMP4)* expression are correlated with deeper beak shapes among Darwin's finches. *BMP4* is a member of the transforming growth factor- β superfamily of proteins, which are ligands involved in many cell-signaling processes. *BMP4* was previously known to promote bone development in vertebrates and Abzhanov et al. (2004) showed that overexpression of *BMP4* in a chick embryo altered beak development in the predicted direction. Therefore, differential expression of *BMP4* provides a reasonable explanation for changes in finch beak shape. This is but one of many examples in which a correlation between expression of a developmental patterning gene and phenotypic variation makes sense in light of the known developmental function of the gene. However, most of these studies have not discriminated between *cis*-regulatory evolution and coding evolution as the cause of observed changes in gene expression patterns. The mutation(s) underlying the beak size difference have not been identified yet. It is possible, for example, that they affect the coding region of a transcription factor regulating the *BMP4* gene. For example, the increase in *scute* expression associated with the production of extra bristles in a Moroccan population of *D. melanogaster* has been shown to result from a coding change in a transcription factor gene regulating *scute* expression (Gibert et al. 2005). We thus consider the prevalence of evolutionary changes in gene expression patterns as good evidence for the regulatory hypothesis, but as weak evidence for the *cis*-regulatory hypothesis.

CONSERVATION OF CODING SEQUENCES ACROSS TAXA

Comparative DNA sequence data have also been used to support the *cis*-regulatory hypothesis. For example, King and Wilson (1975) argued that the 1% protein sequence divergence observed between humans and chimpanzees could not account for the many phenotypic differences between these species. Instead, they suggested, changes in the regulation of gene expression may have played a large role in phenotypic evolution and they explicitly favored *cis*-regulatory mutations as the cause of regulatory evolution.

Nucleotide substitutions in a promoter or operator gene would affect the production, but not the amino acid sequence, of proteins in that operon. Nucleotide substitutions in a structural gene coding for a regulatory protein such as a repressor, hormone or receptor protein, could bring about amino acid substitutions, altering the regulatory properties of the protein. However, we suspect that only a minor fraction of the substitutions which accumulate in regulatory proteins would be likely to alter their regulatory properties. (King and Wilson 1975, p. 114)

Many of the genes encoding transcription factors and signaling molecules display sequence conservation across vast phylogenetic distances. This has lent further support to the idea that *cis*-regulatory changes might be more common than coding changes.

Although gene-coding regions are usually remarkably conserved across taxa as divergent as humans and worms, the total number of coding changes between closely related species is not negligible. There are, for example, about 60,000 nonsynonymous differences in gene-coding regions between humans and chimpanzees (Eyre-Walker 2006). Although it is difficult to estimate the number of phenotypic differences between species, in strict numerical terms these numerous coding changes may be sufficient to explain most phenotypic variation. Thus, by itself, this argument does not offer compelling support for the *cis*-regulatory hypothesis.

DIFFERENT CONSTRAINTS ON CIS-REGULATORY AND CODING REGIONS

Most genes are expressed in multiple tissues at multiple times in development (Tomancak et al. 2002). These complex expression patterns are generated primarily by binding of the transcription factors expressed and active in each cell at a particular time—the regulatory state—to specific sites in the *cis*-regulatory regions of many genes (Wray et al. 2003; Davidson 2006). Often, independent transcription factor binding sites are clustered in regions of several hundreds of base pairs and together they encode a particular transcriptional output.

Because information is encoded differently in *cis*-regulatory and coding regions, these regions may evolve at different rates. First, the redundancy of the genetic code causes about 24% of

ao desenvolvimento de organismos multicelulares (Ptashne e Gann 2002). Salvo algumas exceções, todas as células de um organismo multicelular são geneticamente idênticas, e as diferenças fenotípicas entre as células (músculo cardíaco, neurónios, folículos capilares, etc.) são determinadas pela regulação dos genes. Diferentes conjuntos de genes são ativados e desativados em diferentes partes do corpo e em diferentes fases do desenvolvimento. A partir destes factos, tornou-se razoável extrapolar a hipótese de que grande parte da diversidade fenotípica entre espécies é provocada por alterações na expressão do gene. Este argumento foi essencial na defesa da hipótese da regulação. No entanto, este argumento não proporciona uma defesa clara para a hipótese da regulação *cis*, porque as alterações da regulação tanto podem surgir através da regulação *cis* como de mutações da codificação.

CORRELAÇÃO ENTRE ALTERAÇÃO FENOTÍPICA E ALTERAÇÃO NA EXPRESSÃO DO GENE

Um vasto conjunto de dados comparativos demonstrou que as alterações nos padrões de expressão de genes padronizantes do desenvolvimento estão geralmente correlacionadas com fenótipos desenvolvidos. Estas correlações são convincentes não só devido ao seu elevado número mas também pelo facto de compreendermos como é que as alterações de transcrição destes genes podem alterar os fenótipos estudados, com base nas suas propriedades moleculares. Exemplificando, Abzhanov et al. (2004) descobriram que os níveis elevados de expressão da *proteína morfogenética do osso 4* (PMO4) estão correlacionados com os diferentes tamanhos de bicos dos tentilhões de Darwin. A PMO4 é um membro da superfamília de proteínas do factor- β de transformação de crescimento, que são ligandos envolvidos em vários processos de sinalização celular. A PMO4 era previamente conhecida por promover o desenvolvimento ósseo em vertebrados, e Abzhanov et al. (2004) demonstraram que a superexpressão do PMO4 no embrião de pintainho altera o desenvolvimento do bico na direção prevista. Por conseguinte, o diferencial de expressão da PMO4 fornece uma explicação aceitável para as alterações na forma do bico do tentilhão. Trata-se de apenas um entre vários exemplos nos quais a correlação entre a expressão do gene padronizante do desenvolvimento e a variação fenotípica faz sentido, face à conhecida função de desenvolvimento do gene. No entanto, a maior parte destes estudos não fazem distinção entre a evolução na regulação *cis* e na codificação, como causa das alterações observadas nos padrões de expressão do gene. A(s) mutação(ões) subjacente(s) às diferenças de tamanho do bico não foi(foram) ainda identificada(s). É possível, por exemplo, que afetem a região codificante de um fator de transcrição que regule o gene da PMO4. A título de exemplo, ficou demonstrado que o aumento da expressão do *scute* associado à produção de pelos adicionais numa população marroquina de *D. melanogaster*, resulta de uma alteração na codificação do fator de transcrição que regula a expressão do *scute* (Gibert et al. 2005). Assim, consideramos que a prevalência das alterações evolucionárias nos padrões de expressão do gene é uma prova firme da hipótese da regulação, mas não da hipótese da regulação *cis*.

CONSERVAÇÃO DE SEQUÊNCIAS DE CODIFICAÇÃO ATRAVÉS DA TAXA.

Foram também utilizados dados de sequências de ADN comparativas para apoiar a hipótese da regulação *cis*. Por exemplo, King e Wilson (1975) argumentam que 1% das divergências observadas na sequência de proteínas (entre humanos e chimpanzés) não justifica as várias diferenças fenotípicas entre estas espécies. Sugere-se então que as alterações na regulação da expressão do gene podem ter tido um papel importante na evolução fenotípica e deram clara preferência às mutações na regulação *cis* como causa da evolução na regulação.

As substituições de nucleótidos num promotor ou operão afetaria a produção de proteínas nesse mesmo operão, mas não afetaria a sequência de aminoácidos. As substituições de nucleótidos de um gene estrutural a codificar uma proteína reguladora, como por exemplo um repressor, uma hormona ou proteína recetora, pode gerar substituições de aminoácidos, alterando assim as propriedades de regulação dessa proteína. No entanto, acreditamos que apenas uma pequena fração das substituições que se acumulam nas proteínas reguladoras seriam suscetíveis de alterar as propriedades de regulação. (King and Wilson 1975, p.114)

Muitos dos genes que codificam fatores de transcrição e moléculas de sinalização apresentam uma conservação da sequência através de vastas distâncias filogenéticas, o que apoia a ideia de que as alterações na regulação *cis* poderão ser mais comuns que as alterações na codificação.

Apesar das regiões codificantes dos genes se manterem notavelmente conservadas ao longo de taxa tão divergente como humanos e vermes, o número total de alterações na codificação entre espécies de relação próxima é significativo. Existem, por exemplo, cerca de 60 000 diferenças não-sinónimas entre regiões codificantes dos genes de humanos e chimpanzés (Eyre-Walker 2006). Apesar da dificuldade em calcular o número de diferenças fenotípicas entre espécies, estas numerosas alterações na codificação podem, em termos estritamente numéricos, ser suficientes para explicar a maioria da variação fenotípica. Deste modo, este argumento não fornece por si só uma defesa convincente da hipótese de regulação *cis*.

DIFERENTES CONSTRANGIMENTOS NAS REGIÕES DE REGULAÇÃO CIS E CODIFICAÇÃO

A maioria dos genes expressam-se em diversos tecidos, em diversas fases do desenvolvimento (Tomancak et al. 2002). Estes complexos padrões de expressão são produzidos principalmente pela ligação de fatores de transcrição expressos e ativos em cada célula numa determinada fase (o estado de regulação) em locais específicos nas regiões de regulação *cis* de vários genes (Wray et al. 2003; Davidson 2006). Os locais de ligação do fator de transcrição independentes são, muitas vezes, conjuntos de várias centenas de pares de bases, e juntos codificam um determinado output de transcrição. Devido à informação se codificar de modo diferente na região de regulação *cis* e na de codificação, estas regiões podem evoluir a diferentes níveis. Primeiro, a redundância do código genético faz com que cerca de 24% das

Table 1. Haploid genome sizes and the proportion of coding and noncoding regions for various eukaryotes (modified from tables 3.2 and 3.3 of Lynch 2007).

	Approximate haploid genome size (in Mb)	Proportion coding	Proportion noncoding	Estimated proportion of noncoding DNA that is regulatory ¹
<i>Saccharomyces cerevisiae</i>	12	74.2	25.8	22
<i>Aspergillus nidulans</i>	30	45.9	54.1	2
<i>Plasmodium falciparum</i>	23	52.8	47.2	2
<i>Caenorhabditis elegans</i>	100	26.4	73.6	12
<i>Drosophila melanogaster</i>	137	19.4	80.6	20
<i>Mus musculus</i>	2500	1.4	98.6	2
<i>Homo sapiens</i>	2900	1.4	98.6	2

¹The amount of regulatory DNA was estimated from islands of conserved DNA sequence between closely related species. See Lynch (2007) for details.

mutations in coding regions to be synonymous changes (Wilke 2004). Similarly, a single transcription factor can bind to multiple similar DNA sequences, rather than to only a single specific sequence. For example, the transcription factor SRY, which is involved in sex determination, binds to the DNA sequence WWCAAW, where W can be either T or A (Harley et al. 1994). It is not yet known if this “transcriptional code” is more or less redundant than the genetic code.

Second, most insertions or deletions in protein-coding regions (those that are not multiples of 3 bp) will disrupt the reading frame and cause a premature stop codon or alteration of amino acids in the protein. These changes are likely to have deleterious consequences. In contrast, insertions or deletions in *cis*-regulatory regions are less likely to alter *cis*-regulatory function because clusters of transcription factor binding sites can be separated by nonfunctional DNA of heterogeneous length.

Third, new transcription factor binding sites may evolve more easily than new coding regions. The rate of point mutations is sufficiently high to rapidly generate new transcription factor binding sites (Stone and Wray 2001). For instance, the human genome contains abundant polymorphism in transcription factor binding sites that lead to both loss and gain of expression (Rockman and Wray 2002).

Furthermore, in most cases, the precise arrangement of transcription factor binding sites relative to each other is not critical to function. For example, the *cis*-regulatory region that drives *even-skipped* expression in stripe 2 of the *Drosophila* embryo has gained and lost transcription factor binding sites and retained the same function (Ludwig et al. 1998, 2000). In addition, among different fly species, the *cis*-regulatory region of *hunchback* contains a similar number of binding sites for the transcription factor Bicoid, but the precise number, orientation, and location of the binding sites differs among species (McGregor et al. 2001).

These features cause *cis*-regulatory DNA sequences to evolve faster than coding DNA sequences. However, it is not clear that

these rapid changes in *cis*-regulatory regions have produced more phenotypic evolution than coding mutations. It is possible that a large number of the mutations occurring in *cis*-regulatory regions do not alter the phenotype and fitness. There is currently insufficient data to determine whether the higher flexibility in *cis*-regulatory sequence evolution biases the number of mutations causing phenotypic evolution toward *cis*-regulatory regions.

MUTATIONAL TARGET SIZE

The probability of fixation of a new mutation depends on the fitness effect associated with the mutation and the effective and demographic population sizes (Kimura 1962; Bürger and Ewens 1995). In addition, the probability that a mutation with a particular phenotypic effect arises in the first place is a function of the site-specific mutation rate and the mutational target size. If *cis*-regulatory regions encompass a larger mutational target size than coding regions, then we might expect more evolutionarily relevant mutations to accumulate in *cis*-regulatory regions than in coding regions. It is currently difficult to know whether *cis*-regulatory regions in fact have a larger mutational target size than coding regions.

Eukaryotic genomes are divided unevenly between noncoding and coding DNA (Table 1). In some eukaryotes, such as the yeast *Saccharomyces cerevisiae* and the malaria parasite *Plasmodium falciparum*, genomes contain more coding than noncoding DNA. In most eukaryotes, however, noncoding DNA is present in large excess relative to coding DNA. In *D. melanogaster*, about 80% of the genome is noncoding and in humans and mice more than 98% of the genome is noncoding. However, in humans more than 90% of this noncoding DNA is occasionally transcribed into RNA although the functional roles, if any, of most of this RNA are currently poorly understood (Mattick 2003; Birney et al. 2007).

An unknown fraction of the noncoding DNA encodes a *cis*-regulatory function. Estimates from first principles imply that in an organism like *D. melanogaster* only about 3% of the

Tabela 1 Dimensões do genoma haploide e proporção de regiões de codificação e não-codificação para vários eucariotas (alterado a partir das tabelas 3.2. e 3.3 de Lynch 2007).

	Tamanho aproximado (em Mb) do genoma haploide	Proporção da codificação	Proporção da não-codificação	Proporção estimada de ADN não codificante que é reguladora ¹
<i>Saccharomyces cerevisiae</i>	12	74,2	25,8	22
<i>Aspergillus nidulans</i>	30	45,9	54,1	2
<i>Plasmodium falciparum</i>	23	52,8	47,2	2
<i>Caenorhabditis elegans</i>	100	26,4	73,6	12
<i>Drosophila melanogaster</i>	137	19,4	80,6	20
<i>Mus musculus</i>	2500	1,4	98,6	2
<i>Homo sapiens</i>	2900	1,4	98,6	2

¹A quantidade de ADN regulador foi calculada a partir de ilhas de sequências de ADN conservadas entre espécies com relação próxima. Ver Lynch (2007) para mais detalhes.

mutações nas regiões de codificação sejam alterações sinónimas (Wilke 2004). Da mesma forma, um único fator de transcrição pode ligar-se a várias sequências de ADN similares, em vez de se ligar a uma única sequência específica. Por exemplo, o fator de transcrição SRY, envolvido na determinação do sexo, liga-se à sequência de ADN WWCAAW, no qual W pode ser T ou A (Harley et al. 1994). Ainda não se sabe se este “código de transcrição” é mais ou menos redundante que o código genético.

Em segundo lugar, a maioria das inserções e deleções nas regiões de codificação de proteínas (as que não são múltiplas de 3 bp) irão quebrar a grelha de leitura e provocar uma paragem prematura do codão de finalização ou alterar os aminoácidos na proteína. Estas alterações podem vir a ter consequências deletérias. Já as inserções e deleções nas regiões de regulação *cis* são, pelo contrário, menos suscetíveis de modificar a função de regulação *cis* porque o conjunto de locais de ligação do fator de transcrição pode ser separado por ADN não-funcional de comprimento heterogéneo.

Em terceiro lugar, os novos locais de ligação de fatores de transcrição podem evoluir mais facilmente do que as novas regiões de codificação. A variação de mutações pontuais é suficientemente alta para gerar rapidamente novos locais de ligação do fator de transcrição (Stone e Wray 2001). Por exemplo, o genoma humano contém polimorfismo abundante em locais de ligação do fator de transcrição que conduzem tanto à perda como ao ganho de expressão (Rockman e Wray 2002).

Além disso, na maioria dos casos o arranjo exato dos locais de ligação do fator de transcrição relativos uns aos outros não é crucial para o seu funcionamento. Por exemplo, a região de regulação *cis* que transporta a expressão do gene *even-skipped* na banda 2 do embrião de *Drosophila* ganhou e perdeu locais de ligação do fator de transcrição e manteve o seu funcionamento (Ludwig et al. 1998, 2000). Além disso, a região de regulação *cis* da corcunda em diferentes espécies de moscas contém um número similar de locais de ligação para o fator de transcrição Bicoid, mas o número exato, a orientação e localização dos locais de ligação difere entre espécies (McGregor et al. 2001).

Estas características fazem com que as sequências de ADN de regulação *cis* evoluam mais rapidamente do que as sequências de ADN da codificação. No entanto, não se sabe ao certo se estas rápidas alterações nas regiões da

regulação *cis* produziram uma maior evolução fenotípica que as mutação da codificação. É possível que um grande número de mutações a ocorrerem nas regiões de regulação *cis* não alterem o fenótipo e o fitness: Atualmente não existem dados suficientes para determinar se a elevada flexibilidade na evolução da sequência da regulação *cis* influencia o número de mutações que provocam a evolução fenotípica para com as regiões de regulação *cis*.

DIMENSÃO PREVISTA DA MUTAÇÃO

A probabilidade de fixação de uma nova mutação depende da capacidade de *fitness* associada à mutação e às dimensões efetivas e demográficas das populações (Kimura 1962; Bürger e Ewens 1995). Além disso, a probabilidade de surgir uma mutação com um efeito fenotípico particular é função da taxa de mutação de um local específico e da dimensão prevista da mutação. Se as regiões de regulação *cis* abrangem uma dimensão prevista da mutação superior às regiões de codificação, prevê-se que se acumulem mais mutações evolutivas relevantes em regiões de regulação *cis* do que em regiões de codificação. Atualmente, é difícil prever se as regiões de regulação *cis* têm de facto uma dimensão prevista de mutação superior às regiões de codificação.

Os genomas eucarióticos estão divididos de forma desigual entre o ADN codificante e não-codificante (Tabela 1). Os genomas de alguns eucariotas, como a levedura *Saccharomyces cerevisiae* e o parasita da malária *Plasmodium falciparum*, contêm mais ADN codificante do que não-codificante. No entanto, na maioria dos eucariotas o ADN não codificante encontra-se presente em maior excesso do que no ADN codificante. Na *D. melanogaster*, cerca de 80% do genoma é não-codificante, e em humanos e ratos mais de 98% do genoma é não-codificante. No entanto, nos humanos, mais de 90% deste ADN não-codificante é por vezes transcrito em ARN apesar das funções funcionais, se aplicáveis, de maior parte deste ARN serem, atualmente, de compreensão muito limitada (Mattick 2003; Birney et al. 2007).

Uma fração desconhecida do ADN não-codificante codifica uma função de regulação *cis*. As estimativas dos princípios iniciais sugerem que, num organismo como a *D. melanogaster*, apenas cerca de 3% dos

genome encodes *cis*-regulatory function (Alonso and Wilkins 2005), whereas 19% is coding (Lynch 2007). Estimates of *cis*-regulatory DNA based on islands of sequence conservation among closely related species span from 1% to 22% of the noncoding DNA, depending on species (Table 1). Estimates based on evolutionary comparisons of rates of sequence evolution between closely related *Drosophila* species reveal that 40–70% of nucleotides in noncoding regions are more conserved than synonymous sites, suggesting that they are under functional constraint (Andolfatto 2005). However, regions can be conserved for reasons other than *cis*-regulatory function. They could participate in other functions such as DNA structure, chromosome replication, or they might also encode functional RNA transcripts. Given the wide range of current estimates and the many sources of variation, it is premature to generate confident estimates of the fraction of eukaryotic genomes that contains *cis*-regulatory information.

Not all mutations, whether in coding or *cis*-regulatory regions, will generate a functional change at the molecular level (a change in protein or RNA or a change in gene expression, respectively). For a random coding DNA sequence, the percentage of nonsynonymous substitutions equals 76% if all substitutions are equally likely (Wilke 2004). Functional changes in *cis*-regulatory regions can result from complete loss or gain of binding sites or from quantitative changes in transcription factor binding activity. Our current limited understanding of *cis*-regulatory regions prevents an accurate estimate of the percentage of random substitutions that are likely to alter gene expression.

Finally, only a fraction of these functional changes will cause a change in morphology, metabolism, physiology, or behavior (Kacser and Burns 1981). It is not yet clear whether the larger noncoding DNA regions of many eukaryotic genomes has led to a larger role of *cis*-regulatory mutations in phenotypic variation. Most noncoding DNA may not have a *cis*-regulatory function, and most *cis*-regulatory mutations may not have any phenotypic effects.

A POPULATION GENETICS ARGUMENT

Another argument supporting the *cis*-regulatory hypothesis derives from a detailed understanding of molecular and developmental mechanisms and rests, ultimately, on population genetics reasoning. This argument is that natural selection favors *cis*-regulatory mutations because they may have fewer pleiotropic effects than coding mutations. This argument has two core assumptions. The first assumption is that natural selection should favor mutations with fewer pleiotropic effects over those with more pleiotropic effects. The second assumption is that mutations in *cis*-regulatory regions should have relatively specific, less pleiotropic, effects than mutations in coding regions. We first explore both assumptions, which leads to support for this line of reasoning. Then, we critique this argument.

The theoretical consequences of pleiotropy

Does pleiotropy impact evolution? The consequences of pleiotropy for evolution have been explored within several theoretical frameworks: Fisher's geometric model of adaptation (Orr 1998, 2000; Welch and Waxman 2003); a framework in which the total selection acting on an allele is the sum of the positive direct effects and positive or negative pleiotropic effects (Hill and Keightley 1988; Barton 1990; Otto 2004); and a framework in which a mutation can influence one trait under directional selection and a second trait under stabilizing selection, so called "hidden pleiotropy" (Batz and Wagner 1997). Despite this variety of approaches, all models agree on several results. First, as the degree of pleiotropy increases, the rate of adaptive evolution decreases dramatically. Orr (2000) estimates that this reduction in the rate of adaptation scales as n^{-1} , where n equals the number of phenotypic dimensions along which mutations can move the phenotype. Orr (2000) calls this the "cost of complexity," because universal pleiotropy in complex organisms should greatly reduce the rate of adaptation. Welch and Waxman (2003) show that this result is robust to a wide range of parameter values and to alterations to the model. They show that the reduction in the rate of adaptation is probably even faster than n^{-1} . In other words, mutations that move the phenotype along only one dimension should contribute to adaptation much more often than mutations that move the phenotype along two or more dimensions simultaneously, that is mutations that cause pleiotropic effects. The other modeling approaches reinforce this conclusion. Otto (2004) finds, for example, that under weak selection, pleiotropy reduces the total selection coefficient on an adaptive allele by half on average. She also finds that the probability of fixation for a mutation that improves a trait, but causes pleiotropic effects, is proportional to the square of the selection coefficient, s^2 , where $0 < s < 1$. In contrast, the probability of fixation for an advantageous mutation without pleiotropic effects is approximately $2s$ (Haldane 1927; Fisher 1930; Wright 1931; Kimura 1962). Thus, strong pleiotropy will prevent fixation of most potentially adaptive mutations. Similarly, "hidden pleiotropy" dramatically reduces the rate of adaptation (Batz and Wagner 1997). Thus, pleiotropy does not stop adaptation in its tracks, it just slows it, to a surprisingly large extent.

In real species—where one can get the distinct impression that natural selection has molded dozens, hundreds or thousands of discrete subtle adaptations for each species—the cost of complexity would appear to present a serious hurdle to adaptation. The models seem to imply that complex organisms with universal pleiotropy will have enormous difficulty adapting to novel environments. But complex organisms can evolve quickly both in the laboratory and in the field (Endler 1986). Either the models are wrong or the assumptions are wrong or both (Welch and Waxman 2003).

genomas codificam a função de regulação *cis* (Alonso e Wilkins 2005) enquanto que 19% é de codificação (Lynch 2007). As estimativas do ADN de regulação *cis* baseadas em ilhas de conservação de sequência entre espécies com relação próxima vão desde 1% a 22% do ADN não-codificante, dependendo das espécies (Tabela 1). As estimativas baseadas em comparações evolutivas de taxas de evolução de sequências entre espécies de *Drosophila* com relação próxima, revelam que 40 a 70 % dos nucleótidos em regiões não-codificantes são mais conservados do que os locais sinónimos, o que sugere estarem sob regulação funcional (Andolfatto 2005). No entanto, as regiões podem ser conservadas por outros motivos além da função de regulação *cis*. Podem participar noutras funções, como a estrutura do ADN, a replicação de cromossomas, ou podem também codificar cópias de ARN funcionais. Tendo em conta a vasta gama de estimativas atuais e as várias fontes de variação, é prematuro gerar estimativas seguras sobre a fração de genomas eucarióticos que contêm informação da regulação *cis*.

Nem todas as mutações, quer nas regiões de codificação ou da regulação *cis*, irão gerar uma alteração funcional a nível molecular (uma alteração na proteína ou no ARN, ou uma alteração na expressão do gene, respetivamente). Numa sequência aleatória de ADN codificante, a percentagem de substituições não-sinónimas equivale a 76%, caso todas as substituições sejam igualmente prováveis (Wilke 2004). As alterações funcionais nas regiões de regulação *cis* podem ser o resultado da perda total ou ganho de locais de ligação, ou podem resultar de alterações quantitativas na atividade de ligação do fator de transcrição. O nosso conhecimento atualmente limitado sobre regiões de regulação *cis* impossibilita uma estimativa precisa da percentagem de substituições aleatórias suscetíveis de alterar a expressão do gene.

Por último, apenas uma fração destas alterações funcionais provocará uma alteração na morfologia, metabolismo, fisiologia ou comportamento (Kacser e Burns 1981). Ainda não é claro se as maiores regiões de ADN não-codificante de vários genomas eucarióticos fizeram com que as mutações na regulação *cis* tivessem uma maior função na variação fenotípica. A maioria do ADN não-codificante pode não ter uma função de regulação *cis*, e a maioria das mutações na regulação *cis* podem não ter quaisquer efeitos fenotípicos.

O ARGUMENTO DA GENÉTICA DE POPULAÇÕES

Um outro argumento que apoia a hipótese da regulação *cis* deriva da compreensão detalhada dos mecanismos moleculares e de desenvolvimento e, em última análise, serve-se da argumentação da genética de populações. Estes argumentos indicam que a seleção natural favorece as mutações na regulação *cis* porque podem ter menos efeitos pleiotrópicos do que as mutações na codificação. Este argumento tem duas suposições base. A primeira suposição indica que a seleção natural deve favorecer as mutações com menos efeitos pleiotrópicos sobre as que têm mais efeitos pleiotrópicos. A segunda suposição indica que as mutações nas regiões de regulação *cis* devem ter efeitos relativamente específicos, menos pleiotrópicos, do que as mutações nas regiões de codificação. Começamos por explorar ambas as

suposições, de modo a fornecer um suporte para esta argumentação. A argumentação será, seguidamente, criticada.

As consequências teóricas da pleiotropia.

A pleiotropia tem impacto na evolução? As consequências da pleiotropia para a evolução têm sido exploradas em várias estruturas teóricas: o modelo geométrico de adaptação de Fisher (Orr 1998, 200; Welch e Waxman 2003), uma estrutura na qual toda a seleção que atua num alelo é a soma dos efeitos diretos positivos e efeitos pleiotrópicos positivos ou negativos (Hill e Keightley 1988; Barton 1990; Otto 2004); e também uma estrutura na qual a mutação pode influenciar uma característica sob uma seleção direcional e uma segunda sob uma seleção estabilizadora, designada de “pleiotropia escondida” (Baatz e Wagner 1997). Apesar da variedade de abordagens, todos os modelos estão de acordo em vários resultados. Primeiro, à medida que o nível de pleiotropia aumenta, a taxa de evolução adaptativa diminui drasticamente. Orr (2000) estima que esta redução na taxa de adaptação ajusta-se em n^{-1} , no qual n refere-se ao número de dimensões fenotípicas, nas quais as mutações podem mover o fenótipo. Orr (2000) designa isto de “custo da complexidade”, pois a pleiotropia universal em organismos complexos deve reduzir bastante a taxa de adaptação. Welch e Waxman (2003) demonstram tratar-se de um resultado robusto para uma vasta gama de valores de parâmetro e para alterações ao modelo. Demonstram ainda que a redução da taxa de adaptação é provavelmente ainda mais rápida que n^{-1} . Por outras palavras, as mutações que movem o fenótipo ao longo de uma só dimensão contribuem muito mais frequentemente para a adaptação do que as mutações que movem o fenótipo ao longo de duas ou mais dimensões em simultâneo, ou seja, mutações que provocam efeitos pleiotrópicos. As abordagens dos outros modelos reforçam esta conclusão. Otto (2004) conclui, por exemplo, que, sob seleção fraca, a pleiotropia reduz, em média, o coeficiente de seleção total num alelo adaptativo em metade. Conclui também que a probabilidade de fixação para uma mutação que melhora uma característica mas provoca efeitos pleiotrópicos, é proporcional ao quadrado do coeficiente de seleção s^2 , no qual $0 < s < 1$. Pelo contrário, a probabilidade de fixação para uma mutação vantajosa sem efeitos pleiotrópicos é de aproximadamente $2s$ (Haldane 1927; Fisher 1930; Wright 1931; Kimura 1962). Deste modo, uma forte pleiotropia irá prevenir a fixação das mutações com maior potencial adaptativo. A “pleiotropia escondida” também reduz a taxa de adaptação do mesmo modo (Baatz e Wagner 1997). A pleiotropia não para, portanto, a adaptação, apenas a atrasa, mas a um nível surpreendentemente elevado.

Em espécies reais, nas quais se pode ter a nítida impressão de que a seleção natural moldou dezenas, centenas ou milhares de subtis adaptações discretas em cada espécie, o custo da complexidade parece ser um sério obstáculo à adaptação. Estes modelos aparentam significar que os organismos complexos com pleiotropia universal terão enormes dificuldades na adaptação a novos ambientes. Mas os organismos complexos podem evoluir rapidamente tanto em laboratório como no terreno (Endler 1986). Ou os modelos estão errados, ou as suposições estão erradas, ou então estão ambos (Welch e Waxman 2003).

Given the diverse modeling approaches that have been taken, it seems unlikely that the general results are flawed. It seems more likely that the assumptions are flawed. Because organisms are, rather obviously, complex, it is unlikely that there is a problem with the assumption that organisms must adapt along multiple phenotypic axes simultaneously.

Welch and Waxman (2003) explored one possible way out of this paradox. They recognized the prevalence of developmental modularity in multicellular organisms (Raff and Kaufman 1983; Gerhart and Kirschner 1997) and asked whether modularity might eliminate the cost of complexity imposed by pleiotropy. They divided development into the activity of relatively independent modules. For example, legs develop largely independently from arms and heads. However, modularity at the level of organ development does not eliminate the cost of complexity (Welch and Waxman 2003). This is partly because the effect of pleiotropy on the rate of adaptation is so strong that it kicks in with even modest levels of pleiotropy.

Another possible escape from the cost of complexity is offered by results from several of these models. Strong selection on the main target of selection partially ameliorates the effect of deleterious pleiotropic effects on the rate of adaptation (Batz and Wagner 1997; Otto 2004). Under strong selection, the response to selection more nearly reflects the sum of the direct and pleiotropic fitness effects. This makes intuitive sense. If a mutation arises in a cow that causes a phenotypic effect cherished by a farmer, this trait can be selected even in the face of severe pleiotropic fitness effects. However, strong selection appears to be rare in natural populations (Hoekstra et al. 2001; Kingsolver et al. 2001; Andolfatto 2007), suggesting that the assumption of relatively weak selection is probably realistic.

It seems most likely that there is a problem with the assumption of universal pleiotropy. How common is pleiotropy? Is it universal?

Gene pleiotropy

Gene pleiotropy, the effect of an allelic variant on multiple different phenotypes rather than just one, has historically been considered universal or at least very common. J. B. S. Haldane (1932, p. 62–63) wrote “Since the gene exists in every cell of the body, it may be expected to affect the organism as a whole, even if its most striking effect is on some particular organ or function.” Sewall Wright (1968, p. 61) wrote “The available evidence indicates that pleiotropy is virtually universal.” This idea has been repeated throughout the evolutionary literature and underlies all models of the effects of pleiotropy on evolution.

Although these sentiments were based largely on observations of the phenotypic effects of individual mutations in domesticated races and in laboratory strains, recent studies have confirmed the existence of widespread pleiotropy throughout the

genome. For example, examination of 501 morphological phenotypes in each of 4710 deletion mutants of nonessential genes in yeast found that about 35% of deletion mutants affected at least two morphological characters (Ohya et al. 2005). A second study assayed relative fitness of these deletion mutants in 21 adverse growth conditions versus a control medium (Dudley et al. 2005). About 58% of genes displayed growth reduction in two or more conditions, which the authors considered to reflect pleiotropic effects in different growth conditions. These two studies have explored a limited number of growth conditions and aspects of morphology. In addition, yeast possess far fewer phenotypic features than multicellular eukaryotes. Therefore, deletion of genes in more complex eukaryotes is likely to cause at least as much pleiotropy as observed in yeast. Although gene pleiotropy may be relatively high in yeasts because they have relatively few genes, it seems likely that the classical view of widespread gene pleiotropy is accurate.

Further analysis of these datasets shows that gene pleiotropy is correlated with the number of biological processes and the number of protein–protein interactions that gene products participate in (He and Zhang 2006). Pleiotropy is not correlated with the number of molecular functions or the number of protein domains per gene. That is, pleiotropy results from the participation of a gene product in multiple cellular processes in which it performs the same molecular function. In multicellular organisms, therefore, considerable pleiotropy is expected to result from expression of genes in multiple tissues during development.

In the growth assays, elimination of genes displaying more pleiotropy reduced fitness to a greater degree than elimination of genes displaying less pleiotropy (Cooper et al. 2007). These results fit comfortably within the intuition embodied in Fisher’s geometric model of adaptation (Fisher 1930). In this model, universal gene pleiotropy causes most mutations to move a population away from the fitness optimum.

Finally, the extent of gene pleiotropy discovered in the growth assays is correlated with the level of protein sequence conservation (He and Zhang 2006). This agrees with theoretical expectations that pleiotropy should constrain protein sequences (Waxman and Peck 1998). All together, these observations support the hypothesis that gene pleiotropy constrains protein evolution because gene products participate in multiple cellular functions.

Gene pleiotropy seems, therefore, to be extremely common. How, then, do complex organisms evolve specific adaptations apparently so readily?

Pleiotropic genes versus pleiotropic mutations

It is a striking fact that none of the population genetic models exploring the effects of pleiotropy have incorporated realistic models of gene structure and function. All have assumed that gene pleiotropy is universal and, more importantly, that universal

Tendo em conta as diversas abordagens de modelação, parece pouco provável que os resultados gerais contenham erros. Será mais provável que as suposições estejam erradas, porque como os organismos são, como é óbvio, complexos, é pouco provável existir um problema com o pressuposto de que os organismos têm que se adaptar ao longo de vários eixos fenotípicos em simultâneo.

Welch e Waxman (2003) exploraram uma possível solução para este paradoxo. Reconheceram a prevalência da modularidade do desenvolvimento em organismos multicelulares (Raff e Kaufman 1983; Gerhart e Kirschner 1997) e questionaram se a modularidade pode eliminar o custo da complexidade imposto pela pleiotropia. Dividiram o desenvolvimento na atividade de módulos relativamente independentes. Por exemplo, as pernas desenvolvem-se de forma maioritariamente independente dos braços e da cabeça. No entanto, a modularidade ao nível do desenvolvimento dos órgãos não elimina o custo da complexidade (Welch e Waxman 2003). Isto deve-se em parte ao facto do efeito da pleiotropia na taxa de adaptação ser tão forte ao ponto de ser acionado mesmo em níveis mais baixos de pleiotropia.

Uma outra solução para o custo da complexidade é dada pelos resultados de vários destes modelos. A forte seleção no alvo principal da seleção melhora parcialmente o efeito dos efeitos pleiotrópicos deletérios na taxa de adaptação (Baatz and Wagner 1997; Otto 2004). Sob forte seleção, a resposta à seleção reflete, mais particularmente, a soma dos efeitos de fitness diretos e pleiotrópicos. Tudo isto faz sentido, de forma intuitiva. Se surgir uma mutação numa vaca que provoque efeitos fenotípicos valorizados pelo seu agricultor, poderá selecionar esta característica, mesmo que enfrente graves efeitos pleiotrópicos do fitness. No entanto, a seleção forte parece ser rara em populações naturais (Hoekstra et al. 2001; Kingsolver et al. 2001; Andolfatto 2007), o que sugere que o pressuposto de seleção relativamente fraca seja provavelmente realista.

Parece ser mais provável existir um problema com o pressuposto da pleiotropia universal. Quão comum é a pleiotropia? É universal?

Pleiotropia do gene

A pleiotropia do gene, ou seja, o efeito de uma variante alélica em múltiplos e diferentes fenótipos em vez de apenas um, tem sido historicamente considerada como universal, ou pelo menos muito comum. J. B. S. Haldane (1932, p. 62–63) dizia “É provável que o gene afete o organismo como um todo pois encontra-se em todas as células do corpo, mesmo que o seu efeito mais forte se dê apenas num determinado órgão ou função”. Sewall Wright (1968, p. 61) dizia que “as evidências existentes indicam que a pleiotropia é virtualmente universal”. Esta ideia tem sido repetida ao longo da literatura evolucionária e está na base de todos os modelos dos efeitos da pleiotropia na evolução.

Apesar de estas opiniões serem maioritariamente baseadas em observações dos efeitos fenotípicos de mutações individuais em raças domésticas e em estirpes de laboratório, estudos recentes confirmaram a existência de pleiotropia generalizada em todo o genoma. Por exemplo, a análise de 501 fenótipos morfológicos em cada um dos 4710 mutantes de deleção em genes não-

essenciais de levedura, revelou que 35% dos mutantes de deleção afetaram pelo menos duas características morfológicas (Ohya et al. 2005). Um segundo estudo analisou o *fitness* relativo destes mutantes de deleção em 21 condições adversas de crescimento versus um meio de controlo (Dudley et al. 2005). Cerca de 58% dos genes apresentaram uma redução do crescimento em duas ou mais condições, que os autores consideraram refletir os efeitos pleiotrópicos em diferentes condições de crescimento. Estes dois estudos exploraram um número limitado de condições de crescimento e aspetos da morfologia. Além disso, a levedura possui muitos menos características fenotípicas que os eucariotes multicelulares. Por conseguinte, é provável que a deleção de genes em eucariotas mais complexos provoque, no mínimo, tanta pleiotropia como a observada na levedura. Apesar da pleiotropia do gene poder ser relativamente alta em leveduras, pois estes têm relativamente menos genes, parece provável que a clássica perspetiva da pleiotropia geral de genes esteja correta.

A análise aprofundada deste conjunto de dados demonstrou que a pleiotropia do gene está correlacionada com o número de processos biológicos e o número de interações proteína-proteína nas quais participam os produtos do gene (He e Zhang 2006). A pleiotropia não está correlacionada com o número de funções moleculares nem com o número de domínios proteicos por gene. Isto é, a pleiotropia resulta da participação de um produto do gene em múltiplos processos celulares, nos quais executa a mesma função molecular. Por conseguinte, é provável que a expressão de genes em vários tecidos durante o desenvolvimento de organismos resulte em elevada pleiotropia.

Nas análises de crescimento, a eliminação de genes que exibem mais pleiotropia reduziu a aptidão a um nível mais elevado que a eliminação dos genes que exibem menos pleiotropia (Cooper et al. 2007). Estes resultados ajustam-se pelo modelo geométrico de adaptação de Fisher (Fisher 1930). Neste modelo, a pleiotropia universal do gene faz com que a maioria das mutações distanciem uma população do valor ótimo de fitness.

Por fim, a dimensão da pleiotropia do gene descoberta em análises de crescimento está correlacionada com o nível de conservação da sequência de proteínas (He e Zhang 2006), o que coincide com as expectativas teóricas de que a pleiotropia deverá restringir as sequências de proteínas (Waxman e Peck 1998). O conjunto destas observações apoiam a hipótese que indica que a pleiotropia do gene restringe a evolução de proteínas, porque os produtos do gene participam em múltiplas funções celulares.

Por conseguinte, a pleiotropia do gene aparenta ser extremamente comum. Assim sendo, como é que os organismos complexos fazem evoluir adaptações específicas de forma aparentemente tão rápida?

Genes pleiotrópicos versus mutações pleiotrópicas

É surpreendente o facto de nenhum dos modelos de genética populacional que explore os efeitos da pleiotropia ter incorporado modelos realistas da estrutura e função do gene. Todos os modelos assumiram que a pleiotropia do gene é universal e, mais importante, que a pleiotropia universal do gene

gene pleiotropy equals universal pleiotropy of mutational effects. However, there is no equivalence between the pleiotropic roles of genes and the effects of individual mutations (Stern 2000). Some of the mutations occurring in coding or in *cis*-regulatory regions of genes are likely to have specific effects, potentially without any pleiotropic effects. That is, the conception of pleiotropy that has historically gripped evolutionary biologists and served as the central assumption for recent theoretical treatments is effectively based on the effects of null mutations (Stern 2000). This conception is incomplete and future evolutionary models should be based on more realistic estimates of the developmental effects of individual mutations. In particular, models should incorporate the potentially differential pleiotropic effects of mutations in *cis*-regulatory and coding regions, as discussed below.

Modularity of cis-regulatory regions

In most cases studied in sufficient detail, *cis*-regulatory regions can be divided into small DNA regions, on the order of hundreds of base pairs, each of which encodes a small part of the entire pattern (Davidson 2006). In most cases, each of these *cis*-regulatory modules acts independently of the others. That is, these modules can be experimentally dissected from their native genomic region, artificially coupled to a heterologous promoter and reporter gene, and shown to drive reporter gene expression in a small part of the complete pattern. Each *cis*-regulatory module is a collection of transcription factor binding sites that, together, encode a transcriptional output. Usually, each module contains multiple binding sites for each of several transcription factors.

Cis-regulatory modules appear to evolve as structural features of the genome that are independent of one another (Wray et al. 2003). The simplest observation supporting this contention is that *cis*-regulatory modules are often separated from each other by large DNA regions, thousands to tens of thousands of base pairs long. In addition, although the modules can retain conserved function, the DNA regions separating them often evolve in size. This suggests that the precise distance between these regions is not required for function. This is consistent with the current mechanistic view that the intervening DNA is looped out, allowing the transcription factors bound to the DNA to make contact with the basal transcription apparatus. This looping apparently enables *cis*-regulatory modules to reside very far, sometimes over 100 kbp, from the basal promoter. The modules can reside 5' or 3' of the coding region, in introns, or, in rare instances, in exons (Wray et al. 2003).

The *cis*-regulatory regions of genes expressed in various tissues often contain multiple modules that act independently to drive part of the total expression pattern. The modules apparently have only weak constraints on position relative to each other and on position relative to the basal promoter. These features imply that modification or loss of a single module will normally al-

ter only a small part of the total transcriptional pattern of the gene.

Coding mutations may either eliminate gene function or alter the gene product. Modifying a single domain of a protein or an RNA molecule is likely to affect the molecular properties of this molecule in many of the cellular contexts where it is present. If the gene encodes a transcription factor, then this change can alter the transcription of many other genes in every cell where the transcription factor is expressed. In contrast, mutations in *cis*-regulatory modules will normally alter gene expression in only a small part of the complete expression pattern.

Because *cis*-regulatory regions are usually more modular than coding regions, and because *cis*-regulatory modules are more independent of each other than the domains composing a protein or RNA molecules, mutations in *cis*-regulatory modules are likely to have fewer pleiotropic effects than mutations in coding regions. Together with the population genetics argument, this implies that *cis*-regulatory mutations are more likely to cause phenotypic evolution than coding changes. This argument thus supports the narrow *cis*-regulatory hypothesis, that phenotypic evolution should result more from *cis*-regulatory mutations than coding changes. It does not, however, provide strong support for the broad *cis*-regulatory hypothesis because the relative importance of mutations that are neither *cis*-regulatory nor coding changes is not explicitly addressed by this argument.

Difficulties with the population genetics argument

The population genetics argument was originally formulated for genes with pleiotropic roles in development and in possession of modular *cis*-regulatory regions. This is certainly the simplest way to conceptualize the problem and is consistent with the observation that many genes contain modular *cis*-regulatory regions. However, many genes may not have *cis*-regulatory regions organized into independent modules and it is not yet clear whether mutations in nonmodular *cis*-regulatory regions have nonpleiotropic effects. Although it is true that the majority of genes that have been examined in detail appear to possess at least some modularity in their *cis*-regulatory regions, it is also true that this modularity does not appear to be always complete. For example, experimental dissection of the *cis*-regulatory region of the *runt* gene in *D. melanogaster* has identified a >10 kb region required for embryonic expression in seven stripes. It was not possible, however, to dissect this region into smaller, independently acting modules (Klingler et al. 1996). The transcription factor binding sites driving expression in seven stripes may be dispersed throughout this region, rather than being compacted into individual modules.

In addition, the "file drawer" effect (Scargle 2000) may have caused an excess of modular *cis*-regulatory regions to be reported in the literature. Several of our colleagues have informed us of their unpublished attempts to dissect *cis*-regulatory regions that

igual a pleiotropia universal dos efeitos de mutação. No entanto, não existe equivalência entre as funções pleiotrópicas do gene e os efeitos de mutações individuais (Stern 2000). É provável que algumas das mutações que ocorrem nas regiões de codificação ou de regulação *cis* dos genes tenham efeitos específicos, potencialmente sem quaisquer efeitos pleiotrópicos. Isto é, a concepção de pleiotropia que tem sido historicamente utilizada por biólogos evolutivos e que serve de pressuposto central para recentes abordagens teóricas, é efetivamente baseada nos efeitos de mutações nulas (Stern 2000). Esta concepção está incompleta e os futuros modelos evolutivos deverão ser baseados em estimativas mais realistas dos efeitos do desenvolvimento em mutações individuais. Estes modelos em particular devem incorporar os efeitos pleiotrópicos potencialmente diferenciais das mutações nas regiões de codificação e de regulação *cis*, como discutido de seguida.

A modularidade das regiões de regulação cis

Na maioria dos casos estudados em suficiente detalhe, as regiões de regulação *cis* podem ser divididas em pequenas regiões do ADN, na ordem das centenas de pares de bases, sendo que cada qual codifica uma pequena parte de todo o padrão (Davidson 2006). Na maioria dos casos, cada um destes módulos de regulação *cis* atua de forma independente dos outros. Isto é, estes módulos podem ser dissecados experimentalmente da sua região genómica nativa, conjugada artificialmente a um promotor heterólogo e gene repórter, e podem acionar a expressão do gene repórter numa pequena parte do padrão completo. Cada módulo de regulação *cis* é uma coleção de locais de ligação do fator de transcrição que, em conjunto, codificam um output transcricional. Cada módulo contém, geralmente, múltiplos locais de ligação para cada um dos vários fatores de transcrição.

Os módulos de regulação *cis* parecem evoluir como características estruturais do genoma que são independentes umas das outras (Wray et al. 2003). A observação mais simples que apoia esta polémica indica que os módulos de regulação *cis* são geralmente separados uns dos outros por grandes regiões de ADN, com milhares a dezenas de milhares de pares de bases. Além do mais, apesar dos módulos terem a capacidade de reter a função conservada, as regiões de ADN que os separam evoluem, geralmente, em tamanho. Isto sugere que a distância exata entre estas regiões não é necessária para a função. Isto é consistente com a atual opinião mecanicista que exclui o ADN interveniente, o que permite estabelecer contacto entre os fatores de transcrição ligados ao ADN e o aparelho basal de transcrição. Aparentemente, esta exclusão permite que os módulos de regulação *cis* residam a grande distância, por vezes a mais de 100 kbp, do promotor basal. Estes módulos podem residir na região 5' ou 3' da codificação, em intrões ou, mais raramente, em exões (Wray et al. 2003).

As regiões de regulação *cis* dos genes expressos em vários tecidos contêm regularmente múltiplos módulos que atuam de modo independente para acionar parte do total do padrão de expressão. Os módulos aparentam possuir somente fracas restrições na posição relativa entre si e na posição relativa ao promotor basal. Estas características implicam que a modificação ou perda de um único módulo, normalmente irá modificar apenas uma pequena parte do padrão de transcrição total do gene.

As mutações de codificação tanto podem eliminar a função do gene como modificar o produto do gene. É provável que a modificação de um único domínio de uma proteína ou molécula de ARN afete as propriedades moleculares desta molécula em vários dos contextos celulares nos quais está presente. Se um gene codifica um fator de transcrição, esta mudança pode alterar a transcrição de muitos outros genes em todas as células onde o fator de transcrição está expresso. Em contrapartida, as mutações nos módulos de regulação *cis* irão normalmente modificar a expressão do gene apenas numa pequena parte de todo o padrão de expressão.

Devido ao facto das regiões de regulação *cis* serem mais modulares que as de codificação, e os módulos de regulação *cis* serem mais independentes uns dos outros que os domínios que compõem uma proteína ou moléculas de ARN, é provável que as mutações nos módulos de regulação *cis* tenham menos efeitos pleiotrópicos que as mutações nas regiões de codificação. Este argumento, juntamente com o da genética de populações, sugere que as mutações na regulação *cis* têm mais probabilidades de provocar evolução fenotípica do que as alterações na codificação. Deste modo, este argumento apoia a hipótese limitada da regulação *cis*, no sentido em que a evolução fenotípica deverá resultar maioritariamente de mutações na regulação *cis* e não de alterações na codificação. No entanto, não fornece uma defesa sólida para a hipótese geral da regulação *cis* porque a importância relativa das mutações que não são alterações da regulação *cis* ou da codificação, não está explicitamente indicada neste argumento.

Dificuldades no argumento da genética de populações

O argumento da genética de populações foi originalmente formulado para genes com funções pleiotrópicas no desenvolvimento e controle de regiões modulares da regulação *cis*. Esta é certamente a maneira mais simples de conceptualizar o problema e é também consistente com a observação de que muitos genes contêm regiões modulares de regulação *cis*. No entanto, muitos dos genes podem não ter as regiões de regulação *cis* organizadas em módulos independentes e não está ainda claro se as mutações em regiões não-modulares de regulação *cis* têm efeitos não-pleiotrópicos. Apesar de ser verdade que a maioria dos genes examinados em detalhe parece possuir, no mínimo, alguma modularidade nas suas regiões de regulação *cis*, é também verdade que esta modularidade nem sempre aparenta estar completa. Por exemplo, a dissecação experimental da região de regulação *cis* do gene *runt* na *D. melanogaster*, identificou uma região >10 kb necessária à expressão embrionária em sete bandas. No entanto, não foi possível dissecar essa região em módulos mais pequenos e de ação independente (Klingler et al. 1996). Os locais de ligação do fator de transcrição que acionam a expressão nas sete bandas podem ser dispersos através desta região, em vez de serem compactados em módulos individuais.

Além do mais, o “efeito gaveta” (Scargle 2000) pode ter despoletado excessivas menções na literatura às regiões modulares de regulação *cis*. Fomos informados por vários colegas das suas tentativas não publicadas de dissecar as regiões de regulação *cis* que

Table 2. The distribution of evolutionarily relevant mutations in plants and animals.

	Plants	Animals
Coding ¹	71	163
<i>Cis</i> -regulatory	26	48
Other ²	16	7
Total	113	218
Null ³	67	32

¹Includes mutations altering mRNA splicing.

²Includes gene amplification, gene loss, stable DNA methylation, as well as four cases in plants where the mutations were mapped to a gene but not localized to a coding versus *cis*-regulatory change.

³Number of total alleles that are presumed null based on existence of premature stop codons, altered splice sites that disrupt the protein, and deletions of part or all of the protein-coding sequence.

have not uncovered simple, modular *cis*-regulatory regions. These studies normally remain in the “file drawer” (for an exception, see Davis et al. 2007). Because most such studies have not been published, it is currently impossible to accurately estimate the proportion of genes containing independent *cis*-regulatory modules (Scargle 2000).

The lack of *cis*-regulatory modularity in some genes may not really be a vulnerability of the population genetics argument. These genes might still contain *cis*-regulatory regions in which at least some mutations can have specific, nonpleiotropic effects. That is, the *cis*-regulatory regions may be functionally modular without displaying obvious clusters of transcription factor binding sites. For example, although the *runt* *cis*-regulatory region cannot be dissected into independent modules, it is possible that some individual mutations would influence only one of the seven stripes driven by this enhancer region.

Although the *cis*-regulatory hypothesis was originally formulated as requiring only that mutations in *cis*-regulatory modules have fewer pleiotropic effects than mutations in coding regions, the current cost of complexity theory suggests that the absolute number of pleiotropic effects is the more important parameter. It is possible that individual mutations in *cis*-regulatory modules still have pleiotropic effects. We are aware of no empirical studies of the potential pleiotropic effects of *cis*-regulatory mutations.

Finally, the population genetics argument might not hold when other factors, such as genetic drift in populations of small effective size, strong selection, or differences in mutational target sizes bypass or overcome the cost of complexity.

In conclusion, we cannot find any serious difficulties with the proposal that genes with pleiotropic functions will preferentially accumulate evolutionarily relevant mutations in the *cis*-regulatory region of the gene, especially when selection is weak and other parameter values are equal. However, in some circumstances—such as for genes without pleiotropic roles, for evolution by strong

selection or in small populations, or for genes with a higher mutational target size for coding mutations than for *cis*-regulatory mutations—this prediction might not hold.

EXPERIMENTAL EVIDENCE

The last argument that has been advanced to support the *cis*-regulatory hypothesis is that many *cis*-regulatory changes have been identified as responsible for evolutionary changes in phenotype, or more specifically in morphology. As Hoekstra and Coyne (2007) justly noted, a comprehensive survey of the experimental evidence is required to test this claim. We have compiled a database of published studies that provide compelling evidence for the individual genetic mutations causing evolved phenotypic variation within and between species of multicellular organisms (Appendix 1). We found a total of 234 mutations in coding regions and 74 mutations in *cis*-regulatory regions (Table 2), including 62 coding changes and 43 *cis*-regulatory changes causing morphological evolution (Table 3). The absolute numbers of reported mutations in coding and *cis*-regulatory regions on their own clearly do not provide support for the *cis*-regulatory hypothesis in its simplest formulation. However, as discussed below (in section “The Data: Evidence for a Predictive Theory of Genetic Evolution”), many sources of ascertainment bias strongly inflate the reported contribution of coding changes to evolution. A more detailed analysis of the data is given in the same section.

WHITHER THE *CIS*-REGULATORY HYPOTHESIS?

The arguments reviewed above do not, on their own or combined together, provide definitive support for the *cis*-regulatory hypothesis. The strongest case against the *cis*-regulatory hypothesis is, currently, that more protein-coding changes are known to cause phenotypic evolution than *cis*-regulatory changes. However, the apparent abundance of protein-coding changes has resulted from several layers of ascertainment bias, which we discuss in more detail in the section “The Data: Evidence for a Predictive Theory of Genetic Evolution.” In summary, theoretical arguments provide reasons why phenotypic evolution is likely to be dominated by *cis*-regulatory evolution, but other evolutionary forces such as mutational target size or population demographic history might obscure the expected trend. In any case, the arguments presented above do not prove that phenotypic evolution is dominated by *cis*-regulatory evolution. The question of the importance of *cis*-regulatory evolution will ultimately be settled with empirical data.

The Data: Evidence for a Predictive Theory of Genetic Evolution

We now address what we believe to be the fundamental disagreement between proponents of the *cis*-regulatory hypothesis and

Tabela 2. A distribuição de mutações relevantes para a evolução em plantas e animais.

	Plantas	Animais
Codificação ¹	71	163
Regulação <i>cis</i>	36	48
Outro 2	16	7
Total	113	218
Nulo 3	67	32

¹ Inclui mutações que modificam a ligação ARNm.

² Inclui amplificação genética, perda de genes, metilação do ADN estável, assim como quatro casos de plantas nas quais as mutações foram mapeadas a um gene mas não foram localizadas na alteração de codificação versus regulação *cis*.

³ Número total de alelos que se presumem ser nulos, com base na existência de codões de finalização prematuros, zonas de união modificadas que quebram a proteína, e deleções parciais ou integrais de uma sequência de codificação de proteína.

não revelaram simples regiões modulares de regulação *cis*. Normalmente, estes estudos ficam arquivados na “gaveta” (à exceção de Davis et al.2007). Muitos destes estudos não foram publicados, o que faz com que seja impossível determinar com precisão a proporção de genes que contêm módulos independentes de regulação *cis* (Scargle 2000).

A falta de modularidade da regulação *cis* nalguns genes pode não ser uma vulnerabilidade do argumento da genética de populações. Estes genes podem ainda conter regiões de regulação *cis* nas quais algumas mutações podem ter efeitos específicos não-pleiotrópicos. Isto é, as regiões de regulação *cis* podem ser funcionalmente modulares sem apresentarem conjuntos óbvios de locais de ligação do fator de transcrição.

Por exemplo, apesar da região de regulação *cis* do gene *runt* não poder ser dissecada em módulos independentes, é possível que algumas mutações individuais possam influenciar apenas uma das sete bandas acionadas por esta região intensificadora.

Quando a hipótese da regulação *cis* foi originalmente formulada, requeria apenas que as mutações nos módulos de regulação *cis* tivessem menos efeitos pleiotrópicos que as mutações nas regiões de codificação. No entanto a atual teoria do custo da complexidade sugere que o número absoluto de efeitos pleiotrópicos deverá ser um parâmetro mais importante. É possível que as mutações individuais em módulos da regulação *cis* continuem a ter efeitos pleiotrópicos. Não temos conhecimento de estudos empíricos dos efeitos com potencial pleiotrópico nas mutações de regulação *cis*.

Por fim, o argumento da genética de populações pode não se sustentar, quando comparado com outros fatores, tais como a deriva genética de populações com tamanho efetivo pequeno, ou caso as diferenças nas dimensões previstas da mutação contornem ou ultrapassem o custo da complexidade.

Concluindo, não se encontram sérias dificuldades na proposta de que os genes com funções pleiotrópicas irão acumular mutações preferencialmente relevantes para a evolução na região de regulação *cis* do gene, especialmente quando a seleção é fraca e os valores de outros parâmetros são iguais. No entanto, em certas circunstâncias, esta previsão pode não se sustentar, tais como em genes sem funções pleiotrópicas, na evolução

através de seleção forte ou em pequenas populações, ou em genes com uma dimensão prevista de mutação superior nas mutações da codificação do que nas mutações na regulação *cis*.

EVIDÊNCIAS EXPERIMENTAIS

O último argumento que foi avançado para apoiar hipótese de regulação *cis* indica que muitas das alterações na regulação *cis* foram identificadas como sendo responsáveis por alterações evolutivas no fenótipo ou, mais especificamente, na morfologia. Tal como Hoekstra e Coyne (2007) indicaram, é necessário um levantamento completo das evidências experimentais para testar estas afirmações. Compilamos uma base de dados de estudos publicados que oferecem provas convincentes de mutações genéticas individuais que fazem com que a variação fenotípica evolua dentro e entre espécies de organismos multicelulares (Apêndice 1). Foram descobertos um total de 234 mutações nas regiões de codificação e 74 mutações nas regiões de regulação *cis* (Tabela 2), incluindo 62 alterações na codificação e 43 alterações na regulação *cis* que provocam evolução morfológica (Tabela 3). Os números absolutos de referências a mutações nas regiões de codificação e de regulação *cis* são claramente insuficientes para defender as hipóteses da regulação *cis* na sua formulação mais simples. No entanto, tal como é adiante debatido (na secção “Os Dados: Evidências para uma Teoria Predizível de Evolução Genética”), muitas das fontes de viés de averiguação sobrevalorizam fortemente as contribuições remetidas sobre alterações na codificação para a evolução. Essa mesma secção fornece uma análise mais detalhada dos dados.

QUAL DAS HIPÓTESES DE REGULAÇÃO *CIS*?

Os argumentos acima revistos não fornecem, por si só ou em conjunto, uma defesa conclusiva da hipótese da regulação *cis*. Atualmente, o facto da evolução fenotípica ser provocada por mais alterações na codificação de proteínas do que na regulação *cis* é o caso mais sólido contra a hipótese de regulação *cis*. No entanto, a aparente abundância de alterações na codificação de proteína é resultado de vários vieses de averiguação, como será discutido em maior detalhe na secção: “Os Dados: Evidências para uma Teoria Predizível de Evolução Genética.” Resumindo, os argumentos teóricos justificam por que é provável que a evolução fenotípica seja dominada pela evolução da regulação *cis*, mas as outras forças evolutivas, tais como a dimensão prevista de mutação ou o histórico da demografia da população, podem ofuscar a tendência prevista. Em qualquer dos casos, os argumentos acima apresentados não provam que a evolução fenotípica seja dominada pela evolução da regulação *cis*. A questão da importância da evolução da regulação *cis* será apenas esclarecida com dados empíricos.

Os Dados: Evidências para uma Teoria Predizível da Evolução Genética

Será agora abordado o desacordo que julgamos ser fundamental entre os proponentes da hipótese de regulação *cis* e

Table 3. Distribution of evolutionary relevant mutations among phenotypic classes and among regulatory network levels.

	Morphology	Physiology	Behavior	DGB member ¹	Non-DGB member ²
Coding ³	62	170	2	132	102
<i>Cis</i> -regulatory	43	29	2	34	37
Other ⁴	3	20	0	9	14
Total	108	219	4	175	153
Null ⁵	41	58	0	22	77

¹Gene is a known or presumptive member of a differentiation gene battery (DGB).

²Gene known or presumed to reside upstream of a DGB. Three genes could not be assigned to the DGB or non-DGB category because their function is unknown.

³Includes mutations altering mRNA splicing.

⁴Includes gene duplications, gene losses, stable DNA methylation, and four cases in which the mutations were mapped to a gene but not localized to a coding versus *cis*-regulatory change.

⁵Alleles presumed null based on existence of premature stop codons, altered splice sites, and deletions of part or all of the protein-coding sequence.

Hoekstra and Coyne (Coyne and Hoekstra 2007; Hoekstra and Coyne 2007). Evolutionary developmental biologists have examined the structure of developmental regulatory networks and the structure–function relationship of individual genes and predicted that *cis*-regulatory mutations should play a dominant role in morphological evolution. Hoekstra and Coyne (2007) have argued that these molecular mechanisms might not bias the distribution of evolutionarily relevant mutations and, in any case, that historical contingency might dominate patterns of genetic evolution (see also Appendix 2). We state the problem in stark terms to clarify what is at stake: a theory of genetic evolution. Does the architecture of gene regulatory networks and the structure of genes influence which mutations are favored during evolution? Or, does the historical contingency of the mutational process dominate and cause fundamentally unpredictable patterns of genetic evolution?

So far, debate has focused on the proportion of *cis*-regulatory versus coding mutations causing phenotypic evolution. Empirical data clearly demonstrate that considerable numbers of both types of mutations contribute to phenotypic evolution. We believe that little progress will be made by structuring the debate as an enquiry simply into the proportion of *cis*-regulatory versus coding changes. This superficially attractive dichotomy hides considerable complexity resulting from precisely how development generates the phenotype and in how mutations traverse populations to cause phenotypic variation and population differentiation. It may be more profitable to turn the problem around and ask more specific questions. For example, how do we expect particular kinds of mutations to generate particular kinds of phenotypic variation? How do we expect population genetic parameters to influence the spread and fixation of different kinds of mutations?

In this spirit, we discuss three predictions that derive from the *cis*-regulatory hypothesis and from our current understanding of the molecular basis for development and test them with available

data. We chose to focus on multicellular plants and animals. Many studies indicate that unicellular organisms show predictable patterns of genetic evolution (see for example Boucher et al. 1992; Wichman et al. 1999; Riehle et al. 2001; Dunham et al. 2002; Hittinger et al. 2004; Segre et al. 2006; Woods et al. 2006).

THE DATA

Currently, considerable effort is devoted to identifying the genes and mutations underlying phenotypic evolution, particularly in domesticated races and in natural populations of single species. We have compiled a database of published studies that provide compelling evidence for the individual genetic mutations causing evolved phenotypic variation (Appendix 1). We included variation in domesticated species (99 cases), intraspecific variation in wild species (157 cases), and interspecific differences (75 cases). We included domesticated species because, ever since Darwin, they have been considered as potential models for how evolution might occur in the wild (Price 2002; Andersson and Georges 2004). We did not include variation selected in laboratory experiments. The dataset includes many studies from both plants and animals (Table 2). Although we have almost certainly inadvertently overlooked some relevant studies, we have attempted to be comprehensive.

This dataset includes extensive ascertainment bias, both in the choice of genes studied and in the gene regions examined. Most researchers have focused on candidate genes. Even genome-wide mapping studies usually include a search for candidate genes in the mapped region, rather than functional surveys of all genes in the mapped region. But the most important consequence of investigator bias is that the relative number of coding versus *cis*-regulatory mutations causing phenotypic evolution is almost certainly inflated. Indeed, it is easier to identify potentially important coding changes, especially nonsense mutations, than potentially

Tabela 3. Distribuição de mutações evolutivas relevantes entre classes fenotípicas e entre níveis de redes de regulação.

	Morfologia	Fisiologia	Comportamento	Membro DGB ¹	Não-membro DGB ²
Codificação ³	62	170	2	132	102
Regulação <i>cis</i>	43	29	2	34	37
Outro ⁴	3	20	0	9	14
Total	108	219	4	175	153
Nulo ⁵	41	58	0	22	77

¹ O gene é um membro conhecido ou presumível membro de um conjunto de genes de diferenciação (DGB).

² Gene conhecido por, ou que se presume, residir na corrente superior do DGB. Três genes não puderam ser atribuídos à categoria DGB ou não-DGB porque a sua função é desconhecida.

³ Inclui mutações que modificam a ligação ARNm.

⁴ Inclui duplicação genética, perda de genes, metilação do ADN estável, assim como quatro casos nos quais as mutações foram mapeadas a um gene, mas não foram localizadas na alteração de codificação versus regulação *cis*.

⁵ Alelos que se presumem ser nulos, devido à existência de códons de finalização prematuros, zonas de união modificadas, e deleções parciais ou integrais de uma sequência de codificação de proteína.

e Hoekstra e Coyne (Coyne e Hoekstra 2007; Hoekstra e Coyne 2007). Os biólogos especialistas em evolução e desenvolvimento examinaram a estrutura das redes reguladoras do desenvolvimento e a relação estrutura-função de genes individuais e previram que as mutações na regulação *cis* deverão ter uma função dominante na evolução morfológica. Hoekstra e Coyne (2007) defendem que estes mecanismos moleculares podem não influenciar a distribuição de mutações relevantes para a evolução e, em qualquer caso, este contingente histórico pode dominar os padrões da evolução genética (ver Apêndice 2). Determinamos o problema em termos absolutos para clarificar o que está em jogo: a teoria da evolução genética. A arquitetura das redes reguladoras dos genes e a estrutura dos genes influenciam a escolha de mutações favorecidas durante a evolução? Ou será que a contingência histórica do processo mutacional domina e provoca padrões fundamentalmente imprevisíveis da evolução genética?

Até agora, o debate centrou-se na proporção de mutações na regulação *cis* versus mutações na codificação que provocam a evolução fenotípica. Os dados empíricos demonstram claramente que ambos os tipos de mutações contribuem para a evolução fenotípica. Acreditamos que não se alcançará grande progresso caso se estructure o debate na questão da proporção de alterações na regulação *cis* versus codificação. Apesar de parecer atraente, esta dicotomia é bastante complexa, pois é necessário saber exatamente como é que o desenvolvimento gera o fenótipo e de que modo as mutações atravessam as populações, provocando variação fenotípica e diferenciação de população. Poderá ser mais rentável direcionar o problema noutra sentida e colocar perguntas mais específicas. Por exemplo, como é que esperamos que certos tipos de mutações gerem tipos específicos de variação fenotípica? Como se pode esperar que os parâmetros da genética populacional influenciará a propagação e fixação de diferentes tipos de mutações?

Neste contexto, discutimos as três previsões derivadas da hipótese de regulação *cis* e da nossa atual compreensão da base molecular do desenvolvimento e testámo-las com os dados disponíveis. Optamos por nos concentrar em plantas animais multicelulares. Existem muitos estudos que indicam que os

organismos unicelulares apresentam padrões previsíveis de evolução genética (ver, por exemplo, Boucher et al. 1992; Wichman et al. 1999; Riehle et al. 2001; Dunham et al. 2002; Hittinger et al. 2004; Segre et al. 2006; Woods et al. 2006).

OS DADOS

Atualmente, existem vários esforços dedicados à identificação de genes e mutações subjacentes à evolução fenotípica, em particular nas raças domésticas e em populações naturais de espécies únicas. Compilamos uma base de dados de estudos publicados que oferecem provas convincentes de mutações genéticas individuais que fazem com que a variação fenotípica evolua (Apêndice 1). Incluímos a variação nas espécies domésticas (99 casos), variação intraespecífica em espécies selvagens (157 casos) e diferenças interespecíficas (75 casos). Incluímos espécies domésticas porque, desde Darwin, têm sido consideradas potenciais modelos para compreender como a evolução pode ocorrer no meio selvagem (Price 2002; Andersson e Georges 2004). Não incluímos variação selecionada de experiências laboratoriais. O conjunto de dados inclui vários estudos de plantas e animais (Tabela 2). Tentamos ser abrangentes, mas alguns estudos relevantes foram quase de certeza ignorados, embora sem intenção.

Este conjunto de dados é tendencioso, tanto na escolha de genes estudados como nas regiões de genes examinadas. A maioria dos investigadores focaram-se nos genes candidatos. Até mesmo os estudos do mapeamento do genoma incluem uma pesquisa de genes candidatos na região mapeada, em vez do levantamento funcional de todos os genes na região mapeada. Mas a consequência mais importante do viés do investigador é que o número relativo de mutações de codificação versus mutações na regulação *cis* que provocam a evolução fenotípica é quase certamente exagerado. De facto, é mais fácil identificar alterações na codificação potencialmente mais importantes, especialmente em mutações sem sentido, do que alterações potencialmente

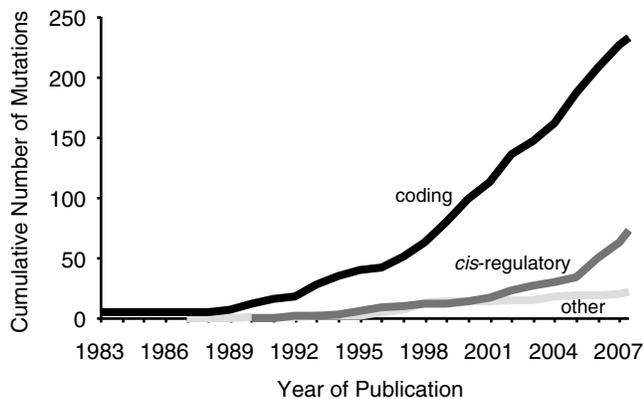


Figure 2. Cumulative number of coding mutations, *cis*-regulatory mutations and other types of mutations (gene amplification, gene loss, etc.) that have been identified over time as responsible for phenotypic evolution. Results are from data in Appendix 1. Note that the slope for *cis*-regulatory mutations has increased in recent years. The current discovery rate of *cis*-regulatory mutations approximately equals the discovery rate of coding mutations. If this reflects the long-term trend, then we expect ultimately to observe approximately equal numbers of *cis*-regulatory and coding mutations.

relevant changes in *cis*-regulatory regions by simple examination of the DNA sequence. The recent surge in examples of *cis*-regulatory evolution (Fig. 2) may be due to the fact that more powerful experimental approaches for identifying *cis*-regulatory mutations have been developed recently (e.g., McGregor et al. 2007).

Therefore, we cannot estimate, based on our dataset, the real overall frequency of *cis*-regulatory versus coding mutations causing phenotypic evolution. Twenty-two percent of mutations in our dataset occurred in *cis*-regulatory regions for all types of phenotypic change (22% with domesticated examples excluded), and 40% for morphological evolution (55% with domesticated examples excluded). These values are almost certainly minimum estimates of the frequency of *cis*-regulatory changes causing phenotypic evolution.

Although we cannot confidently test the *cis*-regulatory hypothesis based on our compilation, it is likely, although not certain, that similar ascertainment bias has been applied to different kinds of traits and to different taxonomic levels. We can therefore use the experimental data to compare the relative importance of *cis*-regulatory evolution between types of traits and between taxonomic levels.

We now examine three predictions derived from the *cis*-regulatory hypothesis and, to the extent possible, test them with available data.

PREDICTION 1: MORPHOLOGICAL EVOLUTION RESULTS FROM MORE *CIS*-REGULATORY CHANGES THAN PHYSIOLOGICAL EVOLUTION

The *cis*-regulatory hypothesis has been applied traditionally to the evolution of morphology. Occasionally, the idea that physiological evolution might involve more coding changes than morphological evolution has been made explicit (Carroll 2005). Hoekstra and Coyne (2007) claim that the *cis*-regulatory hypothesis should apply to all adaptation or not at all. As they argue, and we agree, the division between anatomy and physiology may be a misleading dichotomy. At the very least, there is not a clear boundary between genetic mechanisms generating morphology and physiology. Both result from activity of genes that are embedded within gene regulatory networks. Developmental processes are influenced by physiological processes, and vice versa. For example, development can be regulated by steroid hormones and by sugar and lipid metabolism (Wilkins 2002) whereas endocrine glands are formed through developmental mechanisms.

The distinction between morphological and physiological evolution that has been largely implicit in discussions within the evolutionary developmental biology community is that physiological traits tend to result from genes located at or near the terminal points of regulatory networks, whereas anatomy usually results from the activity of genes embedded deeper in the developmental network. Genes that act at or near the terminal points of regulatory networks, named differentiation gene batteries (Davidson 2006), represent genes expressed in differentiated tissues to fulfill cell-type-specific functions. These gene products build muscle cells, make skeletal biominerals, mediate synaptic transmission, etc. They do not regulate other genes, and they do not control the progressive formation of spatial patterns of gene expression that underlie development. They receive rather than generate developmental instructions. Even though they may be expressed in various tissues, their pleiotropic roles are more limited than gene products that function higher in the regulatory network. For example, a coding mutation in the *D. melanogaster* gene *forked*, a gene involved in terminal differentiation of bristles, can cause every bristle to develop differently, but all in the same way. In contrast, a coding mutation in *wingless*, a signaling molecule involved in diverse regulatory networks in various tissues, will alter the development of segments, legs, wings, genitalia, and eyes in *Drosophila*, all in different ways. Therefore, it should be more difficult to evolve a coding change in *wingless* that somehow provides an advantage than a coding change in *forked*. Because the position of a gene within a gene regulatory network is likely to influence its pleiotropic roles, we might thus predict that genes at terminal positions in regulatory cascades, members of differentiation gene batteries, should result from coding changes more often than genes embedded deeper within regulatory networks.

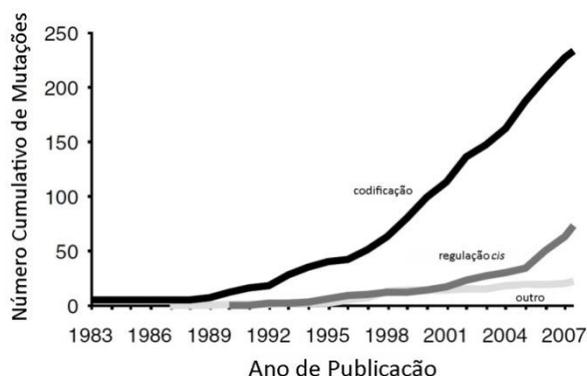


Figura 2. Número cumulativo de mutações de codificação, mutações na regulação *cis* e outros tipos de mutação (amplificação do gene, perde de gene(s), etc.) que foram sendo identificadas como responsáveis pela evolução fenotípica. Os resultados são provenientes dos dados no Apêndice 1. Note-se que a variação de mutações na regulação *cis* aumentou nos anos recentes. A recente quantidade de descobertas de mutações na regulação *cis* equipara-se aproximadamente à quantidade de descobertas de mutações da codificação. Caso isto se torne numa tendência a longo prazo, esperamos observar, em última análise, números muito próximos de mutações na regulação *cis* e na codificação.

relevantes nas regiões de regulação *cis* através da simples examinação da sequência de ADN. A recente vaga de exemplos da evolução na regulação *cis* (Fig.2) pode dever-se ao recente desenvolvimento de abordagens experimentais mais fortes na identificação de mutações na regulação *cis* (por ex. McGregor et al. 2007).

Deste modo, o conjunto de dados utilizado não possibilita uma estimativa real da frequência geral de mutações na regulação *cis* versus mutações na codificação que provocam a evolução fenotípica. Vinte e dois por cento das mutações do nosso conjunto de dados ocorreram nas regiões de regulação *cis* para todos os tipos de alterações fenotípicas (22%, excluindo exemplos domesticados), e 40% para a evolução morfológica (55% com exemplos domesticados excluídos). Estes valores são estimativas quase certamente mínimas da frequência das alterações na regulação *cis* que provocam a evolução fenotípica.

Apesar de não ser possível testar com segurança a hipótese da regulação *cis* baseada na nossa compilação, é provável, embora não seja certo, que tenham sido aplicados vieses de averiguação similares em diferentes tipos de características e em diferentes níveis taxonómicos. Deste modo, é possível utilizar os dados experimentais para comparar a importância relativa da evolução da regulação *cis* entre tipos de características e entre níveis taxonómicos.

Serão agora examinadas as três previsões derivadas da hipótese da regulação *cis* e, na medida do possível, testadas com os dados disponíveis.

PREVISÃO 1: A EVOLUÇÃO MORFOLÓGICA RESULTA DE MAIS ALTERAÇÕES NA REGULAÇÃO *CIS* DO QUE DE EVOLUÇÃO FISIOLÓGICA

A hipótese da regulação *cis* tem sido tradicionalmente aplicada à evolução da morfologia. A ideia de que a evolução fisiológica poderá envolver mais alterações na codificação do que evolução morfológica tem sido ocasionalmente tornada explícita (Carroll 2005). Hoekstra e Coyne (2007) declaram que a hipótese da regulação *cis* deve ser aplicada a toda a adaptação ou a nenhuma. Concordamos com o seu argumento de que a divisão entre anatomia e fisiologia pode ser uma dicotomia enganadora, ou que pelo menos não existe uma divisão clara entre mecanismos genéticos que geram morfologia e fisiologia. Ambos resultam da atividade dos genes que estão inseridos nas redes de regulação do gene. Os processos de desenvolvimento são influenciados por processos fisiológicos, e vice-versa. O desenvolvimento pode ser, por exemplo, regulado por hormonas esteroides e pelo metabolismo de carboidratos e lípidos (Wilkins 2002), enquanto as glândulas endócrinas são formadas através de mecanismos de desenvolvimento.

A distinção entre evolução morfológica e fisiológica, que tem sido largamente implícita em discussões da comunidade da biologia evolutiva do desenvolvimento, indica que as características fisiológicas têm tendência a resultar de genes localizados dentro ou próximos de pontos terminais das redes de regulação, enquanto a anatomia resulta geralmente da atividade de genes inseridos em níveis mais profundos da rede de desenvolvimento. Os genes que atuam dentro ou próximo de redes de regulação, designados de conjunto de genes de diferenciação (Davidson 2006), são genes expressos em tecidos diferenciados que efetuam as funções específicas do tipo de célula. Estes produtos dos genes constroem as células musculares, produzem os biominais do esqueleto, mediam a transmissão sináptica, etc. Não regulam outros genes, e não controlam a formação progressiva de padrões espaciais da expressão do gene subjacentes ao desenvolvimento. Em vez de gerarem instruções de desenvolvimento, recebem-nas. Apesar de se poderem exprimir em vários tecidos, as suas funções pleiotrópicas são mais limitadas que os produtos dos genes que atuam num nível superior da rede de regulação. Por exemplo, a mutação da codificação no gene *forked* da *D. melanogaster*, um gene envolvido na diferenciação terminal dos pelos, pode fazer com que cada pelo se desenvolva de modo diferente, mas todos da mesma maneira, enquanto que uma mutação de codificação na *wingless*, uma molécula de sinalização envolvida em diversas redes de regulação em vários tecidos, irá alterar o desenvolvimento de segmentos, pernas, asas, órgãos sexuais e olhos da *Drosophila*, de maneiras diferentes. Por conseguinte, deverá ser mais difícil evoluir uma alteração da codificação no *wingless*, o que, de certo modo, é mais vantajoso que a alteração na codificação do *forked*. Tendo em conta que é provável que a posição de um gene numa rede de regulação de genes influencie as suas funções pleiotrópicas, é possível prever que os genes em posições terminais nas cascatas reguladoras, membros de conjuntos de genes de diferenciação, podem ser o resultado de alterações na codificação de modo mais frequente que os genes inseridos em níveis mais profundos das redes de regulação.

To test these hypotheses, we have divided the data in several ways. First, we examine the traditional morphology versus physiology hypothesis by classifying mutations, as we felt most biologists would, as contributing to morphology, physiology, or behavior. Just as the “morphology” and “physiology” categories are poorly defined, the “behavior” category also hides considerable diversity in mechanisms. Behavior might evolve because an odorant receptor, a member of a differentiation gene battery (DGB), evolves affinity for a new odorant. Or behavior might evolve because neurons are connected differently as a result of a new pattern of development. To our knowledge, less than 10 alleles causing evolved behavior have been identified so far, so no generalizations about behavioral evolution are possible yet.

Second, we divided the data into genes that are known or are likely to be members of a DGB versus genes that are known or are likely to be embedded more deeply within the developmental regulatory network (non-DGB).

As expected, we found that the proportion of *cis*-regulatory to coding mutations is significantly higher for morphological traits compared to physiological traits (Table 3: Fisher’s exact test, two-tailed, $P < 10^{-6}$). This supports the intuition of evolutionary developmental biologists that *cis*-regulatory mutations have more often been reported for morphological variation than for physiological variation.

However, the proportion of *cis*-regulatory to coding mutations for non-DGB genes is not significantly different from the proportion for DGB genes (Table 3: Fisher’s exact test, two-tailed, $P < 0.23$). This is despite the fact that morphological evolution for the most part involves mutations in non-DGB genes (80 mutations in non-DGB genes vs. 26 in DGB genes) whereas physiological variation mostly involves changes in DGB genes (148 mutations in DGB genes vs. 70 in non-DGB genes). Although we cannot exclude the possibility that our classification of genes into the DGB and non-DGB categories is erroneous, this suggests that factors other than those discussed here influence the proportion of evolutionary relevant *cis*-regulatory mutations that are identified for different kinds of genes. For example, it is possible that evolutionary developmental biologists studying morphological variation are more likely to study evolution of *cis*-regulatory regions than are physiologists. In any case, this observation fails to support the hypothesis that evolution of genes embedded within regulatory networks are more likely to result from changes in *cis*-regulatory regions than are DGB genes. In both cases, about 20% of reported mutations are *cis*-regulatory.

PREDICTION 2: THE STRENGTH OF SELECTION AND THE EVOLUTIONARY TIME SCALE SHOULD INFLUENCE THE SPECTRUM OF EVOLUTIONARILY RELEVANT MUTATIONS

Over the past 50 years, various authors have suggested that population structure and the strength of selection might influence the

kinds of mutations that are selected (Crow 1956; Lande 1983; Liu et al. 1996). Weak selection is expected to bias the spectrum of selected mutations toward those with few or no pleiotropic effects (Otto 2004). In contrast, strong selection, such as that encountered during laboratory selection experiments and perhaps during rapid local adaptation, can overcome pleiotropic deleterious effects of mutations (Baatz and Wagner 1997; Otto 2004). Domestication may also sometimes involve strong selection (Wang et al. 1999). Based on the function of different gene regions, we expect that, on average, *cis*-regulatory mutations will have fewer pleiotropic effects than missense mutations, which in turn should have fewer pleiotropic effects than nonsense mutations or gene deletions.

In addition, short-term evolution may lead to a different spectrum of mutations than long-term evolution. During long-term evolution, mutations may be tested in a variety of environments. Thus, mutations advantageous in one environment, but deleterious in others, may be eliminated over time. In addition, mutations that maintain adaptive plasticity will also be favored over longer time scales when environments vary. Both heterogeneous fitness in multiple environments and loss of plasticity can be considered pleiotropic effects of mutations. *Cis*-regulatory mutations are more likely to limit such pleiotropic effects than coding changes. Thus, *cis*-regulatory mutations may be favored over longer time scales.

We asked whether there is a different distribution of *cis*-regulatory, missense, and nonsense mutations that cause phenotypic differences in domesticated populations, segregating within wild species and between species. These taxonomic categories are expected to vary in the strength and duration of selection and perhaps in other uncontrolled variables. Our compilation reveals that domesticated races and intraspecific variants show striking differences from interspecific comparisons (Table 4).

Thirty-seven percent of the identified mutations underlying intraspecific variation in domesticated species and in a few wild species, especially *Arabidopsis thaliana*, cause the elimination of gene function (Table 4, Appendix 1). The null mutations in domesticated and some wild species often have large phenotypic effects. Many of these traits are closely related to fitness, like flowering time and growth rate, so it is likely that the mutations experienced strong selection. This is a striking result, because most of these genes display strong evolutionary conservation across vast taxonomic distances. Because the elimination of gene function, especially through insertion or deletion events, is largely inconsistent with conservation of these gene sequences through purifying selection, we must conclude that most of these mutations reflect recent selection. This is entirely consistent with the recent origin of domesticated species. It is also consistent with the recent spread of some species, like *A. thaliana*, to novel habitats and the relative rarity of most of these alleles (Le Corre et al. 2002).

Para testar esta hipótese, os dados foram divididos de várias formas. Primeiro, foi examinada a morfologia tradicional versus a hipótese da fisiologia através da classificação das mutações, como julgamos que a maioria dos biólogos faria, contribuindo assim para a morfologia, fisiologia ou comportamento. Da mesma forma que as categorias da “morfologia” e “fisiologia” são parcamente definidas, a categoria do “comportamento” também esconde uma considerável diversidade nos mecanismos. O comportamento pode evoluir porque um recetor de odor, membro de um conjunto de genes de diferenciação (DGB), desenvolve afinidade para com um novo odor. Ou então, o comportamento pode evoluir porque os neurónios estão ligados de forma diferente, como resultado de num novo padrão de desenvolvimento. Tanto quanto sabemos, foram apenas identificados menos de 10 alelos que provocam comportamento desenvolvido, o que não possibilita qualquer generalização sobre a evolução comportamental.

Em segundo lugar, dividimos os dados em genes conhecidos e com probabilidade de serem membros do DGB *versus* genes que são conhecidos e que estarão mais provavelmente inseridos em níveis mais profundos da rede de regulação do desenvolvimento (não-DGB).

Como era esperado, concluímos que a proporção de mutações na regulação *cis* para mutações de codificação é significativamente superior para as características morfológicas, comparando com as características fisiológicas (Tabela 3: teste de Fisher, duas caudas, $P < 10^{-6}$). Isto apoia a intuição dos biólogos especialistas em evolução e desenvolvimento de que as mutações na regulação *cis* estão relacionadas mais vezes com a variação morfológica do que com a variação fisiológica.

No entanto, a proporção de mutações na regulação *cis* para mutações de codificação em genes não-DGB não é significativamente diferente da proporção de genes DGB (Tabela 3: o teste de Fisher, duas caudas, $P < 0,23$). Isto apesar do facto de a evolução morfológica envolver, na sua maioria, mutações em genes não-DGB (80 mutações in não-DGB gene vs. 26 em genes DGB), enquanto a variação fisiológica envolve, na sua maioria, alterações em genes DGB (148 mutações em genes DGB vs. 70 em genes não-DGB). Apesar de não se poder excluir a possibilidade da nossa classificação de genes nas categorias DGB e não-DGB estar incorreta, estes dados sugerem que outros fatores, além dos aqui discutidos, influenciam a proporção de mutações na regulação *cis* relevantes para a evolução que estão identificados em diferentes tipos de genes. Por exemplo, é possível que os biólogos especialistas em evolução e desenvolvimento que estudam a variação morfológica sejam capazes de estudar mais a evolução das regiões na regulação *cis* do que os fisiólogos. Em qualquer dos casos, esta observação não permite a defesa da hipótese de que a evolução dos genes inseridos nas redes de regulação tem maior probabilidade de ser o resultado de alterações nas regiões de regulação *cis* do que os genes DGB. Em ambos os casos, cerca de 20% das mutações reportadas são da regulação *cis*.

PREVISÃO 2: A FORÇA DA SELEÇÃO E A ESCALA DE TEMPO EVOLUTIVA DEVEM INFLUENCIAR O ESPECTRO DE MUTAÇÕES EVOLUTIVAMENTE RELEVANTES

Nos últimos 50 anos, vários autores sugeriram que a estrutura populacional e a força da seleção podem influenciar os tipos de

mutações que são selecionadas (Crow 1956; Lande 1983; Liu et al. 1996). É de prever que a seleção fraca influencie o espectro de mutações selecionadas perante as que têm poucos ou nenhuns efeitos pleiotrópicos (Otto 2004). A seleção forte, pelo contrário, pode ultrapassar os efeitos deletérios pleiotrópicos das mutações, como se verificou em experiências laboratoriais de seleção e talvez durante a adaptação rápida (Baatz and Wagner 1997; Otto 2004). A domesticação pode por vezes também envolver seleção forte (Wang et al. 1999). Com base na função das diferentes regiões do gene, espera-se que as mutações na regulação *cis* venham a ter, em média, menos efeitos pleiotrópicos que as mutações de sentido invertido que, por sua vez, devem ter menos efeitos pleiotrópicos que as mutações sem sentido ou que a supressão de genes.

Além disso, a evolução a curto prazo pode levar a um espectro de mutações diferente da evolução a longo prazo. Durante a evolução a longo prazo, as mutações podem ser testadas numa variedade de ambientes. Deste modo, as mutações vantajosas num determinado ambiente, mas deletérias noutros podem ser eliminadas com o passar do tempo. Para além disso, as mutações que mantêm plasticidade adaptativa serão também favorecidas em maiores escalas de tempo, quando os ambientes variarem. A aptidão heterogénea em múltiplos ambientes e a perda de plasticidade podem ser ambos considerados efeitos pleiotrópicos de mutações. As mutações na regulação *cis* têm mais probabilidades de limitar estes efeitos pleiotrópicos do que as alterações na codificação. Deste modo, as mutações na regulação *cis* podem ser favorecidas ao longo de maiores escalas de tempo.

Questionamos se existe uma distribuição diferente da regulação *cis*, de mutações de sentido inverso e sem sentido que provocam diferenças fenotípicas em populações domésticas, separando-se dentro de espécies selvagens e entre espécies. Espera-se que estas categorias taxonómicas variem na sua força e duração e talvez noutras variáveis absolutas. A nossa compilação revela que as raças domésticas e as variantes intraespecíficas apresentam diferenças surpreendentes das comparações interespecíficas (Tabela 4).

Trinta e sete por cento das mutações identificadas subjacentes à variação intraespecífica em espécies domésticas e nalgumas espécies selvagens, especialmente a *Arabidopsis thaliana*, provocam a eliminação da função do gene (Tabela 4, Apêndice 1). As mutações nulas em espécies domésticas e nalgumas espécies selvagens têm geralmente grandes efeitos fenotípicos. Muitas destas características estão muito relacionadas com o *fitness*, como o tempo de floração e a taxa de crescimento, por isso é provável que as mutações sofram seleção forte. É um resultado surpreendente, porque a maioria dos genes apresentam uma forte conservação evolutiva ao longo de grandes distâncias taxonómicas. A eliminação da função do gene, principalmente devido a processos de inserção e deleção, é muito inconsistente com a conservação destas sequências de gene através de uma seleção purificante, o que nos leva a concluir que a maioria destas mutações reflete seleção diferente, o que consiste integralmente com a origem recente das espécies domésticas. É também consistente com a recente propagação de algumas espécies, como a *A. thaliana*, em novos habitats e a raridade relativa da maioria dos alelos (Le Corre et al. 2002).

Table 4. Distribution of evolutionary relevant mutations among taxonomic levels.

	Domesticated	Intraspecific	Interspecific ¹	Higher taxonomic level ²
Coding ³	65	122	28	19
<i>Cis</i> -regulatory	23	24	24	3
Other ⁴	11	11	1	0
Total	99	157	53	22
Null ⁵	55	39	3	2

¹Includes recently diverged populations that experience reproductive isolation and are often considered different species, such as divergent stickleback populations.

²Comparisons of species that are not sibling species.

³Includes mutations altering mRNA splicing.

⁴Includes gene duplications, gene losses, stable DNA methylation, and four cases in which the mutations were mapped to a gene but not localized to a coding versus *cis*-regulatory change.

⁵Includes alleles presumed null based on existence of premature stop codons and deletions of part or all of the protein-coding sequence.

In contrast, it is striking that only about 7% of comparisons above the species level (five cases out of 75) have so far identified null alleles (Table 4). Instead, studies above the species level have found mostly coding mutations that modify but do not eliminate protein function and *cis*-regulatory mutations that alter only part of the gene's function (Table 4).

The similarities in the mutational spectra of domesticated species and *Arabidopsis* populations suggest that population structure and history and the strength and duration of selection can influence which mutations are selected by natural selection. This implies one of two things, or perhaps a mixture of both. The first possibility is that domesticated species and, to a certain extent, *A. thaliana* are poor proxies for genetic evolution happening in most wild populations. Alternatively, many wild populations may experience selective regimes for loss of function alleles similar to those experienced by domesticated populations and *A. thaliana* on the short time scale, but these mutations are not fixed. Instead, other mutations must arise that are fixed to cause differences between species. In most cases, both the domesticated populations and the *A. thaliana* populations are still segregating for the ancestral, conserved alleles of the genes causing phenotypic variation, in addition to the derived loss-of-function alleles. These loss-of-function alleles may not ultimately contribute to species differences if they carry pleiotropic fitness costs (Scarcelli et al. 2007). More specific alleles—at the same or different loci—that impart the advantageous effect without the pleiotropic consequences may replace the original allele and eventually become fixed.

We can generate a more specific prediction by combining knowledge of gene network positions and the likely history of selection. Weak selection and selection across multiple environments are more likely to have acted upon species differences than on phenotypic variants under domestication and perhaps in some recently evolved populations, like *Arabidopsis*. It is under these conditions that we expect the population genetics argument to

become more important. We thus expect to observe more *cis*-regulatory mutations in regulatory genes (non-DGB genes) for interspecies comparisons than for domesticated races or recently evolved populations. This hypothesis receives significant support from published studies.

The proportion of *cis*-regulatory mutations identified at various taxonomic ranks for morphological and physiological traits and for DGB and non-DGB genes is shown in Figure 3. Morphological differences at the interspecies level or higher involve significantly more *cis*-regulatory changes than coding changes (Fig. 3A; Table 5). Similarly, for non-DGB genes, phenotypic differences between species involve significantly more *cis*-regulatory mutations than coding mutations (Fig. 3C; Table 5). For physiological traits and for DGB genes, all mutations responsible for intergeneric differences have been found in coding regions (Fig. 3). To reduce potential bias introduced by the discovery of multiple mutations in the same genes in studies following an initial report, we also analyzed a restricted dataset, where only one or two mutations were included per gene. (Two mutations, one coding and one *cis*-regulatory, were included in the analysis only if both coding and *cis*-regulatory mutations were found for a single gene [Appendix 1].) The trends reported above are also observed for this restricted dataset. (compare Fig. 3A with 3B and 3C with 3D). The restricted dataset contains relatively few mutations causing phenotypic evolution between species, which highlights the need for more data. A recent analysis has shown that evolutionary variation in gene expression levels is more often caused by *cis*-regulatory changes in the target gene between species than within species (Wittkopp et al. 2008). This observation is consistent with our analysis of phenotypic differences within and between species.

These observations provide a possible explanation for the disagreement between most evolutionary developmental biologists and Hoekstra and Coyne (2007). Studies of physiological traits

Tabela 4. Distribuição de mutações evolutivas relevantes entre níveis taxonômicos

	Domésticas	Intraespecíficas	Interespecíficas ¹	Nível taxonômico superior ²
Codificação ³	65	122	28	19
Regulação <i>cis</i>	23	24	24	3
Outro ⁴	11	11	1	0
Total	99	157	53	22
Nulo ⁵	55	39	3	2

1 Inclui populações recentemente divergidas que passam por isolamento reprodutiva e são geralmente consideradas espécies diferentes, tais como as populações divergentes do esgana-gata.

2 Comparações de espécies que não são irmãs.

3 Inclui mutações que modificam a ligação ARNm.

4 Inclui duplicação genética, perda de genes, metilação do ADN estável, assim como quatro casos nos quais as mutações foram mapeadas a um gene, mas não foram localizadas na alteração de codificação versus regulação *cis*.

5 Inclui alelos que se presumem nulos, devido à existência de códons de finalização prematuros e deleções parciais ou integrais de uma sequência de codificação de proteína.

Mas é, pelo contrário, surpreendente que apenas cerca de 7% das comparações em níveis superiores ao das espécies (cinco casos em 75) foram identificadas como alelos nulos (Tabela 4). Os estudos em níveis superiores ao da espécie encontraram mais mutações da codificação que modificam, mas não eliminam, a função da proteína, e também mutações na regulação *cis*, que alteram apenas parte das funções do gene (Tabela 4).

Estas similaridades nos espectros mutacionais de espécies domesticadas e de populações *Arabidopsis* sugerem que a estrutura da população e a história e força da duração da seleção possa influenciar as mutações selecionadas por seleção natural. Isto implica uma ou duas coisas, ou talvez um pouco de cada. A primeira indica que as espécies domesticadas e, até certo ponto, a *A. thaliana* são fracamente representantes da evolução genética na maioria das populações selvagens. Em alternativa, muitas das populações selvagens podem passar por regimes seletivos devido à perda de alelos funcionais, um processo similar ao das populações domésticas e a *A. thaliana* embora em menor escala, apesar de estas mutações não estarem fixadas. Pelo contrário, devem surgir outras mutações que estão fixadas a provocar diferenças entre espécies. Na maioria dos casos, tanto as populações domésticas como as populações de *A. thaliana* continuam a segregar para os alelos ancestrais e conservados dos genes provocando variação fenotípica, além de perda de alelos daí derivada. Esta perda de função dos alelos pode, afinal, não contribuir para as diferenças das espécies se transportarem os custos do fitness pleiotrópico (Scarcelli et al. 2007). Os alelos mais específicos – no mesmo locus ou em diferentes loci – que transmitem efeitos vantajosos sem consequências pleiotrópicas podem substituir o alelo original e acabam por ficar fixados.

É possível gerar uma previsão mais específica combinando o conhecimento das posições das redes de genes e o histórico provável da seleção. A seleção fraca e a seleção em múltiplos ambientes têm maiores probabilidades de atuarem em diferenças das espécies do que em variantes fenotípicas, ou talvez também em algumas populações que evoluíram recentemente, como a *Arabidopsis*. É por causa destas condições que se espera que o argumento da genética de populações se torne mais importante. Deste modo, espera-se

observar mais mutações de regulação *cis* em genes reguladores (genes não-DGB) nas comparações interespecíficas do que nas raças domésticas ou populações recentemente evoluídas. Esta hipótese obteve grande apoio em estudos publicados.

A proporção de mutações na regulação *cis* identificadas em vários níveis taxonômicos pelas características morfológicas e fisiológicas e genes DGB e não-DGB está presente na Figura 3. As diferenças morfológicas a nível interespecífico ou superior, envolvem um número significativamente maior de alterações na regulação *cis* do que alterações na codificação (Fig. 3 A; Tabela 5). Do mesmo modo, as diferenças fenotípicas entre espécies para genes não-DGB envolvem muitas mais mutações na regulação *cis* do que na codificação (Fig. 3C; Tabela 5). Relativamente às características fisiológicas e genes DGB, todas as mutações responsáveis por diferenças intergenéricas foram encontradas nas regiões de codificação (Fig. 3). Foi também analisado um conjunto de dados mais restrito, nos quais foram incluídas apenas uma ou duas mutações, de modo a reduzir o potencial viés proveniente da descoberta de múltiplas mutações nos mesmos genes, em estudos feitos após uma investigação inicial. (Só foram incluídos na análise duas mutações, uma na codificação e uma na regulação *cis*, caso a codificação e a regulação *cis* surgissem num único gene [Apêndice 1]). As tendências acima referidas são também estudadas neste conjunto de dados restrito. (comparar a Fig. 3A com 3B e a 3C com a 3D). O conjunto de dados restrito contém menos mutações que provocam a evolução fenotípica entre espécies, o que destaca a necessidade de dados adicionais. Uma análise recente demonstrou que a variação evolutiva nos níveis de expressão do gene é provocada mais frequentemente por alterações na regulação *cis* do gene alvo entre espécies do que dentro das espécies (Wittkopp et al. 2008). Esta observação é consistente com a nossa análise das diferenças fenotípicas dentro e entre espécies.

Estas observações fornecem uma possível explicação para o desacordo existente entre a maioria dos biólogos especialistas em evolução e desenvolvimento e Hoekstra e Coyne (2007). Os estudos das características fisiológicas

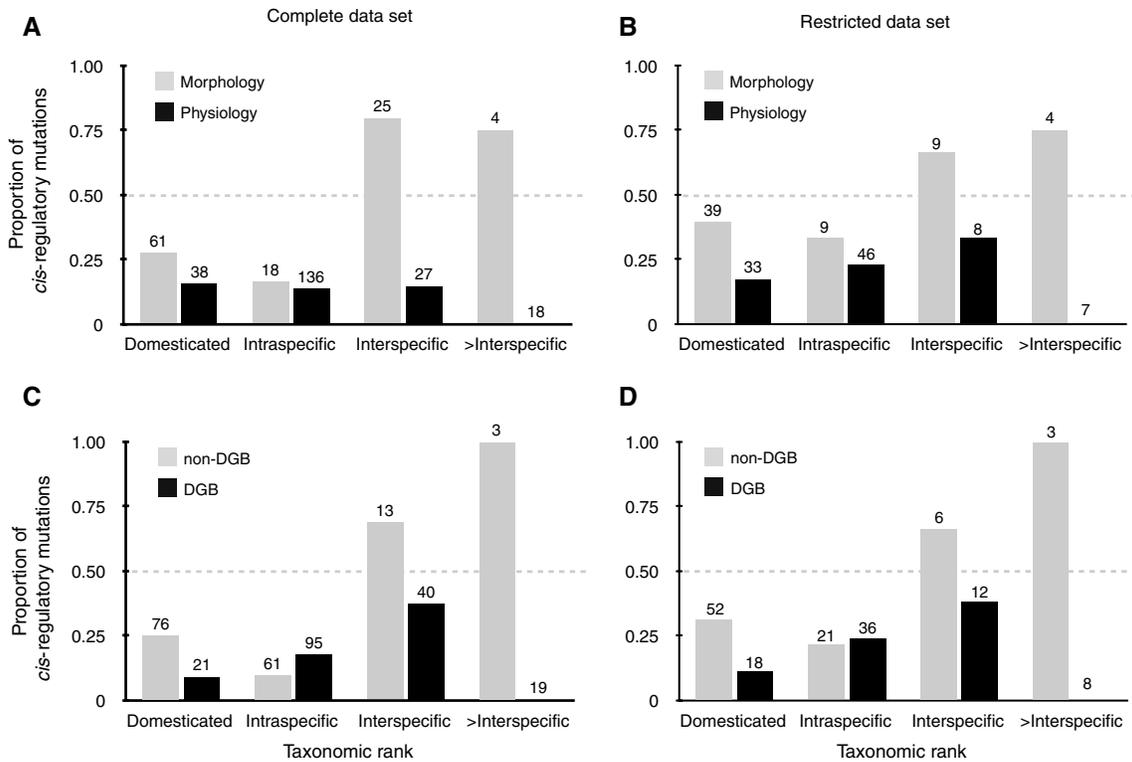


Figure 3. Evolutionarily relevant *cis*-regulatory mutations are more frequently found in interspecific comparisons than in intraspecific comparisons or among domesticated races. (A) The proportion of all mutations that are *cis*-regulatory mutations for morphological and physiological traits in the complete dataset. (B) Proportion of *cis*-regulatory mutations for morphological and physiological traits in the restricted dataset, where only one or two mutations per gene were included. Two mutations were included only if both coding and *cis*-regulatory mutations were found for a single gene. (C) Proportion of *cis*-regulatory mutations for DGB versus non-DGB genes in the complete dataset. (D) Proportion of *cis*-regulatory mutations for DGB versus non-DGB genes in the restricted dataset. The total number of mutations for each category is shown above the bars.

and DGB genes between species and of all genes within species have usually provided evidence for a predominance of coding changes. Conversely, comparisons of developmental regulatory genes across species have provided support for the prevalence of

cis-regulatory changes causing morphological evolution. The data make more sense when both gene function and population genetics are considered simultaneously. The current evidence suggests that strong selection often results in selection of mutations with

Table 5. Statistical comparisons of the frequency of *cis*-regulatory and coding mutations for different phenotypic classes, different gene-network classes and different taxonomic levels.

	Complete dataset			Restricted dataset		
	G ¹	P ²	Intraspecific vs. interspecific ³ Fisher's exact ⁴	G ¹	P ²	Intraspecific vs. interspecific ³ Fisher's exact ⁴
Morphology	25.9	<0.00002	<0.00007	6.2	<0.11	<0.35
Physiology	6.5	<0.09	<0.45	4.0	<0.26	1
DGB	16.8	<0.0008	<0.42	6.2	<0.11	1
Non-DGB	27.0	<0.000006	<0.000002	7.6	<0.06	<0.05

¹Value of G test of independence for the number of *cis*-regulatory versus coding mutations for domesticated, intraspecific, interspecific, and intergeneric taxonomic levels.

²P values for all G tests of independence were calculated using three degrees of freedom.

³A test of the frequency of *cis*-regulatory versus coding mutations in intraspecific versus interspecific populations. Data from domesticated races were excluded and the interspecific and higher taxonomic level data were pooled.

⁴The P value for a Fisher's exact test of independence is reported.

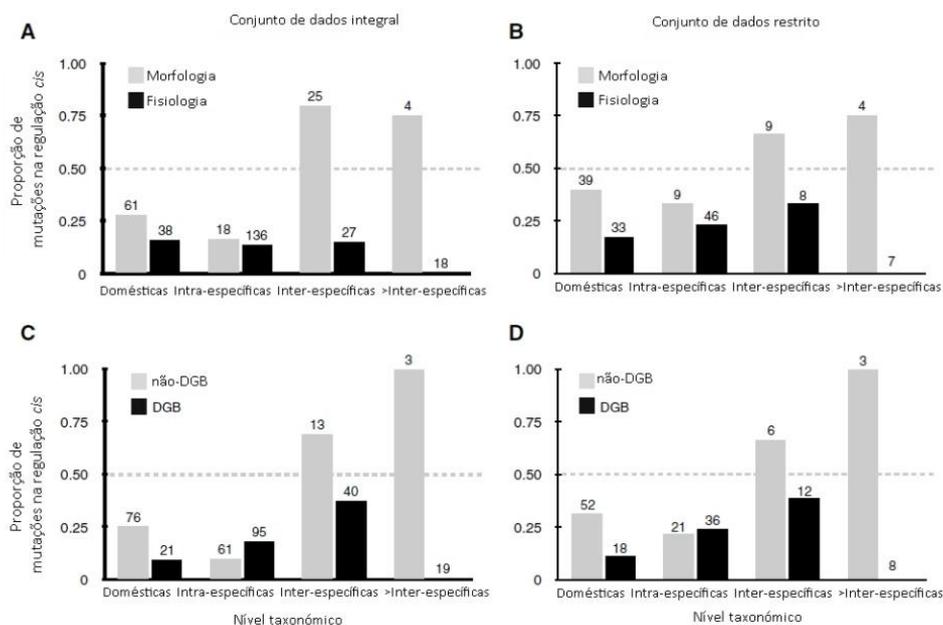


Figura 3. As mutações na regulação *cis*, evolutivamente relevantes, surgem mais frequentemente em comparações interespecíficas do que em comparações intraespecíficas ou entre raças domésticas. (A) Proporção de todas as mutações que são mutações na regulação *cis* nas características morfológicas e fisiológicas no conjunto de dados integral. (B) Proporção de todas as mutações na regulação *cis* nas características morfológicas e fisiológicas no conjunto de dados restrito, no qual foram incluídos apenas uma ou duas mutações por gene. Foram incluídas duas mutações, apenas se as mutações na codificação e na regulação *cis* fossem encontrados num único gene. (C) Proporção de mutações na regulação *cis* em DGB versus genes não-DGB no conjunto de dados integral. (D) Proporção de mutações na regulação *cis* em DGB versus genes não-DGB no conjunto de dados restrito. O número total de mutações em cada categoria é apresentado acima das barras.

e dos genes DGB entre espécies e os estudos de todos os genes dentro das espécies têm geralmente fornecido evidências da predominância das alterações na codificação. Já a comparação entre genes reguladores do desenvolvimento entre espécies ajuda a defender a prevalência das alterações

na regulação *cis* que provocam a evolução morfológica. Os dados fazem mais sentidos se se considerar simultaneamente a função dos genes e a genética de populações. As evidências atuais sugerem que uma seleção forte resulta geralmente na seleção de mutações com

Tabela 5. Comparação estatística da frequência das mutações na regulação *cis* e mutações de codificação em diferentes classes fenotípicas, diferentes classes de redes de genes e diferentes níveis taxonômicos.

	Conjunto de dados integral				Conjunto de dados restrito			
	G^1	P^2	Intraespecíficas vs Interespecíficas ³ Teste Exato de Fisher ⁴		G^1	P^2	Intraespecíficas vs Interespecíficas ³ Teste Exato de Fisher ⁴	
Morfologia	25,9	<0,00002	<0,00007		6,2	<0,11	<0,35	
Fisiologia	6,5	<0,09	<0,45		4,0	<0,26	1	
DGB	16,9	<0,0008	<0,42		6,2	<0,11	1	
Não-DGB	27,0	<0,000006	<0,000002		7,6	<0,06	<0,35	

¹ Valor de teste G de independência para o número de mutações na regulação *cis* versus mutações na codificação a nível doméstico, intraespecífico, interespecífico, e taxonômico intergenérico.

² O valor P para todos os testes G de independência foi calculado utilizando três graus de liberdade.

³ Teste da frequência das mutações na regulação *cis* versus mutações na codificação em populações intraespecíficas versus interespecíficas. Os dados das raças domésticas foram excluídos e os dados relativos aos níveis interespecíficos e taxonômico foram reunidos.

⁴ Indicação do valor P para o Teste Exato de Fisher de independência.

strong and pleiotropic effects, such as those often observed in domesticated populations and in *A. thaliana*. Conversely, evolution over longer time periods apparently leads to fixation of mutations with more subtle and specific effects. Incorporating these observations into a coherent explanation for genetic evolution is an outstanding problem in evolutionary biology.

PREDICTION 3: THE POSITION OF A GENE WITHIN A REGULATORY NETWORK SHOULD IMPACT THE DISTRIBUTION OF EVOLUTIONARILY RELEVANT MUTATIONS AMONG GENES IN THE NETWORK

As discussed above, we believe that the position of a gene in a regulatory network is an important parameter to consider when determining whether *cis*-regulatory or coding mutations are more likely to contribute to phenotypic evolution. In this section we argue that the structure of regulatory networks may also influence which genes in the network are more likely to accumulate evolutionarily relevant mutations. This is a vast topic (Davidson 2006) and we focus here on particular parts of regulatory networks, where a single transcription factor serves as a key regulator of cell differentiation. These cases allow us to see most clearly how developmental regulatory architecture might help us to predict the genetic causes of phenotypic evolution.

This idea is best explained with two examples from *Drosophila*: trichome patterning and bristle patterning. Trichomes are cuticular extensions produced by insect epidermal cells. Larval trichomes may aid movement (Inestrosa et al. 1996). Bristles are pluricellular sensory organs produced through cell division of a single sensory precursor cell selected from a field of epidermal cells (Lai and Orgogozo 2004). Trichomes and bristles are produced during development at specific positions on the fly body (Fig. 4). Summaries of the regulatory networks that generate the final pattern of trichomes and bristles are shown in Figure 5.

For both trichomes and bristles, all of the information from patterning genes is ultimately integrated within each cell by a single gene: *shavenbaby/ovo* (*svb*) for most trichomes and *scute* for most bristles. We can consider *svb* and *scute* as input/output devices (Davidson and Erwin 2006). They integrate an extensive array of inputs, the regulatory state, and they produce an on or off transcriptional output. *Svb* and *scute* are transcription factors that each regulate at least dozens of terminal differentiation genes. Expression of these input/output device genes determines whether a cell differentiates a trichome, a bristle, or smooth cuticle. They therefore occupy bottleneck positions in their respective gene regulatory networks. All patterning information must flow through them and they then regulate multiple downstream genes.

Multiple changes in the pattern of trichomes and bristles have occurred during fly evolution. Based on our current understanding of developmental regulatory networks, we predict that most of these evolutionary changes have probably occurred through *cis*-

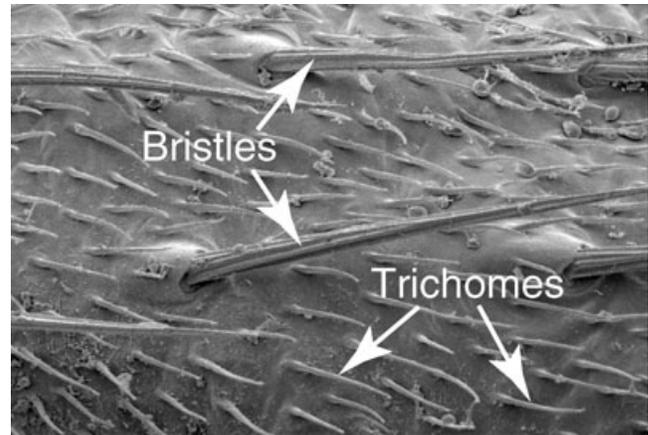


Figure 4. Scanning electron micrograph of trichomes and bristles on a leg of *Drosophila melanogaster*. Trichomes are nonsensory cuticular extensions. Bristles are sensory organs innervated by single neurons.

regulatory mutations at *svb* and *scute*, respectively. The reasoning is as follows. Mutations causing evolutionary changes in trichome or bristle position are likely to result from changes in genes already involved in trichome and bristle development, respectively. Because genes that act downstream of *svb* and *scute* must act in combination with other genes to generate a trichome or bristle (Hartenstein 2004; Chanut-Delalande et al. 2006), mutations in these genes are unlikely to produce a change in trichome or bristle position (mutations in these genes may alter trichome or bristle shape or size). Mutations in patterning genes acting upstream of *svb* or *scute*, whether *cis*-regulatory or coding, are also unlikely to be favored because they will alter development of other structures. For example, the genes regulating *svb* and *scute* expression determine the pattern of multiple epidermal structures and features in addition to trichomes and sensory bristles: muscle attachment cells, oenocytes, epidermal glands, cuticle pigmentation, etc. (Calleja et al. 2002). Therefore, *cis*-regulatory mutations in *svb* and *scute* are likely to have the most specific, least pleiotropic effects of any mutations in any genes that might alter the pattern of trichomes and bristles.

Cis-regulatory changes have been shown to cause a loss of dorsal larval trichomes in *D. sechellia* for *svb* (Sucena and Stern 2000; McGregor et al. 2007) and a gain of thoracic bristles in *D. quadrilineata* for *scute* (Marcellini and Simpson 2006). Furthermore, other changes in trichome and bristle patterns have been shown to correlate with changes in *svb* expression (Sucena et al. 2003) and *scute* expression (Wülbeck and Simpson 2000; Pistillo et al. 2002; Skaer et al. 2002b), respectively, whereas genes that act upstream of *svb* and *scute* show unchanged patterns of expression (Dickinson et al. 1993; Wülbeck and Simpson 2002; Richardson and Simpson 2006; Simpson et al. 2006).

The positions of *svb* and *scute* in their respective networks are like the positions of light switches in an electrical circuit. There are

efeitos fortes e pliotrópicos, tais como os observados em populações domésticas e na *A. thaliana*. A evolução sob longos períodos de tempo, pelo contrário, parece conduzir à fixação das mutações com efeitos mais subtis e específicos. Um grande problema da biologia evolutiva é incorporar estas observações numa explicação coerente.

PREVISÃO 3: A POSIÇÃO DE UM GENE NUMA REDE DE REGULAÇÃO DEVE IMPACTAR A DISTRIBUIÇÃO DE MUTAÇÕES EVOLUTIVAMENTE RELEVANTES ENTRE OS GENES DA REDE.

Tal como discutido anteriormente, acreditamos que a posição dum gene numa rede de regulação é um parâmetro importante a considerar quando se determina se são as mutações na regulação *cis* ou na codificação que têm mais probabilidades de contribuir para a evolução fenotípica. Nesta secção argumentamos que a estrutura das redes de regulação pode também influenciar que genes da rede têm mais probabilidades de acumular mutações evolutivamente relevantes. Trata-se de um tópico vasto (Davidson 2006), mas concentrar-nos-emos em determinadas partes das redes de regulação, nas quais um único fator de transcrição serve de regulação chave para a diferenciação celular. Estes casos permitem perceber mais claramente como é que a arquitetura reguladora do desenvolvimento pode ajudar a prever as causas genéticas da evolução fenotípica.

Este conceito pode ser explicado através de dois exemplos da *Drosophila*: padronização de tricomas e padronização de pelos. Os tricomas são extensões cuticulares produzidas por células epidérmicas de insetos. Nas larvas, os tricomas podem ajudar o movimento (Inestrosa et al. 1996). Os pelos são órgãos sensoriais multicelulares produzidos através da divisão celular de uma só célula sensorial precursora, selecionada a partir de um conjunto de células epidermais (Lai e Orgogozo 2004). Os tricomas e os pelos são produzidos durante o desenvolvimento, em posições específicas do corpo da mosca (Fig. 4). Os sumários das redes de regulação que geram o padrão final dos tricomas e pelos estão presentes na Figura 5.

Nos tricomas e pelos, toda a informação dos genes padronizantes é integrada dentro de cada célula por um único gene: *shavenbaby* / *ovo* (*svb*), na maioria dos tricomas, e *scute*, na maioria dos pelos. O *svb* e o *scute* podem ser considerados aparelhos de input/output (Davidson e Erwin 2006). Estes integram uma vasta seleção de inputs, ou seja, o estado regulador, e produzem um output transcripcional que pode estar ligado ou desligado. O *svb* e o *scute* são fatores de transcrição, e cada um regula no mínimo dezenas de genes de diferenciação terminal. A expressão destes genes com input/output determina se uma célula diferencia um tricoma, um pelo ou uma cutícula lisa. Por conseguinte, ocupam posições de engarrafamento nas respetivas redes de regulação. Toda a informação padronizante deve circular entre eles, para que depois regulem múltiplos genes afastados da fonte de regulação.

Já ocorreram múltiplas alterações no padrão de tricomas e pelos durante a evolução da mosca. Com base na nossa atual compreensão das redes de regulação, prevemos que a maioria

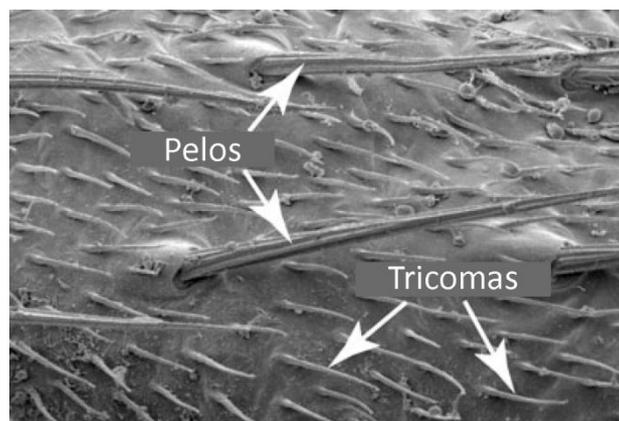


Figura 5. Micrografia eletrônica de tricomas e pelos numa pata da *Drosophila melanogaster*. Os tricomas são extensões cuticulares não-sensoriais. Os pelos são órgãos sensoriais inervados por neurónios individuais.

destas alterações evolutivas tenha ocorrido, provavelmente, através de mutações da regulação *cis* no *svb* e no *scute*, respetivamente. A argumentação é a seguinte. As mutações que provocam alterações evolutivas nas posições dos tricomas ou pelos resultam provavelmente de alterações em genes já envolvidos no desenvolvimento de tricomas e pelos, respetivamente. Devido ao facto de os genes que atuam longe do regulador do *svb* e *scute* terem de atuar em conjunto com outros genes para gerar um tricoma ou um pelo, (Hartenstein 2004; Chanut-Delalande et al. 2006), as mutações nestes genes têm poucas probabilidades de produzir uma alteração na posição do tricoma ou do pelo (as mutações nestes genes podem alterar a forma ou tamanho do tricoma ou pelo). As mutações em genes padronizantes precursores do *svb* ou *scute*, quer se deem na regulação *cis* ou na codificação, têm também poucas probabilidades de serem favorecidas, pois irão alterar o desenvolvimento de outras estruturas. Os genes que regulam a expressão *svb* e *scute*, por exemplo, determinam o padrão das múltiplas estruturas e características epidérmicas e contêm, além dos tricomas e pelos sensoriais, células de ligação muscular, oenócitos, glândulas epidérmicas, pigmentação da cutícula, etc. (Calleja et al. 2002). Por conseguinte, é provável que as mutações na regulação *cis* do *svb* e *scute* tenham efeitos mais específicos e menos pliotrópicos de qualquer mutação em qualquer gene capaz de alterar o padrão dos tricomas e pelos.

As alterações na regulação *cis* já demonstraram provocar uma perda de tricomas dorsais nas larvas da *D. sechellia* no *svb* (Sucena e Stern 2000; McGregor et al. 2007) e um ganho de pelos torácicos na *D. quadrilineata* no *scute* (Marcellini e Simpson 2006). Além disso, foi demonstrado que outras alterações nos padrões de tricomas e pelos correlacionam-se com alterações na expressão do *svb* (Sucena et al. 2003) e na expressão do *scute* (Wülbeck e Simpson 2000; Pistillo et al. 2002; Skaer et al. 2002b), respetivamente, enquanto os genes que próximos da fonte do *svb* e *scute* apresentam padrões de expressão inalterados (Dickinson et al. 1993; Wülbeck e Simpson 2002; Richardson e Simpson 2006; Simpson et al. 2006).

A posição do *svb* e do *scute* nas suas respetivas redes pode ser comparada à posição dos interruptores de luz num circuito elétrico. Existem

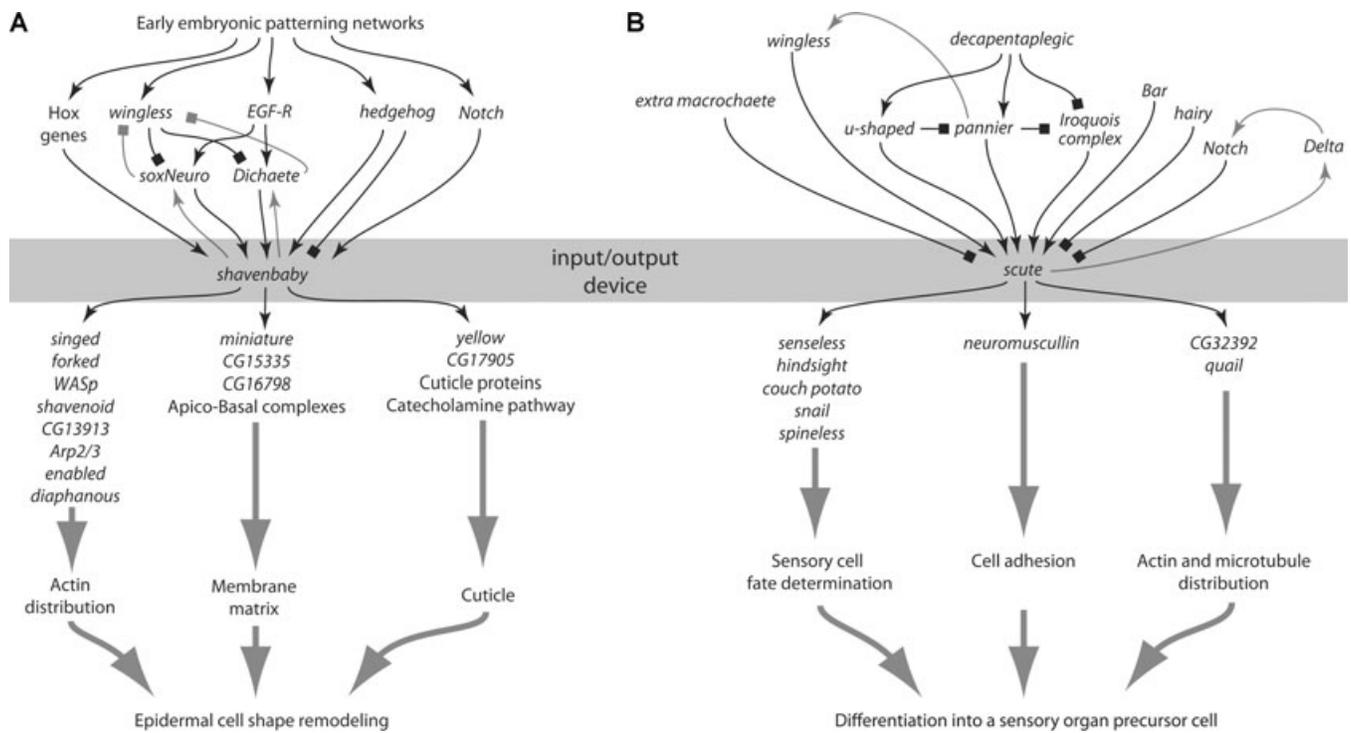


Figure 5. Partial regulatory networks patterning (A) trichomes (modified from results in Chanut-Delalande et al. 2006; Overton et al. 2007) and (B) bristles in *Drosophila melanogaster* (modified from Calleja et al. 2002; Hartenstein 2004).

multiple ways to turn off a light in a room. One could shut down the power generating station, cut the power line to the house, shut off the main breaker in a house, cut wires leading to the switch, flip the light switch, or break the light bulb. Clearly, of all of these options, flipping the light switch is both the most specific and the easiest to reverse. *Svb* is the switch that flips trichomes on or off and *scute* is the switch that flips bristles on or off. The multiple *cis*-regulatory modules of each gene are like the individual light switches in each room of a house. They provide great precision in evolutionary changes with minimal or no pleiotropic effects.

These examples illustrate the explanatory power that arises from a detailed understanding of the molecular mechanisms underlying cell differentiation and development. Of course, our “prediction” based on regulatory networks is really a post hoc explanation developed only after accumulating several pieces of evidence that support the predominance of *cis*-regulatory mutations in *svb* and *scute*. A real test of this hypothesis requires study of additional examples of trichome and bristle pattern evolution. But the true value of a predictive theory of genetic evolution will emerge only when novel predictions are made based on an understanding of other regulatory networks that are then tested with studies of natural variation.

The concept of an input/output gene resembles the concept of “cell-type specific selector gene” (Garcia-Bellido 1975; Mann and Carroll 2002) and is based on a more detailed understanding of development regulatory networks (Davidson 2006). Recogni-

tion of such a gene category was fundamental to the prediction discussed here. New compelling predictions for genetic evolution will probably require the development of new concepts and new gene categories based on a more detailed understanding of developmental biology.

Conclusions

In the absence of a population genetics framework, evolutionary developmental biologists have inferred from (1) our current understanding of gene regulatory networks, (2) our understanding of gene structure and function, and (3) the extensive conservation of developmental genes, that mutations in the *cis*-regulatory regions of developmental patterning genes are likely to underlie most of phenotypic evolution. However, no single argument proposed so far provides definitive proof that *cis*-regulatory mutations constitute the predominant cause of phenotypic evolution. By considering development and population genetics simultaneously, a survey of published data suggests that patterns of genetic evolution are not entirely obscured by historical contingency. Population genetics and development must be considered simultaneously to make sense of the data.

It may be unhelpful to pose the coding versus *cis*-regulatory debate as a quantitative question: do coding changes explain more of phenotypic evolution than *cis*-regulatory changes? It may be more productive to turn the question around and ask what kinds

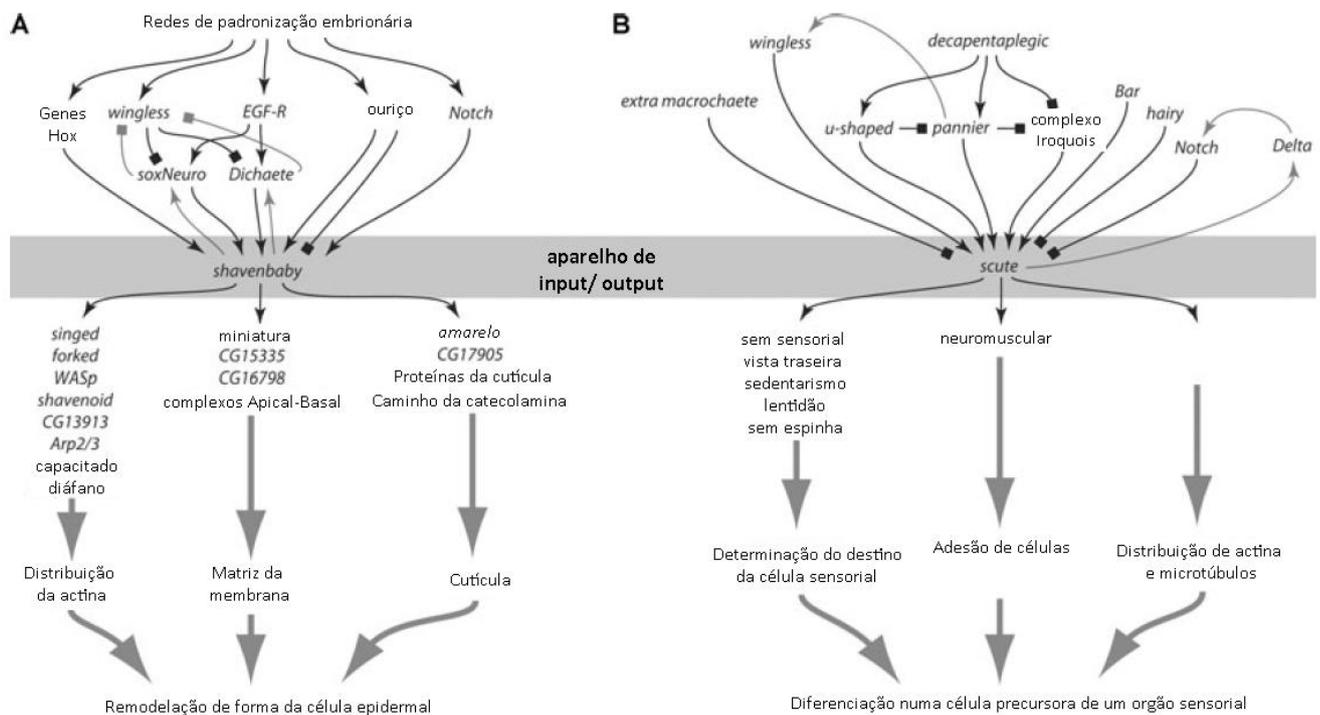


Figura 5. Padronização parcial de redes de regulação (A) tricomas (alteração dos resultados de Chanut-Delalande et al. 2006; Overton et al. 2007) e (B) pelos na *Drosophila melanogaster* (alteração de Callega et al. 2002; Hartenstein 2004).

várias maneiras de desligar a luz num quarto. Pode-se desligar a central elétrica, cortar as linhas elétricas da casa, desligar o disjuntor, cortar os fios dum comutador, desligar o interruptor, ou partir uma lâmpada. De todas estas opções, desligar o interruptor é claramente o método mais específico e mais fácil de inverter. O *svb* é o interruptor que liga ou desliga os tricomas e o *scute* é o interruptor que liga ou desliga os pelos. Os múltiplos módulos de regulação *cis* de cada gene são como interruptores individuais em cada divisão dum casa. Fornecem uma elevada precisão nas alterações evolutivas, com poucos ou nenhuns efeitos pleiotrópicos.

Estes exemplos ilustram o poder explanatório que surge de uma compreensão detalhada dos mecanismos moleculares subjacentes à diferenciação e desenvolvimento celular. É claro que a nossa “previsão” baseada em redes de regulação é, na verdade, uma explicação detalhada, desenvolvida somente depois de se acumularem várias evidências que apoiam a predominância das mutações na regulação *cis* no *svb* e no *scute*. Um teste real deste hipótese requer o estudo adicional de exemplos da evolução de tricomas e pelos. Mas o verdadeiro valor de uma teoria predizível da evolução genética só será comprovado quando forem executadas novas previsões com base na compreensão de outras redes de regulação, que são depois testadas com estudos de variação natural.

O conceito de gene de input/output assemelha-se ao conceito de “gene seletor específico do tipo de célula” (Garcia-Bellido 1975; Mann e Carrol 2002) e é baseado numa compreensão mais detalhada de redes de regulação do desenvolvimento (Davidson 2006). O reconhecimento desta

categoria de genes foi fundamental para a previsão aqui discutida. Novas e estimulantes previsões da evolução genética irão provavelmente necessitar do desenvolvimento de novos conceitos e novas categorias de genes, baseados numa compreensão mais detalhada da biologia do desenvolvimento.

Conclusões

Na ausência de um contexto de genética de populações, os biólogos especialistas em evolução e desenvolvimento demonstraram através da (1) atual compreensão das redes dos genes reguladores, (2) da compreensão da estrutura e função dos genes, e (3) da conservação abrangente de genes do desenvolvimento, as mutações nas regiões de regulação *cis* dos genes padronizantes do desenvolvimento estão, provavelmente, na base de grande parte da evolução fenotípica. No entanto, nenhum argumento proposto até à data fornece uma prova definitiva de que as mutações na regulação *cis* constituem a causa predominante da evolução fenotípica. Ao considerar o desenvolvimento e a genética de populações em simultâneo, um levantamento de dados publicados sugere que os padrões da evolução genética não estão completamente ofuscados pelo contingente histórico. A genética de populações e o desenvolvimento devem ser considerados em simultâneo, para dar sentido aos dados.

Pode ser inútil centrar o debate da codificação versus regulação *cis* na questão quantitativa: as alterações na codificação explicam melhor a evolução fenotípica que as alterações na regulação *cis*? Talvez seja mais produtivo redirecionar a questão no sentido de compreender que tipos

of phenotypic changes are expected from particular coding versus *cis*-regulatory changes in specific genes. As we show, patterns in the currently available data imply that morphological and physiological traits are caused by different frequencies of coding and *cis*-regulatory changes. This is consistent with our molecular understanding of how coding and *cis*-regulatory changes might influence physiology and morphology.

We also found that different spectra of evolutionarily relevant mutations segregate within populations and between species. Interspecific differences in morphology seem to be more often caused by *cis*-regulatory changes than intraspecific variation. This result is not predicted by a traditional neo-Darwinian view of the contribution of intraspecific variation to interspecific differences. Instead, it appears that evolution over longer time scales results in fixation of a specific subset of the genetic variation contributing to intraspecific phenotypic variation.

By fusing developmental and evolutionary genetics, evolutionary biologists may be able to predict, in a probabilistic sense, the mutations underlying phenotypic evolution. Fortunately, scientists are rapidly identifying the genetic causes of phenotypic evolution, providing abundant data for testing new predictions about the genetic basis of evolution.

ACKNOWLEDGMENTS

We thank C. Lee and J. Doebley for bringing several relevant papers to our attention, P. Simpson and F. Schweisguth for their advice on Figure 5 and H. Hoekstra and members of Marie-Anne Felix's laboratory for helpful comments on the manuscript. We are very grateful to the two reviewers of our article, J. Doebley and G. Wagner, who both provided extensive helpful comments and revealed their identity. Our research is supported by grants from NIH (GM063622-06A1), NSF (IOS-0640339), and the David & Lucile Packard Foundation to DLS and from the CNRS to VO.

LITERATURE CITED

- Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant, and C. J. Tabin. 2004. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* 305:1462–1465.
- Adamowicz, S. J., and A. Purvis. 2006. From more to fewer? Testing an allegedly pervasive trend in the evolution of morphological structure. *Evol. Int. J. Org. Evol.* 60:1402–1416.
- Akam, M. 1995. Hox genes and the evolution of diverse body plans. *Philos. Trans. R. Society Lond. B* 349:313–319.
- . 1998. Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *Int. J. Develop. Biol.* 42:445–451.
- Alonso, C. R., and A. S. Wilkins. 2005. The molecular elements that underlie developmental evolution. *Nat. Rev. Genet.* 6:709–715.
- Andersson, L., and M. Georges. 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* 5:202–212.
- Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149–1152.
- . 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res.* 17:1755–1762.
- Averof, M., R. Dawes, and D. Ferrier. 1996. Diversification of arthropod Hox genes as a paradigm for the evolution of gene functions. *Cell Develop. Biol.* 7:539–551.
- Baatz, M., and G. P. Wagner. 1997. Adaptive inertia caused by hidden pleiotropic effects. *Theor. Popul. Biol.* 51:49–66.
- Barton, N. H. 1990. Pleiotropic models of quantitative variation. *Genetics* 124:773–782.
- Birney, E., J. A. Stamatoyannopoulos, A. Dutta, R. Guigo, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.
- Boucher, C. A., E. O'Sullivan, J. W. Mulder, C. Ramautarsing, P. Kellam, G. Darby, J. M. Lange, J. Goudsmit, and B. A. Larder. 1992. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. *J. Infect. Dis.* 165:105–110.
- Britten, R. J., and E. H. Davidson. 1969. Gene regulation for higher cells: a theory. *Science (New York, N.Y.)* 165:349–357.
- . 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46:111–133.
- Bürger, R., and W. J. Ewens. 1995. Fixation probabilities of additive alleles in diploid populations. *J. Math. Biol.* 33:557–575.
- Calleja, M., O. Renaud, K. Usui, D. Pistillo, G. Morata, and P. Simpson. 2002. How to pattern an epithelium: lessons from achaete-scute regulation on the notum of *Drosophila*. *Gene* 292:1–12.
- Carroll, S. B. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–485.
- . 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245.
- Chanut-Delalande, H., I. Fernandes, F. Roch, F. Payre, and S. Plaza. 2006. Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol.* 4:e290.
- Chen, L., A. L. DeVries, and C. H. Cheng. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. USA* 94:3811–3816.
- Cooper, T. F., E. A. Ostrowski, and M. Travisano. 2007. A negative relationship between mutation pleiotropy and fitness effect in yeast. *Evol. Int. J. Org. Evol.* 61:1495–1499.
- Coyne, J. A., and H. E. Hoekstra. 2007. Evolution of protein expression: new genes for a new diet. *Curr. Biol.* 17:R1014–R1016.
- Crow, J. F. 1956. Genetics of DDT resistance in *Drosophila*. International Genetics Symposia. The Organizing Committee, International Genetics Symposia, Tokyo, Japan.
- Darwin, C. 1859. The origin of species by means of natural selection or the preservation of favored races in the struggle for life. The Modern Library, New York.
- Davidson, E. H. 2001. Genomic regulatory systems. Academic Press, San Diego.
- . 2006. The regulatory genome: gene regulatory networks in development and evolution. Academic Press, Burlington.
- Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311:796–800.
- Davis, G. K., D. G. Srinivasan, P. J. Wittkopp, and D. L. Stern. 2007. The function and regulation of *Ultrabithorax* in the legs of *Drosophila melanogaster*. *Dev. Biol.* 308:621–631.
- Dickinson, W. J., Y. Tang, K. Schuske, and M. Akam. 1993. Conservation of molecular prepatterning during the evolution of cuticle morphology in *Drosophila* larvae. *Evolution* 47:1396–1406.
- Dudley, A. M., D. M. Janse, A. Tanay, R. Shamir, and G. M. Church. 2005. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* 1:2005 0001.

de alterações fenotípicas são esperadas em determinadas alterações na codificação versus regulação *cis*, em genes específicos. Como demonstramos, os padrões existentes nos dados atualmente disponíveis sugerem que as características morfológicas e fisiológicas são provocadas por diferentes frequências das alterações na codificação e na regulação *cis*. Isto é coerente com a nossa compreensão a nível molecular de como as alterações na codificação e na regulação *cis* podem influenciar a fisiologia e a morfologia.

Descobrimos também que os diferentes espectros de mutações evolutivamente relevantes são separados dentro de populações e entre espécies. As diferenças interespecíficas na morfologia parecem ser provocadas mais frequentemente pelas alterações na regulação *cis* do que pela variação intraespecífica. Este resultado não está previsto na tradicional abordagem neo-Darwiniana da contribuição de variação intraespecífica para as diferenças interespecíficas. Ao invés disso, a evolução em longos períodos de tempo resulta na fixação de um subconjunto específico da variação genética, que contribui para a variação fenotípica intraespecífica.

Ao unir as genéticas da evolução e do desenvolvimento, os biólogos evolutivos poderão ser capazes de prever, de maneira probabilística, as mutações que estão na base da evolução fenotípica. Felizmente, os cientistas estão rapidamente a identificar as causas genéticas da evolução fenotípica, e fornecem abundantes dados para testar novas previsões sobre a base genética da evolução.

AGRADECIMENTOS

Agradecemos a C. Lee e J. Doebley por chamarem a nossa atenção para vários artigos relevantes; P. Simpson e F. Schweisguth pelo aconselhamento da Figura 5 e H. Hoekstra e membros do laboratório Marie-Anne Felix pelos úteis comentários ao manuscrito. Estamos muito agradecidos aos dois revisores do nosso artigo, J. Doebley e G. Wagner, que nos deram exaustivos comentários úteis e revelaram a sua identidade. A nossa pesquisa foi apoiada por bolsas da NIH (GM063622-06A1), da NSF (IOS-0640339), e da David & Lucille Packard Foundation to DLS Development and Learning Sciences, e ao CNRS através do VO (Virtual Observatory).

Referências Bibliográficas

Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant, and C. J. Tabin. 2004. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* 305:1462–1465.

Adamowicz, S. J., and A. Purvis. 2006. From more to fewer? Testing an allegedly pervasive trend in the evolution of morphological structure. *Evol. Int. J. Org. Evol.* 60:1402–1416.

Akam, M. 1995. Hox genes and the evolution of diverse body plans. *Philos. Trans. R. Society Lond. B* 349:313–319.

———. 1998. Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *Int. J. Develop. Biol.* 42:445–451.

Alonso, C. R., and A. S. Wilkins. 2005. The molecular elements that underlie developmental evolution. *Nat. Rev. Genet.* 6:709–715.

Andersson, L., and M. Georges. 2004. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* 5:202–212.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149–1152.

———. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions

in the *Drosophila melanogaster* genome. *Genome Res.* 17:1755–1762.

Averof, M., R. Dawes, and D. Ferrier. 1996. Diversification of arthropod Hox genes as a paradigm for the evolution of gene functions. *Cell Develop. Biol.* 7:539–551.

Baatz, M., and G. P. Wagner. 1997. Adaptive inertia caused by hidden pleiotropic effects. *Theor. Popul. Biol.* 51:49–66.

Barton, N. H. 1990. Pleiotropic models of quantitative variation. *Genetics* 124:773–782.

Birney, E., J. A. Stamatoyannopoulos, A. Dutta, R. Guigo, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.

Boucher, C. A., E. O'Sullivan, J. W. Mulder, C. Ramautarsing, P. Kellam, G. Darby, J. M. Lange, J. Goudsmit, and B. A. Larder. 1992. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. *J. Infect. Dis.* 165:105–110.

Britten, R. J., and E. H. Davidson. 1969. Gene regulation for higher cells: a theory. *Science (New York, N.Y.)* 165:349–357.

———. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46:111–133.

Bürger, R., and W. J. Ewens. 1995. Fixation probabilities of additive alleles in diploid populations. *J. Math. Biol.* 33:557–575.

Calleja, M., O. Renaud, K. Usui, D. Pistillo, G. Morata, and P. Simpson. 2002. How to pattern an epithelium: lessons from achaete-scute regulation on the notum of *Drosophila*. *Gene* 292:1–12.

Carroll, S. B. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–485.

———. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245.

Chanut-Delalande, H., I. Fernandes, F. Roch, F. Payre, and S. Plaza. 2006. Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol.* 4:e290.

Chen, L., A. L. DeVries, and C. H. Cheng. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. USA* 94:3811–3816.

Cooper, T. F., E. A. Ostrowski, and M. Travisano. 2007. A negative relationship between mutation pleiotropy and fitness effect in yeast. *Evol. Int. J. Org. Evol.* 61:1495–1499.

Coyne, J. A., and H. E. Hoekstra. 2007. Evolution of protein expression: new genes for a new diet. *Curr. Biol.* 17:R1014–R1016.

Crow, J. F. 1956. Genetics of DDT resistance in *Drosophila*. International Genetics Symposia. The Organizing Committee, International Genetics Symposia, Tokyo, Japan.

Darwin, C. 1859. The origin of species by means of natural selection or the preservation of favored races in the struggle for life. The Modern Library, New York.

Davidson, E. H. 2001. Genomic regulatory systems. Academic Press, San Diego.

———. 2006. The regulatory genome: gene regulatory networks in development and evolution. Academic Press, Burlington.

Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311:796–800.

Davis, G. K., D. G. Srinivasan, P. J. Wittkopp, and D. L. Stern. 2007. The function and regulation of *Ultrabithorax* in the legs of *Drosophila melanogaster*. *Dev. Biol.* 308:621–631.

Dickinson, W. J., Y. Tang, K. Schuske, and M. Akam. 1993. Conservation of molecular prepatterning during the evolution of cuticle morphology in *Drosophila* larvae. *Evolution* 47:1396–1406.

Dudley, A. M., D. M. Janse, A. Tanay, R. Shamir, and G. M. Church. 2005. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* 1:2005 0001.

- Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig, and D. Botstein. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 99:16144–16149.
- Endler, J. A. 1986. *Natural selection in the wild*. Princeton Univ. Press, Princeton, NJ.
- Eyre-Walker, A. 2006. The genomic rate of adaptive evolution. *Trends Ecol. Evol.* 21:569–575.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. Longman, Essex.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford Univ. Press, Oxford.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Force, A., W. A. Cresko, F. B. Pickett, S. R. Proulx, C. Amemiya, and M. Lynch. 2005. The origin of subfunctions and modular gene regulation. *Genetics* 170:433–446.
- Garcia-Bellido, A. 1975. Genetic control of wing disc development in *Drosophila*. *Ciba Found Symp.* 29:161–182.
- Gerhart, J., and M. Kirschner. 1997. *Cells, embryos, and evolution*. Blackwell Science, Malden.
- Gibert, J. M., S. Marcellini, J. R. David, C. Schlotterer, and P. Simpson. 2005. A major bristle QTL from a selected population of *Drosophila* uncovers the zinc-finger transcription factor *poils-au-dos*, a repressor of *achaete-scute*. *Dev. Biol.* 288:194–205.
- Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marcos and the Panglossian paradigm: a critic of the adaptationist programme. *Proc. R. Soc. Lond. B* 205:581–598.
- Graveley, B. R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17:100–107.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc. Camb. Philos. Soc.* 28:838–844.
- . 1932. *The causes of evolution*. Princeton Univ. Press, edition (1990), Princeton, NJ.
- Harley, V. R., R. Lovell-Badge, and P. N. Goodfellow. 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.* 22:1500–1501.
- Hartenstein, V. 2004. Developmental of insect sensilla. Pp. 379–419 in L. Gilbert, K. Iatrou, and S. Gill, eds. *Comprehensive molecular insect*, Science Elsevier BV, Oxford.
- He, X., and J. Zhang. 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173:1885–1891.
- Hemingway, J., N. J. Hawkes, L. McCarroll, and H. Ranson. 2004. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem. Mol. Biol.* 34:653–665.
- Herre, E. A. 1985. Sex ratio adjustment in fig wasps. *Science* 228:896–898.
- . 1987. Optimality, plasticity and selective regime in fig wasp sex ratios. *Nature* 329:627–629.
- Hill, W. G., and P. D. Keightley. 1988. Interrelations of mutations, population size, artificial and natural selection. Pp. 55–70 in B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong, eds. *Proceedings of the Second International Conference on Quantitative Genetics*. Sinauer, Sunderland, MA.
- Hittinger, C. T., A. Rokas, and S. B. Carroll. 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl. Acad. Sci. USA* 101:14144–14149.
- Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evol. Int. J. Org. Evol.* 61:995–1016.
- Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. USA* 98:9157–9160.
- Holland, P. W. H., and J. Garcia-Fernandez. 1996. Hox genes and chordate evolution. *Develop. Biol.* 173:382–395.
- Hurley, I., M. E. Hale, and V. E. Prince. 2005. Duplication events and the evolution of segmental identity. *Evol. Dev.* 7:556–567.
- Inestrosa, N. C., C. E. Sunkel, J. Arriagada, J. Garrido, and R. Godoy-Herrera. 1996. Abnormal development of the locomotor activity in yellow larvae of *Drosophila*: a cuticular defect? *Genetica* 97:205–210.
- Jacob, F. 1973. *The logic of life*. Princeton Univ. Press, Princeton, NJ.
- . 1977. Evolution and tinkering. *Science* 196:1161–1166.
- Jacob, F., and J. Monod. 1961. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* 26:193–211.
- Kacser, H., and J. A. Burns. 1981. The molecular basis of dominance. *Genetics* 97:639–666.
- Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- King, M.-C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245–261.
- Klingler, M., J. Soong, B. Butler, and J. P. Gergen. 1996. Disperse versus compact elements for the regulation of runt stripes in *Drosophila*. *Dev. Biol.* 177:73–84.
- Lai, E. C., and V. Orgogozo. 2004. A hidden program in *Drosophila* peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Dev. Biol.* 269:1–17.
- Lande, R. 1983. The response to selection on major and minor mutations affecting a metrical trait. *Heredity* 50:47–65.
- Le Corre, V., F. Roux, and X. Reboud. 2002. DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.* 19:1261–1271.
- Li, X., and M. Noll. 1994. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* 367:83–87.
- Liu, J., J. M. Nerker, L. F. Stam, G. C. Gibson, Z.-B. Zeng, and C. C. Laurie. 1996. Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. *Genetics* 142:1129–1145.
- Lopez, A. J. 1998. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* 32:279–305.
- Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403:564–567.
- Ludwig, M. Z., N. H. Patel, and M. Kreitman. 1998. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125:949–958.
- Lynch, M. 2007. *The origins of genome architecture*. Sinauer Associates, Inc., Sunderland, MA.
- Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.)* 290:1151–1155.
- Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.
- Lynch, M., M. O’Hely, B. Walsh, and A. Force. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–1804.

- Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig, and D. Botstein. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 99:16144–16149.
- Endler, J. A. 1986. *Natural selection in the wild*. Princeton Univ. Press, Princeton, NJ.
- Eyre-Walker, A. 2006. The genomic rate of adaptive evolution. *Trends Ecol. Evol.* 21:569–575.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. Longman, Essex.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford Univ. Press, Oxford.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Force, A., W. A. Cresko, F. B. Pickett, S. R. Proulx, C. Amemiya, and M. Lynch. 2005. The origin of subfunctions and modular gene regulation. *Genetics* 170:433–446.
- Garcia-Bellido, A. 1975. Genetic control of wing disc development in *Drosophila*. *Ciba Found Symp.* 29:161–182.
- Gerhart, J., and M. Kirschner. 1997. *Cells, embryos, and evolution*. Blackwell Science, Malden.
- Gibert, J. M., S. Marcellini, J. R. David, C. Schlotterer, and P. Simpson. 2005. A major bristle QTL from a selected population of *Drosophila* uncovers the zinc-finger transcription factor *poils-au-dos*, a repressor of *achaete-scute*. *Dev. Biol.* 288:194–205.
- Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marcos and the Panglossian paradigm: a critic of the adaptationist programme. *Proc. R. Soc. Lond. B* 205:581–598.
- Graveley, B. R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17:100–107.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc. Camb. Philos. Soc.* 28:838–844.
- . 1932. *The causes of evolution*. Princeton Univ. Press, edition (1990), Princeton, NJ.
- Harley, V. R., R. Lovell-Badge, and P. N. Goodfellow. 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.* 22:1500–1501.
- Hartenstein, V. 2004. Developmental of insect sensilla. Pp. 379–419 in L. Gilbert, K. Iatrou, and S. Gill, eds. *Comprehensive molecular insect*. Science Elsevier BV, Oxford.
- He, X., and J. Zhang. 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173:1885–1891.
- Hemingway, J., N. J. Hawkes, L. McCarroll, and H. Ranson. 2004. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem. Mol. Biol.* 34:653–665.
- Herre, E. A. 1985. Sex ratio adjustment in fig wasps. *Science* 228:896–898.
- . 1987. Optimality, plasticity and selective regime in fig wasp sex ratios. *Nature* 329:627–629.
- Hill, W. G., and P. D. Keightley. 1988. Interrelations of mutations, population size, artificial and natural selection. Pp. 55–70 in B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong, eds. *Proceedings of the Second International Conference on Quantitative Genetics*. Sinauer, Sunderland, MA.
- Hittinger, C. T., A. Rokas, and S. B. Carroll. 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl. Acad. Sci. USA* 101:14144–14149.
- Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evol. Int. J. Org. Evol.* 61:995–1016.
- Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. USA* 98:9157–9160.
- Holland, P. W. H., and J. Garcia-Fernandez. 1996. Hox genes and chordate evolution. *Develop. Biol.* 173:382–395.
- Hurley, I., M. E. Hale, and V. E. Prince. 2005. Duplication events and the evolution of segmental identity. *Evol. Dev.* 7:556–567.
- Inestrosa, N. C., C. E. Sunkel, J. Arriagada, J. Garrido, and R. Godoy-Herrera. 1996. Abnormal development of the locomotor activity in yellow larvae of *Drosophila*: a cuticular defect? *Genetica* 97:205–210.
- Jacob, F. 1973. *The logic of life*. Princeton Univ. Press, Princeton, NJ.
- . 1977. Evolution and tinkering. *Science* 196:1161–1166.
- Jacob, F., and J. Monod. 1961. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* 26:193–211.
- Kacser, H., and J. A. Burns. 1981. The molecular basis of dominance. *Genetics* 97:639–666.
- Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- King, M.-C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245–261.
- Klingler, M., J. Soong, B. Butler, and J. P. Gergen. 1996. Disperse versus compact elements for the regulation of runt stripes in *Drosophila*. *Dev. Biol.* 177:73–84.
- Lai, E. C., and V. Orgogozo. 2004. A hidden program in *Drosophila* peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Dev. Biol.* 269:1–17.
- Lande, R. 1983. The response to selection on major and minor mutations affecting a metrical trait. *Heredity* 50:47–65.
- Le Corre, V., F. Roux, and X. Reboud. 2002. DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.* 19:1261–1271.
- Li, X., and M. Noll. 1994. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* 367:83–87.
- Liu, J., J. M. Nercer, L. F. Stam, G. C. Gibson, Z.-B. Zeng, and C. C. Laurie. 1996. Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. *Genetics* 142:1129–1145.
- Lopez, A. J. 1998. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* 32:279–305.
- Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403:564–567.
- Ludwig, M. Z., N. H. Patel, and M. Kreitman. 1998. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125:949–958.
- Lynch, M. 2007. *The origins of genome architecture*. Sinauer Associates, Inc., Sunderland, MA.
- Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.)* 290:1151–1155.
- Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.
- Lynch, M., M. O'Hely, B. Walsh, and A. Force. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–1804.

- Mann, R. S., and S. B. Carroll. 2002. Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* 12:592–600.
- Marcellini, S., and P. Simpson. 2006. Two or four bristles: functional evolution of an enhancer of scute in drosophilidae. *PLoS Biol.* 4:e386.
- Mattick, J. S. 2003. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25:930–939.
- Mayr, E. 1963. *Animal species and evolution*. The Belknap Press of Harvard Univ., Cambridge.
- McGregor, A. P. 2005. How to get ahead: the origin, evolution and function of bicoid. *Bioessays* 27:904–913.
- McGregor, A. P., P. J. Shaw, J. M. Hancock, D. Bopp, M. Hediger, N. S. Wratten, and G. A. Dover. 2001. Rapid restructuring of bicoid-dependent hunchback promoters within and between Dipteran species: implications for molecular coevolution. *Evol. Develop.* 3:397–407.
- McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, and D. L. Stern. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* 448:587–590.
- Mombaerts, P. 2001. The human repertoire of odorant receptor genes and pseudogenes. *Annu. Rev. Genomics Hum. Genet.* 2:493–510.
- Nackley, A. G., S. A. Shabalina, I. E. Tchivileva, K. Satterfield, O. Kochynskyi, S. S. Makarov, W. Maixner, and L. Diatchenko. 2006. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science (New York, N.Y.)* 314:1930–1933.
- Nei, M. 2007. The new mutation theory of phenotypic evolution. *Proc. Natl. Acad. Sci. USA* 104:12235–12242.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, New York.
- Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T. L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, et al. 2005. High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. USA* 102:19015–19020.
- Orr, H. A. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- . 2000. Adaptation and the cost of complexity. *Evolution* 54:13–20.
- Otto, S. P. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. Biol. Sci.* 271:705–714.
- Overton, P. M., W. Chia, and M. Buescher. 2007. The *Drosophila* HMG-domain proteins SoxNeuro and Dichaete direct trichome formation via the activation of shavenbaby and the restriction of Wingless pathway activity. *Development* 134:2807–2813.
- Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39:1256–1260.
- Piatigorsky, J., and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science (New York, N.Y.)* 252:1078–1079.
- Pistillo, D., N. Skaer, and P. Simpson. 2002. Scute expression in *Calliphora vicina* reveals an ancestral pattern of longitudinal stripes on the thorax of higher Diptera. *Development* 129:563–572.
- Price, T. D. 2002. Domesticated birds as a model for the genetics of speciation by sexual selection. *Genetica* 116:311–327.
- Ptashne, M., and A. Gann. 2002. *Genes & Signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Raff, R. A., and T. C. Kaufman. 1983. *Embryos, genes, and evolution*. Indiana Univ. Press, Bloomington.
- Richardson, J., and P. Simpson. 2006. A conserved trans-regulatory landscape for scute expression on the notum of cyclorhaphous Diptera. *Dev. Genes Evol.* 216:29–38.
- Riehle, M. M., A. F. Bennett, and A. D. Long. 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 98:525–530.
- Robertson, F. W. 1959. Studies in quantitative inheritance. XII. Cell size and number in relation to genetic and environmental variation of body size in *Drosophila*. *Genetics* 44:869–896.
- Robertson, H. M., and J. H. Thomas. 2006. The putative chemoreceptor families of *C. elegans*. *WormBook* 1–12.
- Rockman, M. V., and G. A. Wray. 2002. Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.* 19:1991–2004.
- Scarcelli, N., J. M. Cheverud, B. A. Schaal, and P. X. Kover. 2007. Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. *Proc. Natl. Acad. Sci. USA* 104:16986–16991.
- Scargle, J. D. 2000. Publication bias: the “File-Drawer” problem in scientific inference. *J. Sci. Explor.* 14:91–106.
- Segre, A. V., A. W. Murray, and J. Y. Leu. 2006. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol.* 4:e256.
- Simpson, P., M. Lewis, and J. Richardson. 2006. Conservation of upstream regulators of scute on the notum of cyclorhaphous Diptera. *Dev. Genes Evol.* 216:363–371.
- Skaer, N., D. Pistillo, J. M. Gibert, P. Lio, C. Wulbeck, and P. Simpson. 2002a. Gene duplication at the achaete-scute complex and morphological complexity of the peripheral nervous system in Diptera. *Trends Genet.* 18:399–405.
- Skaer, N., D. Pistillo, and P. Simpson. 2002b. Transcriptional heterochrony of scute and changes in bristle pattern between two closely related species of blowfly. *Dev. Biol.* 252:31–45.
- Stam, L. F., and C. C. Laurie. 1996. Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* 144:1559–1564.
- Stern, D. L. 2000. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* 54:1079–1091.
- . 2003. Gene regulation. Pp. 145–151 in B. Hall, and W. Olson, eds. *Keywords and concepts in evolutionary developmental biology*. Harvard Univ. Press, Cambridge.
- Stone, J. R., and G. A. Wray. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.* 18:1764–1770.
- Sucena, E., and D. L. Stern. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc. Natl. Acad. Sci. USA* 97:4530–4534.
- Sucena, E., I. Delon, I. Jones, F. Payre, and D. L. Stern. 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* 424:935–938.
- Tomancak, P., A. Beaton, R. Weiszmam, E. Kwan, S. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, et al. 2002. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 3:RESEARCH0088.
- Wallace, B. 1963. Genetic diversity, genetic uniformity, and heterosis. *Can. J. Genet. Cytol.* 5:239–253.
- Wang, R.-L., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. *Nature* 398:236–239.
- Waxman, D., and J. R. Peck. 1998. Pleiotropy and the preservation of perfection. *Science (New York, N.Y.)* 279:1210–1213.
- Welch, J. J., and D. Waxman. 2003. Modularity and the cost of complexity. *Evol. Int. J. Org. Evol.* 57:1723–1734.
- Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne, and J. J. Bull. 1999. Different trajectories of parallel evolution during viral adaptation. *Science (New York, N.Y.)* 285:422–424.

- Mann, R. S., and S. B. Carroll. 2002. Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* 12:592–600.
- Marcellini, S., and P. Simpson. 2006. Two or four bristles: functional evolution of an enhancer of scute in drosophilidae. *PLoS Biol.* 4:e386.
- Mattick, J. S. 2003. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25:930–939.
- Mayr, E. 1963. *Animal species and evolution*. The Belknap Press of Harvard Univ., Cambridge.
- McGregor, A. P. 2005. How to get ahead: the origin, evolution and function of bicoid. *Bioessays* 27:904–913.
- McGregor, A. P., P. J. Shaw, J. M. Hancock, D. Bopp, M. Hediger, N. S. Wratten, and G. A. Dover. 2001. Rapid restructuring of bicoid-dependent hunchback promoters within and between Dipteran species: implications for molecular coevolution. *Evol. Develop.* 3:397–407.
- McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, and D. L. Stern. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* 448:587–590.
- Mombaerts, P. 2001. The human repertoire of odorant receptor genes and pseudogenes. *Annu. Rev. Genomics Hum. Genet.* 2:493–510.
- Nackley, A. G., S. A. Shabalina, I. E. Tchivileva, K. Satterfield, O. Kochynskyi, S. S. Makarov, W. Maitner, and L. Diatchenko. 2006. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science (New York, N.Y.)* 314:1930–1933.
- Nei, M. 2007. The new mutation theory of phenotypic evolution. *Proc. Natl. Acad. Sci. USA* 104:12235–12242.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, New York.
- Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T. L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, et al. 2005. High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. USA* 102:19015–19020.
- Orr, H. A. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- . 2000. Adaptation and the cost of complexity. *Evolution* 54:13–20.
- Otto, S. P. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. Biol. Sci.* 271:705–714.
- Overton, P. M., W. Chia, and M. Buescher. 2007. The *Drosophila* HMG-domain proteins SoxNeuro and Dichaete direct trichome formation via the activation of shavenbaby and the restriction of Wingless pathway activity. *Development* 134:2807–2813.
- Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39:1256–1260.
- Piatigorsky, J., and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science (New York, N.Y.)* 252:1078–1079.
- Pistillo, D., N. Skaer, and P. Simpson. 2002. Scute expression in *Calliphora vicina* reveals an ancestral pattern of longitudinal stripes on the thorax of higher Diptera. *Development* 129:563–572.
- Price, T. D. 2002. Domesticated birds as a model for the genetics of speciation by sexual selection. *Genetica* 116:311–327.
- Ptashne, M., and A. Gann. 2002. *Genes & Signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Raff, R. A., and T. C. Kaufman. 1983. *Embryos, genes, and evolution*. Indiana Univ. Press, Bloomington.
- Richardson, J., and P. Simpson. 2006. A conserved trans-regulatory landscape for scute expression on the notum of cyclorhaphous Diptera. *Dev. Genes Evol.* 216:29–38.
- Riehle, M. M., A. F. Bennett, and A. D. Long. 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 98:525–530.
- Robertson, F. W. 1959. Studies in quantitative inheritance. XII. Cell size and number in relation to genetic and environmental variation of body size in *Drosophila*. *Genetics* 44:869–896.
- Robertson, H. M., and J. H. Thomas. 2006. The putative chemoreceptor families of *C. elegans*. *WormBook* 1–12.
- Rockman, M. V., and G. A. Wray. 2002. Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.* 19:1991–2004.
- Scarcellì, N., J. M. Cheverud, B. A. Schaal, and P. X. Kover. 2007. Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. *Proc. Natl. Acad. Sci. USA* 104:16986–16991.
- Scargle, J. D. 2000. Publication bias: the “File-Drawer” problem in scientific inference. *J. Sci. Explor.* 14:91–106.
- Segre, A. V., A. W. Murray, and J. Y. Leu. 2006. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol.* 4:e256.
- Simpson, P., M. Lewis, and J. Richardson. 2006. Conservation of upstream regulators of scute on the notum of cyclorhaphous Diptera. *Dev. Genes Evol.* 216:363–371.
- Skaer, N., D. Pistillo, J. M. Gibert, P. Lio, C. Wulbeck, and P. Simpson. 2002a. Gene duplication at the achaete-scute complex and morphological complexity of the peripheral nervous system in Diptera. *Trends Genet.* 18:399–405.
- Skaer, N., D. Pistillo, and P. Simpson. 2002b. Transcriptional heterochrony of scute and changes in bristle pattern between two closely related species of blowfly. *Dev. Biol.* 252:31–45.
- Stam, L. F., and C. C. Laurie. 1996. Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* 144:1559–1564.
- Stern, D. L. 2000. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* 54:1079–1091.
- . 2003. Gene regulation. Pp. 145–151 in B. Hall, and W. Olson, eds. *Keywords and concepts in evolutionary developmental biology*. Harvard Univ. Press, Cambridge.
- Stone, J. R., and G. A. Wray. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.* 18:1764–1770.
- Sucena, E., and D. L. Stern. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of *ovo/shaven-baby*. *Proc. Natl. Acad. Sci. USA* 97:4530–4534.
- Sucena, E., I. Delon, I. Jones, F. Payre, and D. L. Stern. 2003. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* 424:935–938.
- Tomancak, P., A. Beaton, R. Weiszmam, E. Kwan, S. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, et al. 2002. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 3:RESEARCH0088.
- Wallace, B. 1963. Genetic diversity, genetic uniformity, and heterosis. *Can. J. Genet. Cytol.* 5:239–253.
- Wang, R.-L., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. *Nature* 398:236–239.
- Waxman, D., and J. R. Peck. 1998. Pleiotropy and the preservation of perfection. *Science (New York, N.Y.)* 279:1210–1213.
- Welch, J. J., and D. Waxman. 2003. Modularity and the cost of complexity. *Evol. Int. J. Org. Evol.* 57:1723–1734.
- Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne, and J. J. Bull. 1999. Different trajectories of parallel evolution during viral adaptation. *Science (New York, N.Y.)* 285:422–424.

- Wilke, C. O. 2004. Molecular clock in neutral protein evolution. *BMC Genet.* 5:25.
- Wilkens, H. 1971. Genetic interpretation of regressive evolutionary processes: studies of hybrid eyes of two *Astyanax* cave populations (Characidae: Pisces). *Evolution* 25:530–544.
- Wilkins, A. S. 2002. The evolution of developmental pathways. Sinauer Associates, Sunderland.
- Wilson, A. C. 1975. Evolutionary importance of gene regulation. *Stadler Symp.* 7:117–134.
- Wistow, G. J., and J. Piatigorsky. 1988. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu. Rev. Biochem.* 57:479–504.
- Wittkopp, P. J., B. K. Haerum, and A. G. Clark. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* 40:346–350.
- Woods, R., D. Schneider, C. L. Winkworth, M. A. Riley, and R. E. Lenski. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 103:9107–9112.
- Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8:206–216.
- Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, and L. A. Romano. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20:1377–1419.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
- . 1968. Evolution and the genetics of populations, vol. 1 (Genetics and biometric foundations). Univ. of Chicago Press, Chicago.
- Wülbeck, C., and P. Simpson. 2000. Expression of achaete-scute homologues in discrete proneural clusters on the developing notum of the medfly *Ceratitis capitata*, suggests a common origin for the stereotyped bristle patterns of higher Diptera. *Development* 127:1411–1420.
- . 2002. The expression of panner and achaete-scute homologues in a mosquito suggests an ancient role of panner as a selector gene in the regulation of the dorsal body pattern. *Development* 129:3861–3871.
- Xing, Y., and C. Lee. 2006. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* 7:499–509.
- Yang, Y., and A. Cvekl. 2005. Tissue-specific regulation of the mouse alphaA-crystallin gene in lens via recruitment of Pax6 and c-Maf to its promoter. *J. Mol. Biol.* 351:453–469.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18:292–298.
- Zhang, X., and S. Firestein. 2002. The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* 5:124–133.
- Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* 30:411–415.
- Zuckerkandl, E. 1963. Perspectives in molecular anthropology. Pp. 256–274 in A. Rich, and N. Davidson, eds. *Structural chemistry and molecular biology*. W. H. Freeman, San Francisco.

Associate Editor: M. Rausher

Appendix 1

A database of studies providing compelling evidence for genetic changes contributing to evolution in domesticated races, within species and between species is provided as Supplementary Material on the Evolution website. We did not include studies that identified mutations resulting from selection experiments. We

included studies that provided compelling genetic and functional evidence for the role of gene regions or individual mutations in generating an evolved phenotypic difference. We excluded most association studies that did not provide further genetic or functional evidence implicating gene regions or individual mutations. We included association studies that identified mutations—such as deletions or nonsense mutations—that are likely to generate complete protein loss of function alleles with obvious phenotypic consequences. The database includes all studies that we examined, together with information about whether they were included or excluded from analysis and why. We would appreciate learning about studies that we inadvertently overlooked.

Appendix 2

The main goal of Hoekstra and Coyne's (2007) commentary was to sow doubt about whether *cis*-regulatory evolution is really the dominant mode of genetic evolution. They attacked the *cis*-regulatory hypothesis from many directions. We address here their most important arguments that were not discussed in the body of our article.

GENE DUPLICATION AND THE *CIS*-REGULATORY HYPOTHESIS

Hoekstra and Coyne (2007) argued that gene duplication has been important for phenotypic evolution and that for two reasons this diminishes the importance of *cis*-regulatory evolution. They argued first that gene duplication seems to be an important source of phenotypic evolution, challenging the broad *cis*-regulatory hypothesis that states that most phenotypic evolution is caused by *cis*-regulatory changes. Second, they argued that gene duplication allows one gene copy to retain an ancestral function and the other to evolve a new function. This may reduce pleiotropic effects associated with coding mutations and thus lessen the importance of the population genetics argument.

How important is gene duplication? Gene duplication and the broad cis-regulatory hypothesis

In 1970, Ohno suggested that gene duplication provided important material for the evolution of novel phenotypes (Ohno 1970). Over the entire sweep of evolution, duplication has created many new genes. At particular periods during life history, gene duplication may have played important roles in phenotypic evolution (Lynch 2007). Particular families of genes have sometimes been duplicated as an apparent adaptive response to novel ecological challenges. In mammals, the olfactory receptor family has expanded to include about 1000 genes (Mombaerts 2001; Zhang and Firestein 2002). The nematode, *Caenorhabditis elegans*, contains about 1000 genes that may serve as chemoreceptors (Robertson and Thomas 2006).

Wilke, C. O. 2004. Molecular clock in neutral protein evolution. *BMC Genet.*5:25.

Wilkens, H. 1971. Genetic interpretation of regressive evolutionary processes: studies of hybrid eyes of two *Astyanax* cave populations (Characidae: Pisces). *Evolution* 25:530–544.

Wilkins, A. S. 2002. The evolution of developmental pathways. Sinauer Associates, Sunderland.

Wilson, A. C. 1975. Evolutionary importance of gene regulation. *Stadler Symp.* 7:117–134.

Wistow, G. J., and J. Piatigorsky. 1988. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu. Rev. Biochem.*57:479–504.

Wittkopp, P. J., B. K. Haerum, and A. G. Clark. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* 40:346–350.

Woods, R., D. Schneider, C. L. Winkworth, M. A. Riley, and R. E. Lenski. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 103:9107–9112.

Wray, G. A. 2007. The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* 8:206–216.

Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, and L. A. Romano. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20:1377–1419.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

———. 1968. *Evolution and the genetics of populations*, vol. 1 (Genetics and biometric foundations). Univ. of Chicago Press, Chicago.

Wülbeck, C., and P. Simpson. 2000. Expression of achaete-scute homologues in discrete proneural clusters on the developing notum of the medfly *Ceratitis capitata*, suggests a common origin for the stereotyped bristle patterns of higher Diptera. *Development* 127:1411–1420.

———. 2002. The expression of *pannier* and *achaete-scute* homologues in a mosquito suggests an ancient role of *pannier* as a selector gene in the regulation of the dorsal body pattern. *Development* 129:3861–3871.

Xing, Y., and C. Lee. 2006. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.*7:499–509.

Yang, Y., and A. Cvekl. 2005. Tissue-specific regulation of the mouse alphaCrystallin gene in lens via recruitment of Pax6 and c-Maf to its promoter. *J. Mol. Biol.* 351:453–469.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18:292–298.

Zhang, X., and S. Firestein. 2002. The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* 5:124–133.

Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* 30:411–415.

Zuckerlandl, E. 1963. Perspectives in molecular anthropology. Pp. 256–274 in A. Rich, and N. Davidson, eds. *Structural chemistry and molecular biology*. W. H. Freeman, San Francisco.

Editor Associado: M. Rausher

Apêndice 1:

No website da Evolution poderá obter, como Material Suplementar, uma base de dados de estudos que fornece evidências convincentes de alterações genéticas que contribuem para a evolução de raças domésticas, dentro de espécies e entre espécies. Não incluímos estudos que identifiquem mutações resultantes de experiências com seleção. Incluímos estudos que fornecem evidências convincentes, genéticas e funcionais, para a função das regiões dos genes ou mutações individuais na geração de diferenças fenotípicas evoluídas. Excluímos a maioria de

estudos associados que não forneciam mais evidências genéticas ou funcionais que implicavam regiões de genes ou mutações individuais. Incluímos estudos associados que identificaram mutações - tais como deleções ou mutações sem sentido - com probabilidade de gerar a perda total de proteínas dos alelos funcionais sem consequências fenotípicas óbvias. A base de dados inclui todos os estudos que examinamos, juntamente com a informação sobre se foram incluídos ou excluídos da análise e porquê. Gostaríamos de ficar a conhecer estudos que possamos, inadvertidamente, ter ignorado.

Apêndice 2

O objetivo principal do comentário de Hoekstra e Coyne (2007) foi disseminar a dúvida se a evolução da regulação *cis* é de facto o método dominante da evolução genética. Criticaram a hipótese da regulação *cis* de várias perspectivas. Abordamos de seguida os argumentos mais importantes que não foram discutidos no corpo do artigo.

DUPLICAÇÃO DE GENES E A HIPÓTESE DA *cis*-REGULAÇÃO

Hoekstra e Coyne (2007) argumentam que a duplicação de genes tem sido importante para a evolução fenotípica e que, por dois motivos, este facto diminui a importância da evolução da regulação *cis*. Começam por argumentar que a duplicação de genes parece ser uma importante fonte de evolução fenotípica, desafiando a hipótese geral da regulação *cis*, que diz que a maioria da evolução fenotípica é provocada por alterações na regulação *cis*. O segundo argumento indica que a duplicação de genes permite que um gene copiado mantenha uma função antiga e o outro evolua até obter uma nova função. Isto pode reduzir os efeitos pleiotrópicos associados às mutações da codificação e, neste sentido, minora a importância do argumento da genética de populações.

Quão importante é a duplicação de genes? Duplicação de genes e a hipótese geral da regulação cis

Em 1970, Ohno sugeriu que a duplicação de genes fornece material importante para a evolução de novos fenótipos (Ohno 1970). Na evolução em geral, a duplicação criou muitos genes novos. Em determinados períodos da história da vida, a duplicação de genes pode ter tido funções importantes na evolução fenotípica (Lynch 2007).

Certas famílias de genes foram por vezes duplicadas como aparente resposta adaptativa aos novos desafios ecológicos. Nos mamíferos, a família do receptor olfativo expandiu-se e inclui cerca de 1000 genes (Mombaerts 2001; Zhang e Firestein 2002). O nemátodo *Caenorhabditis elegans* contém cerca de 1000 genes que podem ter a função de quimiorreceptores (Robertson e Thomas 2006).

Over long time scales, there is clear evidence that gene duplication has provided new material for developmental evolution. For example, the different genes of the Hox complex all originated by gene duplication from a single ancestral Hox gene and each of the extant genes has different DNA-binding specificity derived from the different amino acid sequences in the homeodomain (Akam 1995, Akam 1998; Carroll 1995; Averof et al. 1996; Holland and Garcia-Fernandez 1996; Hurley et al. 2005). Presumably such events occur rarely because it is difficult for new genes with pleiotropic roles to establish novel regulatory linkages without altering multiple processes simultaneously. Nonetheless, these events have occurred, albeit rarely. These rare events may have been more important to generating novel patterns of development than their rarity would imply. For example, evolution of the *bicoid* gene through duplication in higher dipterans may have promoted rapid early embryonic patterning (McGregor 2005) and duplication of the *achaete-scute* genes during insect evolution might have allowed the development of stereotyped bristle patterns (Skaer et al. 2002a).

Gene duplicates are fixed in populations at a rate of about one per gene per 100 million years (Lynch and Conery 2000). However, many of these duplicates eventually become pseudogenes and are eliminated from genomes. In general, the rate of gene duplication appears to be lower than the rate of coding or *cis*-regulatory mutations (Carroll 2005). For example, since the divergence of chimpanzees and humans, the human lineage has accumulated about 720 genes by gene duplication (Zhang 2003) whereas approximately 60,000 nonsynonymous differences and even more noncoding differences are found between humans and chimpanzees (Eyre-Walker 2006). Nevertheless, it is difficult to estimate how many of these duplicated genes and coding and *cis*-regulatory changes have contributed to phenotypic evolution.

Gene duplication versus cis-regulatory and coding changes

How relevant is gene duplication to the *cis*-regulatory hypothesis? Gene duplication usually causes simply an increase in gene expression. Indeed, few cases of gene duplications have been reported to cause phenotypic evolution. They include gene amplification of esterase genes in mosquitoes and aphids that are resistant to organophosphate and carbamate insecticides (Hemingway et al. 2004) and *amylase* gene amplification in humans associated with increased starch consumption (Perry et al. 2007).

Perhaps more importantly, gene duplication often contributes to phenotypic evolution through the novel *cis*-regulatory and coding mutations that occur subsequently or concomitantly. Here are three examples.

- (1) *RNase*—Colobine monkeys eat leaves, rather than fruit and insects, and have evolved a fermenting foregut, similar to ru-

minants. Symbiotic bacteria in the foregut digest the leaves and the monkeys digest the bacteria. Colobine monkeys have evolved a duplicated *RNase1* gene that is expressed specifically in the pancreas and secreted into the small intestine, where it operates in a relatively acidic environment. The duplicated gene has evolved multiple amino-acid substitutions, each of which increases its efficiency in an acidic environment (Zhang et al. 2002). The duplicated gene has also lost the ability to digest double-stranded RNA, which the original gene retains. Thus, this adaptation required gene duplication together with altered expression pattern (expression in the pancreas) and multiple amino-acid substitutions.

- (2) *Eye lens crystallins*—Vertebrate eye lens crystallins have evolved repeatedly from enzymes and heat shock proteins (Wistow and Piatigorsky 1988; Piatigorsky and Wistow 1991; Yang and Cvekl 2005). These proteins have lost their enzymatic activity and are expressed specifically in lens cells at very high levels. Although the loss of enzymatic activity is probably relevant to their function in lens, the regulatory change—expression specifically in the lens and at very high levels—was critical. In most cases, it is thought that the regulatory change occurred first, followed by changes in protein structure and gene duplication (Piatigorsky and Wistow 1991).
- (3) *Antifreeze proteins*—Fish that live in subzero waters have evolved proteins that capture ice in the bloodstream to prevent ice crystals from puncturing cells. In Antarctic notothenioids, this antifreeze protein evolved from a pancreatic trypsinogen (Chen et al. 1997). The protein has undergone extensive modification resulting from selection for binding ice. At some point, the protein became expressed in liver to allow secretion into the bloodstream. Thus, this evolutionary innovation required changes in both protein structure and expression pattern.

In these three examples, the novel gene functions probably evolved after gene duplication and they required changes in both the protein-coding regions and in the expression patterns. Because both duplicated genes are found in the same *trans*-regulatory environment, their change in expression pattern must result, at least in part, from *cis*-regulatory changes. The alternative—a change only in upstream regulator(s)—would affect the expression pattern of both duplicated genes.

Gene duplication and cis-regulatory architecture

Although duplicated genes can evolve novel functions through both protein-coding and *cis*-regulatory changes, it is currently thought that the key feature that allows gene duplicates to persist long enough to evolve novel functions is the fact that genes possess multiple modular *cis*-regulatory elements (Averof et al.

Durante longos períodos de tempo, existem evidências claras de que a duplicação de genes tem fornecido novos materiais para a evolução do desenvolvimento. Por exemplo, os diferentes genes do complexo Hox têm origem na duplicação de genes de um único gene Hox ancestral, e cada um dos genes ainda existentes têm diferentes especificidades de ligação ao ADN, derivadas de sequências de aminoácidos diferentes no homeodomínio (Akam 1995, Akam 1998; Carroll 1995; Averof et al. 1996; Holland e Garcia-Fernandez 1996; Hurley et al. 2005). Estes eventos ocorreram, provavelmente, raras vezes, porque os novos genes com funções pliotrópicas têm dificuldade em estabelecer novos sistemas de ligação de regulação sem alterar simultaneamente os múltiplos processos. Contudo apesar de raros, aconteceram. Estes eventos raros têm sido mais importantes para a geração de novos padrões de desenvolvimento do que a sua raridade implica. Por exemplo, a evolução do gene *bicoid* através da duplicação em dípteros maiores, pode ter promovido uma rápida padronização embrionária prematura (McGregor 2005), e a duplicação de genes *achaete-scute* durante a evolução de insetos pode ter permitido o desenvolvimento de padrões de pelos estereotipados (Skaer et al. 2002a).

Os genes duplicados estão fixados em populações numa proporção de um por gene a cada 100 milhões de anos (Lynch e Conery 2000). No entanto, muitos destes duplicados acabaram por se tornar pseudogenes e foram eliminados dos genomas. Regra geral, a proporção de duplicação de genes parece ser inferior à proporção de mutações na codificação ou na regulação *cis* (Carroll 2005). Por exemplo, desde a divergência de chimpanzés e humanos, a estirpe humana acumulou cerca de 720 genes através da duplicação de genes (Zhang 2003), no entanto foram encontradas mais de 60.000 diferenças não-sinónimas e ainda mais diferenças não-codificantes entre humanos e chimpanzés (Eyre-Walker 2006). Contudo, é difícil prever quantos destes genes duplicados e alterações na codificação e na regulação *cis* contribuíram para a evolução fenotípica.

Duplicação de genes versus alterações na regulação cis e na codificação

Até que ponto a duplicação de genes e a hipótese da regulação *cis* são relevantes? A duplicação de genes geralmente provoca apenas um aumento na expressão do gene. Na verdade, existem poucos casos estudados de duplicação de genes que provoquem evolução fenotípica. Estes incluem a amplificação de genes esterases em mosquitos e pulgões resistentes aos inseticidas com organofosfato e carbamato (Hemingway et al. 2004), e a amplificação de genes da amilase em humanos, associada ao consumo aumentado de amido (Perry et al. 2007).

Talvez mais importante seja o facto de a duplicação dos genes contribuir frequentemente para a evolução fenotípica através de novas mutações na regulação *cis* e na codificação, que ocorrem subsequentemente ou concomitantemente. Seguem-se três exemplos.

(1) *RNase* - Os macacos colobos comem folhas, em vez de frutas e insetos, e desenvolveram um músculo dianteiro de fermentação, semelhante aos ruminantes. As bactérias simbiotes neste músculo digerem as folhas e os macacos

digerem as bactérias. Os macacos colobos desenvolveram um gene *RNase1* duplicado que se exprime especificamente no pâncreas e é segregado para o intestino pequeno, onde trabalha num ambiente relativamente ácido. O gene duplicado desenvolveu múltiplas substituições de aminoácidos, sendo que cada um aumentava a eficiência num ambiente ácido (Zhang et al. 2002). O gene duplicado perdeu também a capacidade de digerir ARN de dupla cadeia, que se mantém no gene original. Assim, esta adaptação necessita de duplicação de genes, juntamente com um padrão de expressão modificado (a expressão no pâncreas) e múltiplas substituições de aminoácidos.

(2) *Cristalino do olho* - Os cristalinos dos olhos de vertebrados têm evoluído repetidamente a partir de enzimas e proteínas de choque térmico (Wistow and Piatigorsky 1988; Piatigorsky e Wistow 1991; Yang e Cvekl 2005). Estas proteínas perderam toda a sua atividade enzimática e estão expressas especificamente nas células da lente, a níveis muito elevados. Apesar de a perda de atividade enzimática ser provavelmente relevante para a função da lente, as alterações reguladoras (a expressão especificamente na lente e a níveis muito elevados) foram críticas. Na maioria dos casos, considera-se que as alterações na regulação ocorrem primeiro, seguidas de alterações na estrutura da proteína e duplicação de genes (Piatigorsky e Wistow 1991).

(3) *Proteínas anticongelantes* - Os peixes que vivem em águas com temperatura negativas desenvolveram proteínas que capturam o gelo na corrente sanguínea para evitar que os cristais de gelo furem as células. Nos *notothenioidei* antárticos, esta proteína anticongelante evoluiu a partir de um tripsinogénio pancreático. A proteína sofreu profundas modificações resultantes da seleção para se isolar do gelo. A certa altura, esta proteína ficou expressa no fígado, o que permitiu a sua segregação para a corrente sanguínea. Deste modo, a inovação evolutiva necessitou de alterações, tanto na estrutura das proteínas como no padrão de expressão.

Nestes três exemplos, as novas funções de genes evoluíram provavelmente após a duplicação de genes e necessitaram de alterações tanto nas regiões de codificação como nos padrões de expressão. Devido ao facto de ambos os genes duplicados serem encontrados no mesmo ambiente trans-regulador, a sua alteração no padrão de expressão deve ser resultado, pelo menos em parte, de alterações na regulação *cis*. A alternativa - uma alteração somente no(s) regulador(es) próximo(s) da fonte - iria afetar o padrão de expressão de ambos os genes duplicados.

Duplicação de genes e a arquitetura da regulação cis

Apesar dos genes duplicados conseguirem desenvolver novas funções de codificação de proteínas e de alterações na regulação *cis*, atualmente considera-se que a característica-chave para a persistência dos genes duplicados, tão duradoura ao ponto de desenvolver novas funções, é o facto de os genes possuírem múltiplos elementos modulares de regulação *cis* (Averof et al. 1996; Force et al. 1999; Lynch and Force 2000; Lynch et al. 2001; Force et al. 2005). Alguns destes módulos podem-se perder durante o próprio processo de duplicação (Averof et al. 1996)

1996; Force et al. 1999; Lynch and Force 2000; Lynch et al. 2001; Force et al. 2005). Some of these modules can be lost either during the duplication process itself (Averof et al. 1996) or subsequently through accumulation of neutral mutations that eliminate complementary *cis*-regulatory enhancers in the original and the duplicated gene (Force et al. 1999; Lynch and Force 2000; Force et al. 2005). Complementary loss of these enhancer modules then generates selection to maintain both duplicates, setting the stage for future functional divergence.

The three examples mentioned above are of genes expressed near or at the end of terminal differentiation. The combination of three facts—the relatively terminal position of these genes in the regulatory hierarchy, the derived genes are duplicates that are not required to maintain the original gene function, and the derived genes have no obvious deleterious effects in the cellular context of the original gene—may provide extensive flexibility in the mutational changes that caused the evolution of novel functions. In contrast, genes that are not at terminal positions within regulatory hierarchies may undergo fewer changes in coding regions following gene duplication. For example, after duplication of the *achaete-scute* genes and of the *paired-gooseberry-gooseberry neuro* genes, the new genes acquired different expression patterns but, in *D. melanogaster*, their gene products can still substitute for one another (Li and Noll 1994; Skaer et al. 2002a). This suggests that in these cases, few or no evolutionarily relevant mutations have occurred in the coding regions of the duplicated genes; they have occurred mostly in their *cis*-regulatory regions.

In conclusion, gene duplication is not really a competitor with the *cis*-regulatory hypothesis because gene duplication simply generates an increase in gene expression. With current data it seems premature to judge whether mutations that follow gene duplication are more likely to occur in coding or in *cis*-regulatory regions. Most examples of gene duplication providing adaptive new functions have involved important changes in both coding and *cis*-regulatory regions.

ADAPTATION AND EVOLUTION

In their critique of the *cis*-regulatory hypothesis, Hoekstra and Coyne (2007) sometimes implicitly equated morphological evolution and adaptation and other times argued that the focus of genetic studies should be on adaptations rather than morphological evolution per se. Thus, their compilation of genes and mutations that contribute to phenotypic evolution comprises almost exclusively traits that are generally recognized to increase fitness or that are maintained by selection.

There are several problems here. First, Hoekstra and Coyne (2007) excluded a few traits from consideration not because they have been shown to be neutral or deleterious, but because they have not been proven to be adaptations. Imagine that we required that only traits with clear adaptive significance were suitable sub-

jects for genetic and developmental analysis. We would greatly reduce the spectrum of admissible traits because the vast majority of tests for natural selection in the wild find very small selection coefficients (Hoekstra et al. 2001; Kingsolver et al. 2001). Unfortunately, most studies are unable to demonstrate that these small selection coefficients are significantly different from zero. This is because either the phenotypic variants are neutral or because sample sizes or selection coefficients are too low to allow experimental detection. The latter model is consistent with recent molecular population genomics analyses (Eyre-Walker 2006; Andolfatto 2007).

The temptation to tell adaptive stories about phenotypic traits has a long and checkered history in evolutionary biology (Gould and Lewontin 1979). In addition, the desire to tell adaptive stories leads to the equally nefarious tendency to dismiss adaptive explanations for traits when our imagination fails us. Darwin warned us of this problem (Darwin 1859). In Chapter 6 of the *Origin of Species*, in a section entitled “Organs of little apparent Importance, as affected by Natural Selection” he wrote

In the first place, we are much too ignorant in regard to the whole economy of any one organic being, to say what slight modifications would be of importance or not.

This remains true today. Therefore, traits with no obvious adaptive value should still be considered when investigating the molecular basis of evolutionary changes in phenotype (Nei 2007).

But there is an even deeper problem with studying only clearly adaptive traits. This might bias our understanding of the genetic causes of variation toward traits under strong selection. As we showed in the section “The Data: Evidence for a Predictive Theory of Genetic Evolution,” traits under strong selection may be caused by an unusual distribution of mutations. For example, some natural genetic variation underlying apparently adaptive polymorphisms is caused by mutations that eliminate or severely disrupt the gene product. Many of these genes are otherwise strongly conserved in other species, suggesting that these recent adaptive polymorphisms have poor long-term prospects. It is thus not entirely clear that focusing only on adaptive traits, particularly traits under strong selection—which are, of course, the easiest to identify—necessarily identifies alleles that are most relevant to long-term evolution.

There is currently a fundamental disconnect between population genomics approaches to studying adaptation and genetic studies of “obviously” adaptive traits, especially of polymorphisms maintained in populations. Population genomics approaches generate estimates of very small selection coefficients, on the order of 10^{-5} for most adaptive fixations in *Drosophila* (Andolfatto 2007). In contrast, when measured, selection in the wild is often about four orders of magnitude greater (Hoekstra et al. 2001; Kingsolver et al. 2001). Studies of clearly adaptive traits are,

ou, subseqüentemente, através da acumulação de mutações neutras que eliminam os intensificadores complementares da regulação *cis* no gene original e no duplicado (Force et al. 1999; Lynch and Force 2000; Force et al. 2005). A perda complementar destes módulos intensificadores dá então origem à seleção, para manter ambos os duplicados, criando assim condições para uma futura divergência funcional.

Os três exemplos mencionados acima são de genes expressos proximamente ou perto do final da diferenciação terminal. A combinação destes três factos (a posição relativamente terminal destes genes na hierarquia de regulação, os genes derivados que não são necessários para manter a função original do gene, e os genes derivados que não têm efeitos deletérios óbvios no contexto celular do gene original) pode fornecer uma flexibilidade abrangente nas alterações mutacionais que provocam a evolução de funções novas. Os genes que não se encontram em posições terminais dentro das hierarquias de regulação, podem, pelo contrário, sofrer menos alterações nas regiões de codificação, seguindo-se de duplicação de genes. Por exemplo, após a duplicação dos genes *achaete-scute* e do par de neuro genes *gooseberry-gooseberry*, os novos genes adquiriram diferentes padrões de expressão mas na *D. melanogaster* os seus produtos do gene podem ainda substituir-se uns pelos outros (Li e Noll 1994; Skaer et al. 2002a). Isto sugere, nestes casos, terem ocorrido poucas ou nenhuma mutações evolutivamente relevantes nas regiões de codificação dos genes duplicados; as mutações ocorreram maioritariamente nas suas regiões de regulação *cis*.

Concluindo, a duplicação de genes não compete realmente com a hipótese da regulação *cis* porque a duplicação de genes gera apenas um aumento da expressão do gene. Com os dados atuais existentes, é prematuro estimar se as mutações que seguem a duplicação de genes têm mais probabilidade de ocorrer na região de codificação ou de regulação *cis*. A maioria dos exemplos de duplicação de genes que fornecem novas funções adaptativas implica alterações importantes, tanto na codificação como nas regiões de regulação *cis*.

ADAPTAÇÃO E EVOLUÇÃO

Por vezes, na sua crítica à hipótese da regulação *cis*, Hoekstra e Coyne (2007) compararam implicitamente a evolução morfológica à adaptação. Também defenderam que o estudo da genética devia centrar-se nas adaptações e não na evolução morfológica em si. Deste modo, a sua compilação de genes e mutações que contribuem para a evolução fenotípica abrange quase exclusivamente as características conhecidas por aumentarem o fitness ou que são preservadas pela seleção.

Isto traz vários problemas. Para começar, Hoekstra e Coyne (2007) deixaram de parte algumas características, não por serem neutras ou deletérias, mas porque não se provou serem adaptações. Imagine-se se apenas fossem consideradas adequadas as características com um significado adaptativo na análise genética e do desenvolvimento. Iríamos reduzir fortemente o espectro de características admissíveis porque a grande maioria dos testes de seleção natural na natureza revelam coeficientes de seleção muito pequenos (Hoekstra et al. 2001; Kingsolver et al. 2001). Infelizmente, a maioria dos

estudos não são capazes de demonstrar que estas curtas seleções de coeficientes são significativamente diferentes de zero. Isto porque, ou as variantes fenotípicas são neutras, ou a dimensão das amostras ou coeficientes de seleção são demasiado pequenos para permitir uma deteção experimental. Este último exemplo é consistente com as recentes análises moleculares da genómica populacional (Eyre-Walker 2006; Andolfatto 2007).

A tentação de contar histórias adaptativas sobre características fenotípicas tem um longo e variado historial na biologia evolutiva (Gould e Lewontin 1979). Além disso, o desejo de contar estas histórias adaptativas conduz à tendência igualmente perversa de rejeitar esclarecimentos adaptativos das características, quando a imaginação nos falha.

Darwin preveniu-nos deste problema (Darwin 1859). No Capítulo 6 da Origem das Espécies, numa secção intitulada “Ação da Seleção Natural sobre órgãos aparentemente pouco importantes”, Darwin refere o seguinte:

Em primeiro lugar, somos demasiado ignorantes relativamente à economia de qualquer ser vivo para podermos dizer quais são as modificações que têm ou não importância.

É algo que continua a ser uma verdade hoje em dia. Por conseguinte, as características sem valor adaptativo explícito devem mesmo assim ser tidas em conta na investigação da base molecular das alterações evolutivas do fenótipo (Nei 2007).

Mas existe um problema ainda maior no estudo de características exclusivamente adaptativas. Tal pode enviesar a nossa compreensão das causas genéticas da variação, relativamente às características sob seleção forte. Como demonstramos na secção “Os Dados: Provas para uma Teoria Predizível da Evolução Genética”, as características sob seleção forte podem ser causadas pela distribuição irregular de mutações. Por exemplo, uma determinada variação genética natural subjacente a polimorfismos aparentemente adaptativos, é causada por mutações que eliminam ou quebram gravemente o produto do gene. Fora isto, muitos destes genes são fortemente conservados noutras espécies, o que sugere que estes recentes polimorfismos adaptativos têm fracas perspectivas, a longo prazo. Não está, portanto, totalmente claro se o foco exclusivo nas características adaptativas, em especial as características sob seleção forte - que são obviamente as mais fáceis de identificar - identificam necessariamente os alelos que são mais relevantes para a evolução a longo prazo.

Existe atualmente uma discrepância fundamental entre as abordagens da genómica populacional ao estudo da adaptação, e os estudos genéticos das características adaptativas “óbvias”, especialmente os polimorfismos preservados nas populações. As abordagens da genómica populacional geram estimativas de coeficientes de seleção muito pequenos, na ordem de 10 para 5 na maioria das fixações adaptativas na *Drosophila* (Andolfatto 2007). Pelo contrário, quando a seleção, no estado selvagem, é avaliada, apresenta resultados cerca de quatro vezes superiores (Hoekstra et al. 2001; Kingsolver et al. 2001). Os estudos de características claramente adaptativas não são, deste modo, representativos

therefore, clearly unrepresentative of the vast majority of substitutions fixed by selection.

Despite our objection to Hoekstra and Coyne's focus on clearly adaptive traits, we do think that there would be enormous value in determining the selection coefficients associated with phenotypic evolution. This would then allow a robust test of whether phenotypic variants with different selection coefficients are caused by different kinds of mutations. However, given the great difficulty of measuring, let alone detecting, weak selection in natural conditions, we are not convinced that in practice it will be possible to gather the required data. Nonetheless, the population genetics argument supporting the *cis*-regulatory hypothesis is, at heart, an argument about the action of natural selection, as shown in the section "Arguments for the *Cis*-regulatory hypothesis."

ARE OBSERVED EXAMPLES OF *CIS*-REGULATORY EVOLUTION BIASED TOWARD TRAIT LOSSES?

Hoekstra and Coyne (2007) parenthetically suggested that the available genetic data might be biased toward trait losses and that trait loss may be easier via *cis*-regulatory changes than through protein-coding changes (p. 1004). They did not provide a molecular explanation for their hypothesis and it is not obvious how

cis-regulatory evolution would more easily generate phenotypic loss than coding changes. For example, loss of a transcription factor binding site for an activator can lead to loss of expression, but loss of a repressor binding site can cause gain of expression. In addition, a loss of gene expression might be associated with a trait gain and vice versa.

In addition, it is not always clear that a human subjective assignment of trait "gain" or "loss" is appropriate. For example, "loss" of trichomes might just as easily be called "gain" of naked cuticle. Most insect epidermal cells differentiate into one of these two alternative states and both states require the regulation of a large set of genes in a regulatory network. It is not at all clear that one represents a gain or loss relative to the other in any objective sense.

Finally, the abundance of studies of trait loss observed by Hoekstra and Coyne (2007) may reflect the more mundane fact that when comparing closely related species, where analysis is more straightforward, phenotypic loss is more common than gain (Adamowicz and Purvis 2006). That is, the apparent bias noted by Hoekstra and Coyne (2007) may reflect the frequency of trait gain versus loss during evolution, rather than a fundamental bias in the way *cis*-regulatory regions evolve.

Supporting Information

The following supporting information is available for this article:

Database for Appendix 1.

References for Appendix 1 Database.

Supporting information may be found in the online version of this article.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

da grande maioria das substituições fixadas pela seleção.

Apesar das nossas objeções ao enfoque de Hoekstra e Coyne em características claramente adaptativas, julgamos que a determinação de coeficientes de seleção associados à evolução fenotípica seria de grande valor. Isto permitiria um teste robusto para saber se as variantes fenotípicas com diferentes coeficientes de seleção são provocadas por diferentes tipos de mutações. No entanto, tendo em conta a elevada dificuldade de avaliação, e mais ainda de deteção, de seleção fraca em condições naturais, não estamos convencidos que será possível, na prática, recolher os dados necessários. Contudo, o argumento da genética de populações que defende a hipótese da regulação *cis* é, no fundo, um argumento sobre a ação da seleção natural, como demonstrado na secção “Argumentos para a hipótese da regulação *cis*”.

OS EXEMPLOS OBSERVADOS DE EVOLUÇÃO DA REGULAÇÃO CIS SÃO ENVIESADOS PARA PERDAS DE CARACTERÍSTICAS?

Hoekstra e Coyne (2007) sugeriram parenteticamente que os dados genéticos disponíveis podem estar enviesados para a influência da perda de características e que a perda da característica pode ser mais facilitada através das alterações na regulação *cis* do que através das alterações na codificação das proteínas (p. 1004). Os autores não fornecem uma explicação molecular para a sua hipótese e não clarificam como é que a evolução da regulação *cis* pode gerar mais

facilmente perdas fenotípicas do que as alterações na codificação. Por exemplo, a perda de um local de ligação de um fator de transcrição para um ativador pode levar à perda de expressão, mas a perda de um local de ligação de um repressor pode provocar o ganho de expressão.

Além do mais, a perda de expressão do gene pode estar associada a um ganho de características, e vice-versa. É também importante referir que a definição humana subjetiva de “ganho” ou “perda” nem sempre é adequada. Por exemplo, a “perda” de tricomas pode muito bem ser designada de “ganho” de cutículas nuas. A maioria das células epidérmicas de insetos distinguem-se num destes dois estados alternativos, e ambos os estados necessitam da regulação de um grande conjunto de genes numa rede de regulação. Não está claro se um destes representa um ganho ou uma perda relativamente ao outro, num sentido objetivo.

Por fim, a abundância de estudos sobre a perda de características observada por Hoekstra e Coyne (2007) pode demonstrar factos mais banais do que a comparação de espécies de relação próxima, na qual a análise é mais direta e a perda fenotípica é mais comum do que o ganho (Adamowicz e Purvis 2006). Isto é, a aparente influência de Hoekstra e Coyne (2007) pode refletir a frequência de ganho de características versus perda, durante a evolução, em vez de uma influência fundamental no modo como as regiões de regulação *cis* se desenvolvem.

Informações de Apoio

Encontra-se disponível a seguinte informação de apoio para este artigo:

Base de Dados para o Apêndice 1

Referências para a Base de Dados do Apêndice 1

A informação de apoio poderá ser obtida na versão em linha deste artigo.

Nota: A Blackwell Publishing não se responsabiliza pelo conteúdo ou funcionalidade de qualquer material de apoio fornecido pelos autores. Quaisquer dúvidas, que não relativas a material em falta, deverão ser colocadas ao respetivo autor do artigo.