Diogo Amaral Rebouças Melo

Evolução Morfológica e Modularidade

São Paulo 2012

## Diogo Amaral Rebouças Melo

# Evolução Morfológica e Modularidade

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na Área de Biologia Evolutiva.

Orientador: Gabriel Marroig

São Paulo 2012 Diogo Amaral Rebouças Melo Evolução Morfológica e Modularidade 87 páginas Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.
1. Genética quantitativa;
2. Modularidade;
3. Matriz G;
4. Simulação computacional;
5. Seleção direcional.
1. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

## Comissão Julgadora:

Prof. Dr. Nome Prof. Dr. Nome

Prof. Dr. Gabriel Marroig

Para Nancy e Nelcivone

## Epígrafe

No man is an island entire of itself; every man is a piece of the continent, a part of the main; if a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as a manor of thy friends or of thine own were; any man's death diminishes me, because I am involved in mankind. And therefore never send to know for whom the bell tolls; it tolls for thee.

John Donne

## Agradecimentos

Primeiro ao Gabriel, esse orientador fantástico, desafiador, profundamente interessado, com uma dedicação comovente, que arranca o melhor de todos nós, caso nós queiramos ou não.

Todos nós, no caso, são os membros do LEM, que absolutamente fazem jus ao orientador que tem. É o melhor ambiente de trabalho que eu poderia pedir, e obrigado a todos por tudo. Harley, Paulinha, Aninha, Dani, Tafinha, Alex, Lugar, Roberta, Fino, Arthur, Leila, Monique, Wally, Pato, Edgar e Anna, vocês são pessoas incríveis, elevam o nível de qualquer lugar que estejam e me deixam feliz de fazer o que eu faço. Este trabalho deve especialmente aos três últimos, Pato por ouvir minhas besteiras e pegar meus erros ao longo de todo o trabalho; Edgar, por me ensinar mais sobre programação e matemática que quase todos os meus professores; e a Anna, que é a melhor amiga que eu já tive e me manteve lúcido em mais de uma ocasião.

Meu pai, que me apoia de um jeito indescritível, e que me definiu como pessoa de um jeito que acho que nem ele sabe. Se eu tenho algum modelo, alguma meta, é ser alguém tão bom e incrível e generoso como ele, que faz isso como se fosse a coisa mais fácil do mundo desde sempre. Minha mãe, essa companhia fantástica, tão dedicada e carinhosa, que foi quem me inspirou a ser um cientista, que me ajuda e me apoia e me faz querer ser sempre melhor. Cassola, que me ensina e me mostra como é ser um cientista de verdade desde que eu me lembro, Júnia, Rachel, Carol, Haroldo, vocês são minha família, sem meias palavras. Rita, meu amor, que mesmo longe faz parte de mim.

E claro, a FAPESP e a Capes, que pagaram a conta.

#### Resumo

Sistemas morfológicos quantitativos são descritos por medidas contínuas. A relação genética entre essas características dos indivíduos é representada pela matriz de covariância genética aditiva, a matriz G. Entender a evolução da matriz G, portanto, é de suma importância para compreender os padrões de diversificação encontrados na natureza. Neste trabalho estudamos modelos computacionais para a evolução de traços contínuos em populações naturais, sujeitas a variados tipos de seleção e condições internas, focando no problema da evolução dos padrões de integração e modularidade nessas populações. Testamos dois modelos com diferentes combinações de parâmetros em sua capacidade de reproduzir e elucidar padrões naturais. Seleção direcional correlacionada se mostrou uma força importante na criação desses padrões de covariação e a seleção estabilizadora correlacionada se mostrou fundamental para a manutenção desses padrões.

**Palavras-chave:** genética quantitativa, modularidade, matriz G, simulação computacional, seleção direcional

#### Abstract

Quantitative morphological systems are described by continuous measurements. The genetic relation between these characteristics of the individuals is represented by the genetic additive co-variance matrix, the G matrix. Understanding the evolution of the G matrix is, therefore, of paramount importance for proper interpretation of the patterns of diversification we observe in nature. In this work we study computational models for the evolution of quantitative traits in natural populations, subject to different natural selection and internal conditions, focusing on the problem of the evolution of the patterns of morphological integration and modularity. We test two models with different sets of parameters in their ability to reproduce and elucidate natural patterns. Directional correlated selection was necessary for the shaping of the patterns of morphological integration, and correlated stabilizing selection was fundamental to the maintenance of these patterns.

**Keywords:** quantitative genetics, modularity, G matrix, computational simulation, directional selection

# Lista de Figuras

1.1	Representação esquemática da organização modular dos seres vivos. As	
	setes representam qualquer tipo de relação entre as partes de um indiví-	
	duo. Adaptado de Klingenberg (2008)	3

- 3.2 Evolução da correlação entre três caracteres ligados por efeitos mutacionais pleiotrópicos sobre seleção estabilizadora correlacionada. A seleção propicia a integração entre dois caracteres (linha vermelha) e a desintegração desses dois com o terceiro (linhas azul e verde). Nesta simulação  $r_{\omega}$  como na equação 3.1,  $Ne = 5.000, p = 3, m = 50, \ldots 35$
- 3.3 Comparação entre a matriz de correlação final para 3 caracteres após 10.000 gerações de seleção e a matriz  $\omega$  da superfície de seleção. . . . . . 36
- 3.4 Evolução da correlação entre dez caracteres ligados por efeitos mutacionais pleiotrópicos sobre seleção estabilizadora correlacionada. Apesar da seleção promover a integração entre os 5 primeiros e os 5 últimos caracteres e a desintegração entre esse módulos, isso não se transmite à matriz de correlação. Nesta simulação  $r_{\omega}$  como na equação 3.3, Ne = 5.000,  $p = 10, m = 50. \dots 337$

3.5	Evolução da modularidade $L$ e AVG-Ratio para as matrizes de correlação representadas na figura $3.4$	37
3.6	Comparação entre a matriz de correlação final para 10 caracteres após 10.000 gerações de seleção representadas na figura 3.4 e a matriz $\boldsymbol{\omega}$ da superfície de seleção	38
3.7	Evolução da correlação entre dez caracteres ligados por efeitos mutaci- onais pleiotrópicos sobre seleção estabilizadora correlacionada e seleção direcional intensa para mudança correlacionada dentro dos módulos. No-	
	vamente, isso não se traduz na matriz de correlação genética	39
3.8	Evolução da modularidade $L$ e AVG-Ratio para as matrizes de correlação representadas na figura $3.7$	39
3.9	Comparação entre a matriz de correlação final para 10 caracteres após 10.000 gerações de seleção direcional e estabilizadora representadas na	
	figura 3.7 com a matriz $\pmb{\omega}$ da superfície de seleção	40
3.10	Matriz final média de 10 corridas para diferentes razões de $\mu$ e $\mu_B$ , com seleção estabilizadora correlacionada com 2 módulos	43
3.11	Evolução típica de AVG-Ratio e modularidade $L$ para diferentes razões de $\mu$ e $\mu_B$ , com seleção estabilizadora correlacionada com 2 módulos	44
3.12	Comparação entre a matriz de correlação final média de 10 corridas para o 10 caracteres após 10.000 gerações de seleção direcional e estabilizadora com a matriz $\boldsymbol{\omega}$ da superfície de seleção e $\mu/\mu_{P} = 0.1$ , $Ne = 2.500$ .	
	$\Delta_S = 0.2, \ m/p = 2  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots $	45
3.13	Matriz final média de 10 corridas para diferentes razões de $\frac{\mu}{\mu_B}$ , com seleção estabilizadora correlacionada com 2 módulos e seleção direcional correlacionada favorecendo os mesmos módulos. ( $N_e = 2.500$ , $\Delta_S = 0.2$ ,	
	m/p=2)	46

- 3.14 Evolução típica de AVG-Ratio e modularidade L para diferentes razões  $\mu/\mu_B$ , com seleção estabilizadora correlacionada com 2 módulos e seleção direcional correlacionada favorecendo os mesmos módulos. Atenção para a mudança de escala nos eixos verticais. ( $N_e = 2.500$ ,  $\Delta_S = 0.2$ , m/p = 2) 47

- 3.21 Evolução de modularidade L e AVG-Ratio para as correlações apresentadas na figura 3.20 ( $\Delta_S = 0.2$  por 5.000 gerações,  $\mu/\mu_B = 5$ , m/p = 2). 55
- 3.23 Evolução de modularidade L e AVG-Ratio para as correlações apresentadas na figura 3.22 ( $\Delta_S = 0.2$  por 5.000 gerações,  $\mu/\mu_B = 5$ , m/p = 2). 57
- 3.24 Evolução da correlação entre dez caracteres ligados por efeitos pleiotrópicos controlados por uma função ontogenética linear binária variável sofrendo seleção estabilizadora correlacionada e seleção direcional intensa para mudança correlacionada dentro dos módulos, para diferentes valores do número de loci. ( $Ne = 5.000, \mu/\mu_B = 5, \Delta_S = 0.2$ ) . . . . . . . . . 60

- 4.3 Evolução da média das correlações dentro de cada módulo e entre módulos de uma população sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .  $\mu/\mu_B = 5$ , Ne = 5.000, m/p = 5. Seleção direcional a partir da geração 20.000. . . 69

# Lista de Tabelas

2.1	Tabela de abreviações	32
3.1	Intensidades de seleção, representadas pelo diferencial de seleção por caráter e pelo valor do gradiente de seleção por caráter $ \beta_i $	49
4.1	Intensidades de seleção, representadas pelo diferencial de seleção por ca- ráter e pelo valor do gradiente de seleção por caráter $ \beta_i $	66
4.2	Intensidades de seleção, representadas pelo diferencial de seleção por ca- ráter e pelo valor do gradiente de seleção por caráter $ \beta_i $ para as simu-	
	lações com 1.000 gerações de seleção direcional correlacionada	74

# Sumário

1	Intr	Introdução			
	1.1	Modularidade	1		
	1.2	Genética Quantitativa e Modularidade	4		
	1.3	Evolução da matriz ${\bf G}$ e a resposta à seleção	7		
	1.4	Consequências Evolutivas	16		
	1.5	Evolução da Modularidade	18		
2	Moo	delagem	21		
	2.1	Introdução	21		
	2.2	Modelos	21		
		2.2.1 Modelo com Matriz $\mathbf{M}$	21		
		2.2.2 Modelo com Matriz <b>B</b>	25		
	2.3	Parâmetros	27		
	2.4	Estatísticas de Modularidade e Integração	29		
	2.5	Resumo de Parâmetros	31		
3	Teste e Validação dos Modelos		33		
	3.1	Introdução	33		
	3.2	Resultados e Discussão · Matriz $M$	33		
		3.2.1 Seleção Estabilizadora	33		

		3.2.2	Seleção Direcional	36
	3.3 Resultados e Discussão · Matriz <b>B</b>			41
		3.3.1	Seleção Estabilizadora	41
		3.3.2	Seleção Direcional	42
		3.3.3	Intensidade de Seleção Direcional	45
		3.3.4	Seleção direcional seguida de seleção estabilizadora	49
		3.3.5	Intensidade de recombinação	58
4	Evo	lução	de Modularidade em Diferentes Regimes de Seleção	63
4	<b>Evo</b> 4.1	olução Introd	de Modularidade em Diferentes Regimes de Seleção	<b>63</b> 63
4	<b>Evo</b> 4.1 4.2	lução Introd Estabo	de Modularidade em Diferentes Regimes de Seleção         ução	<b>63</b> 63 63
4	<b>Evo</b> 4.1 4.2 4.3	lução Introd Estabo Seleçã	de Modularidade em Diferentes Regimes de Seleção         ução	<ul> <li>63</li> <li>63</li> <li>63</li> <li>64</li> </ul>
4	<b>Evo</b> 4.1 4.2 4.3 4.4	lução Introd Estabo Seleçã Seleçã	de Modularidade em Diferentes Regimes de Seleção         ução         ução         elecimento de equilíbrio         o Direcional         o direcional "corredor"	<ul> <li>63</li> <li>63</li> <li>63</li> <li>64</li> <li>71</li> </ul>
4	Evo 4.1 4.2 4.3 4.4 Con	lução Introd Estabo Seleçã Seleçã nsidera	de Modularidade em Diferentes Regimes de Seleção ução	<ul> <li>63</li> <li>63</li> <li>64</li> <li>71</li> <li>77</li> </ul>

# Capítulo 1

# Introdução

### 1.1 Modularidade

Na imensa maioria dos organismos, conseguimos identificar partes relativamente discretas e separadas, frequentemente envolvidas no desempenho de alguma função. Em organismos unicelulares podemos distinguir organelas desempenhando funções específicas, bem como regiões internas ou na membrana responsáveis por processos distintos. Já nos multicelulares, tipos celulares são organizados em tecidos espacialmente separados, formando órgãos de funções distintas, que por sua vez são organizados em sistemas responsáveis por funções distintas. Modularidade se refere a esse padrão de organização dos seres vivos, onde algumas partes são mais relacionadas entre si do que com outras partes do mesmo organismo. Podemos descrever, e entender, a organização entre partes constituintes dos organismos através das relações entre elas, sendo cada tipo de relação adequada a um nível de complexidade ou organização. As partes do organismo as quais nos referimos podem ser as bases nitrogenadas de uma molécula de RNA (Ancel e Fontana, 2000), genes (Costanzo et al., 2010), proteínas (Han et al., 2004), caracteres morfológicos como peso, altura ou marcos anatômicos (Klingenberg, 2008; Porto et al., 2009; Marroig et al., 2009). Essas relações podem ser medidas de diversas formas, como interação física entre proteínas, padrões de expressão conjunta entre genes, ou, como é o caso no presente trabalho, correlação entre caracteres quantitativos. Esse grupo de características muito relacionadas entre si constituem um módulo, como esquematizado na figura 1.1. Módulos, então, são caracterizados por uma alta conectividade interna e relativa independência de outros módulos.

A formação desses módulos reflete uma complexa interação entre genes, desenvolvimento e ambiente, e tem consequências diversas para o funcionamento e evolução dos seres vivos (Wagner, 1996). Talvez o aspecto funcional seja o ponto de partida mais simples para entender a origem da organização modular. Interação com o ambiente e a necessidade de realizar determinada função provoca respostas evolutivas nas populações, privilegiando o surgimento de estruturas e sistemas adaptados para realizar essas funções. Para que essas estruturas possam se adequar às diferentes pressões evolutivas que cada parte do organismo está sujeita, uma relativa independência entre elas se estabelece (Wagner et al., 2007; Klingenberg, 2008). Essa independência se estabelece em vários níveis. Através de experimentos controlados de cruzamento e uso de marcadores genéticos, é possível mapear regiões no genoma envolvidas com determinação de características macroscópicas nos organismos. Com essas metodologias foi verificado que loci tendem a influenciar regiões discretas, sendo poucos os genes com efeitos entre regiões funcionais diferentes (Cheverud et al., 1997). Genes que influenciam mais de uma característica de um indivíduo são denominados pleiotrópicos. Ou seja, seleção privilegia pleiotropia restrita a regiões funcionais distintas (Cheverud, 1984). Além disso, as estruturas e processos ontogenéticos responsáveis pela formação dos indivíduos apresentam uma independência relativa, sendo possível alterar parte do processo ou estrutura sem um efeito cascata sobre outros processos ou estruturas (Klingenberg, 2008). Talvez o exemplo mais simples seja imaginar um organismo onde todos os caracteres ou processos ontogenéticos estejam integrados (altamente correlacionados) entre si, ou, dito de outra forma, pertencentes ao mesmo módulo. Nesse organismo qualquer mudança em qualquer caráter ou processo levaria a alterações em todos os outros. Uma alteração pode ser benéfica em um caráter, mas deletéria em vários outros. A modularidade evita essa cenário, permitindo com que partes independentes se alterem de forma relativamente independente. O outro lado da moeda são os processos de integração, resultantes da necessidade do organismo de manter a coerência interna dos seus módulos e do indivíduo como um todo. Integração leva a aumento da coesão entre as características do organismo. Evolutivamente, integração e modularidade permitem que grupos de características funcionalmente ou ontogeneticamente ligadas se modifiquem de forma harmoniosa; e que características em módulos diferentes possam se alterar de forma relativamente independente.



Figura 1.1: Representação esquemática da organização modular dos seres vivos. As setes representam qualquer tipo de relação entre as partes de um indivíduo. Adaptado de Klingenberg (2008)

Em caracteres quantitativos, descritos em detalhe na seção 1.2, a expressão de todos os processos genéticos, ontogenéticos e ambientais que formam a organização modular do organismo se dá na forma de covariação (ou correlação) entre os caracteres. Ou seja, o conflito entre os processos de modularidade e integração em todos os níveis de organização do indivíduo podem ser percebidos na estrutura de covariação final da população (Klingenberg, 2008). Módulos caracterizados por alta correlação entre caracteres dentro do módulo e baixa correlação entre caracteres de módulos diferentes são chamados de módulos variacionais (Wagner et al., 2007). Estes são frequentemente identificados como relacionados a funções específicas, como mastigação, inserção muscular ou proteção (Cheverud et al., 1997). Novamente, seleção privilegia interações pleiotrópicas e ontogenéticas restritas a características ligadas a uma função comum, que resulta em covariação entre essas características na população (Wagner, 1996).

#### 1.2 Genética Quantitativa e Modularidade

O estudo de caracteres quantitativos se baseia amplamente na teoria da genética quantitativa. A genética quantitativa estuda a herança de características contínuas nos indivíduos de uma população, como peso, altura ou taxas de crescimento (Falconer e Mackay, 1996). A partir desse formalismo, originário do melhoramento genético de animais e plantas, podemos prever como a média e a distribuição de características fenotípicas varia em uma população de uma geração para outra.

Caracteres contínuos são usualmente determinados geneticamente por muitos loci, sendo portanto denominados caracteres poligênicos. A influência exata de cada um desses loci no fenótipo do individuo é desconhecida, porém a combinação de todos esses efeitos e dos efeitos ontogenéticos e ambientais resultam no valor fenotípico observado. Em Crow e Kimura (1964) e Kimura (1965) os autores descrevem um modelo para os efeitos alélicos agindo sobre caracteres contínuos. Neste modelo, chamado modelo do continuo de alelos, cada locus em principio pode ter qualquer efeito sobre o tamanho de um caráter, aumentando ou diminuindo seu valor fenotípico, e qualquer um desses efeitos pode ser atingido via mutação. A enorme variabilidade possível em sequências genéticas e o número de loci controlando cada caráter garante que essa suposição seja verossímil. Nesse contexto, o valor fenotípico de um caráter p é dado pela soma dos efeitos genéticos (g) e ambientais (e) atuando sobre ele:

$$p = g + e \tag{1.1}$$

A variação total de uma dada característica é resumida na sua variância fenotípica, denominada  $V_p$ . Esse valor de variância total é fruto da variação genética  $(V_g)$  presente na população e da variação devido ao ambiente no qual a população se encontra  $(V_e)$ , além da possível interação entre elas. Como as variâncias são aditivas, podemos escrever:

$$V_p = V_g + V_e + V_{g \cdot e} \tag{1.2}$$

Em organismos com sistemas de desenvolvimento estáveis, relativamente tamponados, como é o caso dos mamíferos, podemos em geral ignorar o termo de interação entre genótipo e ambiente ( $V_{g\cdot e} = 0$ ). Podemos continuar particionando as variâncias. A variância genética é resultado de toda a variação no efeitos aditivos dos alelos ( $V_a$ , variância aditiva), mais os efeitos não aditivos, efeitos de dominância ( $V_d$ ) e efeitos epistáticos ( $V_i$ , representando interações entre diferentes loci). Ou seja:

$$V_q = V_a + V_d + V_i \tag{1.3}$$

O termo aditivo,  $V_a$ , representa a porção da variância genética responsável pela semelhança por parentesco. Isso se reflete na equação de resposta à seleção direcional univariada. A pressão seletiva sobre alguma característica fenotípica pode ser descrita pela diferença na média da população parental antes ( $\overline{z}$ ) e depois ( $\overline{z}^*$ ) do evento de seleção. Essa quantidade é denominada diferencial de seleção e é representada por S. Esse tipo de seleção, que afeta média de um caráter na população, é chamado de seleção direcional. A resposta observada na média na geração seguinte ( $\Delta \overline{z}$ ) é dada por:

$$\Delta \overline{z} = \frac{V_a}{V_p} (\overline{z}^* - \overline{z}) = h^2 S \tag{1.4}$$

Ou seja, a parcela da variância fenotípica devido à variação genética aditiva define o quão eficiente será a resposta da população à seleção direcional. Essa parcela é denominada herdabilidade, representada por  $h^2$ . Mas essa equação ignora o efeito de seleção indireta em outras características do indivíduo. Para levar em conta esses efeitos, devemos lançar mão de um formalismo multivariado.

A equação multivariada de resposta à seleção, proposta por Lande (1979), permite ligar a mudança na média de um conjunto de caracteres ( $\Delta \mathbf{z}$ ) ao diferencial de seleção imposto à população ( $\mathbf{S}$ ) e à suas matrizes de covariação genética aditiva e fenotípica ( $\mathbf{G} \in \mathbf{P}$ ).

$$\Delta \mathbf{z} = \mathbf{G} \mathbf{P}^{-1} \mathbf{S} \tag{1.5}$$

 $\Delta \mathbf{z}$  e  $\mathbf{S}$  agora são vetores de mesma dimensão que o número de características consideradas na analise. As variâncias são expressas nas matrizes  $\mathbf{G}$  e  $\mathbf{P}$ , que são os análogos multivariados das quantidades  $V_a$  e  $V_p$ . O caso multivariado apresenta um comportamento qualitativamente diferente do caso univariado. Como os caracteres são ligados por efeitos genéticos correlacionados, representados pela matriz  $\mathbf{G}$ , a seleção direcional não é capaz de atuar de forma isolada em cada caráter. Seleção direcional, ainda que restrita a um caráter, provoca respostas evolutivas em todos os caracteres que covariam com o caráter sobre seleção. e Utilizando o conceito de superfície de seleção, proposto por Wright (1932), podemos interpretar o termo de seleção  $\mathbf{P}^{-1}\mathbf{S}$ de forma geométrica. A superfície de seleção associa cada valor fenotípico para todos os caracteres a um valor de aptidão ou *fitness*. Caso essa superfície seja gaussiana, Lande e Arnold (1983) mostraram que  $\mathbf{P}^{-1}\mathbf{S}$  é equivalente ao gradiente da média da superfície de seleção na população. O gradiente de um campo escalar, como é caso da superfície de seleção, é definido como o campo vetorial que aponta para a direção de maior variação do campo escalar, com magnitude igual a taxa de variação do campo escalar. Ou seja, o gradiente em cada ponto da superfície de seleção indica qual é a direção de mudança fenotípica que mais aumenta a aptidão da população que se encontra nesse ponto do morfoespaço. Por esse motivo, a quantidade  $\mathbf{P}^{-1}\mathbf{S}$  é chamada de gradiente de seleção, e é representada pelo vetor  $\beta$ . Nas próximas seções vamos nos referir também a seleção estabilizadora. Esse tipo de seleção age sobre a distribuição de um caráter, diminuindo sua variabilidade, mas não sobre a média do caráter. Com respeito à superfície de seleção, a seleção estabilizadora é resultado da curvatura da superfície, ou da sua Hessiana (Lande e Arnold, 1983). No caso de seleção direcional, podemos escrever a equação de resposta a seleção como:

$$\Delta \overline{\mathbf{z}} = \mathbf{G} \mathbf{P}^{-1} S = \mathbf{G} \beta \tag{1.6}$$

A evolução dos caracteres por seleção direcional toma a forma de uma subida de gradiente da superfície de seleção, alterado pela estrutura de covariação genética, representada pela matriz **G**.

### 1.3 Evolução da matriz G e a resposta à seleção

Na seção 1.1, nós descrevemos como os processos seletivos interagem com o desenvolvimento e arquitetura genética das populações, gerando modularidade. Dissemos também que a modularidade presente nos processos internos dos organismos se manifesta na população na sua estrutura de covariação. Essa estrutura de covariação é representada pela matriz **G**. Ela descreve como a variação aditiva, herdável, da população está particionada entre os seus caracteres, e como estes se relacionam entre si. Caracteres associados por efeitos pleiotrópicos e ontogenéticos irão apresentar covariação alta, enquanto caracteres independentes, não relacionados, irão apresentar covariação baixa. Como a população responde à seleção direcional, portanto, depende da sua estrutura de modularidade e de como os caracteres estão relacionados entre si.

A equação de Lande (1979) também pode ser invertida e usada de forma retrospectiva. Ou seja, partindo da matriz  $\mathbf{G}$  e da diferença observada entre as médias antes de depois do evento de seleção, podemos inferir qual foi o gradiente de seleção que gerou essa resposta. Mais claramente:

$$\beta = \mathbf{G}^{-1} \Delta \mathbf{z} \tag{1.7}$$

Além disso, esse mesmo raciocínio pode ser usado para investigar padrões macroevolutivos de resposta ao longo de muitas gerações (Lande e Arnold, 1983; Marroig e Cheverud, 2004, 2005, veja também a equação 1.9).

Porém, nossa habilidade de inferir a resposta evolutiva a muitas gerações de seleção depende fundamentalmente da constância da matriz G ao longo do tempo, e a teoria de Lande não explora detalhadamente a dinâmica da própria matriz G, assumindo implicitamente nas equações sua constância. Dizer que a distribuição das características na população é gaussiana implica que essa distribuição não é afetada pela seleção direcional (Barton e Turelli, 1987). Isso se reflete na equação de resposta a seleção direcional, que envolve apenas mudanças na média dos caracteres. Essa suposição simplificadora de Lande se baseia em um resultado derivado do modelo do continuo de alelos, demonstrado por Kimura (1965): sobre seleção estabilizadora, a distribuição de equilíbrio dos efeitos alélicos sobre um determinado caráter é gaussiana, supondo que os efeitos mutacionais sejam pequenos frente à variação existente na população. Ou seja, assumindo que os efeitos alélicos sejam efetivamente gaussianos, a dinâmica da média da população pode ser completamente descrita apenas pela média e pela matriz de covariação desses efeitos (Barton e Turelli, 1987).

Notamos então que uma série de suposições não triviais devem ser feitas para que esse modelo gaussiano quadrático de Lande seja teoricamente plausível. Esses problemas foram apresentados ao longo da década de 80 por uma série de artigos, principalmente e pioneiramente por Turelli (1984, 1985, 1986); Barton e Turelli (1987, 1989). Talvez a questão mais importante seja a suposição de que os alelos tenham distribuição gaussiana proposta por Kimura (1965). Para que isso seja realista, a taxa de mutação por locus deve ser suficientemente alta para manter a distribuição de alelos por locus gaussiana frente à deriva e seleção estabilizadora, que removem variabilidade (Falconer e Mackay, 1996). Além disso, podemos nos perguntar se mutações novas realmente tem efeitos pequenos em relação a variância original da população nas direções morfológicas afetadas, ou se novas mutações introduzem variabilidade maior que a originalmente presente na população. O modelo proposto por Turelli (1984) utiliza taxas de mutação baixas e efeitos mutacionais altos, mas ainda mantendo o espaço de alelos contínuo. Neste modelo, a maior parte da variação fenotípica acaba sendo atribuída a alelos raros, de baixa frequência, mas com efeitos aditivos grandes, e não a vários alelos polimórficos de efeitos pequenos e distribuição gaussiana. Nessas condições, a matriz  ${\bf G}$  não seria constante ao longo do tempo e fatores de correção deveriam ser acrescidos à equação de resposta de Lande para descrições macro evolutivas. Esses fatores dependem do desvio por geração da matriz G em relação à matriz G média no período de seleção (Jones et al., 2004). A estimativa desse "efeito Turelli" é difícil de ser realizada praticamente.

Em Barton e Turelli (1987), os autores, interessados no problema da dinâmica das variâncias de forma geral, apresentam um formalismo relativamente completo para lidar com o problema da evolução de todos os momentos da distribuição fenotípica de um caráter quantitativo, abordando tanto a aproximação gaussiana quanto a aproximação de alelos raros, proposta por Turelli (1984). O procedimento seguido pelos autores é escrever a dinâmica de mudança de todos o momentos da distribuição de efeitos alélicos da população sujeita a uma pressão de seleção na forma de uma subida de gradiente imposta por uma superfície de seleção arbitrária (mas com todas as derivadas parciais bem definidas no espaço dos momentos) (Arnold et al., 2001). Isso permite abarcar o formalismo de Lande, tomando apenas os dois primeiros momentos (média e variância), supondo o segundo momento fixo e desprezando os momentos mais altos; ou incluir mudanças até uma certa ordem e obter os resultados de Turelli. Por outro lado, a descrição do sistema baseada nos seus momentos é analiticamente bastante complicada, e não permite a inclusão de epistasia (interação entre loci) de forma simples. Posteriormente, buscando um formalismo mais natural e genérico para a evolução de caracteres quantitativos associados por correlações genéticas, os mesmo autores em Turelli e Barton (1994) apresentam um formalismo baseado nos cumulantes da distribuição de efeitos alélicos. Apesar de menos intuitivo, o formalismo baseado em cumulantes é mais simples no caso de alelos aditivos, e fornece equações recursivas para a evolução dos caracteres, que podem ser calculadas computacionalmente. Apesar disso, ainda temos que ignorar efeitos epistáticos, que podem ter efeitos importantes na evolução.

Epistasia pode ser definida de duas formas: epistasia fisiológica e estatística (Cheverud e Routman, 1995). A epistasia estatística, cuja variância  $V_i$  aparece na equação 1.3, se refere a desvios nos valores aditivos individuais de cada alelo quando em condições multi locus, ou seja, desvios populacionais nos valores aditivos de um alelo devido a interações entre locus (Falconer e Mackay, 1996). Já a epistasia fisiológica se refere a interações na expressão de alelos em locus diferentes, dentro de cada indivíduo, ou seja, a expressão de um dado alelo depende do estado de outro locus. A epistasia fisiológica é um pré-requisito para a existência da epistasia estatística. Enquanto a epistasia estatística afeta apenas o termo de interação ( $V_i$ ) na partição de variância genética, a epistasia fisiológica pode ser bastante comum e afetar também os termos aditivos e de dominância na partição de variâncias, influindo, portanto, na semelhança por parentesco (Crow e Kimura, 1970; Falconer e Mackay, 1996; Cheverud e Routman, 1995, 1996). Além disso, epistasia fisiológica fornece uma fonte de variação críptica importante, introduzindo não linearidade na resposta à seleção. Em Cheverud e Routman (1996) vemos um exemplo disso, com interações epistáticas se mostrando capazes de aumentar a variância aditiva de uma população após um evento de gargalo populacional, mesmo com uma queda na diversidade alélica da população. Isso se deve a uma mudança na relação entre os loci decorrente da alteração nas suas frequências alélicas causadas pelo gargalo populacional. Na ausência de epistasia e dominância, um gargalo populacional sempre causa declínio da variância aditiva (Falconer e Mackay, 1996). O aumento da variação aditiva pode ter consequências drásticas para a evolução de uma população. Em Whitlock (1995) o cenário de mudanças entre picos adaptativos é explorado juntamente com aumento de variância de uma população. Whitlock mostra como a mudanças entre picos adaptativos pode ser feita de forma trivial, sem passar por períodos de baixo valor adaptativo, quando aliada a aumento de variação fenotípica. Dessa forma, novos nichos adaptativos podem ser ocupados e mudanças grandes no fenótipo podem se estabelecer com a ação da epistasia.

Dada a complexidade do problema da dinâmica das variâncias em um caso multivariado, o estudo da estabilidade da matriz  $\mathbf{G}$  e suas consequências para o estudo da evolução de caracteres quantitativos se tornou um problema eminentemente experimental, sejam esses experimentos conduzidos de forma retrospectiva, amostrando a diversidade dos seres vivos, seja em experimentos manipulativos feitos com colônias de animais em cativeiro ou, como é o caso do presente trabalho, por via de simulações computacionais (para uma revisão sobre estabilidade da matriz  $\mathbf{G}$  veja Arnold et al., 2008). Vamos agora revisar as principais abordagens computacionais.

A sequência de artigos Jones et al. (2003, 2004, 2007) foi pioneira nos estudos computacionais de estabilidade da matriz **G** num contexto multivariado moderno, abordando o tema da evolução de dois caracteres correlacionados em diferentes situações. O estilo de simulação apresentado em Jones et al. (2003) e utilizado em todos os artigos subsequentes serviu de ponto de partida para nossas simulações e será descrito em detalhes nas próximas seções.

O problema mais simples, abordado em Jones et al. (2003), consiste no estudo de estabilidade da matriz  $\mathbf{G}$  associada a dois caracteres quantitativos com interações pleiotrópicas fixas em uma população finita sofrendo mutação, deriva e seleção estabilizadora gaussiana. Os principais parâmetros que influem na estabilidade da matriz G nessas simulações são o padrão de mutação pleiotrópico e a superfície de seleção. Ambos são descritos por matrizes de covariância. Os valores na diagonal dessas matrizes representam a intensidade das mutações e a intensidade da seleção estabilizadora em cada caráter separadamente. Como o problema é bidimensional, essas matrizes tem apenas um valor fora da diagonal. Esse valor é apresentado como uma correlação de mutação no caso da matriz mutacional  $(r_{\mu})$  e uma correlação de seleção na matriz que descreve a superfície de aptidão ou fitness  $(r_{\omega})$ . O valor de  $r_{\mu}$  define o quão correlacionados serão os valores das mutações que afetam ambos os caracteres. Quanto mais alto, mais correlacionados eles serão, com sua magnitude dada pelos valores diagonais da matriz de mutação. Já o valor de  $r_{\omega}$  define o quão forte é a seleção estabilizadora no sentido de manter os dois caracteres correlacionados. Nas condições das simulações, o fator mais importante para a estabilidade da matriz G foi a existência de mutações correlacionadas. Quando  $r_{\mu}$  e  $r_{\omega}$  são similares, a matriz tende a ser bastante estável. Além disso, tamanhos populacionais grandes também contribuem para a estabilidade da matriz. Um resultado interessante é que existem tipos diferentes de estabilidade para matrizes de covariância. Por exemplo: uma matriz pode manter seus autovalores estáveis, não sofrendo alterações na quantidade de variância genética disponível, mas ter autovetores variáveis, alterando assim em qual direção do morfoespaço a variabilidade populacional está disponível para responder a seleção. Os autores ressaltam que matrizes com um autovalor dominante tendem a ser mais estáveis. Ou seja, correlação alta entre os dois caracteres leva à estabilidade da matriz  $\mathbf{G}$ . Isso pode ser interpretado como um primeiro indício para a importância da integração e modularidade nesses sistemas, apesar de ser difícil definir esses conceitos de maneira convincente com apenas dois caracteres. Em resumo, esse artigo mostra que é possível observarmos matrizes estáveis em condições plausivelmente naturais, e dá o primeiro passo para a quantificação dessas condições.

O próximo passo foi a inclusão de seleção direcional em Jones et al. (2004). O estilo de simulação é basicamente idêntico ao artigo anterior, mas agora a posição do pico adaptativo gaussiano é variado, alterando assim a média da população e criando uma seleção direcional. Além de permitir abordar a questão da estabilidade sob novas condições, essa nova simulação permite também explorar a reconstrução da seleção direcional via equação de Lande (1979) e quantificar o efeito de Turelli devido a flutuações na matriz **G**. Para entender o efeito Turelli, vamos explorar como a extensão da equação de Lande deve ser feita para abarcar a mudança evolutiva ao longo de várias gerações. Começamos com a equação de resposta à seleção nas médias dos caracteres ( $\overline{z}$ ) para uma única geração de uma população sujeita ao gradiente de seleção  $\beta_t$ :

$$\Delta \overline{\mathbf{z}} = \overline{\mathbf{z}}_{t+1} - \overline{\mathbf{z}}_t = G\beta_t \tag{1.8}$$

Com essa equação, e de posse das médias antes e depois da seleção e da matriz **G** dessa população, podemos estimar o valor de  $\beta_t$ . Além disso, podemos pensar em definir um gradiente médio  $\beta_T$  ao longo de várias gerações de seleção (Lande, 1979), como:

$$\beta_T \equiv \sum_{t=0}^{T-1} \beta_t = \overline{\mathbf{G}}^{-1} \Delta \overline{\mathbf{z}}_T \tag{1.9}$$

Ou seja, a soma dos gradientes individuais, onde  $\overline{\mathbf{G}}$  representa a matriz  $\mathbf{G}$  média e  $\Delta \overline{\mathbf{z}}_T$  a mudança global na média dos caracteres. Turelli (1988) mostrou que essa equação se torna incorreta caso a flutuação da matriz  $\mathbf{G}$  em torno da média seja grande entre uma geração e outra. Isso fica claro expandindo essa equação de forma a mostrar os termos de flutuação:

$$\beta_T = \overline{\mathbf{G}}^{-1} \left[ \Delta \overline{\mathbf{z}}_T - \sum_{t=0}^{T-1} (\mathbf{G}_t - \overline{\mathbf{G}}) \beta_t \right]$$
(1.10)

Se o termo de Turelli ( $\sum_{t=0}^{T-1} ({\bf G}_t - \overline{{\bf G}}) \beta_t$ ) for grande, a estimativa de  $\beta_T$  será precária. É necessário então avaliar se a estabilidade da matriz G é suficiente para manter esse termo pequeno o bastante para possibilitar a estimativa de gradientes de seleção realistas. Para isso, foram realizadas simulações com 3 tipos de movimentos do pico adaptativo: ao longo da direção de aumento de um caráter, mantendo o outro estável; aumentando simultaneamente os dois caracteres, na direção usualmente descrita como direção de tamanho, por representar aumento simultâneo de todas os caracteres do indivíduo (Marroig e Cheverud, 2005); e na direção de aumento de um caráter e diminuição do outro. Estas situações são representadas pelos símbolos  $\rightarrow,$   $\nearrow$ e $\searrow,$  respectivamente, fazendo uma alusão à representação bidimensional cartesiana dos caracteres. Elas são interessantes por representarem tipos de seleção que interagem de forma diferente com a presença da correlação mutacional. Quando  $r_{\mu}$  for positivo, a seleção  $\nearrow$ está alinhada com o eixo de maior variação da matriz M. Esta direção é denominada eixo de menor resistência evolutiva (Schluter, 1996), e representa a direção do morfoespaço com maior quantidade de variação para responder a seleção natural. Como neste modelo mutações são geradas preferencialmente ao longo dos eixos de maior variação da matriz  $\mathbf{M}$ , a direção do primeiro componente principal da matriz mutacional tende ser aquele com maior variação disponível para resposta à seleção. Apesar disso, alterações na matriz G devido a seleção e deriva podem alterar essa situação e criar variação maior em alguma outra direção do morfoespaço, Já as outras duas direções de movimento do pico se dão em eixos diferentes, com o $\searrow$ sendo o mais antagônico ao eixo de menor resistência. Sob essas condições, foi verificado que todos os fatores promotores de estabilidade no caso

do ótimo fenotípico fixo continuam válidos. Além disso, com  $r_{\mu}$  e  $r_{\omega}$  de mesmo sinal, ou seja, com alinhamento da matriz mutacional com a matriz de seleção, o movimento do pico adaptativo na direção da linha de menor resistência evolutiva cria uma matriz G ainda mais estável. Em contrapartida, seleção direcional em outras direções tende a desestabilizar a matriz e, para algumas combinações de parâmetros, criar um efeito de maladaptação permanente, onde a média da população não coincide com o ótimo da superfície de seleção. Quanto ao efeito Turelli, os resultados sugerem que mesmo em populações relativamente pequenas  $(N_e = 342)$  a magnitude do efeito é entorno de 5-7% na norma de  $\beta_T$ , diminuindo para até 1-2% em populações maiores ( $N_e = 2731$ ). O efeito Turelli então pode ser pequeno em situações naturais, mas sempre existem condições onde ele pode ser relevante, como quando existe correlação entre  $\beta_t$  e  $\mathbf{G}_t$  por muitas gerações seguidas. A mensagem então é que, apesar de existirem situações onde o efeito Turelli é importante, não é de se esperar que estas sejam a regra. Apesar disso, os autores alertam que na prática estimativas fidedignas de  $\overline{\mathbf{G}}$  são bastante difíceis de serem obtidas, tanto pelas flutuações entre gerações quanto pela dificuldade em se obter tamanhos amostrais adequados para estimativas de matriz G (Marroig et al., 2012).

O mais recente artigo nessa sequência (Jones et al., 2007) trata da evolução da própria matriz mutacional e das consequências das mudanças nessa matriz para o confronto entre as visões de Lande (gaussiana) e de Turelli (alelos raros) na questão das distribuições dos efeito alélicos atuando em caracteres quantitativos. Para tal foi definido um terceiro caráter, que controla o valor de  $r_{\mu}$  e está sujeito a mudanças via mutação, de forma análoga aos outros caracteres (veja seção 2.2.1 para mais detalhes). Isso permitiu aos autores estudar como o padrão de pleiotropia e epistasia, representado na matriz mutacional, pode variar e como isso afeta o fenótipo e o genótipo de uma população sujeita a seleção estabilizadora gaussiana, mutação e deriva. Essa é a abordagem mais interessante para nossos objetivos neste trabalho, que visa entender em quais condições sistemas modulares podem evoluir. Para isso uma estrutura pleiotrópica plástica é imprescindível (Wagner, 1996; Pavlicev e Hansen, 2011). Já vimos que matrizes mutacionais e seletivas alinhadas promovem estabilidade da matriz  $\mathbf{G}$ . A pergunta que mais nos interessa no artigo de Jones et al. (2007) é a que se refere ao tipo de seleção que atua sobre o valor de  $r_{\mu}$ : seria a seleção estabilizadora gaussiana, descrita por um valor de correlação  $r_{\omega}$ , capaz de moldar o valor de  $r_{\mu}$ ? Ou seja, será que ocorre o alinhamento da matriz mutacional (e portanto, da matriz G) com a matriz da superfície de seleção? A seleção atua somente sobre caracteres fenotípicos, logo sua influência sobre a matriz mutacional é indireta. Será que essa força de seleção indireta é suficiente para superar flutuações aleatórias na matriz mutacional e proporcionar seu alinhamento? Apesar de fraca, a seleção indireta sobre a matriz mutacional se mostrou capaz de promover alinhamento com a superfície adaptativa, e portanto promover uma estabilidade maior da matriz G (Jones et al., 2007). Esse resultado é interessante por misturar a atuação de duas forças seletivas importantes: seleção estabilizadora clássica, ambiental, externa ao organismo; e restrições internas aos organismos, representadas pelos seu padrão de pleiotropia e codificados aqui na matriz mutacional. Dois tipos de restrições diferentes se influenciando mutuamente e afetando o padrão de covariação e expressão fenotípica dos sistemas biológicos. Além desse resultado, os autores mostram que nem o modelo gaussiano nem o de alelos raros são capazes de explicar totalmente os resultados, levando a uma visão intermediaria para a real distribuição dos efeitos

### 1.4 Consequências Evolutivas

alélicos nas populações.

A interação entre restrições externas e internas pode ser fundamental para entendermos padrões naturais de covariação. É devido a essa combinação das restrições internas, derivadas do desenvolvimento, e de restrições funcionais, seletivas, que o padrão de integração e modularidade se estabelece e se manisfesta na estrutura de covariação das populações. Vamos abordar agora evidências empíricas ao problema da estabilidade da matriz  $\mathbf{G}$  e suas consequências evolutivas, tema de trabalho há mais de 10 anos do Laboratório de Evolução de Mamíferos.

Em Marroig e Cheverud (2001), os autores utilizam o clado Platyrrhini, de primatas neotropicais como modelo, obtendo estimativas e comparando de forma filogeneticamente estruturada os 16 gêneros, quanto às suas matrizes de covariação fenotípica para 39 caracteres cranianos. Essa comparação revelou similaridades altas entre todos os gêneros. Além disso, os níveis de similaridade presentes eram pouco correlacionadas com a filogenia e sim com fatores ambientais, como a dieta de cada espécie. Além disso, foram testadas hipóteses de modularidade baseadas em desenvolvimento e função. A existência desses módulos específicos nos grupos aponta para seleção estabilizadora advinda de restrições internas como uma força importante na manutenção dos padrões de covariação. Isso reforça a ideia de alinhamento das matrizes de covariação com o ambiente seletivo ao qual cada população é sujeita. Além disso, a similaridade alta dos padrões, mesmo num clado com 30 milhões de anos de divergência e com variação morfológica notável, aponta para uma estabilidade marcante da matriz G. Além da comparação de padrões, em Marroig e Cheverud (2005, 2010) os autores abordam a influência desses padrões genéticos estáveis na evolução dos caracteres fenotípicos dentro dos primatas neotropicais. Segundo eles, a influência de restrições genéticas é clara e deve ser levada em conta nas tentativas de entender os processos evolutivos que formaram os grupos atuais.

Expandindo o padrão de estabilidade da matriz  $\mathbf{G}$ , em Porto et al. (2009) e Marroig et al. (2009) comparações entre matrizes obtidas para todas as ordens de mamíferos confirmam novamente o padrão de estabilidade da matriz  $\mathbf{G}$ , além de evidenciar a importância da modularidade para compreensão das respostas evolutivas possíveis e observadas.

A relação entre a organização modular e a evolução pode ser descrita por algumas

estatísticas associadas ao padrão de covariação das populações (Hansen e Houle, 2008). Uma dessas estatísticas é a flexibilidade evolutiva, definida como a correlação média entre a resposta observada numa população e o gradiente de seleção que gerou essa resposta (Marroig et al., 2009). Quanto mais alta a flexibilidade, mais capaz a população é de responder na direção morfológica privilegiada pela seleção natural. A evolvabilidade, complementar à flexibilidade, mede o quão grande foi a mudança na direção do gradiente de seleção, ou, mais claramente, mede quanta variação havia disponível na direção privilegiada pela seleção. Essas estatísticas permitem descrever a capacidade de mudanças evolutivas nas populações, e podem ser usadas de forma preditiva (veja, por exemplo, Marroig e Cheverud, 2010). Modularidade é frequentemente associada à evolvabilidade de uma população, pois uma estrutura modular permite que partes do organismo sejam modificadas pela ação da seleção natural sem simultaneamente causar alterações deletérias a alguma outra parte não envolvida no processo de seleção em questão. Apesar disso, um balanço entre modularidade e integração deve se estabelecer no organismo para que este forme um todo coeso. Uma população de indivíduos constituídos unicamente de caracteres completamente independentes poderia em teoria ter evolvabilidade e flexibilidade máximas, mas a ausência de interrelação entre suas partes torna tais organismos impossíveis. Nesta população hipotética, respostas evolutivas em um caráter não implicariam em mudanças correlacionadas nos caracteres ao qual o caráter sobre seleção está associado funcionalmente ou estruturalmente. Como o organismo não se modifica de forma harmoniosa, praticamente qualquer alteração seria

#### 1.5 Evolução da Modularidade

deletéria por romper com relações funcionais e estruturais.

Além de interferir em como os organismos respondem à seleção natural, a modularidade também é fruto da seleção (Wagner, 1996; Wagner et al., 2007). Como o padrão modular
é praticamente ubíquo em sistemas biológicos, o problema de como ele surgiu é difícil de ser abordado de forma experimental. Algumas evidências de condições para emergência da modularidade são encontradas em experimentos com sistemas muito simples, como cadeias de RNA (Ancel e Fontana, 2000) ou redes metabólicas simplificadas (Espinosa-Soto e Wagner, 2010). Em ambos os casos, a seleção é fundamental para a emergência de modularidade no nível estudado.

No caso de sistemas quantitativos, Pavlicev et al. (2010) mostraram que em um organismo contando com variação epistática no padrão de correlação de dois caracteres, a seleção direcional é capaz de promover a fixação do alelo responsável pela manutenção de alta correlação entre os caracteres sobre seleção direcional simultânea (seleção do tipo  $\nearrow$  em Jones et al. (2004)). Esse artigo, porém, faz uso direto da equação de seleção de Lande (1979) e não simula explicitamente os alelos responsáveis pelo fenótipo sobre seleção direcional, se restringindo aos alelos responsáveis pela correlação entre os caracteres na matriz **G**. A existência desses alelos responsáveis pela covariação entre caracteres foi comprovada experimentalmente, e estes foram chamados de loci quantitativos de relação, ou rQTLs (Pavlicev et al., 2008). De qualquer forma, essa é mais uma evidência do poder da seleção direcional de criar associação entre caracteres e portanto promover a modularidade em condições naturais.

\* \* \*

Nosso objetivo neste trabalho foi estabelecer uma técnica de modelagem que permita simular a evolução de caracteres quantitativos controlados por efeitos aditivos expressos em populações sujeitas a seleção natural. Com isso esperamos adquirir um melhor entendimento dos padrões de diversificação e integração morfológica observados na natureza, além de testar a viabilidade de possíveis cenários evolutivos que moldaram essas populações e seus padrões de covariação. Não pretendemos aqui criar um formalismo totalmente quantitativo e preditivo para a simulação de populações naturais, e sim capturar alguma parte dos complexos mecanismos que atuam na evolução de populações e de suas características fenotípicas. Com isso em mente, alguns dos parâmetros usados nas simulações são de difícil acesso experimental, apesar de terem interpretações claras. Na ausência de estimativas reais para alguns dos parâmetros utilizados, nós optamos por testar vários valores possíveis desses parâmetros, de forma a evidenciar (1) se esses parâmetros são relevantes ou não; e (2) em que aspecto eles são relevantes na evolução dos padrões de covariação das populações. Além disso, nossas simulações podem limitar o espectro de variação plausível para quantidades desconhecidas e ajudar nas suas eventuais estimativas. No capítulo 2 vamos apresentar dois modelos possíveis para a simulação de caracteres quantitativos. No capítulo 3 vamos estudar o comportamento de populações simuladas pelos dois modelos e avaliar sua capacidade de reproduzir padrões modulares encontrados na natureza sobre diferentes condições de seleção e parâmetros dos modelos. No capítulo 4 vamos utilizar uma determinada combinação de parâmetros para estudar possíveis cenários de evolução de padrões modulares.

# Capítulo 2

# Modelagem

## 2.1 Introdução

O primeiro passo do projeto foi escolher o modelo a ser implementado. Foi necessário encontrar uma técnica computacional que permitisse simular explicitamente cada individuo de uma população com alelos de efeitos contínuos em caracteres morfológicos. Esses caracteres seriam então apresentados à seleção natural. As populações são diploides, hermafroditas e se reproduzem sexuadamente, com gerações discretas e tamanho populacional fixo. Existem algumas possibilidades de como implementar as mutações e a transição de efeitos genéticos para caracteres fenotípicos, que serão apresentadas a seguir.

### 2.2 Modelos

### 2.2.1 Modelo com Matriz M

O modelo usado em Jones et al. (2003, 2004, 2007) se baseia em um conjunto de loci pleiotrópicos. Nesse sistema, todos os loci afetam todos os caracteres. Num sistema com p caracteres, podemos pensar em cada um dos m loci diploides abrigando dois alelos,

representados por vetores em  $\mathbb{R}^p$ . O conjunto de todos os alelos  $\mathbf{x}^i$  em um indivíduo seria escrito como:

$$\mathbf{x}^{i} = (\mathbf{x}_{1}^{i}, \mathbf{x}_{2}^{i}, \dots, \mathbf{x}_{p}^{i}), i \in \{1, \dots, 2m\}, \mathbf{x}^{i} \in \mathbb{R}^{p}$$

$$(2.1)$$

Onde cada um dos  $\mathbf{x}_{j}^{i}$  com  $j \in \{1, ..., p\}$  é um número real representando o efeito aditivo do alelo *i* no caráter *j*. Para gerar a próxima geração, sorteamos um  $\mathbf{x}^{i}$  a cada dois, sequencialmente, representando a formação de gametas nos indivíduos. Ou seja, para o locus *k*, do par ( $\mathbf{x}^{k}, \mathbf{x}^{k+1}$ ), apenas um deles será passado para a próxima geração.

Cada um dos p caracteres é formado pela soma dos efeitos aditivos de cada um dos 2m alelos agindo sobre esse caráter, além de um efeito ambiental e, com distribuição normal e média zero  $(N(0,V_e))$ . Um caráter  $\mathbf{z}_j$  é dado por:

$$\mathbf{z}_j = \sum_{i=1}^{2m} \mathbf{x}_j^i + e, \mathbf{z}_j \in \mathbb{R}, e \sim N(0, V_e)$$
(2.2)

A seleção nesse modelo é gaussiana. Logo, a aptidão do indivíduo com fenótipo  ${\bf z}$  é dada por:

$$W(\mathbf{z}) = exp\left[-\frac{1}{2}((\mathbf{z}-\theta)^T \boldsymbol{\omega}^{-1}(\mathbf{z}-\theta))\right]$$
(2.3)

Onde  $\boldsymbol{\omega}$  representa a matriz de covariação da superfície de seleção gaussiana e  $\theta$  o pico adaptativo, ou seja, o fenótipo com maior aptidão. Alterando  $\theta$  podemos impor regimes de seleção direcional. A cada geração a função  $W(\mathbf{z})$  representa a probabilidade de reprodução de um indivíduo com fenótipo  $\mathbf{z}$ . Quanto maior  $W(\mathbf{z})$  em relação aos outros indivíduos, mais descendentes o indivíduo com fenótipo  $\mathbf{z}$  terá na próxima geração. A seleção estabilizadora é controlada pela matriz  $\boldsymbol{\omega}$ , e pode ser de dois tipos: correlacionada e não correlacionada. A seleção estabilizadora mais simples, não correlacionada, é codificada em uma matriz  $\boldsymbol{\omega}$  diagonal e não influencia a correlação entre

os caracteres, apenas sua média e variância. Já a seleção estabilizadora correlacionada é dada por uma matriz  $\boldsymbol{\omega}$  com valores não nulos fora da diagonal e afeta a correlação entre os caracteres.

As mutações nesse modelo são gaussianas e afetam um alelo por vez. A frequência com que acontecem mutações é dada por uma probabilidade de mutação por locus por geração, definida *a priori*. O valor de um alelo  $\mathbf{x}^i$  é alterado por uma variação  $\delta \mathbf{x}$ . Após a mutação,  $\mathbf{x}^i$  assume um certo valor  $\mathbf{x}'^i$ :

$$\mathbf{x}^{\prime i} = \mathbf{x}^{i} + \delta \mathbf{x} = (\mathbf{x}_{1}^{i} + \delta \mathbf{x}_{1}, \mathbf{x}_{2}^{i} + \delta \mathbf{x}_{2}, \dots, \mathbf{x}_{p}^{i} + \delta \mathbf{x}_{p})$$
(2.4)

O valor de  $\delta \mathbf{x}$  é tomado de uma distribuição gaussiana com média zero e matriz de covariância  $\mathbf{M}$  (N(0,  $\mathbf{M}$ )), chamada matriz mutacional:

$$\delta \mathbf{x} \sim N(0, \mathbf{M}) = exp\left[-\frac{1}{2}(\delta \mathbf{x}^T \mathbf{M}^{-1} \delta \mathbf{x})\right], \delta \mathbf{x} \in \mathbb{R}^p$$
(2.5)

A matriz **M** resume todo o conjunto de interações pleiotrópicas e epistáticas em cada individuo, definindo como os efeitos de cada alelo se correlacionam no que diz respeito aos diferentes caracteres fenotípicos. A diagonal da matriz **M**, ou seja, as variâncias mutacionais, que definem a escala dos efeitos mutacionais, é definida *a priori*. Já os valores fora da diagonal, as covariâncias mutacionais, podem ser definidas a priori (Jones et al., 2003, 2004) ou podem ser definidas por um segundo conjunto de loci com seus próprios efeitos aditivos (Jones et al., 2007). Uma matriz mutacional variável permite explorar como mudanças nos efeitos pleiotrópicos se comportam em diferentes condições.

A matriz mutacional é controlada por um conjunto de loci mutacionais diploides, com alelos  $\mu_k^{ij}$  (com  $k \in \{1, \dots, 2m_\mu\}$ ), que codificam valores de correlações mutacionais  $r_{\mu}^{ij}$ , que por sua vez são usados para gerar os valores das covariâncias entre os caracteres  $i \in j$  na matriz **M**:

$$\mathbf{M}_{ij} = r^{ij}_{\mu} \sqrt{\mathbf{M}_{ii} \mathbf{M}_{jj}} \tag{2.6}$$

Com

$$r_{\mu}^{ij} = \Phi\left(\sum_{k=1}^{2m_{\mu}} \mu_k^{ij}\right) \tag{2.7}$$

Como todos os  $r_{\mu}^{ij}$  são correlações, o valor deles deve ir de -1 a 1. A função  $\Phi$  faz esse papel, levando a soma de valores aditivos dos  $\mu_k^{ij}$  de valores reais a valores entre -1 e 1. A escolha da função  $\Phi$  é arbitrária. Um exemplo de função  $\Phi$  (usado em Jones et al., 2007) é:

$$\Phi(x) = \frac{2e^{2x}}{1+e^{2x}} - 1 \tag{2.8}$$

Essa função garante que os valores de  $r^{ij}_{\mu}$  permaneçam em intervalos possíveis para valores de correlação.

Quando temos apenas dois caracteres, a correção dada pela função  $\Phi$  é suficiente para garantir que a matriz **M** seja bem definida. Porém, quando vamos tratar de sistemas com mais caracteres precisamos tomar cuidados adicionais. Isso se deve ao fato da matriz **M** ser uma matriz de covariância, e portanto necessariamente positiva definida (Anderson, 1984). Existem várias condições equivalentes para garantir que uma matriz seja positiva definida, por exemplo ter todos os autovalores estritamente positivos ou a existência da sua descomposição de Cholesky<sup>1</sup>. Na extensão para mais de dois caracteres, nós optamos por verificar se a matriz **M** é positiva definida aplicando a decomposição de Cholesky: caso a decomposição exista, ela pode ser usada para gerar os valores de  $\delta \mathbf{x}$  de forma eficiente; caso a decomposição não exista, um procedimento de extensão (Marroig et al., 2012) é realizado, substituindo todos os autovalores não

<sup>&</sup>lt;sup>1</sup>A decomposição de Cholesky de uma matriz positiva definida simétrica  $\mathbf{C}$  é dada por  $\mathbf{C} = \mathbf{LL}^*$ , onde  $\mathbf{L}$  é uma matriz triangular inferior e o asterisco representa a transposição complexa conjugada.

positivos pelo último positivo. Esse procedimento garante uma matriz ainda próxima à matriz  $\mathbf{M}$  original, mas positiva definida. A escolha desse procedimento é relativamente arbitrária e pode ser modificada em simulações futuras.

### 2.2.2 Modelo com Matriz B

Além de testar o uso do modelo utilizando a matriz **M** para mais de dois caracteres, nós optamos por criar também um modelo próprio, inspirado pela descrição feita em Wagner (1984). O modelo descrido por Wagner também faz uso do valor dos efeitos alélicos. Novamente temos p caracteres sendo controlados por m loci diploides. Cada indivíduo possui um conjunto de alelos **y** da forma:

$$\mathbf{y} = (y_1, y_2, \dots, y_{2m}) \tag{2.9}$$

Cada posição  $y_i$  representa algum produto gênico do alelo no locus i, com efeito fisiológico relevante para as características fenotípicas em questão. Para a formação da próxima geração, novamente sorteamos um alelo de cada dois, sequencialmente, formando gametas haploides.

A transição de um conjunto de efeitos fisiológicos dos 2m alelos **y** para o efeito genético  $\mathbf{x} \in \mathbb{R}^p$  no fenótipo do indivíduo é feita via uma função ontogenética f

$$\mathbf{x} = f(\mathbf{y}), f : \mathbb{R}^{2m} \to \mathbb{R}^p \tag{2.10}$$

A função ontogenética resume todas as interações pleiotrópicas e epistáticas entre os produtos gênicos, e pode ser altamente não linear e complexa. Supondo que f seja diferenciável, podemos aproximá-la por uma série de Taylor:

$$\mathbf{x} = f(\mathbf{y}) = f(0) + D[f(0)]\mathbf{y} + \frac{1}{2!}\mathbf{y}^T D^2[f(0)]\mathbf{y} + \dots$$
(2.11)

26

Tomando f(0) = 0 e mantendo apenas o termo linear, podemos considerar a aproximação linear da função f como:

$$\mathbf{x} = f(\mathbf{y}) = D[f(0)]\mathbf{y} = \mathbf{B}\mathbf{y}$$
(2.12)

Onde **B** é definida como uma "matriz ontogenética", que descreve de forma simplificada como os efeitos genéticos são convertidos em valores aditivos, resumindo, de forma linear, todos os efeitos pleiotrópicos e epistáticos entre os alelos. Essa aproximação é equivalente a supor que toda a variância genética é aditiva (Wagner, 1989). Como no modelo anterior, acrescentamos também um valor de ruido ambiental **e**, com distribuição multivariada normal e média zero. O indivíduo com alelos **y** e matriz ontogenética **B** tem o seu fenótipo **z** dado por:

$$\mathbf{z} = \mathbf{x} + \mathbf{e} = \mathbf{B}\mathbf{y} + \mathbf{e} \tag{2.13}$$

A princípio as entradas da matriz **B** poderiam tomar qualquer valor real, porém, como uma simplificação adicional, nós optamos por simular a matriz **B** como sendo apenas uma matriz de influência, ou seja, no nosso modelo ela codifica se um determinado loci tem seu valor genético adicionado ao valor aditivo de um caráter ou não. Dessa forma, as entradas da matriz **B** no nosso modelo podem tomar os valores 0 ou 1. Caso  $\mathbf{B}_{ij} = 1$  o caráter  $\mathbf{z}_j$  tem seu valor aditivo incrementado por  $\mathbf{y}_i$  (que pode ser positivo ou negativo), caso contrário ( $\mathbf{B}_{ij} = 0$ ) não.

A razão entre a dimensão de loci m e o número de caracteres p é relevante , pois ela define a grosso modo a quantidade de recombinação presente no modelo (Wagner, 1984). Quando maior for a razão m/p, mais recombinação teremos, pois cada um dos 2m alelos é herdado de forma independente. Quando mais loci tivermos no modelo, mais unidades recombinantes independentes, afetando potencialmente todos os caracteres, estarão segregando na população. Nesse sentido, os loci podem também ser interpretados como conjuntos de muitos genes, formando cromossomos, que afetam vários caracteres e são herdados independentemente.

Mutações são consideravelmente mais simples que no caso da matriz  $\mathbf{M}$ , porém são divididas em duas classes: as genéticas e as ontogenéticas. As genéticas afetam o valor de cada  $\mathbf{y}_i$  individualmente, e possuem uma taxa de mutação  $\mu$  por geração por locus definida *a priori*. Uma mutação no locus *i* altera o valor de  $\mathbf{y}_i$  por um valor real tomado de uma distribuição normal com média zero e variância pré-definida. Já as mutações ontogenéticas alteram a matriz  $\mathbf{B}$ , afetando cada casela da matriz com probabilidade  $\mu_B$  por geração por casela. Uma mutação na posição *ij* altera o valor de  $\mathbf{B}_{ij}$  de 0 para 1 ou vice-versa. Como esses dois tipos de mutação impõem dinâmicas temporais em escalas diferentes para os dois níveis de complexidade, nós testamos razões variadas para  $\mu \in \mu_B$ .

A seleção nesse modelo toma a mesma forma que no caso anterior, expressa na equação 2.3. Novamente a seleção estabilizadora pode ser correlacionada ou não correlacionada, dependendo da matriz  $\boldsymbol{\omega}$  e a seleção direcional é aplicada utilizando um ótimo variável na superfície gaussiana.

### 2.3 Parâmetros

Alguns parâmetros foram mantidos constantes em todas as simulações. O valor da variância mutacional de todos os caracteres foi fixada em 0.5 no modelo com matriz  $\mathbf{M}$  e 0.2 no modelo com matriz  $\mathbf{B}$ . No caso do modelo com matriz  $\mathbf{M}$ , o valor da variância das mutações nos alelos mutacionais foi de 0.05. A variância dos efeitos ambientais foi fixada em 0.4 nas simulações com matriz M e 0.8 nas simulações com matriz  $\mathbf{B}$ . O valor da variância ambiental define a escala dos caracteres e é importante na determinação de herdabilidade dos caracteres. Todos esses parâmetros foram escolhidos de forma a manter a herdabilidade dos caracteres por volta de 0.6 (no regime de seleção estabilizadora), um valor biologicamente plausível e capaz de responder a seleção direcional de forma eficiente (Cheverud, 1988). Escolhas iniciais foram baseadas nos valores utilizados em Jones et al. (2003, 2004, 2007).

A taxa de mutação nos alelos aditivos ( $\mu$ ) foi mantida em 0.0005 em todas simulações. Estimativas da taxa de mutação aditiva são relativamente simples, porém dependem da força de seleção estabilizadora para cada população e caráter (Lynch e Walsh., 1998). Nosso valor para  $\mu$  está dentro dos valores esperados para aumento de variância fenotípica causado por mutação (Bürger e Lande, 1994). Por outro lado, o valor de  $\mu_B$  é mais difícil de ser estimado empiricamente. A matriz **B** representa basicamente a influência de uma região genômica na porção aditiva de um caráter fenotípico. Alterações na relações entre genes e fenótipo podem ser reguladas de diversas formas, como alterações nas relações epistáticas entre genes, alterações nas pleiotropia ou mesmo ganho e perda de genes influenciando características fenotípicas. Todas essas interações são resumidas na matriz  $\mathbf{B}$ , e suas alterações representadas pela taxa de mutação  $\mu_B$ . O valor numérico de  $\mu_B$  está relacionado, então, com a taxa de ganho e perda de regiões genômicas que afetam os caracteres fenotípicos em questão, os chamados loci de caráteres quantitativos (QTL). Para estimar  $\mu_B$ , seria necessário acompanhar uma população durante várias gerações, possivelmente com a aplicação de seleção direcional para gerar mudanças observáveis em escalas de tempo curtas, medir caracteres fenotípicos na população e sequenciar todos os indivíduos no início e no fim do processo, afim de detectar ganho e perda de QTL afetando os caracteres de interesse. Aumento da eficiência de sequenciamento e técnicas de inferência de QTL avançadas podem permitir esse tipo estimativa.

Nas simulações nós alteramos o tamanho populacional (Ne), o número de caracteres (p), o número de alelos (m), a taxa de mutação nas caselas da matriz **B**  $(\mu_B)$ , e o tipo de seleção (estabilizadora e direcional). Alteramos também a intensidade da seleção direcional. Para isso, modificamos o quanto o pico adaptativo era alterado para cada

caráter a cada geração, de 0.01 até 0.2. Esse deslocamento por caráter é chamado de  $\Delta_S$  e equivale às entradas do vetor de diferencial de seleção da equação 1.5. A variância fenotípica das populações em equilíbrio mutação-seleção direcional era por volta de 1.6. Então, nossa seleção mais intensa implicava um movimento de pico de cerca de 0.125 desvios padrões por geração por caráter ( $\Delta_S = 0.2$ ), resultando num movimento total de cerca de 125 desvios padrões por caráter após 10.000 gerações de seleção direcional; enquanto a seleção mais fraca é dada por movimento de 0.006 desvios padrões por geração por caráter ( $\Delta_S = 0.01$ ), resultando num movimento total de cerca

Na linguagem da equação de resposta à seleção (equação 1.6), os valores de  $\Delta_S$ equivalem às entradas do vetor do diferencial de seleção **S** e podem ser usados para calcular o valor do vetor do gradiente de seleção. Kingsolver et al. (2001), revisando a literatura de experimentos de seleção, calculou uma distribuição para os valores de intensidades de seleção na natureza, utilizando a norma das entradas do gradiente de seleção ( $|\beta_i|$ ) como métrica, ou seja, os valores por caráter do vetor do gradiente de seleção direcional. Kingsolver et al. (2001) coloca os valores de  $|\beta_i|$  como variando entre zero e um, com uma mediana de 0.08 para caracteres de estratégias bionômicas e mediana de 0.17 para caracteres morfológicos. O valor de  $|\beta_i|$  típico varia ao longo das simulações, pois ele depende da matriz de variância fenotípica **P**.

### 2.4 Estatísticas de Modularidade e Integração

Quantificar a evolução das relações entre os caracteres quantificadas na matriz de covariância é fundamental para a interpretação dos resultados das simulações. Podemos simplesmente inspecionar visualmente as matrizes, mas medidas quantitativas são importantes quando o volume de dados se torna suficientemente grande. Para isso, vamos explorar algumas estatísticas descritivas já utilizadas em estudos de biologia tradicionais e apresentar técnicas novas para a quantificação de modularidade baseadas nas matrizes de correlação.

Quando conhecemos de antemão a estrutura modular apresentada na matriz de correlação, podemos quantificar o nível de modularidade medindo apenas a médias das correlações dentro de e entre módulos, quantidade conhecida como razão das médias, ou AVG-Ratio do inglês *AVeraGe Ratio*. Certamente essa medida é informativa em quantificar o quão evidentes são os módulos variacionais expressos na população por quantificar a diferença no nível de correlação dentro de e entre módulos. Podemos também observar a evolução das médias das correlações dentro de módulos (AVG+) e entre módulos (AVG-).

Contudo, essas medidas exigem conhecimento prévio da estrutura de modularidade presente na população. Esse certamente não é o caso quando amostrando uma população natural, embora hipóteses a priori possam ser utilizadas quando temos conhecimento sobre a função ou desenvolvimento dos caracteres em um grupo de organismos (Cheverud e Leamy, 1985; Marroig e Cheverud, 2001). Para sanar esse problema, propomos também uma estatística de modularidade inspirada em técnicas de detecção de comunidades em redes (Newman, 2006b,a; Reichardt e Bornholdt, 2006). O princípio dessa técnica é alocar cada caráter em um módulo de modo que algum parâmetro de modularidade seja máximo. Para isso, definimos a modularidade L de uma matriz de correlação **A** como sendo:

$$L = \sum_{i \neq j} \left[ \mathbf{A}_{ij} - \frac{k_i k_j}{2m} \right] \delta(g_i, g_j)$$
(2.14)

Os termos  $g_i \in g_j$  representam a partição dos caracteres, ou seja, em qual grupo os caracteres  $i \in j$  se encontram. A função  $\delta(\cdot, \cdot)$  é a função delta de Kronecker, ou seja:

$$\delta(x,y) = \begin{cases} 1 & \text{se } x = y \\ 0 & \text{se } x \neq y \end{cases}$$
(2.15)

Portanto apenas caracteres no mesmo módulo contribuem para o valor de L. Já os  $k_i$  e  $k_j$  representam o valor total de correlação atribuído a cada caráter, ou, mais claramente, a soma da correlação de cada caráter com todos os outros:

$$k_i = \sum_{j \neq i} A_{ij} \tag{2.16}$$

E *m* representa a soma de todos os k ( $m = \sum_i k_i$ ). O termo  $\frac{k_i k_j}{2m}$  funciona como um tipo de expectativa nula para a distribuição das correlações entre os caracteres *i* e *j* dado que cada em deles possui, na matriz **A**, um valor de correlação total dado por  $k_i$ e  $k_j$ . Essa escolha para a expectativa nula aparece naturalmente quando impomos que ela deva ser simétrica entre os caracteres e que dependa unicamente de  $k_i$  e  $k_j$  (Para detalhes veja Newman, 2006a). Portanto, caracteres no mesmo módulo que possuam correlação maior que a esperada ao acaso contribuem para aumentar a modularidade L, enquanto caracteres no mesmo módulo com correlação abaixo da esperada ao acaso contribuem para diminuir L.

Definida a função L de modularidade, temos que encontrar a partição dos caracteres (os valores de  $g_i$ ) que maximizem a função L. Isso é feito com a técnica de Monte Carlo utilizando um algoritmo do tipo Metropolis-Hasting (Metropolis et al., 1953). A partição que maximize L é tomada como a estrutura de modularidade do sistema e o valor máximo de L como um índice de modularidade, comparável ao AVG-Ratio. Neste trabalho, pela estrutura de modularidade ser conhecida, o único propósito de usar essa técnica de detecção de módulos é *quantificar* a modularidade das matrizes sem utilizar o conhecimento prévio de sua estrutura. Em situações naturais, onde não conhecemos a estrutura de modularidade, essa técnica poderia ser usada para inferir a estrutura de modularidade utilizando apenas a matriz de correlação.

## 2.5 Resumo de Parâmetros

Para facilitar a leitura do texto, apresentamos aqui um resumo dos parâmetros relevantes aos modelos.

Simbolo	Interpretação	Modelo
p	número de caracteres	$\mathbf{M} \in \mathbf{B}$
m	número de alelos aditivos	$\mathbf{M} \in \mathbf{B}$
Ne	tamanho populacional	$\mathbf{M} \in \mathbf{B}$
$\omega$	matriz de seleção estabilizadora	$\mathbf{M} \in \mathbf{B}$
$\mu$	taxa de mutação por alelo aditivo por geração	$\mathbf{M} \in \mathbf{B}$
$\mu_B$	taxa de mutação por casela da matriz ${\bf B}$	В
$\Delta_S$	taxa de mudança do pico adaptativo	$\mathbf{M} \in \mathbf{B}$
	por geração de seleção direcional	
$ \beta_i $	valor do gradiente de seleção	$\mathbf{M} \in \mathbf{B}$
	por caráter por geração	

Tabela 2.1: Tabela com símbolos e abreviações usadas nas simulações

# Capítulo 3

# Teste e Validação dos Modelos

## 3.1 Introdução

Nosso objetivo nesse capitulo é avaliar o comportamento de cada um dos modelos com diferentes parâmetros, tanto internos quanto externos às populações simuladas. Alguns exemplos são número de caracteres, tamanho populacional, taxas de mutação, recombinação, força e tipo de seleção, etc. Vamos analisar os resultados sempre do ponto de vista da estrutura de covariação, seu caráter modular e sua estabilidade ao longo das gerações.

## **3.2** Resultados e Discussão · Matriz M

#### 3.2.1 Seleção Estabilizadora

Para implementar o modelo com matriz **M** inicialmente procuramos reproduzir os resultados obtidos em Jones et al. (2007), focando no alinhamento entre matriz **G** e matriz  $\boldsymbol{\omega}$  pela seleção indireta na matriz **M**. Esse efeito pode ser observado na figura 3.1, onde vemos a correlação entre dois caracteres aumentando em resposta à atuação da seleção estabilizadora correlacionada.



Figura 3.1: Evolução da correlação entre dois caracteres ligados por efeitos mutacionais pleiotrópicos sobre seleção estabilizadora correlacionada. Nesta simulação  $r_{\omega} = 0.8$ , Ne = 5.000, p = 2, m = 50

Como nosso objetivo é expandir a modelagem para mais caracteres, em seguida adicionamos um caráter ao sistema. Nesse caso, as considerações sobre a dificuldade de manter a matriz  $\mathbf{M}$  positiva definida já se aplicam, sendo necessário fazer uso da correção de autovalores. Mesmo assim, o resultado é bastante semelhante, ainda com a correlação entre os 3 caracteres se alinhando com a matriz  $\boldsymbol{\omega}$  (figuras 3.2 e 3.3).

Na figura 3.3 vemos representações gráficas da matriz de correlação fenotípica ao final da simulação e da matriz  $\boldsymbol{\omega}$  de seleção estabilizadora correlacionada. Caselas mais claras indicam correlação maior. Nesse caso, a matriz de correlação da superfície de seleção foi:

$$Corr(\boldsymbol{\omega}) = \begin{pmatrix} 1 & 0.9 & 0.1 \\ 0.9 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{pmatrix}$$
(3.1)

(3.2)



E a matriz de covariância da superfície de seleção:

Figura 3.2: Evolução da correlação entre três caracteres ligados por efeitos mutacionais pleiotrópicos sobre seleção estabilizadora correlacionada. A seleção propicia a integração entre dois caracteres (linha vermelha) e a desintegração desses dois com o terceiro (linhas azul e verde). Nesta simulação  $r_\omega$  como na equação 3.1, Ne = 5.000, p = 3, m = 50.

Quando procuramos ampliar o número de caracteres, porém, o alinhamento da matriz G com a matriz  $\boldsymbol{\omega}$  se mostrou problemático. Nesse caso a matriz de correlação da superfície de seleção continha dois módulos, e era dada por:

Na figura 3.4 vemos a trajetória típica das correlações fenotípicas com uma matriz  $\boldsymbol{\omega}$  com dois módulos (ver figura 3.6). Aqui não existe mais a distinção entre correlações



Figura 3.3: Comparação entre a matriz de correlação final para 3 caracteres após 10.000 gerações de seleção e a matriz  $\boldsymbol{\omega}$  da superfície de seleção.

dentro de módulos (altas) e entre módulos (baixas). Os resultados da evolução da modularidade L e do AVG-Ratio confirmam a ausência de modularidade no sistema. A modularidade L apresenta uma leve subida ao longo da simulação, mas o AVG-Ratio permanece estável (Figura 3.5). Isso se deve à modularidade L capturar qualquer tipo de modularidade na matriz, não só aquela associada a superfície de seleção. Na figura 3.6 fica claro que não existe semelhança entre as matrizes  $\mathbf{G} \in \boldsymbol{\omega}$ . Realizamos essas simulações com uma gama grande de intensidades de seleção e chegando até 100.000 gerações de seleção, sem nenhum indício de alinhamento estável entre as matrizes  $\mathbf{G} \in \boldsymbol{\omega}$ .

#### 3.2.2 Seleção Direcional

Em seguida, procuramos explorar como a inclusão de seleção direcional afetaria o alinhamento das matrizes  $\mathbf{G} \in \boldsymbol{\omega}$ . Utilizamos novamente uma matriz  $\boldsymbol{\omega}$  modular e acrescentamos seleção direcional correlacionada, de mudança conjunta na média dos caracteres



Figura 3.4: Evolução da correlação entre dez caracteres ligados por efeitos mutacionais pleiotrópicos sobre seleção estabilizadora correlacionada. Apesar da seleção promover a integração entre os 5 primeiros e os 5 últimos caracteres e a desintegração entre esse módulos, isso não se transmite à matriz de correlação. Nesta simulação  $r_{\omega}$  como na equação 3.3, Ne = 5.000, p = 10, m = 50.



Figura 3.5: Evolução da modularidade Le AVG-Ratio para as matrizes de correlação representadas na figura 3.4





Figura 3.6: Comparação entre a matriz de correlação final para 10 caracteres após 10.000 gerações de seleção representadas na figura 3.4 e a matriz  $\boldsymbol{\omega}$  da superfície de seleção.

dentro de cada um dos módulos, e mudanças antagônicas entre módulos. Ou seja, os 5 primeiros caracteres tiveram seu ótimo fenotípico aumentado a uma taxa constante e os 5 últimos caracteres tiveram seu ótimo reduzido de forma simétrica ( $\Delta_S = 0.2$ ). As populações respondem a essa seleção de forma eficiente, alterando suas médias de acordo com a posição do ótimo. Após 10.000 gerações de seleção direcional, observamos a matriz final e a evolução de todas as correlações genéticas nesse período. Os resultados podem ser vistos nas figuras 3.7, 3.8 e 3.9. Mesmo nessas condições, não observamos a emergência de módulos variacionais na matriz **G**, que indicaria semelhança com a matriz  $\boldsymbol{\omega}$  e o padrão de seleção direcional correlacionada.

Acreditamos que essa dificuldade em reproduzir os resultados encontrados em sistemas com 2 e 3 caracteres no sistema com 10 caracteres se deve ao aumento considerável do espaço possível de variação que a inclusão de cada novo caráter representa. O morfoespaço cresce rapidamente com a inclusão de cada novo caráter, e o espaço permitido



Figura 3.7: Evolução da correlação entre dez caracteres ligados por efeitos mutacionais pleiotrópicos sobre seleção estabilizadora correlacionada e seleção direcional intensa para mudança correlacionada dentro dos módulos. Novamente, isso não se traduz na matriz de correlação genética.



Figura 3.8: Evolução da modularidade Le AVG-Ratio para as matrizes de correlação representadas na figura 3.7



Figura 3.9: Comparação entre a matriz de correlação final para 10 caracteres após 10.000 gerações de seleção direcional e estabilizadora representadas na figura 3.7 com a matriz  $\boldsymbol{\omega}$  da superfície de seleção.

para a matriz **M** também. A seleção indireta, então, se torna muito fraca para promover o alinhamento entre as matrizes, provavelmente porque as mutações nos alelos mutacionais não exploram de forma adequada (ergódica) o espaço da matriz mutacional, levando a uma situação de não alinhamento mesmo com uma pressão seletiva considerável. Outro problema é a técnica de extensão utilizada para gerar matrizes positivas definidas. Apesar de gerar matrizes similares às criadas diretamente dos alelos mutacionais, qualquer modificação na matriz diminui a semelhança entre matrizes de uma geração para outra, e introduz ruido na herança. Em outras palavras, a semelhança da própria matriz **M** de um geração para a outra, definida através de alelos mutacionais aditivos, diminui com o aumento da dimensionalidade do sistema e impossibilita que seleção atue de forma eficiente na sua estrutura, tanto pelo aumento da dimensionalidade quanto pelas correções necessárias para tornar a matriz bem definida. Talvez outros esquemas de simulação para matrizes de mutação herdáveis, definidas para

Matriz ω

cada indivíduo separadamente, variáveis e sujeitas à seleção natural, venham a sanar essa limitação. O uso de correlações parciais para gerar as matrizes mutacionais, por exemplo, evita o problema da geração de matrizes não positivas definidas (Joe, 2006; Lewandowski et al., 2009), e pode ser uma saída interessante para o modelo com matriz **M** em muitas dimensões. Pretendemos explorar essa solução no futuro.

### 3.3 Resultados e Discussão · Matriz B

Todas as simulações dessa seção começam com o mesmo estado inicial, com a matriz **B** totalmente formada por 1, ou seja, integração total, e todos os valores aditivos 0, ou seja, todos os caracteres quantitativos determinados apenas por efeitos ambientais e com sua porção aditiva igual a zero. As primeiras gerações são dependentes desse estado inicial e não trazem informação. Após cerca de 2.000 gerações o sistema se encontra em equilíbrio seleção-mutação-deriva e pode ser analisado independentemente do estado inicial. Isso se reflete em todos os gráficos como um transiente instável, seguido de uma região estável.

#### 3.3.1 Seleção Estabilizadora

No segundo tipo de modelagem, começamos já com 10 caracteres sofrendo seleção estabilizadora correlacionada dada pela matriz apresentada na equação 3.3. Exploramos então a influência de várias razões entre a taxa de mutação nos efeitos genéticos ( $\mu$ ) e a taxa de mutação nas caselas da matriz **B** ( $\mu_B$ ). Essa razão  $\mu/\mu_B$  define como será a dinâmica temporal de mudança entre dois níveis: o dos efeitos genéticos e o da atribuição desses efeitos à porção aditiva dos caracteres quantitativos, que chamamos aqui de efeito *ontogenético*, incluindo também todos os efeitos pleiotrópicos e epistáticos dos loci sobre o fenótipo. Conforme vemos nas figuras 3.10 e 3.11, apesar de mudanças na magnitude das correlações com o aumento da razão  $\mu/\mu_B$ , não observamos o surgimento de módulos variacionais apenas com seleção estabilizadora correlacionada.

### 3.3.2 Seleção Direcional

Com a inclusão de seleção direcional divergente, ou seja, para aumento dos 5 primeiros caracteres e diminuição dos 5 últimos, como na figura 3.9, e  $\mu/\mu_B = 0.1$ , ainda não observamos a formação de módulos na matriz **G** (figura 3.12). Porém, como vemos na figura 3.13, o aumento da razão  $\mu/\mu_B$  aliada à seleção direcional é capaz de alterar esse comportamento. Esse aumento parece promover um sistema capaz de responder de forma eficiente às pressões seletivas impostas e gerar uma matriz **G** com a estrutura modular privilegiada pela seleção estabilizadora e direcional. Lembramos que na ausência de seleção direcional esse efeito não é observado, mesmo com as mesmas razões  $\mu/\mu_B$ . Na figura 3.14 vemos que razões baixas de  $\mu/\mu_B$  levam a sistemas bastante instáveis, apesar de modulares. Isso pode ser interpretado como uma consequência de taxas de mudança em cada nível da simulação incompatíveis com populações estáveis, pois as mudanças nos efeitos genéticos, ou o sistema não tem tempo de responder à seleção de forma estável.

Para  $\mu/\mu_B = 1$ , ou seja, com a taxa de mutação igual nos níveis genético e ontogenético, o valor de modularidade L flutua violentamente entre as gerações, sugerindo uma modularidade muito instável. Além disso, para valores de baixos de  $\mu/\mu_B$  as correlações entre os caracteres são extremamente baixas, em torno de 0 entre módulos e por volta de 0.2 dentro de módulos. A partir de  $\mu/\mu_B = 4$  o sistema se torna muito mais estável e progressivamente menos modular. Nós optamos por utilizar o valor  $\mu/\mu_B = 5$ nas simulações subsequentes. Essa opção se deu por uma série de fatores: nesse valor o sistema ainda apresenta um aumento de modularidade ao longo das gerações; as correlações aumentam para cerca de 0.2 entre módulos e 0.4 dentro de módulos ao final da simulação; os valores de modularidade L e AVG-Ratio são bastante concordantes,



Figura 3.10: Matriz final média de 10 corridas para diferentes razões de  $\mu$  e  $\mu_B$ , com seleção estabilizadora correlacionada com 2 módulos.



Figura 3.11: Evolução típica de AVG-Ratio e modularidade L para diferentes razões de  $\mu \in \mu_B$ , com seleção estabilizadora correlacionada com 2 módulos.

sugerindo que os módulos favorecidos pela seleção estão efetivamente sendo expressos na população; e, por último, o sistema possui uma estabilidade razoável sem se tornar incapaz de responder à seleção direcional.



Figura 3.12: Comparação entre a matriz de correlação final média de 10 corridas para o 10 caracteres após 10.000 gerações de seleção direcional e estabilizadora com a matriz  $\boldsymbol{\omega}$  da superfície de seleção e  $\mu/\mu_B = 0.1$ , Ne = 2.500,  $\Delta_S = 0.2$ , m/p = 2

Na figura 3.15 vemos um exemplo de corrida com  $N_e = 2.500$  e  $\mu/\mu_B = 5$ , mostrando claramente a separação de grupos de correlação dentro de e entre módulos. Posteriormente vamos explorar a influência do tamanho populacional na estabilidade das matrizes geradas pelas nossas simulações (figura 3.20).

#### 3.3.3 Intensidade de Seleção Direcional

Estamos interessados também na força de seleção direcional necessária para gerar a estrutura modular nas matrizes de covariação. Nas figuras 3.16 e 3.18 vemos matrizes médias finais para 10 corridas com  $N_e = 1.000$ ,  $\mu/\mu_B = 5$  e intensidade de seleção variável. Para isso, modificamos o quanto o pico adaptativo era alterado para cada



Figura 3.13: Matriz final média de 10 corridas para diferentes razões de  $\frac{\mu}{\mu_B}$ , com seleção estabilizadora correlacionada com 2 módulos e seleção direcional correlacionada favorecendo os mesmos módulos. ( $N_e = 2.500$ ,  $\Delta_S = 0.2$ , m/p = 2)



Figura 3.14: Evolução típica de AVG-Ratio e modularidade L para diferentes razões  $\mu/\mu_B$ , com seleção estabilizadora correlacionada com 2 módulos e seleção direcional correlacionada favorecendo os mesmos módulos. Atenção para a mudança de escala nos eixos verticais. ( $N_e = 2.500$ ,  $\Delta_S = 0.2$ , m/p = 2)



Figura 3.15: Evolução da correlação entre dez caracteres ligados por efeitos pleiotrópicos controlados por uma função ontogenética linear binária variável, sofrendo seleção estabilizadora correlacionada e seleção direcional intensa para mudança correlacionada dentro dos módulos durante as 10.000 gerações. ( $N_e = 2.500$ ,  $\mu/\mu_B = 5$ ,  $\Delta_S = 0.2$ , m/p = 2)

caráter a cada geração, de 0.01 até 0.2 por geração durante 10.000 gerações.

Podemos calcular os valores de  $|\beta_i|$  para essas simulações, obtendo uma estimativa de seleção direcional comparável com dados empíricos. No caso de seleção mais fraca, com  $\Delta_S = 0.01$ , o  $|\beta_i|$  médio para as primeiras gerações foi de 0.002 aumentando para 0.004 nas últimas, caracterizando seleção fraca. O  $|\beta_i|$  total média por caráter para seleção fraca foi de 2.9. Já para seleção forte, com  $\Delta_S = 0.2$ , o  $|\beta_i|$  médio começa em um valor de cerca de 0.02 e rapidamente, após cerca de 1500 gerações, atinge um valor estável de 0.04. Para uma geração, esse valor de  $|\beta_i|$  não seria considerado uma seleção forte nos critérios delimitados por Kingsolver et al. (2001). Porém, essa seleção é mantida ao longo de muitas gerações. Para caráteres morfológicos, o valor típico de  $|\beta_i|$  seria estabelecido após cerca de 4 gerações ( $|\beta_i| = 0.17$ ), e para caráteres de história de vida, cerca de 2 gerações ( $|\beta_i| = 0.08$ ). A tabela 3.1 resume os valores de  $|\beta_i|$  médio e total por caráter para todos os valores de seleção.

Nas figuras 3.17 e 3.19 vemos evoluções típicas de AVG-Ratio e modularidade L para corridas com várias intensidades de seleção, mostrando o quão estáveis e o quão rápido

$\Delta_S$	0.01	0.02	0.03	0.04	0.05
$ \beta_i $ médio	0.003	0.005	0.008	0.010	0.013
$ \beta_i $ total médio	2.969	5.326	7.974	9.847	12.196
$\Delta_S$	0.06	0.07	0.08	0.09	0.10
$ \beta_i $ médio	0.015	0.017	0.019	0.020	0.022
$ \beta_i $ total médio	13.856	16.160	17.381	18.881	20.502
$\Delta_S$	0.11	0.12	0.13	0.14	0.15
$ \beta_i $ médio	0.024	0.027	0.028	0.030	0.030
$ \beta_i $ total médio	21.774	24.571	25.207	27.209	27.561
$\Delta_S$	0.16	0.17	0.18	0.19	0.20
$ \beta_i $ médio	0.031	0.032	0.035	0.037	0.039
$ \beta_i $ total médio	28.709	29.473	32.301	33.914	35.528

Tabela 3.1: Intensidades de seleção, representadas pelo diferencial de seleção por caráter e pelo valor do gradiente de seleção por caráter  $|\beta_i|$ 

os padrões modulares se estabelecem nessas condições. Vemos que mesmo com um tamanho populacional considerável (Ne = 2.500) ainda observamos flutuações grandes.

### 3.3.4 Seleção direcional seguida de seleção estabilizadora

Outra questão relevante é o quão estáveis são os padrões de covariação modulares gerados pela seleção direcional e estabilizadora. Para responder a essa pergunta nós sujeitamos populações de tamanhos diferentes a uma mesma pressão seletiva intensa durante 5.000 gerações (período suficiente para o surgimento de módulos), seguidas de 5.000 gerações de seleção estabilizadora correlacionada. Apesar da diferença entre as correlações dentro e entre módulos diminuir, ainda observamos uma distinção estável entre elas (figura 3.20 e 3.21). O efeito do tamanho populacional é especialmente marcante nessas condições, com populações grandes se mostrando mais eficientes na manutenção prolongada dos padrões estabelecidos pela seleção direcional (figuras 3.20 e 3.21).

Ainda na questão da estabilidade dos padrões de covariação, o papel da seleção esta-



Figura 3.16: Matriz final média de 10 corridas com  $\mu/\mu_B = 5$ , Ne = 1.000, m/p = 2, sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .



Figura 3.17: AVG-Ratio e modularidade L para corridas com  $\mu/\mu_B = 5$ , Ne = 2.500, m/p = 2, sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .



Figura 3.18: Matriz final média de 10 corridas com  $\mu/\mu_B = 5$ , Ne = 1.000, m/p = 2, sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .



Figura 3.19: AVG-Ratio e modularidade L para corridas com  $\mu/\mu_B = 5$ , Ne = 2.500, m/p = 2, sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .



Figura 3.20: Evolução da correlação entre dez caracteres ligados por efeitos pleiotrópicos controlados por uma matriz **B** variável, sofrendo seleção estabilizadora correlacionada e seleção direcional para mudança correlacionada dentro dos módulos durante as 5.000 primeiras gerações, seguidos de 5.000 gerações de seleção estabilizadora correlacionada  $(\Delta_S = 0.2, \mu/\mu_B = 5, m/p = 2).$


Figura 3.21: Evolução de modularidade L e AVG-Ratio para as correlações apresentadas na figura 3.20 ( $\Delta_S = 0.2$  por 5.000 gerações,  $\mu/\mu_B = 5$ , m/p = 2).



Figura 3.22: Evolução da correlação entre dez caracteres ligados por efeitos pleiotrópicos controlados por uma matriz **B** variável, sofrendo seleção estabilizadora correlacionada e seleção direcional para mudança correlacionada dentro dos módulos durante as 5.000 primeiras gerações, seguidos de 5.000 gerações de seleção estabilizadora não correlacionada ( $\Delta_S = 0.2, \ \mu/\mu_B = 5, \ m/p = 2$ ).



Figura 3.23: Evolução de modularidade L e AVG-Ratio para as correlações apresentadas na figura 3.22 ( $\Delta_S = 0.2$  por 5.000 gerações,  $\mu/\mu_B = 5$ , m/p = 2).

bilizadora correlacionada pode ser avaliado sujeitando as populações à seleção direcional e estabilizadora correlacionada por 5.000 gerações, seguidas de 5.000 gerações de seleção estabilizadora não correlacionada, apenas mantendo a média de cada caráter individualmente. Nas figura 3.22 e 3.23 podemos ver a importância da seleção estabilizadora correlacionada na manutenção dos padrões modulares gerados pela seleção direcional. Comparando cada tamanho populacional mostrado nas figuras 3.20, 3.21 e 3.22, 3.23 podemos ver uma degradação muito mais acentuada dos padrões modulares na ausência da seleção estabilizadora correlacionada. Apesar da diminuição da modularidade, alguma diferenciação entre correlações dentro e entre módulos se mantém, especialmente nas populações maiores. Para  $N_e$  acima de 2.500, e mais marcadamente acima de 5.000, o valor de AVG-Ratio se mantém acima de 1.5 de maneira relativamente estável, porém o valor de modularidade L decai para 1. Para populações suficientemente grandes, apenas seleção estabilizadora não correlacionada pode ser suficiente para manter modularidade variacional detectável e com consequências evolutivas, mesmo na ausência de correlação na superfície de seleção.

A importância da seleção estabilizadora para explicar padrões de estase no registro fóssil foi amplamente discutida em Charlesworth et al. (1982). No contexto de padrões de covariação, seleção estabilizadora correlacionada também foi apontada como uma provável explicação para a estabilidade de padrões de covariação observados na natureza (Cheverud, 1984; Marroig e Cheverud, 2001; Porto et al., 2009). Nossos resultados dão credibilidade a essa hipótese. A origem dessa seleção estabilizadora pode ser atribuída a fatores internos dos organismos, como restrições ontogenéticas responsáveis pela coesão entre as partes do indivíduo. Como essas restrições estão sempre presentes, a seleção correlacionada está sempre presente e os padrões tendem a se manter estáveis.

#### 3.3.5 Intensidade de recombinação

Um parâmetro ainda não mencionado é o número de loci m envolvidos na determinação do valor fenotípico de cada um dos p caracteres. No nosso esquema de simulação, aumentar o valor da razão m/p é análogo a aumentar a força de recombinação, pois o valor de m define o número de unidades recombinantes independentes segregando na população. Podemos pensar em m como o número de cromossomos no genoma do indivíduo (na ausência de recombinação homologa), em vez de pensar nele como número de loci. Até agora a razão m/p foi mantida em 2, um valor suficiente para promover respostas complexas de todos os caracteres mas ainda computacionalmente viável. Razões maiores de m/p teoricamente aumentam o espaço de variação possível para população, pois mais genes estão disponíveis para atuar sobre cada caráter, separadamente ou de forma coordenada. A figura 3.24 mostra como o aumento na razão m/p afeta a evolução das correlações fenotípicas nas populações sujeitas à seleção direcional intensa e seleção estabilizadora correlacionada. Vemos na figura 3.24 que o aumento do número de alelos torna o sistema mais lento em adquirir um caráter modular. Isso pode ser associado com o aumento na recombinação, que age na direção oposta à das forças seletivas direcionais e estabilizadora correlacionada. Isso nos levou a crer que tempos mais longos de seleção poderiam gerar o mesmo resultado que o obtido com valores baixos de m/p. Na figura 3.25 vemos duas corridas longas, de 20.000 gerações, com m/p = 9 e 10, mostrando o tempo de estabilização mais longo e a modularidade mais discreta, com menor diferença nas correlações dentro de e entre módulos. Isso pode ser explicado pela força de recombinação aumentada.

Praticamente todos os exemplos de surgimento de modularidade mostrados neste capítulo podem ser interpretados como seleção natural privilegiando alelos e interações entre alelos que formem de fenótipos com aptidão maior. Associações modulares na estrutura genética e aditiva são privilegiadas por responderem melhor a pressão seletiva imposta. Porém, essas associações não são herdáveis, apenas os alelos individuais são



Figura 3.24: Evolução da correlação entre dez caracteres ligados por efeitos pleiotrópicos controlados por uma função ontogenética linear binária variável sofrendo seleção estabilizadora correlacionada e seleção direcional intensa para mudança correlacionada dentro dos módulos, para diferentes valores do número de loci. (Ne = 5.000,  $\mu/\mu_B = 5$ ,  $\Delta_S = 0.2$ )



Figura 3.25: Evolução da correlação entre dez caracteres ligados por efeitos pleiotrópicos controlados por uma função ontogenética linear binária variável sofrendo seleção estabilizadora correlacionada e seleção direcional intensa para mudança correlacionada dentro dos módulos, para diferentes valores do número de loci. (Ne = 5.000,  $\mu/\mu_B = 5$ ,  $\Delta_S = 0.2$ )

passados para a próxima geração. Isso significa que recombinação pode quebrar as associações criadas pela seleção, dificultando o estabelecimento da estrutura modular. Dessa forma, recombinação maior leva a uma menor eficiência na resposta à seleção natural. Essa discussão se encaixa no contexto de alvo e unidade de seleção (Templeton, 2003). Alvo de seleção é definido como o nível de organização que apresenta o fenótipo sendo selecionado. No nosso caso o indivíduo é o único alvo de seleção. Já a unidade de seleção se refere ao nível de organização genética que permite prever a resposta à seleção. Para tal, a unidade de seleção deve ter continuidade entre uma geração e outra. Essa definição pode abarcar diferentes níveis, dependendo da estrutura genética da população. O alvo de seleção pode ser um único locus ou o genoma todo. As associações criadas pela seleção natural são exemplos de unidades de seleção, onde diversos loci interagem para formar um indivíduo com determinado valor de aptidão. As associações modulares observadas nos nossos resultados são um exemplo de unidade de seleção maiores que um locus. É nesse ponto que a recombinação se torna relevante para definir a unidade de seleção, pois recombinação age no sentido de quebrar associação entre loci ao longo das gerações, reduzindo a unidade de seleção e embaralhando associações modulares.

\* \* \*

Com isso exploramos algumas das combinações de parâmetros possíveis nesses modelos, além das suas interpretações e consequências na evolução do padrão de correlação nas populações simuladas. No próximo capitulo vamos utilizar combinações plausíveis de parâmetros no nosso modelo para testar a viabilidade de trajetórias evolutivas que levem a modularidade variacional.

### Capítulo 4

# Evolução de Modularidade em Diferentes Regimes de Seleção

#### 4.1 Introdução

Neste capitulo vamos utilizar o modelo com matriz **B** apresentado no capítulo 2 e parametrizado no capítulo 3 para estudar algumas possibilidades de trajetórias evolutivas responsáveis por moldar o padrão de integração e modularidade de populações naturais. Novamente estamos interessados principalmente na evolução dos padrões de covariação e surgimento de modularidade. Os parâmetros gerais usados nessas simulações são  $p = 10, m = 50, \mu/\mu_B = 5, Ne = 5.000.$ 

#### 4.2 Estabelecimento de equilíbrio

Afim de minimizar ao máximo a influência de estados iniciais das populações, todas as simulações nesse capítulo são baseadas na mesma população inicial, que por sua vez passou por 10.000 gerações de deriva pura, sem seleção de qualquer tipo, seguidos de 10.000 gerações de seleção estabilizadora correlacionada com matriz  $\omega$  com dois módulos, como na equação 3.3. No primeiro período, de deriva, garantimos que o estado inicial é totalmente perdido por ação de mutação e deriva. Depois, durante a seleção estabilizadora, garantimos que a população esteja em equilíbrio seleção-mutação-deriva, mas ainda com variação suficiente para responder à seleção direcional. Padronizando as populações iniciais podemos isolar de forma mais sistemática efeitos de diferentes regimes de seleção. Na figura 4.1 observamos a evolução das variâncias fenotípicas, variâncias genéticas dos valores aditivos e da herdabilidade para todos os 10 caracteres da população. Notamos que inicialmente as variâncias crescem rapidamente, pois sem seleção a mutação introduz nova variação a cada geração que só pode ser removida por deriva. Como a população é bastante grande (Ne = 5.000), a remoção é lenta e o equilíbrio mutação-deriva se dá num valor de variância muito alto (comparado com a variância ambiental ou variância do erro e, veja capítulo 2, equação 2.13). Nas mesmas figuras, após 10.000 gerações de deriva, foi imposto um regime de seleção estabilizadora correlacionada. Imediatamente as variâncias genéticas e fenotípicas caem bruscamente para os valores de equilíbrio seleção-mutação-deriva, e o sistema se estabiliza nesses valores. Podemos ver no painel 4.1d o efeito sobre as correlações, que também diminuem com a remoção seletiva da variabilidade dessa população. A seleção estabilizadora é mantida por 10.000 gerações.

#### 4.3 Seleção Direcional

Tomados os cuidados com a inicialização da população, iniciamos o regime de seleção direcional com intensidade variáveis, como na seção 3.3.2. Novamente nossa seleção é medida por mudanças no pico adaptativo, com  $\Delta_S$  de 0.01 até 0.2. Todas os caracteres sofrem seleção direcional simultânea correlacionada, de aumento nos 5 primeiros e diminuição nos 5 últimos, de forma concordante com estrutura modular da seleção estabilizadora correlacionada. Calculamos os valores de  $|\beta_i|$ , afim de comparar nossa



Figura 4.1: Evolução de variâncias fenotípicas e genéticas, herdabilidades e correlações para uma população passando por 10.000 gerações de deriva, seguidas de 10.000 gerações de seleção estabilizadora correlacionada (Ne = 5.000, m/p = 5,  $\mu/\mu_B = 5$ ).

intensidade de seleção com aquela medida em populações naturais. Os resultados estão na tabela 4.1. Valores de AVG-Ratio e modularidade L para essas corridas podem ser vistos na figura 4.2. Incluímos também, por clareza, as últimas gerações de seleção estabilizadora, a seleção direcional começa a atuar a partir da geração 20.000.

Tabela 4.1:	Intensidades	de seleção,	representadas	s pelo d	liferencial	de seleção	por	caráter
e pelo valo	r do gradiente	e de seleção	por caráter $ _{i}$	$ \beta_i $				

$\Delta_S$	0.01	0.02	0.03	0.04	0.05
$ \beta_i $ médio	0.0016	0.0032	0.0048	0.0062	0.0076
$ \beta_i $ total médio	1.466	2.908	4.325	5.588	6.872
$\Delta_S$	0.06	0.07	0.08	0.09	0.10
$ \beta_i $ médio	0.0088	0.0103	0.0115	0.0128	0.0143
$ \beta_i $ total médio	7.975	9.290	10.398	11.583	12.913
$\Delta_S$	0.11	0.12	0.13	0.14	0.15
$ \beta_i $ médio	0.0152	0.0165	0.0175	0.0186	0.0200
$ \beta_i $ total médio	13.726	14.920	15.759	16.746	18.054
$\Delta_S$	0.16	0.17	0.18	0.19	0.20
$\frac{\Delta_S}{ \beta_i  \text{ médio}}$	0.16	0.17 0.0224	0.18 0.0238	0.19 0.0242	0.20 0.0254

Nessa condições, o sistema se modulariza rapidamente a partir de uma dada intensidade de seleção. O aumento tanto da modularidade L quanto do AVG-Ratio, indica presença de modularidade variacional na estrutura de covariação da população simulada. A concordância desses duas estatísticas nos leva a crer que os módulos presentes na superfície de seleção estão realmente se manifestando na estrutura de covariação da população. Podemos ainda dissecar o efeito da seleção direcional nas correlações dentro e entre módulos, olhando para a evolução de cada uma separadamente. Na figura 4.3 podemos ver essa separação para diferentes valores de  $\Delta_S$ . O padrão de aumento das correlações dentro módulos é esperado. Além disso, observamos uma concomitante diminuição das correlações entre módulos. Novamente, isso é característico de uma estrutura variacional modular, com duas classes de correlação correspondendo a relação de caracteres dentro e entre módulos. Como cada módulo responde a forças seletivas antagônicas, a correlação entre eles tente a diminuir, permitindo que os caracteres dentro de cada módulo evoluam independentemente dos outros módulos. Isso é coerente com simulações em sistemas simples feitas por Pavlicev et al. (2010) utilizando a equação de reposta a seleção de Lande (1979) e a existência de rQTLs (Pavlicev et al., 2008). Neste artigo os autores descrevem como seleção direcional pode favorecer ativamente a criação de linhas de menor resistência evolutiva (direções do espaço morfológico ricas em variação, veja Schluter, 1996), alinhadas com a seleção. Reorganização de estrutura variacional relacionada a eventos de seleção prolongada também já foram descritos na literatura (Berg, 1960; Young e Hallgrímsson, 2005; Young et al., 2010; Grabowski et al., 2011). Nossos resultados mostram que isso pode acontecer em sistemas bastante complexos, via seleção indireta nos padrões de covariação, gerando variação nas direções privilegiadas pela seleção direcional e aumentando a evolvabilidade das populações. Evolvabilidade é definida como a projeção da resposta evolutiva na direção privilegiada pela seleção natural (Hansen e Houle, 2008). Esse valor é controlado por dois fatores: a intensidade da pressão seletiva, seleção direcional mais intensa provoca mudanças maiores; e pela quantidade de variação disponível na população na direção da seleção direcional, caso a população não possua variação herdável na direção da seleção, não existe resposta evolutiva. O alinhamento da estrutura de covariação da população com o vetor de seleção basicamente representa o aumento de variação fenotípica da população na direção da seleção. Dessa forma, esse tipo de plasticidade na estrutura de covariação pode ser muito vantajosa, permitindo com que populações sofrendo seleção direcional aumentem sua variabilidade na direção da seleção, aumentando sua resposta adaptativa.

\* \* \*

Nossos valores de seleção direcional, apesar de promoverem diferenciações extremas ao longo de milhares de gerações, são relativamente baixos se comparados com estima-



Figura 4.2: AVG-Ratio e modularidade L para corridas com  $\mu/\mu_B = 5$ , Ne = 5.000, m/p = 5, sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ . Seleção direcional a partir da geração 20.000.



Figura 4.3: Evolução da média das correlações dentro de cada módulo e entre módulos de uma população sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .  $\mu/\mu_B = 5$ , Ne = 5.000, m/p = 5. Seleção direcional a partir da geração 20.000.

tivas em populações naturais. Nesse seção, o valor mais alto de diferencial de seleção direcional ( $\Delta_S = 0.2$ ), levando em conta a estrutura de covariação emergente da população, resultou em um gradiente de seleção médio de apenas  $|\beta_i| = 0.0254$ . Com esse gradiente baixo, apenas 15% de um gradiente típico em sistemas morfológicos e apenas 2% de um gradiente extremo (Kingsolver et al., 2001), nós observamos modularidade aparecendo de forma inequívoca em apenas alguns milhares de gerações, dependendo da força da seleção direcional aplicada (figuras 4.2 e 4.3). Partindo das mesmas populações iniciais, já em equilíbrio seleção-mutação-deriva, decidimos então testar gradientes de seleção mais extremos, com corridas durando apenas 1.000 gerações. Essas populações foram submetidas a forças de seleção  $10 \times$  maiores que as anteriores, com  $\Delta_S$  variando de 0.1 até 2.0. Dessa forma, a quantidade de modificação fenotípica total permaneceu constante, mas os gradientes de seleção foram aumentados sensivelmente. A relação dos  $|\beta_i|$  calculados para uma corrida aparece na tabela 4.2. Os  $|\beta_i|$  totais são semelhantes aos da tabela 4.1, porém os  $|\beta_i|$  médios são cerca de 10× maiores, como esperado. Podemos comparar a evolução das correlações para as corridas de 10.000 gerações de seleção com as de 1.000 gerações. Vemos esses dados na figura 4.4, para alguns valores de  $\Delta_S$ . Ao final das corridas, vemos valores bastante semelhantes para os valores de correlações médias dentro e entre módulos. Isso sugere que o tempo em gerações para o sistema assumir um nível determinado de modularidade depende linearmente da força de seleção aplicada.

Podemos quantificar o tempo de convergência da estrutura variacional da população para a estrutura de covariação da seleção direcional medindo diretamente a correlação da matriz de correlação da população com a matriz de correlação da superfície de seleção. Vemos esses resultados na figura 4.5, onde são mostrados a correlação entre a matriz de correlação da população a cada geração com a matriz  $\omega$  da superfície de seleção, para os extremos de força de seleção direcional nas corridas curtas, de 1.000 gerações, e longas, de 10.000, além de todo o periodo de inicialização descrito anteriormente, que serviu de estado inicial para ambas as corridas seletivas. Aqui vemos uma diferença entre os dois tipos de simulação. Para as corridas longas, sobre seleção fraca ( $\Delta_S = 0.01$ ) a correlação entre as matrizes não aumenta, enquanto para seleção forte ( $\Delta_S = 0.2$ ), após cerca de 4.000 gerações a correlação chega num patamar estável mais alto. Nas corridas curtas isso não acontece, e a semelhança da matriz não se estabiliza após um período correspondente, só se tornando consistentemente mais alto que na corrida com seleção mais baixa após cerca de 750 gerações. Com isso vemos que a dinâmica da resposta à seleção direcional não é totalmente linear, e existem complicações associadas à interação entre a quantidade de variação introduzida na população por mutação e a variação removida pela seleção natural. Quando aumentamos a intensidade da seleção direcional, mais variação é removida a cada geração pela seleção, e o número de indivíduos que compõem a próxima geração diminui, pelo aumento na variância dos valores de aptidão causado pela distância maior entre o pico adaptativo e a média da população. Uma consequência desse "gargalo seletivo" é o aumento da deriva genética, que implica numa variação maior da estrutura de covariação. Todos esses fatores contribuem para que a população demore mais para se estabilizar na estrutura de covariação privilegiada pela seleção, e a correlação entre a matriz de correlação da população e da superfície de seleção acaba flutuando muito e não aumentando tanto quanto no caso de seleção mais fraca e continuada.

#### 4.4 Seleção direcional "corredor"

Seleção direcional parece então ser fundamental para o surgimento de módulos variacionais nas nossas simulações. Porém, a situação onde todos os caracteres de um complexo morfológico estão sobre seleção simultânea divergente não parece ser uma regra na natureza, ou pelo menos é difícil acreditar que todas as instâncias de modularidade variacional sejam devido a eventos de seleção direcional divergente em todos



Figura 4.4: Evolução da média das correlações dentro de cada módulo e entre módulos de uma população sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional com diferentes valores de  $\Delta_S$ .  $\mu/\mu_B = 5$ , Ne = 5.000, m/p = 5. Todas as populações partem da mesma população inicial, e cada linha tem o mesmo  $|\beta_i|$  total médio ao final da corrida. São mostradas apenas as gerações de seleção direcional.



Figura 4.5: Evolução da correlação entre a matriz de correlação da população a cada geração com a matriz de correlação da superfície de seleção  $\omega$ , para corridas de diferentes números de gerações e intensidades de seleção direcional divergente entre módulos. O painel (a) mostra a evolução da correlação antes da seleção direcional, com 10000 gerações de deriva e 10.000 gerações de seleção estabilizadora correlacionada.  $\mu/\mu_B = 5$ , Ne = 5.000, m/p = 5

$\Delta_S$	0.1	0.2	0.3	0.4	0.5
$ \beta_i $ médio	0.014	0.025	0.036	0.048	0.057
$ \beta_i $ total médio	1.431	2.564	3.640	4.824	5.786
$\Delta_S$	0.6	0.7	0.8	0.9	1.0
$ \beta_i $ médio	0.066	0.077	0.087	0.098	0.104
$ \beta_i $ total médio	6.629	7.782	8.792	9.898	10.454
$\Delta_S$	1.1	1.2	1.3	1.4	1.5
$\frac{\Delta_S}{ \beta_i  \text{ médio}}$	1.1 0.113	1.2 0.129	1.3 0.138	$\frac{1.4}{0.157}$	$1.5 \\ 0.154$
$\begin{array}{c} \Delta_S \\  \beta_i  \text{ médio} \\  \beta_i  \text{ total médio} \end{array}$	1.1       0.113       11.348	$     \begin{array}{r}       1.2 \\       0.129 \\       12.937     \end{array} $	$     \begin{array}{r}       1.3 \\       0.138 \\       13.875     \end{array} $	$     \begin{array}{r}       1.4 \\       0.157 \\       15.733     \end{array} $	$     \begin{array}{r}       1.5 \\       0.154 \\       15.499     \end{array} $
$\begin{array}{c} \Delta_S \\  \beta_i  \text{ médio} \\  \beta_i  \text{ total médio} \\ \end{array}$	$ \begin{array}{c c} 1.1 \\ 0.113 \\ 11.348 \\ \hline 1.6 \\ \end{array} $	$ \begin{array}{c c} 1.2 \\ 0.129 \\ 12.937 \\ \hline 1.7 \\ \end{array} $	1.3 0.138 13.875 1.8	$     \begin{array}{r}       1.4 \\       0.157 \\       15.733 \\       1.9 \\       \end{array} $	$     \begin{array}{r}       1.5 \\       0.154 \\       15.499 \\       2.0 \\       \end{array} $
$\begin{array}{c} \Delta_S \\  \beta_i  \text{ médio} \\  \beta_i  \text{ total médio} \\ \hline \\ \Delta_S \\ \hline  \beta_i  \text{ médio} \end{array}$	$ \begin{array}{c} 1.1 \\ 0.113 \\ 11.348 \\ \hline 1.6 \\ 0.163 \\ \end{array} $	$ \begin{array}{r} 1.2\\ 0.129\\ 12.937\\ \hline 1.7\\ 0.177\\ \end{array} $	$ \begin{array}{r} 1.3 \\ 0.138 \\ 13.875 \\ \hline 1.8 \\ 0.184 \\ \end{array} $	$     \begin{array}{r}       1.4 \\       0.157 \\       15.733 \\       \hline       1.9 \\       0.207 \\       \end{array} $	$ \begin{array}{r} 1.5 \\ 0.154 \\ 15.499 \\ \hline 2.0 \\ 0.207 \\ \end{array} $

Tabela 4.2: Intensidades de seleção, representadas pelo diferencial de seleção por caráter e pelo valor do gradiente de seleção por caráter  $|\beta_i|$  para as simulações com 1.000 gerações de seleção direcional correlacionada.

os caracteres dos módulos envolvidos simultaneamente. Seria a seleção natural sobre apenas um conjuntos de caracteres capaz de gerar modularidade do mesmo tipo? Para responder essa pergunta nós começamos com a mesma população inicial, mas agora apenas um dos módulos era sujeito a seleção direcional, com o outro conjunto de caracteres mantido constante pela seleção estabilizadora. Novamente a matriz de seleção da superfície adaptativa era não diagonal, impondo seleção estabilizadora correlacionada na população. Esse cenário evolutivo é conhecido como evolução "corredor", pois uma parte dos caracteres se altera por ação da seleção direcional (ao longo do corredor) enquanto o outro conjunto está restrito por ação de seleção estabilizadora (entre as paredes do corredor).

Na figura 4.6 vemos a evolução das correlações dentro e entre módulos para esse cenário evolutivo. Primeiro destacamos que novamente as correlações entre módulo decaem, de forma semelhante ao caso de seleção direcional divergente nos dois módulos. O módulo sobre seleção direcional também se comporta de forma semelhante ao caso anterior, com correlações aumentando ao longo da simulação. Porém, nessas simulações são geradas 3 classes de correlação. As correlações dentro do módulo sofrendo apenas seleção estabilizadora também aumentam, ainda que de forma menos acentuada que as do módulo que sofre seleção direcional. Em alguns casos as correlações dentro do módulo com seleção estabilizadora chegam ao mesmo nível médio que o módulo sofrendo seleção direcional, como no painel 4.6d. À medida que a seleção direcional aumenta, a distinção entre os dois módulos se torna mais acentuada. Esses resultados apontam para um mecanismo plausível de formação de módulos variacionais em populações naturais. Como a seleção em um grupo de caracteres é suficiente para gerar modularidade na estrutura de covariação completa na população, incluindo caracteres não diretamente envolvidos na seleção direcional, eventos subsequentes de seleção direcional em grupos distintos de caracteres morfológicos podem gerar o padrão intrincado de módulos que observamos na natureza, sem para isso ser necessária a atuação de seleção direcional divergente simultânea sobre todos os módulos.



Figura 4.6: Evolução da média das correlações dentro de cada módulo e entre módulos de uma população sofrendo seleção estabilizadora correlacionada com 2 módulos e seleção direcional em apenas um dos módulos, com diferentes valores de  $\Delta_S$ .  $\mu/\mu_B = 5$ , Ne = 5.000, m/p = 5. Seleção direcional a partir da geração 20.000.

### Capítulo 5

# Considerações Finais

Neste trabalho procuramos trazer uma abordagem matemática/computacional tipicamente utilizada em sistemas físicos para o estudo de modularidade. O simples ato de tentar reproduzir o comportamento de um sistema natural complexo através de interações simples das suas partes nos leva a um melhor entendimento desse sistema, no sentido de evidenciar quais os ingredientes necessários para que o sistema se comporte de determinada maneira, mesmo que esses ingredientes sejam difíceis de ser medidos ou observados na natureza. Além disso, os parâmetros utilizados na criação do modelo tem interpretações no mundo natural. O sistema computacional permite com que estudemos as consequências da variação desses parâmetros em diversas situações inacessíveis ao experimento ou à medição retrospectiva. Podemos também restringir o espaço de variação desses parâmetros, facilitando sua estimativa por vias indiretas ou limitando sua influência em outros processos de interesse.

Nosso trabalho também se coloca como a primeira tentativa de simulação de sistemas morfológicos complexos o suficiente, com mais de dois caracteres, para permitir o surgimento de estruturas de modularidade variacional complexas. Tentativas anteriores se limitam a apenas simular um, ou um número pequeno, de caracteres. Por isso também as exigências computacionais são consideravelmente maiores no nosso caso. Assim, nos limitamos a um tratamento basicamente qualitativo do problema de evolução da modularidade. Nesse contexto, mostramos que seleção direcional correlacionada é necessária para o surgimento de módulos variacionais. Além disso, para a manutenção desses módulos precisamos de seleção estabilizadora correlacionada, mantendo o padrão de variação estabelecido pela seleção direcional. Mostramos ainda que a ação conjunta de seleção direcional e estabilizadora é suficiente para criar padrões covariação complexos, com varias classes de correlação diferentes surgindo pela interação entre os dois tipos de seleção. Com uma implementação mais sofisticada dos nossos algoritmos, aliada a um poder computacional maior, acreditamos que uma extensão quantitativa dos resultados seja não só possível como bastante proveitosa. Pretendemos implementar nosso modelo em linguagens de programação acessíveis, possibilitando seu uso pela comunidade em geral, com suporte para paralelização, manipulação de parâmetros com interface amigável e até criar um projeto integrado de varias possibilidades de simulações com caracteres quantitativos.

Varias questões interessantes permanecem abertas, e pretendemos continuar nesta linha de pesquisa. Já de principio, acreditamos que o modelo nulo, de mutação e deriva, sem seleção, merece ser extensamente replicado, criando um distribuição nula para testes estatísticos rigorosos e estudando que tipos de estruturas variacionais podem surgir por mero acaso. Seriam essas estruturas estáveis por um tempo apreciável? Com que frequência o nível de modularidade observado em populações naturais é observado nesse modelo nulo? Além disso, a interação entre tamanho populacional, recombinação, mutação e seleção se mostrou relevante tanto para o nível de modularidade observado quanto para a velocidade de estabelecimento do padrão modular. Explorar de forma sistemática a relação entre esses parâmetros é um caminho natural para a continuidade deste trabalho.

Pretendemos também abordar possíveis soluções analíticas para o modelo proposto. Existe algum trabalho em esquemas semelhantes (Wagner, 1984, 1989; Jones et al., 2007), mas com abordagens diferentes, não tratando de modularidade ou regiões fora do equilíbrio, como nas nossas simulações.

Extensões metodológicas também são possíveis. Em população naturais, a organização genômica em forma de cromossomos traz uma série de consequências interessantes relacionadas a desequilíbrio de ligação e recombinação, tema frequentemente negligenciado mesmo em estudos teóricos complexos (Barton e Turelli, 1987; Turelli e Barton, 1994). A ação da seleção cria associações entre regiões cromossômicas baseadas em características funcionais, aproximando genes ligados a uma mesma função espacialmente e gerando desequilíbrio de ligação e herança conjunta de blocos funcionais (Templeton, 2003). Esse processo pode ser importante em definir a velocidade de criação e reorganização da estruturas genômicas modulares. A formação dessas associações, que nada mais são que unidades de seleção acima do nível de genes, é contrabalanceado pela ação da recombinação, que embaralha as regiões genômicas e quebra ligações formadas pela ação da seleção. Incluir estruturas cromossômicas de fato, possibilitando desequilíbrio de ligação e reestruturações na ligação entre loci traria mais um camada de complexidade importante na compreensão da evolução da modularidade, permitindo estudar a dinâmica de formação de unidades de seleção ao longo da evolução, e sua relação com forças de seleção e recombinação. Da mesma forma, estudar várias populações simultaneamente, permitindo migração entre elas, pode trazer resultados interessantes na teoria de especiação e diversificação de populações. Atualmente a teoria de genética quantitativa é relativamente pobre para tratar do problema de fluxo gênico. Abordagens computacionais podem mudar esse cenário.

Eventos populacionais são reconhecidamente importantes na evolução de populações naturais, como gargalos, efeito fundador ou mesmo ou episódios consecutivos de seleção ou mesmo seleção dependente de frequência podem ter consequências interessantes.

Outra extensão clara é a de utilizar o nosso modelo para reproduzir padrões específicos de diversificação observados na natureza. Por exemplo, um padrão claro em mamíferos é a diferença entre os níveis de integração total, característica extremamente variável entre os grupos, apesar do padrão de integração ser relativamente estável Porto et al. (2009). Quais cenários evolutivos e tipos de seleção e populações podem levar a extrema diferença na magnitude de integração observada entre babuínos e grandes primatas? A mesma pergunta pode ser feita em relação aos valores altíssimos de integração presentes em marsupiais. Seriam esses valores de altos de integração causados por seleção direcional correlacionada a todos os caracteres? Para abordar esses problemas, poderíamos estudar sistemas já com padrões modulares estabelecidos e simular situações capazes de alterar esses padrões nas diferentes populações. No momento nossas populações sempre são inciadas totalmente integradas ou desintegradas, seria interessante estudar a interação da seleção com padrões já estabelecidos. Consequências evolutivas claras, na forma de seleção indireta, são previstas pela teoria de genética quantitativa quantitativa (Lande e Arnold, 1983; Barton e Turelli, 1987), mas essas situações não foram abordadas computacionalmente em sistemas de muitos caracteres.

## Referências Bibliográficas

- Ancel, L. W. e Fontana, W. (2000). Plasticity, evolvability, and modularity in RNA. The Journal of experimental zoology, 288(3):242–83.
- Anderson, T. W. (1984). Introduction to Multivariate Statistical Analysis. Wiley.
- Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., e Jones, A. G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution*, 62(10):2451–61.
- Arnold, S. J., Pfrender, M. E., e Jones, A. G. (2001). The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica*, 112/113:9–32.
- Barton, N. H. e Turelli, M. (1987). Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genetical research*, 49(2):157–73.
- Barton, N. H. e Turelli, M. (1989). Evolutionary Quantitative Genetics: How Little Do We Know. Annual Review of Genetics, 23(1):337–370.
- Berg, R. L. (1960). The Ecological Significance of Correlation Pleiades. *Evolution*, 14(2):171.
- Bürger, R. e Lande, R. (1994). On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. *Genetics*, 138(3):901–12.
- Charlesworth, B., Lande, R., e Slatkin, M. (1982). A Neo-Darwinian Commentary on Macroevolution. *Evolution*, 36(3):474.

- Cheverud, J. M. (1984). Quantitative genetics and developmental constraints on evolution by selection. *Journal of theoretical biology*, 110(2):155–71.
- Cheverud, J. M. (1988). A Comparison of Genetic and Phenotypic Correlations. *Evolution*, 42(5):958.
- Cheverud, J. M. e Leamy, L. J. (1985). Quantitative genetics and the evolution of ontogeny. III. Ontogenetic changes in correlation structure among live-body traits in randombred mice. *Genetical Research*, 46(3):325–335.
- Cheverud, J. M. e Routman, E. J. (1995). Epistasis and its contribution to genetic variance components. *Genetics*, 139(3):1455–61.
- Cheverud, J. M. e Routman, E. J. (1996). Epistasis as a Source of Increased Additive Genetic Variance at Population Bottlenecks. *Evolution*, 50(3):1042.
- Cheverud, J. M., Routman, E. J., e Irschick, D. J. (1997). Pleiotropic Effects of Individual Gene Loci on Mandibular Morphology. *Evolution*, 51(6):2006.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R. P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.-Y., Liang, W., Marback, M., Paw, J., San Luis, B.-J., Shuteriqi, E., Tong, A. H. Y., van Dyk, N., Wallace, I. M., Whitney, J. a., Weirauch, M. T., Zhong, G., Zhu, H., Houry, W. a., Brudno, M., Ragibizadeh, S., Papp, B., Pál, C., Roth, F. P., Giaever, G., Nislow, C., Troyanskaya, O. G., Bussey, H., Bader, G. D., Gingras, A.-C., Morris, Q. D., Kim, P. M., Kaiser, C. a., Myers, C. L., Andrews, B. J., e Boone, C. (2010). The genetic landscape of a cell. Science (New York, N.Y.), 327(5964):425–31.
- Crow, J. F. e Kimura, M. (1964). The theory of genetic loads. In Proceedings of the XI international congress of genetics., pages 495–505.

- Crow, J. F. e Kimura, M. (1970). An Introduction to Population Genetics Theory. Harper and Row.
- Espinosa-Soto, C. e Wagner, A. (2010). Specialization can drive the evolution of modularity. *PLoS computational biology*, 6(3):e1000719.
- Falconer, D. S. e Mackay, T. F. (1996). Introduction to Quantitative Genetics. Benjamin Cummings, London, 4th edition.
- Grabowski, M. W., Polk, J. D., e Roseman, C. C. (2011). Divergent patterns of integration and reduced constraint in the human hip and the origins of bipedalism. *Evolution: International Journal of Organic Evolution*, 65(5):1336–1356.
- Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P., e Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430(6995):88–93.
- Hansen, T. F. e Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of evolutionary biology*, 21(5):1201–19.
- Joe, H. (2006). Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis, 97(10):2177–2189.
- Jones, A. G., Arnold, S. J., e Bürger, R. (2003). Stability of the g-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution*, 57(8):1747–1760.
- Jones, A. G., Arnold, S. J., e Bürger, R. (2004). Evolution and stability of the g-matrix on a landscape with a moving optimum. *Evolution*, 58(8):1639–1654.
- Jones, A. G., Arnold, S. J., e Bürger, R. (2007). The mutation matrix and the evolution of evolvability. *Evolution*, 61(4):727–45.

- Kimura, M. (1965). A Stochastic Model concerning the Maintenance of Genetic Variability in Quantitative Characters. Proceedings of the National Academy of Sciences, 54(3):731–736.
- Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P., e Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American naturalist*, 157(3):245–61.
- Klingenberg, C. P. (2008). Morphological Integration and Developmental Modularity. Annual Review of Ecology, Evolution, and Systematics, 39:115–132.
- Lande, R. (1979). Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. *Evolution*, 33(1):402.
- Lande, R. e Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution*, 37(6):1210.
- Lewandowski, D., Kurowicka, D., e Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001.
- Lynch, M. e Walsh., B. (1998). Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.
- Marroig, G. e Cheverud, J. (2010). Size as a line of least resistance II: direct selection on size or correlated response due to constraints? *Evolution; international journal* of organic evolution, 64(5):1470–88.
- Marroig, G. e Cheverud, J. M. (2001). A comparison of phenotipic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys. *Evolution*, 55(12):2576–2600.

- Marroig, G. e Cheverud, J. M. (2004). Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys? *The American naturalist*, 163(3):417– 428.
- Marroig, G. e Cheverud, J. M. (2005). Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. *Evolution; international journal of organic evolution*, 59(5):1128–42.
- Marroig, G., Melo, D. A. R., e Garcia, G. (2012). Modularity, noise, and natural selection. Evolution; international journal of organic evolution, 66(5):1506–24.
- Marroig, G., Shirai, L. T., Porto, A., Oliveira, F. B., e Conto, V. (2009). The Evolution of Modularity in the Mammalian Skull II: Evolutionary Consequences. *Evolutionary Biology*, 36(1):136–148.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., e Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal* of Chemical Physics, 21(6):1087.
- Newman, M. E. J. (2006a). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):22.
- Newman, M. E. J. (2006b). Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America, 103(23):8577–82.
- Pavlicev, M., Cheverud, J. M., e Wagner, G. P. (2010). Evolution of adaptive phenotypic variation patterns by direct selection for evolvability. *Proceedings. Biological sciences* / The Royal Society.
- Pavlicev, M. e Hansen, T. F. (2011). Genotype-Phenotype Maps Maximizing Evolvability: Modularity Revisited. *Evolutionary Biology*.

- Pavlicev, M., Kenney-Hunt, J. P., Norgard, E. A., Roseman, C. C., Wolf, J. B., e Cheverud, J. M. (2008). Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. *Evolution*, 62(1):199–213.
- Porto, A., Oliveira, F. B., Shirai, L. T., Conto, V., e Marroig, G. (2009). The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes. *Evolutionary Biology*, 36(1):118–135.
- Reichardt, J. e Bornholdt, S. (2006). Statistical mechanics of community detection. Physical Review E, 74(1):1–16.
- Schluter, D. (1996). Adaptive Radiation Along Genetic Lines of Least Resistance. Evolution, 50(5):1766.
- Templeton, A. (2003). Population Genetics and Microevolutionary Theory. In Population Genetics and Microevolutionary Theory. Wiley & Sons.
- Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical population theory*, 25:138–193.
- Turelli, M. (1985). Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics*, 111(1):165–95.
- Turelli, M. (1986). Gaussian versus non-Gaussian genetic analyses of polygenic mutation-selection balance. In Nevo, E. e Karlin, S., editors, *Evolutionary Processes* and Theory, pages 607–628. Academic Press, New York.
- Turelli, M. (1988). Phenotypic Evolution, Constant Covariances, and the Maintenance of Additive Variance. *Evolution*, 42(6):1342.
- Turelli, M. e Barton, N. H. (1994). Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics*, 138(3):913–41.

- Wagner, G. P. (1984). On the eigenvalue distribution of genetic and phenotypic dispersion matrices: Evidence for a nonrandom organization of quantitative character variation. *Journal of Mathematical Biology*, 21(1):77–95.
- Wagner, G. P. (1989). Multivariate mutation-selection balance with constrained pleiotropic effects. *Genetics*, 122(1):223–34.
- Wagner, G. P. (1996). Homologues, Natural Kinds and the Evolution of Modularity. Integrative and Comparative Biology, 36(1):36–43.
- Wagner, G. P., Pavlicev, M., e Cheverud, J. M. (2007). The road to modularity. Nature reviews. Genetics, 8(12):921–31.
- Whitlock, M. C. (1995). Variance-Induced Peak Shifts. Evolution, 49(2):252.
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc of the 6th International Congress of Genetics, 1(6):356–366.
- Young, N. M. e Hallgrímsson, B. (2005). Serial homology and the evolution of mammalian limb covariation structure. *Evolution; international journal of organic evolution*, 59(12):2691–704.
- Young, N. M., Wagner, G. P., e Hallgrímsson, B. (2010). Development and the evolvability of human limbs. Proceedings of the National Academy of Sciences of the United States of America, 107(8):3400–5.