Guilherme Garcia

**Perspectivas sobre o reconhecimento de
padrões de modularidade e suas
implicações para a evolução de
morfologias complexas**

On the recognition of modularity patterns and its implications
for the evolution of morphological systems

São Paulo
2016

Guilherme Garcia

**Perspectivas sobre o reconhecimento de
padrões de modularidade e suas
implicações para a evolução de
morfologias complexas**

Tese apresentada ao Instituto de Biociências da
Universidade de São Paulo, para a obtenção de
Título de Doutor em Ciências, na Área de Gené-
tica e Biologia Evolutiva.

Orientador: Gabriel Marroig

**São Paulo
2016**

Ficha Catalográfica

Garcia, Guilherme

Perspectivas sobre o reconhecimento de padrões de modularidade e suas implicações para a evolução de morfologias complexas.

160 páginas.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Morfometria;
2. Genética Quantitativa;
3. Integração Morfológica;
4. Crânio;
5. Primatas Antropóides.

I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr.

Prof. Dr. Gabriel Marroig

*Ao Vô Moisés,
por me ensinar que o conhecimento
vai muito além de livros e palavras,
ou das paredes de uma sala de aula;*

*e ao Miro,
que, se tudo der certo,
vai aprender comigo coisas
que eu nem sonhava em ensinar.*

If only I could
Clear my eyes
Then I might breathe once more
Then I might breathe again

Old sun and stars,
And oceans below me
Guide my strides over
Jagged shards, under foot

To slash to the sound
How many sit on woe or peril?
How many walk on their own?

Into the truth
Let myself burn
Now it's written
1000 shards

Isis - 1000 Shards

Black
then
white are
all I see
in my infancy.
Red and yellow then came to be
reaching out to me.
Lets me see.

There is
so
much more
and beckons me
to look through to these
infinite possibilities.
As below
so above and beyond
I imagine.
Drawn outside the lines of reason.
Push the envelope.
Watch it bend.

Over thinking, over analyzing
separates the body from the mind.
Withering my intuition
leaving opportunities behind.

Feed my will to feel this moment
urging me to cross the line.
Reaching out to embrace the random.
Reaching out to embrace whatever may come.

I embrace my desire to
feel the rhythm, to feel connected
enough to step aside and weep like a widow
to feel inspired, to fathom the power,
to witness the beauty, to bathe in the fountain,
to swing on the spiral, to swing on the spiral,
to swing on the spiral of our divinity
and still be a human.

With my feet upon the ground
I lose myself between the sounds
and open wide to suck it in.
I feel it move across my skin.
I'm reaching up and reaching out.
I'm reaching for the random
or whatever will bewilder me.
Whatever will bewilder me.
And following our will and wind
we may just go where no one's been.
We'll ride the spiral to the end
and may just go where no one's been.

Spiral out. Keep going.

Tool - Lateralus

Agradecimentos

Em primeiro lugar, agradeço ao meu orientador, Gabriel Marroig, por ter me dado a oportunidade de trabalhar neste tema e por sempre contribuir para me devolver ao chão com as nossas discussões e o rigor com que ele encara a ciência. Espero que nós tenhamos muitos e muitos anos de colaboração pela frente.

Agradeço também aos meus pares, os demais alunos do LEM, por prover um ambiente fantástico para se fazer ciência, onde as discussões sempre acontecem, inclusive agora, aqui ao lado, enquanto escrevo estes agradecimentos. E claro, muitas vezes nossas discussões vão para além da ciência, mas nunca em direção ao senso comum, e estas outras conversas também contribuem ao seu modo para a confecção de um doutor em ciências.

À Aninha, por ser diversas vezes interrompida por um fluxo caótico de ideias inacabadas e ter a paciência de escutá-las, desvantagem de sentar ao meu lado. À Paulinha, por nunca ter recusado ir tomar um café comigo, mesmo que ela precisasse muito ir embora, e por toda a empolgação que ela traz em relação a tudo que a gente faz. À Dani, pelo suprimento infinito de polenguinho light, por ler vários pedaços desta tese ainda em formação, e por todos os sorrisos e abraços que ajudaram nas horas mais escuras. À Tafinha (ou Bárbara), por manter o bom humor mesmo nestas horas escuras e por sempre estar disposta a discutir essencialmente qualquer coisa, independente de quanto tempo isso vai levar. À Monique, pelas divagações aleatórias a respeito de ciência, que contribuem muito pra esta tese, sem dúvida, e também pelas conversas a respeito das nossas crianças lindas. Ao Ogro (ou Diogo), pela parceria agora de longa data nas partes mais cabeludas das análises e por ser uma fonte inesgotável de boa música, boa comida, e bom gosto em geral. Ao Lugar (ou Fábio), pelo humor *nonsense* afiadíssimo que sempre alegra o dia e pelas nossas discussões edificantes a respeito do mundo mágico da morfometria geométrica. Ao Wally (ou Thiago), pela companhia nas tarefas mais divertidas, por exemplo levar computadores pro conserto, e claro, pelas infinitas discussões sobre ciência, sociedade e muito, muito mais. E à Papete (ou Anna), por ser uma boa companhia pra todas as horas, da mesinha até a volta pra casa, por escutar o que eu tenho a dizer, independente do que seja, e pelas parcerias que a gente ainda deve levar em frente a respeito desses primatas aí. Também agradeço aqueles que já deixaram o LEM, em especial Alex, Fino, Edgar e Harley, por todas as contribuições pessoais e profissionais nos anos que passaram. Nesse embalo, agradeço também o Aríete (ou Gustavo) pelas nossas discussões regadas à café ou cerveja, que com certeza contribuem muito, e também por ler pedaços desta tese em formação.

Agradeço também aos colegas do grupo-irmão do Diogo Meyer, que oferecem diversas

perspectivas diferentes às nossas; em especial à Bárbara e ao Limão (ou Luiz Carlos), pelas conversas muito boas que tivemos recentemente. Também agradeço ao Rui Murrieta, pela oportunidade de ser monitor na disciplina de Filosofia, o que certamente contribuiu para o desenvolvimento desta tese, e ao Sérgio Mاتيoli, por conta dos comentários durante a minha qualificação que foram o ponto de partida para o Capítulo 4.

Agradeço à FAPESP pela concessão de uma bolsa de doutorado.

Devo também agradecer a minha família. Sem todos vocês, seria impossível chegar até aqui. Em especial, aos meus pais, Eduardo e Teresa, pelo apoio incondicional durante todos esses anos. Ao meu irmão Gustavo, que nunca tem receios em me dizer o que eu tenho que ouvir. E à vó Carlota, pela dieta de banana, cabeça de peixe e taboada na infância.

Finalmente, agradeço a minha companheira, a Giuliana, por estar aqui ao meu lado, mesmo nos piores momentos, e por me oferecer aquele abraço perfeito. E ao Ramiro, nosso menininho, por conta daquele sorriso lindo que ele deu ontem quando eu cheguei em casa, e por dar tanto sentido a tudo.

Resumo

A modularidade é uma propriedade característica que sistemas biológicos exibem em relação à distribuição de interações entre seus elementos constituintes; neste contexto, um módulo é um subconjunto de elementos que interagem entre si mais do que com outros subconjuntos. Em relação aos sistemas morfológicos, tais propriedades referem-se geralmente à estrutura do componente linear do mapa genótipo/fenótipo; no entanto, as interações genéticas, ontogenéticas e funcionais que produzem fenótipos são descritas de forma adequada através de dinâmicas não-lineares, e uma apreciação completa da complexidade destas interações é necessária para a compreensão das propriedades variacionais do fenótipo. Ademais, dados avanços metodológicos na área da morfometria, é possível escolher diferentes maneiras de representar a variação morfológica, e as diferenças entre as representações podem impactar inferências feitas sobre estas propriedades variacionais. A presente tese tem como objetivo explorar a relação entre representações morfométricas e a caracterização das propriedades variacionais, focada na análise comparativa de tais propriedades em uma escala macroevolutiva; Primatas Antropóides são utilizados como modelo, dada a disponibilidade de uma grande base de dados de mensurações cranianas destes organismo. Esta relação foi avaliada sob três perspectivas diferentes. Em primeiro lugar, estima-se taxas de erro associadas aos testes de hipótese que descrevem padrões de modularidade, relacionadas com três representações morfométricas distintas; tal avaliação é também associada à exploração de um subconjunto da base de dados utilizada aqui, levando-se em consideração a dinâmica de interações ontogenéticas que produzem o crânio dos Antropóides. Os resultados deste capítulo implicam que uma dessas representações, resíduos de Procrustes, não são capazes de detectar padrões de modularidade neste contexto, considerando suas propriedades matemáticas específicas. Outras duas representações, distâncias entre marcos anatômicos e variáveis locais de forma, produzem resultados semelhantes, que estão diretamente associados à dinâmica de desenvolvimento, e as diferenças que elas apresentam são consistentes com suas diferenças principais; taxas de erro para os testes sobre as duas representações também são aceitáveis. O próximo capítulo trata da comparação entre estas duas representações no que diz respeito a estas diferentes propriedades, focado em estimar relações alométricas associadas às variáveis locais de forma e a relação entre estas estimativas e os padrões de modularidade estimados para distâncias entre marcos anatômicos. Os resultados encontrados enfatizam que os padrões de modularidade observados em distâncias entre marcos são consequência da alometria; linhagens como *Homo* e *Gorilla*, que apresentam padrões distintos de modularidade para as distâncias entre marcos estão associados a mudanças substanciais nas relações alométricas dos caracteres cranianos. O último capítulo explora a estrutura filogenética de mudanças nas propriedades variacionais fenotípicas na diversificação de Anthroipoidea, considerando apenas variáveis de forma locais, uma vez que este capítulo também visa reforçar os resultados anteriores obtidos a partir de distâncias entre marcos, considerando-se um tipo diferente de representação morfométrica. Este capítulo muda o foco de testes a respeito de padrões de modularidade definidos *a priori* em direção a estimar a incerteza relacionada à estrutura de matrizes de covariância, decomposta sobre a filogenia de Anthroipoidea. Os resultados obtidos demonstram que as mudanças na estrutura de covariância nesta linhagem são localizadas nas mesmas regiões do crânio ao longo de toda a história evolutiva do grupo, enquanto outras regiões mantêm associações estáveis. Assim, quando se considera as diferentes propriedades de representações morfométricas cuidadosamente, inferências feitas a partir de tais representações sobre propriedades variacionais são de fato compatíveis.

Abstract

Modularity is a characteristic property biological systems exhibit regarding the distribution of interactions between their composing elements; in this context, a module is a subset of elements which interact more among themselves than with other subsets. Regarding morphological systems, such property usually refers to the structure of the linear component of the genotype/phenotype map; however, the genetic, developmental, and functional interactions that produce phenotypes are often best described by non-linear dynamics, and a full appreciation of the complexity of such interactions is necessary for understanding phenotypic variational properties. Furthermore, given methodological advances in the field of morphometrics, one may choose different ways to represent morphological variation, and differences between representations may impact inferences made regarding variational properties. The present dissertation aims at exploring the relationship between morphometric representations and the characterization of variational properties, focusing on the comparative analysis of such properties on a macroevolutionary timeframe; Anthropoid Primates are used as a model lineage, given the availability of a large database of skull measurements. This relationship was evaluated under three different perspectives. First, an estimation of the error rates associated with tests for hypothesis that describe modularity patterns related to three different morphometric representations; such evaluation is also associated with an exploration of a subset of the database used here, considering the dynamical properties of developmental interactions that produce the Anthropoid skull. The results of this chapter imply that one of such representations, Procrustes residuals, fails to capture modularity patterns in this setting, considering its particular mathematical underpinnings. Other two representations, interlandmark distances and local shape variables, produce similar results which are directly associated with developmental dynamics, and the differences they exhibit are consistent with their different properties; error rates for tests over both representations are also acceptable. The next chapter deals with comparing these two representations with respect to these different properties, focusing on estimating allometric relationships over local shape variables and the relationship between such estimates and modularity patterns estimated for interlandmark distances. The results found stress out that modularity patterns observed in interlandmark distances are a consequence of allometry; lineages such as *Homo* and *Gorilla*, which exhibit distinct modularity patterns in interlandmark distances are associated with substantial changes in allometric relationships for skull traits. The last chapter explores the phylogenetic structure of changes in phenotypic variational properties across Anthropoid diversification, considering local shape variables alone, since this chapter also aims at reinforcing previous results obtained from interlandmark distances, considering a different type of morphometric representation. This chapter shifts the focus from testing *a priori*-defined modularity patterns to estimating the uncertainty related to covariance matrix structure decomposed over the Anthropoid phylogeny. The results obtained demonstrate that changes in covariance structure on this lineage are localized in the same skull regions across the entire evolutionary history of Anthroidea, while other regions maintain stable associations. Thus, when one considers the different properties of morphometric representations carefully, inferences made from such representations regarding variational properties are in fact compatible.

Keywords: morphometrics, morphological integration, quantitative genetics, skull, anthropoid primates

Sumário

1	Introdução	1
	Objetivos	9
2	Modularity and Morphometrics: Error Rates in Hypothesis Testing	19
	Introduction	20
	Methods	24
	Results	31
	Discussion	40
3	The Evolution of Skull Allometry in Anthropoid Primates	48
	Introduction	49
	Methods	50
	Results	55
	Discussion	58
4	A Phylogenetic Analysis of Shape Covariance Structure	65
	Introduction	66
	Methods	70
	Results	77
	Discussion	84
5	Considerações Finais	91
	Referências Bibliográficas	94
	Apêndice: Artigos Publicados	110

Lista de Figuras

1.1	Tamanhos amostrais e efeitos fixos controlados para as 109 OTUs de primatas antropóides . .	13
1.2	Configuração de Marcos Anatômicos, representada sobre um crânio de <i>Callithrix</i>	15
2.1	Distribution of within-set and between-set mean correlations	30
2.2	MHI (a) and RV (b) values for localized trait sets	33
2.3	MHI (a) and RV (b) values for regional trait sets	34
2.4	Distribution of Modularity Hypothesis Index and RV Coefficient for theoretical correlation matrices	36
2.5	Type I error rates	37
2.6	Power for MHI and RV statistics for local shape variables	38
2.7	Power for MHI and RV statistics for interlandmark distances	39
2.8	Power for MHI and RV statistics for Procrustes residuals	41
3.1	Allometric shape variation	56
3.2	Deviations from ancestral allometry	57
3.3	Allometry and Integration	59
3.4	Modularity Hypothesis Indexes (MHI) <i>vs.</i> allometric slope deviations (b_s)	60
3.5	Direction of Multivariate Allometry in <i>Homo</i> and <i>Gorilla</i>	62
4.1	The relationship between Riemannian Distances and Random Skewers among pairs of P -matrices	72
4.2	Representation of the steps used to characterize covariance matrix variation	75
4.3	Phylogenetic decomposition of P -matrix variation	78
4.4	Posterior distribution of eigenvalues obtained for pPCs	80
4.5	Posterior bivariate distribution of variances explained by each phylogenetic PC <i>versus</i> Moran's Indexes estimated for each axis	81
4.6	Covariance structure variation associated with the first three and last pPCs (Global 1-3 and Local 1, respectively), represented using posterior mean SRD scores	82
4.7	Relationship between sample sizes and the first Local phylogenetic Principal Component . . .	83
4.8	Distribution of mean SRD scores for each trait along the entire ordered set of phylogenetic PCs	85

Lista de Tabelas

1.1	Vinte e dois marcos anatômicos registrados no crânio de antropóides	14
1.2	Trinta e oito distâncias euclidianas tomadas sobre os marcos anatômicos . .	17
2.1	Twenty-one species used in the present work	25
3.1	Model selection for phylogenetic linear models of AVG indexes over allo- metric parameters	58
4.1	Phylogenetic decomposition of covariance matrix structure diversity	79

Capítulo 1

Introdução

Modularidade é uma propriedade característica que sistemas biológicos apresentam em relação à distribuição de interações entre seus elementos constituintes; isto é, em um dado sistema, subconjuntos de elementos, denominados módulos, interagem entre si mais do que com outros conjuntos similares (Newman, 2006; Mitteroecker & Bookstein, 2007; Wagner *et al.*, 2007). Esta propriedade foi documentada em diferentes níveis de organização biológica, por exemplo em relação à dinâmica de redes metabólicas (e.g. Ravasz *et al.*, 2002; Andrade *et al.*, 2011) e em relação à estrutura de interações entre os indivíduos em populações (e.g. Fortuna *et al.*, 2008) e entre espécies em comunidades ecológicas (e.g. Genini *et al.*, 2010).

Em relação a sistemas morfológicos, o conceito de modularidade está associado ao quadro conceitual da integração morfológica (Olson & Miller, 1958; Cheverud, 1996a), que se refere à organização de covariâncias ou correlações entre elementos morfológicos e as hipóteses a respeito de suas relações. Neste contexto, a modularidade está relacionada com a distribuição desigual de efeitos genéticos sobre a variação fenotípica, articulada através do desenvolvimento (mapa genótipo/fenótipo; Wagner, 1996); sob a perspectiva da genética quantitativa, estes efeitos genéticos são o resultado de pleiotropia e desequilíbrio de ligação (Falconer & Mackay, 1996; Lynch & Walsh, 1998). Um mapa genótipo/fenótipo composto por grupos de genes que afetam grupos de caracteres apresenta portanto uma organização modular; neste contexto, tal estrutura de interações pleiotrópicas emerge como resultado de seleção para tarefas conflitantes (Wagner & Altenberg, 1996; Espinosa-Soto & Wagner, 2010; Rueffler *et al.*, 2012; Melo & Marroig, 2015).

De modo mais geral, caracteres fenotípicos em uma população evoluem conjuntamente seja devido a processos genéticos e de desenvolvimento compartilhados ou a efeitos comuns sobre a aptidão (Gould & Lewontin, 1979; Lande, 1979; Felsenstein, 1988; Zeng, 1988) e

a resposta à seleção natural é máxima quando ambos os efeitos coincidem, enquanto a discordância entre eles pode dificultar uma resposta direta (Schluter, 1996). Lande (1979) demonstrou que, assumindo uma distribuição gaussiana de efeitos alélicos, a diferença no fenótipo médio entre gerações em uma população sujeita a seleção direcional ($\Delta\bar{z}$) pode ser representada como

$$\Delta\bar{z} = \mathbf{G}\beta \quad (1.1)$$

onde $\mathbf{G}$ representa a matriz de covariâncias aditivas genéticas entre caracteres e β representa a direção de incremento máximo na aptidão, ou seja, o gradiente da paisagem adaptativa média estimado sobre o fenótipo médio ($\beta = \nabla \ln \bar{W}$); portanto, esta equação representa a resposta à seleção devido ao componente linear da paisagem adaptativa. Da mesma forma, as covariâncias aditivas expressas em $\mathbf{G}$ representam a dependência entre caracteres devido a pleiotropia e desequilíbrio de ligação (Cheverud, 1996a; Wagner, 1996), de modo que o mapa genótipo/fenótipo, nesta equação, é reduzido à sua aproximação linear centrada no fenótipo médio da população (Wagner, 1984).

A estrutura de interações pleiotrópicas depende da curvatura local do mapa genótipo/fenótipo, tradicionalmente representada em genética quantitativa na forma de dominância ou epistasia (Rice, 1998, 2004; Wolf *et al.*, 2001), ou seja, interações não lineares dentre alelos em um mesmo *locus* ou entre alelos em *loci* distintos, respectivamente. O efeito de epistasia na modulação de interações pleiotrópicas foi identificado em contextos experimentais (e.g. Cheverud *et al.*, 2004; Wolf *et al.*, 2005; Pavlicev *et al.*, 2008), indicando assim que populações podem abrigar variação genética para os componentes lineares do mapa genótipo/fenótipo. Covariâncias genéticas entre caracteres fenotípicos portanto podem evoluir como consequência de mudanças nas frequências alélicas nestes *loci*, por exemplo em resposta à deriva genética e efeito de fundador (Goodnight, 2000; Brito *et al.*, 2005). Características locais da paisagem adaptativa também podem afetar as covariâncias genéticas entre caracteres, visto que $\mathbf{G}$ reflete os padrões impostos por seleção estabilizadora e efeitos mutacionais (Lande, 1980; Cheverud, 1984; Jones, 2007). Vale ressaltar que mudanças no padrão de desequilíbrio de ligação entre diferentes *loci* são uma explicação suficiente para caracterizar mudanças nas covariâncias genéticas associadas à estrutura não-linear das paisagens adaptativas (Turelli, 1988), sem a necessidade de considerar a existência de pleiotropia epistática. No entanto, dados experimentais recentes (Careau *et al.*, 2015) e modelos baseados em simulação (Jones *et al.*, 2012; Melo & Marroig, 2015) demonstraram os efeitos da seleção direcional sobre a estrutura descrita por $\mathbf{G}$, indicando que o componente linear da paisagem adaptativa também pode afetar as covariâncias

genéticas dada a existência de variação genética nas interações pleiotrópicas.

A relação entre paisagens adaptativas e estrutura de covariância intrapopulacional é mediada através do desempenho (Arnold, 1983), que pode ser entendido como uma propriedade dinâmica do fenótipo; dessa forma, torna-se difícil a distinção entre interações ontogenéticas e funcionais na estrutura descrita por covariâncias fenotípicas (Cheverud, 1996a; Zelditch & Swiderski, 2011). Em relação ao crânio de mamíferos, tecidos precursores (mesoderme paraxial e crista neural) originam duas regiões distintas, Face e Neurocrânio, que exibem contrastes marcantes em termos das funções que executam, interações com tecidos moles e respostas a marcos de desenvolvimento, tais como o nascimento ou o desmame (Hallgrímsson & Lieberman, 2008; Lieberman, 2011). Assim, as regiões Oral, Nasal, e Zigomática estão associadas à Face e são responsáveis pelas funções de mastigação, respiração/olfação e fixação do aparato muscular envolvido na articulação mandibular, respectivamente. As regiões da Abóbada, Órbita e Basicrânio estão associadas ao Neurocrânio e são responsáveis por envolver e proteger cérebro e olhos, pelo suporte do crânio e por suas conexões com os demais sistemas do organismo; a Abóbada também abriga locais de ancoramento da musculatura associada à articulação da mandíbula, o que indica que, até certo grau, cada região pode estar envolvida em mais do que uma função e algumas funções são compartilhadas entre diferentes regiões.

Dessa forma, o reconhecimento de módulos variacionais (Wagner & Altenberg, 1996; Wagner *et al.*, 2007) utilizando padrões de covariância ou de correlação em populações adultas envolve um entendimento da dinâmica ontogénética/funcional subjacente aos caracteres morfológicos (Polly, 2008; Zelditch & Swiderski, 2011). O desenvolvimento do crânio em mamíferos ocorre em uma série de passos, tais como o crescimento do Neurocrânio induzido pelo crescimento do cérebro e o crescimento mediado por interações entre músculos e ossos, com uma relativa sobreposição espaço-temporal entre diferentes processos (Hallgrímsson & Lieberman, 2008; Herring, 2011; Cardini & Polly, 2013). Tais processos específicos resultam de padrões distintos de expressão gênica, os quais são subsequentemente alterados pela difusão de fatores de transcrição, gerando assim um processo de retroalimentação que produz diferenciação entre células e tecidos (Turing, 1952; Marcucio *et al.*, 2005; Meinhardt, 2008; Franz-Odenaal, 2011; Xu *et al.*, 2015). Esses fatores de transcrição afetam linhagens de células específicas, de forma que cada processo pode ser entendido como modular; no entanto, o contato entre tecidos adjacentes podem produzir alterações correlacionadas devido a interações mecânicas e também através de cascatas de sinalização mútua que induzem comportamentos específicos em tecidos vizinhos. Estas interações são necessárias para o desenvolvimento correto de ambos tecidos (Cheverud *et al.*, 1992; Ravosa *et al.*,

2000; Jiang *et al.*, 2002; Marcucio *et al.*, 2011).

Por exemplo, o crescimento da Abóbada, que ocorre nas fases finais de desenvolvimento pré-natal, é um resultado da tensão exercida pelo cérebro crescendo sobre suas membranas envoltórias, induzindo-as a secretar fatores de sinalização que promovem um padrão
5 complexo de deposição de tecido ósseo na porção dorsal da Abóbada e reabsorção dos mesmos na porção ventral (Opperman, 2000; Rice *et al.*, 2003; Lieberman, 2011). Ademais, o crescimento pós-natal da Face é induzido pela atividade muscular relacionada à mastigação, efeito que domina o crescimento pós-natal do crânio (Zelditch *et al.*, 1992; Herring, 2011). Tal atividade muscular afeta principalmente a região Oral em que são
10 exercidas forças musculares, mas outras regiões do crânio em que estes músculos estão ancorados, como o Zigomático e a Abóbada, também são afetadas em menor grau. Portanto, pode ser difícil isolar o efeito de cada processo individual tem sobre a estrutura de covariância, considerando-se tal sobreposição espacial (Hallgrímsson *et al.*, 2009); assim, tanto a hierarquia temporal quanto a organização espacial das interações ontogenéticas e
15 funcionais podem influenciar os padrões embutidos na estrutura de covariância genética. Ademais, estes aspectos moleculares de regulação do desenvolvimento motivam distintas definições de pleiotropia (Zelditch & Swiderski, 2011): uma em que o *locus* considerado afeta uma série de processos de desenvolvimento dentro de uma única linhagem de células (*efeito genético intrínscico*), e outra em que processos em uma linhagem de células dependem da
20 expressão do *locus* considerado em outra linhagem (*efeito epigenético extrínscico*); tal distinção enfatiza a estrutura espacial associada às interações pleiotrópicas e epistáticas.

A variação das taxas de crescimento pós-natal implica na emergência de variação em tamanho corporal (Pélabon *et al.*, 2013; Porto *et al.*, 2013) e tal fonte de variação tem uma importância particular no contexto dos sistemas morfológicos em mamíferos, afetando a
25 magnitude geral de correlações entre caracteres (Wagner, 1984; Young & Hallgrímsson, 2005). Tal magnitude de integração tem consequências importantes tanto para a evolução dos fenótipos médios (Schluter, 1996; Marroig & Cheverud, 2005, 2010; Cardini & Polly, 2013) quanto para a evolução da própria integração morfológica (Oliveira *et al.*, 2009b; Porto *et al.*, 2009, 2013; Shirai & Marroig, 2010). Variação no tamanho corporal entre diferentes
30 linhagens implicam em diferenças substanciais em aspectos fisiológicos e ecológicos de suas história de vida, visto que mudanças no tamanho corporal afetam a forma como os organismos interagem com seus ambientes (Gould, 1966; West *et al.*, 1997).

Relações de escala distintas entre diferentes caracteres e tamanho do corpo são denominadas alometria (Huxley, 1932; Gould, 1966; Cheverud, 1982b; Lande, 1985); tais relações podem
35 ser representadas como leis de potência, que se tornam lineares em escala logarítmica

(Huxley, 1932; mas ver Nijhout & German, 2012). No contexto explorado até o momento, alometria pode ser entendida como um tipo particular de relação entre caracteres entre duas escalas distintas de organização, visto que tamanho é um propriedade emergente do organismo como um todo, enquanto o impacto diferencial desta propriedade sobre diferentes partes do organismo produzem aquilo que reconhecemos como alometria (Young & Hallgrímsson, 2005), de modo que a variação de tamanho em conjunto com as relações alométricas é o componente que abriga grande parte da variação morfológica, especialmente em mamíferos (Wagner, 1984; Marroig & Cheverud, 2001).

Neste sentido, relações alométricas podem agir como restrição à evolução de sistemas morfológicos, desviando trajetórias evolutivas em populações naturais (Gould & Lewontin, 1979; Lande, 1979; Schluter, 1996). Por exemplo, seleção direcional para mudanças no tamanho corporal impulsionada por mudanças em hábitos alimentares dominou grande parte da diversificação de Platyrrhini, e diferenças na morfologia craniana entre as espécies deste grupo provavelmente são uma consequência indireta das relações alométricas entre caracteres cranianos e tamanho corporal (Marroig & Cheverud, 2005, 2010; Marroig *et al.*, 2012). Dentre as linhagens do grupo irmão Catarrhini, um cenário similar poderia explicar a diversidade morfológica no crânio em Cercopithecidae (Cardini *et al.*, 2007; Cardini & Elton, 2008). Em relação aos homínídeos, seleção direcional para a redução de elementos da Face e aumento da capacidade do Neurocrânio desempenhou um papel limitado na divergência entre *Homo* e *Australopithecus* (Ackermann & Cheverud, 2004; Schroeder *et al.*, 2014). No entanto, relações alométricas em uma população são dependentes da taxa de crescimento médio de seus indivíduos e também da covariância entre estas taxas de crescimento e o tamanho corporal atingido na idade adulta; assim, seleção sobre o tamanho corporal é capaz de alterar tais relações (Pélabon *et al.*, 2013; Voje *et al.*, 2013). Evidências experimentais sobre a evolução das relações alométricas (e.g. Frankino *et al.*, 2005; Tobler & Nijhout, 2010; Egset *et al.*, 2012; Bolstad *et al.*, 2015) indicam que estes parâmetros de crescimento abrigam variação genética e estão portanto sujeito aos efeitos da seleção natural; mudanças nestes parâmetros de crescimento então levam à mudanças nas relações alométricas na população.

A evidência empírica a respeito de mudanças a longo prazo na estrutura de covariância genética (genética quantitativa comparativa; Steppan *et al.*, 2002) leva em consideração a correspondência entre G e sua equivalente fenotípica P , uma vez que estimativas de G requerem tamanhos amostrais elevados e, mesmo assim, tal estimativa é propensa a erros, visto que as unidades amostrais destas estimativas são famílias em vez de indivíduos, o que diminui consideravelmente os tamanhos amostrais efetivos de estimativas de G (Meyer, 1991;

Houle & Meyer, 2015). A semelhança na estrutura de covariância entre **P** e **G** (“Conjectura de Cheverud”; Cheverud, 1988; Roff, 1995), foi corroborada em diferentes sistemas de caracteres e organismos (e.g.: Waitt & Levin, 1998; Dochtermann, 2011; Garcia *et al.*, 2014), indicando que os efeitos ambientais (**E**) possivelmente apresentam uma estrutura de correlações semelhantes a **G**, dado que estes efeitos exercem sua influência sobre fenótipos através dos mesmos processos de desenvolvimento sobre os quais a variação genética atua (Rice, 2002, 2004).

A análise comparativa da estrutura de covariância fenotípica mostra que **P** pode permanecer estável em escalas macroevolutivas (e.g: Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b; Kolbe *et al.*, 2011), e essa estabilidade pode ser uma consequência do alinhamento entre a estrutura de covariância fenotípica e as características locais da paisagem adaptativa que atua sobre diferentes linhagens. Por outro lado, a divergência em aspectos ecológicos entre linhagens irmãs pode reorganizar os padrões expressos em **P** em certos contextos (Young & Hallgrímsson, 2005; Young *et al.*, 2010; Monteiro & Nogueira, 2010; Haber, 2015), indicando mudanças na arquitetura subjacente do mapa genótipo/fenótipo. Portanto, os padrões incorporados na paisagem adaptativa podem superar as restrições impostas pelo desenvolvimento (Jamniczky & Hallgrímsson, 2009).

Por exemplo, a radiação adaptativa de morcegos da família Phyllostomidae possivelmente envolveu rearranjos na estrutura de covariância fenotípica mandibular (Monteiro & Nogueira, 2010), visto que a divergência em dieta neste grupo a partir de um ancestral insetívoro em direção a regimes alimentares mais especializados, tais como frugivoria ou hematofagia, implicam relações funcionais distintas entre os componentes mandibulares; assim, essa reorganização pode indicar heterogeneidade nos aspectos não-lineares da paisagem adaptativa dentre as diferentes linhagens deste grupo. Em morcegos em geral, a dissociação entre a função dos membros anteriores e posteriores está associada à modularização do mapa genótipo/fenótipo destas estruturas, evidenciada pela redução das correlações fenotípicas entre membros anteriores e posteriores e um aumento das correlações entre elementos constituintes de ambos os membros, em relação às demais linhagens de mamíferos (Young & Hallgrímsson, 2005). A linhagem dos homínídeos apresenta um padrão similar (Young *et al.*, 2010), na qual uma redução geral das correlações entre todos os elementos é observada em relação ao padrão observado nas demais linhagens de Anthroipoidea. Outro exemplo de divergência na estrutura de covariância entre linhagens devido a mudanças em exigências funcionais é aquela observada em ruminantes (Haber, 2015), relacionada ao aumento nas demandas metabólicas associada à ocupação de habitats abertos, visto que a divergência de Bovidae em relação à Cervidae está associada a uma maior integração na região Nasal.

Assim, a divergência entre linhagens irmãs pode impor a reorganização de ambas estruturas de covariância (genética e fenotípica) se tal divergência está associada a relações funcionais distintas, representadas pelos componentes não-lineares de paisagem adaptativa. Portanto, estabilidade ou divergência na estrutura de covariância em escalas macroevolutivas são ambos cenários possíveis, e a questão de qual cenário é regra ou exceção permanece aberta, 5
embora algumas diferenças teóricas e metodológicas entre diferentes autores no que diz respeito à representação de caracteres morfológicos precisam ser consideradas de modo a avaliar-se a plausibilidade destes cenários.

Morfometria

Tradicionalmente, caracteres morfológicos são medidos utilizando distâncias entre elementos definidos em termos gerais, como “comprimento do rostro” ou “largura do crânio”. 10
Pearson & Davin (1924) introduziu a noção de que as medidas devem ser restritas aos elementos osteológicos individuais, de preferência como distâncias entre elementos homólogos que poderiam ser identificados em linhagens distintas. Cheverud (1982a) associa tal noção ao âmbito da integração morfológica, considerando-se assim as medições sobre os 15
ossos individuais como representações locais de fenômenos regionais, isto é, as interações funcionais, ontogênicas e genéticas que produzem covariâncias entre esses elementos. A influência de tais interações sobre matrizes de covariância ou correlação podem ser avaliadas através da definição de um subconjunto de medidas que estão associadas a processos específicos, estimando-se uma métrica que resume tal particionamento; a hipótese 20
nula de que esta partição é indistinguível de partições aleatoriamente definidas pode então ser testada utilizando-se métodos de Monte Carlo, por exemplo (Mantel, 1967; Cheverud *et al.*, 1989).

Nas últimas três décadas, a morfometria geométrica (Bookstein, 1982, 1991; Kendall, 1984; Rohlf & Slice, 1990; Goodall, 1991) se estabelece como um quadro conceitual para a 25
representação da forma biológica como configurações geométricas de marcos anatômicos homólogos (*landmarks*; Bookstein, 1991). Dois princípios são fundamentais nesta abordagem: em primeiro lugar, a separação conceitual e estatística do tamanho e da forma como componentes distintos da variação morfológica; em segundo lugar, a utilização de métodos baseados em superposição (GPA: Análise Generalizada de Procrustes; Rohlf & Slice, 1990; 30
Goodall, 1991) para a estimativa de parâmetros representativos de tal variação, como média e matriz de covariância. Estimadores de Procrustes foram propostos como uma solução para o problema de identificabilidade (Neyman & Scott, 1948), isto é, tanto a posição e os ângulos de rotação entre diferentes configurações são impossíveis de estimar na ausência

de quaisquer premissas (Goodall, 1991; Lele & McCulloch, 2002).

Embora o uso de estimadores de Procrustes seja amplamente difundido na morfometria geométrica, parte devido à distribuição de pacotes computacionais dedicados ao tratamento estatístico deste tipo de representação (e.g., Klingenberg, 2011; Adams & Otárola-Castillo, 2013), críticas foram formuladas tanto em relação à estimativa de formas médias (por exemplo, Lele, 1993; Kent & Mardia, 1997; Huckemann, 2012) e matrizes de covariância entre coordenadas (Walker, 2000; Adams *et al.*, 2004; Linde & Houle, 2009; Márquez *et al.*, 2012). Para configurações no espaço bidimensional, estimadores de Procrustes para a forma média apresentam um bom desempenho sob a premissa de estrutura de covariância isotrópica (Kent & Mardia, 1997), em que considera-se que covariâncias entre coordenadas ou marcos distintos são nulas; no entanto, quando esta premissa não se sustenta, estimadores de Procrustes para a forma média perdem tanto precisão quanto acurácia, especialmente no caso em que a variação de forma é alta (Huckemann, 2011). O exemplo apresentado por Linde & Houle (2009) demonstra que, quando tal suposição é quebrada, o padrão de covariâncias entre coordenadas também são mal estimados; se a matriz de covariância de forma desconhecida está estruturada devido a diferenças regionais em processos responsáveis pela geração de covariâncias, essa variação pode se distribuir por toda a configuração, efetivamente apagando o sinal dos processos funcionais e ontogenéticos subjacentes.

Diversas alternativas para a estimativa de matrizes de covariância representativas da variação de forma já foram sugeridas; algumas destas alternativas (e.g. Monteiro *et al.*, 2005; Theobald & Wuttke, 2006; Linde & Houle, 2009; Zelditch *et al.*, 2009), propõem alterações à Análise Generalizada de Procrustes, a fim de lidar com a heterogeneidade na estrutura de covariância dentre marcos anatômicos. Márquez *et al.* (2012) propôs uma outra definição de descritores de forma usando técnicas baseadas em interpolação (Cheverud & Richtsmeier, 1986; Bookstein, 1989) como ponto de partida. Estes descritores (denominados na presente tese como variáveis locais de forma) referem-se a expansões ou retrações infinitesimais em referência a uma forma média desconhecida (Woods, 2003), estimada em locais definidos em meio aos marcos anatômicos representados na amostra. Os autores argumentam que tais descritores são mensurações adequadas de variação local de forma, de modo que estes podem ser diretamente ligados aos processos biológicos que geram covariância entre elementos morfológicos.

Apesar destes problemas, o uso de estimadores de Procrustes para investigar aspectos da integração morfológica aumentou nos últimos dez anos (e.g. Klingenberg *et al.*, 2004; Drake & Klingenberg, 2010; Goswami & Polly, 2010; Martínez-Abadías *et al.*, 2011; Sanger

et al., 2012), e os resultados encontrados por alguns destes autores se contrastam de maneira gritante com trabalhos similares utilizando distâncias entre marcos anatômicos (e.g, Cheverud *et al.*, 1997; Oliveira *et al.*, 2009b; Porto *et al.*, 2009). Por exemplo, Martínez-Abadías *et al.* (2011) encontrou um padrão de integração forte entre partições nos padrões de covariância do crânio de *Homo*, enquanto Oliveira *et al.* (2009b) e Porto *et al.* (2009) 5 demonstraram que humanos estão entre os exemplos mais modulares de padrões de covariância no crânio de mamíferos. Da mesma forma, enquanto Cheverud *et al.* (1997) mostrou que 70% dos efeitos pleiotrópicos estão confinados aos componentes mandibulares anteriores ou posteriores, Klingenberg *et al.* (2004) não encontrou nenhuma evidência 10 para uma distribuição modular de efeitos pleiotrópicos entre as partições da mandíbula, utilizando a mesma amostra.

Objetivos

A presente tese tem como objetivo inicial investigar a relação entre as representações utilizadas para caracterizar variação morfológica e as inferências feitas a partir destas representações a respeito de propriedades variacionais em escala macroevolutiva. Em 15 particular, tal relação é investigada considerando-se a morfologia craniana em Primatas da infraordem Anthropoidea, dada a disponibilidade de uma ampla base de dados (Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b), sob a forma de coordenadas tridimensionais de marcos anatômicos, das quais podem ser derivadas outras representações. Explicitamente, consideram-se três representações distintas: distâncias euclidianas entre marcos anatômi- 20 cos, configurações geométricas destes marcos (utilizando-se estimadores de Procrustes), e variáveis locais de forma, estimadas como expansões ou retrações infinitesimais a partir de deformações de placa fina. Incidentalmente, a presente tese também versa sobre a evolução de propriedades variacionais em Anthropoidea, a partir de um entendimento de como estas diferentes representações articulam diferentes aspectos da estrutura de covariância 25 subjacente aos caracteres morfológicos.

Perfil da Tese

O Capítulo 2 (*Modularity and Morphometrics: Error Rates in Hypothesis Testing*) compara três tipos de representações distintas no que diz respeito a testes de hipótese de detecção de módulos variacionais a partir de matrizes de covariância construídas, de modo a
5 estimar taxas de falsos positivos e negativos. Tais estimativas balizam as comparações entre matrizes de covariância obtidas a partir dos dados empíricos, representados de diferentes maneiras, considerando-se também a relação entre estes padrões obtidos para as diferentes espécies e o conhecimento atual a respeito do desenvolvimento craniano em primatas e em mamíferos em geral. Este capítulo demonstra que, considerando-se
10 as diferenças inerentes às representações, distâncias entre marcos e variáveis locais de forma apresentam resultados similares em relação à detecção de módulos variacionais; tais resultados também são condizentes com o entendimento atual a respeito da ontogênese craniana em mamíferos, enquanto estimadores de Procrustes não capturam o padrão de estruturação modular presente na variação morfológica craniana. Testes de hipótese
15 sobre esta última representação apresentam taxas elevadas de falsos negativos em relação às outras, indicando uma possível razão para estes problemas de detecção de módulos variacionais.

O Capítulo 3 (*The Evolution of Skull Allometry in Anthropoid Primates*) se dedica a explorar as mudanças nas relações alométricas em Anthropoidea, considerando um tratamento
20 explicitamente filogenético baseado em modelos mistos para as estimativas de parâmetros alométricos dentre as diferentes espécies desta linhagem. Considerando-se os resultados obtidos no capítulo anterior e as diferenças entre distâncias entre marcos anatômicos e variáveis locais de forma, este capítulo demonstra que os padrões de modularidade observados para distâncias entre marcos são uma consequência da estruturação imposta
25 pelas relações alométricas em Anthropoidea. Em particular, os exemplos mais extremos de modularidade variacional nesta linhagem, *Homo* e *Gorilla*, estão associados a mudanças significativas nas relações alométricas nestas duas linhagem em relação ao padrão ancestral de associação entre forma e tamanho.

O Capítulo 4 (*A Phylogenetic Analysis of Shape Covariance Structure*) tenta abordar a questão
30 da estabilidade dos padrões de covariância fenotípica em Anthropoidea sob uma perspectiva distinta daquela baseada em testes de hipótese, focada na distribuição filogenética dos padrões de covariância de tamanho e forma, esta última representada por variáveis locais de forma, de modo a prover uma reavaliação dos resultados obtidos anteriormente para Platyrrhini (Marroig & Cheverud, 2001) e Catarrhini (Oliveira *et al.*, 2009b). Neste contexto,
35 a estruturação da variação morfológica craniana nas regiões avaliadas no primeiro capítulo

como módulos variacionais através de testes de hipótese *a priori* é incorporada *a posteriori*, como um quadro de referência sobre a qual a estabilidade da estrutura de covariância destes descritores é interpretada. A partir de uma análise baseada na decomposição da variação em matrizes de covariância dentre as diferentes linhagens de Anthroipoidea, considerando-se a estruturação filogenética subjacente ao grupo, observa-se que as principais diferenças no padrão de covariância entre linhagens em Anthroipoidea estão confinadas a um subconjunto dos descritores de forma, localizados na regiões do Basicrânio e Órbita; ademais, a relação entre tamanho e forma craniana, representativa das relações alométricas, também se mostra instável em escala macroevolutiva. Este padrão repete-se nos diferentes eventos de cladogênese que compõem a diversidade de Anthroipoidea, e a instabilidade associada a estas regiões reflete o padrão de interações entre tecidos no desenvolvimento craniano. Os resultados deste capítulo refletem em certo grau aqueles observados no primeiro capítulo, onde aquelas regiões detectadas como módulos variacionais (Oral, Nasal e Abóbada) apresentam uma estrutura de covariância estável, e atribui-se tal estabilidade à manutenção de interações funcionais e ontogenéticas ao longo da diversificação de Anthroipoidea.

Objeto de Estudo: Anthroipoidea

A infraordem Anthroipoidea originou-se no Eoceno da Ásia, com uma rápida expansão e subsequente diversificação na África (Marivaux *et al.*, 2014) e na América do Sul (Bond *et al.*, 2015), ainda durante o Eoceno. Este grupo diferencia-se em relação aos clados mais basais de Primatas pela fusão completa dos frontais bilateralmente simétricos em um único osso, pela fusão das hemimandíbulas na região da sínfise mandibular, e pela formação de uma cavidade pós-orbital em contrapartida à barra pós-orbital encontrada nos Strepsirrhini, que se encontra parcialmente formada no grupo irmão vivente de Anthroipoidea, o gênero *Tarsius*. O surgimento desta última estrutura está provavelmente associado à manutenção de um campo visual estável e direcionado à frente, mesmo durante a alimentação, considerando seu papel na distribuição das forças de mastigação ao redor dos olhos (Ravosa *et al.*, 2000; Ravosa & Savakova, 2004).

A diversidade atual de antropóides se encontra dividida em Platyrrhini, os Macacos do Novo Mundo, presentes atualmente na região intertropical das Américas, e Catarrhini, um grupo que se distribui ao longo da África Subsaariana, no Subcontinente Indiano e também no Sudeste Asiático (Fleagle, 2013). A ocupação destas regiões por ambas as linhagens se deu de forma bastante distinta: enquanto hipotetiza-se que a colonização da América do Sul pelos Platyrrhini envolveu um único evento, possivelmente a partir da África, através da sucessiva colonização de ilhas distribuídas ao longo do Atlântico Sul durante o Paleógeno

(Oliveira *et al.*, 2009a), a diversidade atual de Catarrhini se origina na África, a partir da qual diferentes linhagens colonizaram a Ásia em eventos independentes (Springer *et al.*, 2012).

De modo geral, antropóides apresentam dietas generalistas, apesar da preferência por itens alimentares distintos ser variável dentre as diferentes linhagens (Fleagle, 2013). Dentre os Macacos do Novo Mundo, mudanças na preferência por certos itens alimentares na dieta parece ter sido a razão principal por trás da evolução de tamanho corporal nesta linhagem (Marroig & Cheverud, 2005, 2010), considerando-se que uma dieta predominantemente folívora impõe restrições fisiológicas distintas daquelas impostas por uma dieta frugívora ou insetívora. Apenas um clado particular, Colobinae, possui especializações para uma dieta estritamente folívora, na forma de um trato digestório similar aquele encontrado em ruminantes. À exclusão de um gênero Neotropical (*Aotus*), todas as espécies desta linhagem apresentam hábitos diurnos; aliada à acuidade visual característica do grupo, tal prevalência de hábitos diurnos permite aos antropóides apresentarem uma ampla diversidade de estruturação social, tanto em relação à composição dos grupos quanto à dinâmica espacial e temporal da associação entre indivíduos (Kappeler & Schaik, 2002; Mitani *et al.*, 2012; Kamilar & Baden, 2014).

Amostra

A amostra utilizada aqui é composta por 5108 indivíduos pertencentes a 109 espécies, distribuídas em todas as linhagens de Anthropeidea acima do nível do gênero, abrangendo todos os gêneros de Platyrrhini e uma parte substancial dos gêneros de Catarrhini. Associa-se a este banco de dados uma hipótese filogenética ultramétrica para Anthropeidea (Figura 1.1), derivada de Springer *et al.* (2012).

Os indivíduos nesta amostra são representados por 36 marcos anatômicos, registrados utilizando um digitalizador de marca Polhemus 3Draw para os representantes de Platyrrhini, e um digitalizador de marca MicroScribe 3DS para os representantes de Catarrhini. Vinte e dois marcos únicos representam cada indivíduo (Figura 1.2, Tabela 1.1), uma vez que 14 dos 36 marcos registrados são bilateralmente simétricos; para mais detalhes sobre o registro de marcos anatômicos, ver Marroig & Cheverud (2001) e Oliveira *et al.* (2009b). As bases de dados destes dois trabalhos anteriores foram reunidas em uma única base de dados, mantendo apenas aqueles indivíduos em que todos os marcos anatômicos de ambos os lados se encontravam presentes.

Figura 1.1: Tamanhos amostrais e efeitos fixos controlados para as 109 OTUs de primatas antropóides.

Tabela 1.1: Vinte e dois marcos anatômicos registrados no crânio de antropóides. As regiões e sub-regiões cranianas às quais cada marco anatômico pertence também estão indicadas. A vista na qual cada marco foi registrado está indicada por A (anterior), P (posterior) ou AP (ambas). As descrições em itálico correspondem a marcos anatômicos tomados sobre a linha sagital.

Marco	Posição	Região	Sub-região	Descrição
IS	A	Face	Oral/Nasal	<i>interincisivo superior</i>
PM	A	Face	Oral	sutura maxilar-premaxilar
NSL	A	Face	Nasal/Oral	<i>extremidade rostral do osso nasal</i>
NA	A	Face	Nasal/Órbita	<i>extremidade caudal do osso nasal (junção com o osso frontal)</i>
BR	AP	Neurocrânio	Abóbada	<i>sutura entre o frontal e o parietal (posição medial)</i>
PT	AP	Neurocrânio	Abóbada	sutura entre o frontal e o parietal
FM	A	Face	Zigomático/Órbita	fronto-malar
ZS	A	Face	Oral/Zigomático/Órbita	sutura superior entre o zigomático e o maxilar
ZI	A	Face	Zigomático	sutura inferior entre o zigomático e o maxilar
MT	A	Face	Oral	tuberosidade maxilar caudal ao terceiro molar
PNS	A	Face	Oral	<i>espinha caudal do nasal</i>
APET	A	Neurocrânio	Base	petrous temporal
BA	AP	Neurocrânio	Base	<i>ponto ventral do forame magno</i>
OPI	AP	Neurocrânio	Base/Abóbada	<i>ponto dorsal do forame magno</i>
EAM	A	Neurocrânio	Zigomático/Abóbada	meato auditivo externo rostral
PEAM	A	Neurocrânio	Abóbada	meato auditivo externo caudal
ZYGO	A	Face	Zigomático	sutura entre o zigomático e o temporal
TSP	A	Neurocrânio	Zigomático/Base/Abóbada	sutura entre o temporal o esfenoidal e o frontal
TS	AP	Neurocrânio	Base	sutura entre o temporal e o esfenoidal
JP	AP	Neurocrânio	Base	processo jugular
LD	P	Neurocrânio	Abóbada	<i>sutura entre o occipital e o interparietal na linha média</i>
AS	P	Neurocrânio	Abóbada	sutura entre o parietal e o occipital

Representações de Variação Morfológica

Considera-se ao longo da presente tese três tipos distintos de variáveis descritoras de forma: resíduos de espaço tangente, estimados a partir de uma superposição de Procrustes para toda a amostra, utilizando-se o conjunto de marcos anatômicos (doravante denominados
5 resíduos de Procrustes; Tabela 1.1 e Figura 1.2); distâncias euclidianas entre marcos anatômicos (Tabela 1.2 e Figura 1.2); e variáveis locais de forma (Márquez *et al.*, 2012), que representam medidas de transformações infinitesimais de volume em escala logarítmica entre cada unidade de amostra e uma forma de referência, com base em uma função de interpolação que estima a variação de forma entre os marcos anatômicos amostrados.

10 Utiliza-se como função de interpolação para a estimativa deste tipo de variável deformações de placa fina (*thin plate splines*; Bookstein, 1989), que considera o comportamento dos tecidos adjacentes aos marcos anatômicos similar ao comportamento de uma placa metálica de espessura infinitesimal sob tensão. Assim sendo, a variação de cada indivíduo é descrita por esta função como uma deformação de toda a configuração em relação à forma de
15 referência, de modo que esta função associa qualquer posição arbitrária, representada por suas coordenadas tridimensionais na forma de referência à uma posição equivalente na unidade amostral considerada, também representada por três coordenadas. As variáveis locais de forma então referem-se às expansões ou retrações locais, estimadas como sendo o logaritmo natural do determinante do jacobiano da função de interpolação, estimado
20 em pontos de referência escolhidos a partir de algum critério; jacobianos são matrizes que estendem o conceito de derivada às funções multivariadas, ou seja, descrevem a taxa de variação nas diferentes direções definidas pelo domínio de tal função, de modo que determinantes destas matrizes descrevem a taxa de variação no volume em uma posição arbitrária dentre os marcos anatômicos amostrados.

25 Para cada unidade amostral foram obtidas 38 destas variáveis, correspondentes às localizações dos pontos médios entre os pares de marcos anatômicos utilizados para definir as distâncias entre marcos (Figura 1.2 e Tabela 1.2), a fim de produzir um conjunto de dados que representa forma (isto é, desprovido de variação isométrica; Bookstein, 1991; Zelditch *et al.*, 2004) enquanto certas propriedades globais do conjunto de dados de distâncias entre
30 marcos são mantidas, como a dimensionalidade e a correspondência destes descritores com as hipóteses de relacionamento derivadas de interações funcionais e ontogenéticas, por exemplo.

Ambas representações de forma (resíduos de Procrustes e variáveis locais de forma) são associadas a uma mensuração de tamanho isométrico, na forma do logaritmo do tamanho

Tabela 1.2: Trinta e oito distâncias euclidianas tomadas sobre os marcos anatômicos. As regiões e sub-regiões cranianas às quais cada caráter pertence também são indicadas.

Distância	Região	Sub-região
IS.PM	Face	Oral
IS.NSL	Face	Nasal
IS.PNS	Face	Oral/Nasal
PM.ZS	Face	Oral
PM.ZI	Face	Oral
PM.MT	Face	Oral
NSL.NA	Face	Nasal
NSL.ZS	Face	Nasal
NSL.ZI	Face	Oral/Nasal
NA.BR	Neurocrânio	Abóbada
NA.FM	Neurocrânio	Órbita
NA.PNS	Face	Nasal
BR.PT	Neurocrânio	Abóbada
BR.APET	Neurocrânio	Abóbada
PT.FM	Neurocrânio	Órbita
PT.APET	Neurocrânio	Abóbada
PT.BA	Neurocrânio	Abóbada
PT.EAM	Neurocrânio	Abóbada
PT.ZYGO	Face	Zigomático
FM.ZS	Neurocrânio	Órbita
FM.MT	Face	Zigomático
ZS.ZI	Face	Oral
ZI.MT	Face	Oral
ZI.ZYGO	Face	Zigomático
ZI.TSP	Face	Zigomático
MT.PNS	Face	Oral
PNS.APET	Neurocrânio	Base
APET.BA	Neurocrânio	Base
APET.TS	Neurocrânio	Base
BA.EAM	Neurocrânio	Base
EAM.ZYGO	Face	Zigomático
ZYGO.TSP	Face	Zigomático
LD.AS	Neurocrânio	Abóbada
BR.LD	Neurocrânio	Abóbada
OPI.LD	Neurocrânio	Abóbada
PT.AS	Neurocrânio	Abóbada
JP.AS	Neurocrânio	Base
BA.OPI	Neurocrânio	Base

do centróide (logCS), que possui a propriedade de ser uma mensuração de tamanho que é de fato independente de variação de forma (Bookstein, 1991).

Na presente tese, considerou-se apenas aquelas propriedades associadas ao componente simétrico de variação; portanto, antes de qualquer análise, os efeitos de variação na
5 assimetria foram controlados seguindo-se o procedimento descrito em Klingenberg et al. (2002) para estruturas bilaterais, através da obtenção de uma configuração simétrica de marcos anatômicos para cada indivíduo, estimando-se uma média entre cada forma real e sua reflexão ao longo do plano sagital. Para cada conjunto de dados, também foram removidos efeitos de pouco interesse no contexto desta tese, como dimorfismo sexual,
10 por exemplo. Para distâncias euclidianas entre marcos anatômicos, variáveis locais de forma e também para o logCS estes efeitos foram removidos através de modelos lineares multivariados ajustados separadamente para cada espécie, de acordo com Tabela 2.1; para os resíduos de Procrustes, os mesmos efeitos foram removidos centralizando-se todas as médias dos fatores considerado sobre a forma média de cada espécie, visto que a perda
15 de graus de liberdade imposta pela GPA impede o uso de modelos lineares multivariados sobre esse tipo de representação com o intuito de remover efeitos fixos.

Chapter 2

Modularity and Morphometrics: Error Rates in Hypothesis Testing

Abstract

The study of modularity in morphological systems has increased in the past twenty years, 5
parallel to the popularization of geometric morphometrics, which has led to the emergence
of different criteria for detecting modularity on landmark data. However, compared to
usual covariance matrix estimators, Procrustes estimators have properties that hinder their
use. Here, we compare different representations of form, focusing on detecting modularity
patterns defined *a priori*; we also compare two metrics: one derived from traditional 10
morphometrics (MHI) and another that emerged in the context of landmark data (RV).
Using Anthropoid skulls, we compare these metrics over three representations of form:
interlandmark distances, Procrustes residuals, and local shape variables. Over Procrustes
residuals, both tests fail to detect modularity patterns, while in remaining representations
they show the distinction between early and late development in skull ontogeny. To estimate 15
type I and II error rates, we built covariance matrices of known structure; these tests
indicate that, considering both effect and sample sizes, tests using MHI are more robust
than those using RV. However, both metrics have low statistical power when used on
Procrustes residuals; thus, we conclude that the influence of development and function is
poorly represented on Procrustes estimators for covariance matrices. 20

Introduction

Modularity is a characteristic property that biological systems exhibit regarding the distribution of interactions between their composing elements; that is, in a given system, subsets of elements, denominated modules, interact more among themselves than with other such subsets (Newman, 2006; Mitteroecker & Bookstein, 2007; Wagner *et al.*, 2007). This property has been well documented at different levels of biological organization, from the dynamics of metabolic networks (e.g. Ravasz *et al.*, 2002; Andrade *et al.*, 2011) to the structure of interactions among individuals in populations (e.g. Fortuna *et al.*, 2008) and among species in ecological communities (e.g. Genini *et al.*, 2010).

Regarding morphological systems, the concept of modularity is associated with the framework of morphological integration (Olson & Miller, 1958; Cheverud, 1996a), which refers to the organization of covariances or correlations between morphological elements and the hypotheses concerning their relationships. In this context, modularity is related to the uneven distribution of genetic effects over phenotypic variation articulated through development (genotype/phenotype map; Wagner, 1996); in a classical quantitative genetics view, these genetic effects are the result of pleiotropy and linkage disequilibrium (Falconer & Mackay, 1996; Lynch & Walsh, 1998). A genotype/phenotype map composed of clusters of genes that affect clusters of traits (with little overlap) exhibits a modular organization; this structure is thought to emerge as the result of selection for distinct demands (Wagner & Altenberg, 1996; Espinosa-Soto & Wagner, 2010; Rueffler *et al.*, 2012; Melo & Marroig, 2015). For instance, the decoupling between fore- and hindlimb function in certain mammalian lineages such as bats (Young & Hallgrímsson, 2005) and apes (Young *et al.*, 2010) is associated with the modularization of both structures, as shown by reduced phenotypic correlations between fore- and hindlimbs and increased correlations between elements within these limbs.

The recognition of variational modules (Wagner & Altenberg, 1996; Wagner *et al.*, 2007) using covariance or correlation patterns in adult populations involves an understanding of the underlying developmental and functional dynamics among morphological traits (Polly, 2008; Zelditch & Swiderski, 2011). Skull development in mammals occurs through a series of steps, such as neurocranial growth induced by brain development, and growth mediated by muscle-bone interactions, with spatiotemporal overlapping between such steps (Hallgrímsson & Lieberman, 2008; Herring, 2011; Cardini & Polly, 2013). Both timing and scope of each step is regulated by different profiles of genetic expression exhibited by cells originated from different embryonic precursors and their response to signaling factors expressed at the regional level; the response to these signaling factors further changes

cell expression profiles, thus generating a feedback loop of differentiation (Turing, 1952; Marcucio *et al.*, 2005; Meinhardt, 2008; Hallgrímsson *et al.*, 2009; Franz-Odenaal, 2011; Minelli, 2011). Each step in this temporal hierarchy may be regarded as modular, since they affect a coherent subset of tissues more so than others (Hallgrímsson & Lieberman, 2008), although each step affect adjoining regions through interactions among developing tissues (Cheverud *et al.*, 1992; Lieberman, 2011; Esteve-Altava & Rasskin-Gutman, 2014); thus, the overlapping of such processes throughout development may complicate their association with correlation patterns (Hallgrímsson *et al.*, 2009).

Furthermore, variation in growth rates, which emerges into size variation (Pélabon *et al.*, 2013; Porto *et al.*, 2013), has a particular importance in the context of mammalian morphological systems. Here size variation refers to variation in both scale (isometric variation) and scale relationships (allometric variation). This source of variation affects the overall level of correlations between morphological traits (Wagner, 1984; Young & Hallgrímsson, 2005), and the magnitude of integration has important consequences for both the evolution of mean phenotypes (Schluter, 1996; Marroig & Cheverud, 2005, 2010; Cardini & Polly, 2013) and the evolution of morphological integration itself (Oliveira *et al.*, 2009b; Porto *et al.*, 2009, 2013; Shirai & Marroig, 2010).

In this context, adaptive landscapes may be the central component governing both the stability and divergence in integration patterns, as both stabilizing and directional selection have been shown to produce changes in integration patterns (Jones, 2007; Arnold *et al.*, 2008; Jones *et al.*, 2012; Melo & Marroig, 2015). The empirical evidence available demonstrates that both stability (e.g. Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b; Porto *et al.*, 2009; Willmore *et al.*, 2009) and divergence (e.g. Monteiro & Nogueira, 2010; Grabowski *et al.*, 2011; Sanger *et al.*, 2012; Haber, 2015) of integration patterns are possible outcomes of the evolutionary process. Therefore, the question of whether any of these two scenarios is the rule or exception at macroevolutionary scales remains open, although some theoretical and methodological differences between these works with respect to the representation of morphological features need to be considered.

Morphometrics

Traditionally, morphological features have been measured using distances among elements defined in general terms, such as “cranial length” or “cranial width”. Pearson & Davin (1924) introduced the notion that measurements should be restricted to single osteological elements, preferably as distances between homologous features that could be identified in a wide taxonomic coverage. Cheverud (1982a) accomodates this notion into the framework

of morphological integration, thus considering individual measurements over single bones as local representations of regional phenomena, that is, the functional, developmental and genetic interactions that produce covariances among these elements. The influence of such interactions over covariance or correlation matrices can be accessed by defining a subset of measurements which are associated with particular processes and estimating a metric that summarizes such partitioning; the null hypothesis that this partition is undistinguishable from randomly-defined partitions can then be tested using Monte Carlo methods (Mantel, 1967; Cheverud *et al.*, 1989).

In the past three decades, geometric morphometrics (Bookstein, 1982, 1991; Kendall, 1984; Rohlf & Slice, 1990; Goodall, 1991) has been consolidated as a quantitative framework for the representation of biological shape as geometric configurations of homologous features (landmarks; Bookstein, 1991). Two principles are central here: first, the conceptual and statistical separation of size and shape as components of biological form; second, the use of superimposition-based methods (GPA: Generalized Procrustes Analysis; Rohlf & Slice, 1990; Goodall, 1991) for the estimation of shape statistical parameters, such as mean shape and shape covariance structure. Procrustes estimators were proposed as a solution to the problem that landmark configurations are arbitrarily rotated and translated; such nuisance parameters are impossible to estimate without any assumptions (Goodall, 1991; Lele & McCulloch, 2002), a situation known as the identifiability problem (Neyman & Scott, 1948).

Although the use of Procrustes estimators is currently widespread in geometric morphometrics toolboxes (e.g. Klingenberg, 2011; Adams & Otárola-Castillo, 2013), critics have been made regarding the estimation of both mean shape configurations (e.g. Lele, 1993; Kent & Mardia, 1997; Huckemann, 2012) and shape covariance matrices (Walker, 2000; Adams *et al.*, 2004; Linde & Houle, 2009; Márquez *et al.*, 2012). For configurations in two-dimensional space, Procrustes estimators for mean shape perform well under isotropic landmark covariance structure (Kent & Mardia, 1997), a situation of null covariances among landmarks and coordinates; however, when this assumption does not hold, Procrustes estimates for mean shape may lose either precision and accuracy, especially when shape variation is high (Huckemann, 2011). The example provided by Linde & Houle (2009) demonstrates that when such assumption is broken shape covariance patterns are also poorly estimated; if the unknown landmark covariance matrix is structured due to regional differences in covariance-generating processes, such variation might be displaced and effectively spread out through the entire landmark configuration.

A number of alternatives for estimating shape covariance matrices have already been suggested; some of these alternatives (e.g. Monteiro *et al.*, 2005; Theobald & Wuttke,

2006; Linde & Houle, 2009; Zelditch *et al.*, 2009) propose modifications to the Procrustes analysis in order to deal with heterogeneity in landmark covariance structure. Márquez *et al.* (2012) propose another definition of shape descriptors using interpolation-based techniques (Cheverud & Richtsmeier, 1986; Bookstein, 1989) as a starting point. These descriptors refer to infinitesimal expansions or retractions in reference to a unknown mean shape (Woods, 2003) estimated at definite locations amidst sampled landmarks. The authors argue that such descriptors are proper local measurements of shape variation, as they can be directly linked to biological processes that generate covariation among morphological elements. 5

Despite these caveats, the use of Procrustes estimators for investigating aspects of morphological integration has increased in the past ten years (e.g. Klingenberg *et al.*, 2004; Drake & Klingenberg, 2010; Goswami & Polly, 2010; Martínez-Abadías *et al.*, 2011; Sanger *et al.*, 2012). The results found by these authors are sometimes in stark contrast with similar works using interlandmark distances (e.g. Cheverud *et al.*, 1997; Oliveira *et al.*, 2009b; Porto *et al.*, 2009). For instance, Martínez-Abadías *et al.* (2011) found a pattern of strong integration among partitions in human skull covariance patterns, while Oliveira *et al.* (2009b) and Porto *et al.* (2009) demonstrated that humans are among the most modular examples of mammalian skull covariance patterns. Likewise, while Cheverud *et al.* (1997) had shown that 70% of the pleiotropic effects are confined to either anterior or posterior mandibular components, Klingenberg *et al.* (2004) has found no evidence for a modular distribution of pleiotropic effects among the partitions of the mouse mandible using the same strain of intercrossed mice at the same generation. Furthermore, aside from the issues regarding Procrustes estimators, these works also propose different methods to quantify the effects of interactions over covariance patterns. For example, Martínez-Abadías *et al.* (2011) and Sanger *et al.* (2012) use the RV coefficient, a multivariate correlation coefficient defined by Escoufier (1973) which has been used to quantify modular relationships over landmark covariance patterns since Klingenberg (2009) has proposed its use in this context. 10 15 20 25

Objectives

In the present work, we compare the methods described by Cheverud *et al.* (1989) and Klingenberg (2009) to test *a priori* defined modularity patterns using Anthropoid Primates as a model organism. In order to compare the performance of these methods with respect to different representations of form, individuals in our sample are represented both as interlandmark distances and shape variables. Furthermore, we used an approach based on the construction of theoretical covariance matrices; such matrices are used in order to estimate Type I and Type II error rates for both methods. This work aims to analyze both 30

methods, which have been proposed in different contexts, under the same conceptual and statistical framework, in order to produce meaningful comparisons.

Methods

Sample

5 The database we used here (Table 2.1) consists of 21 species, distributed across all major Anthropoid clades above the genus level. We selected these operational taxonomic units (OTU) from a broader database (Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b) in order to reduce the effects of low sample sizes over estimates of modularity patterns. Individuals in our sample are represented by 36 registered landmarks, measured using a Polhemus 3Draw
10 (for Platyrrhini) and a Microscribe 3DS (for Catarrhini). Twenty-two unique landmarks represent each individual (Figure 1.2, Table 1.1), since 14 of the 36 registered landmarks are bilaterally symmetrical. For more details on landmark registration, see Marroig & Cheverud (2001) and Oliveira *et al.* (2009b).

For each OTU, we estimated phenotypic covariance and correlation matrices for three
15 different types of variables: tangent space residuals, estimated from a Procrustes superimposition for the entire sample, using the set of landmarks described on both Table 1.1 and Figure 1.2 (henceforth Procrustes residuals); interlandmark distances, described in Table 1.2; and local shape variables (Márquez *et al.*, 2012), which are measurements of infinitesimal log volume transformations between each sample unit and a reference (mean) shape, based
20 upon an interpolation function that describes shape variation between sampled landmarks. In this context, we used thin plate splines as interpolating functions (Bookstein, 1989). We obtained 38 transformations corresponding to the locations of the midpoints between pairs of landmarks used to define interlandmark distances, in order to produce a dataset that represents shape (i.e., form without isometric variation; Bookstein, 1991; Zelditch *et al.*,
25 2004) while retaining the overall properties of the interlandmark distance dataset, such as dimensionality for example. Furthermore, we were able to use the same hypotheses of trait associations for both types of variables since the position of local shape variables through the skull mirrors the position of interlandmark distances, although they are conceptually different types of measurements.

30 Here we considered only covariance or correlation structure for the symmetrical component of variation; therefore, prior to any analysis, we controlled the effects of variation in asymmetry. For interlandmark distances, we averaged bilateral measurements within

Table 2.1: Twenty-one species used in the present work. Sample sizes and linear models adjusted over each OTU are also indicated.

Species	Group ^a	<i>n</i>	Model ^b
<i>Alouatta belzebul</i>	P	109	X
<i>Ateles geoffroyi</i>	P	78	-
<i>Cacajao calvus</i>	P	48	S + X
<i>Callicebus moloch</i>	P	93	X
<i>Callithrix kuhlii</i>	P	129	-
<i>Cebus apella</i>	P	110	X
<i>Cercopithecus ascanius</i>	C	61	X
<i>Chiropotes chiropotes</i>	P	56	X
<i>Chlorocebus pygerythrus</i>	C	110	X
<i>Colobus guereza</i>	C	140	X
<i>Gorilla gorilla</i>	C	115	X
<i>Homo sapiens</i>	C	160	S * X
<i>Hylobates lar</i>	C	66	X
<i>Macaca fascicularis</i>	C	69	X
<i>Pan troglodytes</i>	C	61	X
<i>Papio anubis</i>	C	46	X
<i>Ptilocolobus foai</i>	C	83	X
<i>Pithecia pithecia</i>	P	69	S + X
<i>Procolobus verus</i>	C	88	X
<i>Saguinus midas</i>	P	50	S
<i>Saimiri sciureus</i>	P	87	X

^a C: Catarrhini; P: Platyrrhini

^b S: subspecies/population; X: sex.

each individual. For both Procrustes residuals and local shape variables, we followed the procedure outlined in Klingenberg et al. (2002) for bilateral structures by obtaining for each individual a symmetrical landmark configuration, averaging each actual shape with its reflection along the sagittal plane; we estimate local shape variables afterwards. With respect to Procrustes residuals, landmarks placed along the sagittal plane will have zero variation in the direction normal to this plane; we aligned all specimens' sagittal plane to the xz plane, thus removing the y component for each of these landmarks from covariance/correlation matrices.

For each dataset, we estimated covariance and correlation matrices after removing effects of little interest in the present context, such as sexual dimorphism, for example. For interlandmark distances and local shape variables these effects were removed through a multivariate linear model adjusted for each species, according to Table 2.1; for Procrustes

residuals, the same effects were removed by centering all group means to each species' mean shape since the loss of degrees of freedom imposed by the GPA prohibits the use of a full multivariate linear model over this kind of data to remove fixed effects.

In order to consider the effects of size variation on modularity patterns we used different procedures to remove the influence of size from each type of variable. For interlandmark distances we used the approach established by Bookstein et al. (1985); if $\mathbf{C}$ is a correlation matrix, we obtained a correlation matrix $\mathbf{R}$ without the effect of size using the equation

$$\mathbf{R} = \mathbf{C} - \lambda_1 v_1 v_1^t \quad (2.1)$$

where λ_1 and v_1 refer respectively to the first eigenvalue and eigenvector of the spectral decomposition of $\mathbf{C}$, since this eigenvector commonly represents size variation in mammals, especially when interlandmark distances are considered (Wagner, 1984; Mitteroecker *et al.*, 2004; Mitteroecker & Bookstein, 2007); t denotes matrix transpose.

For Procrustes residuals and local shape variables the effects of isometric variation were removed by normalizing each individual to unit centroid size. However, allometric relationships still influence covariance or correlation structure. In order to remove this effect we used a procedure based upon Mitteroecker et al. (2004), which relies on the estimation of an allometric component a for each OTU, composed of normalized regression coefficients for each of the m shape variables (either Procrustes residuals or local shape variables) over log Centroid Size. If $\mathbf{S}$ is a covariance matrix, we obtained a covariance matrix $\mathbf{R}$ without the influence of allometric relationships using the equation

$$\mathbf{R} = (\mathbf{I}_m - aa^t)\mathbf{S}(\mathbf{I}_m - aa^t) \quad (2.2)$$

where $\mathbf{I}_m$ represents the identity matrix of size m . Therefore, our empirical dataset consists of six sets of covariance/correlation matrices, corresponding to each type of morphometric variables considering the presence or absence of size variation.

Empirical Tests

Using these six sets of covariance/correlation matrices, we tested the hypotheses of trait associations described in Table 1.1 for Procrustes residuals and Table 1.2 for interlandmark distances and local shape variables. These trait sets are grouped with respect to their scope;

two regional sets (Face and Neurocranium), each divided into three localized trait sets (Oral, Nasal and Zygomatic for the Face; Orbit, Base and Vault for the Neurocranium).

For all hypotheses, we estimated Modularity Hypothesis Indexes (MHI; Porto *et al.*, 2013) and the RV coefficient (Klingenberg, 2009). Both statistics are estimated by partitioning covariance or correlation matrices into blocks; if $\mathbf{A}$ is a covariance or correlation matrix, the partition

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_h & \mathbf{A}_b \\ \mathbf{A}_b^t & \mathbf{A}_c \end{bmatrix}$$

indicates that the block $\mathbf{A}_h$ contains covariances or correlations between traits that belong to the trait set being considered, while $\mathbf{A}_c$ represents the complementary trait set; $\mathbf{A}_b$ represents the block of covariances or correlations between the two sets. Thus, covariance (S) or correlation (C) matrices can be partitioned into a similar scheme. We estimated MHI values using the equation

$$MHI = \frac{\bar{\rho}_+ - \bar{\rho}_-}{ICV} \quad (2.3)$$

where $\bar{\rho}_+$ represents the mean correlation in $\mathbf{C}_h$, $\bar{\rho}_-$ represents the mean correlation in the remaining sets (both $\mathbf{C}_b$ and $\mathbf{C}_c$), and ICV is the coefficient of variation of eigenvalues of the associated covariance matrix, which is a measurement of the overall integration between all traits (Shirai & Marroig, 2010). We estimated RV coefficients for each hypothesis using the relationship

$$RV = \frac{tr(\mathbf{S}_b \mathbf{S}_b^t)}{\sqrt{tr(\mathbf{S}_h \mathbf{S}_h) tr(\mathbf{S}_c \mathbf{S}_c)}} \quad (2.4)$$

where tr represents the sum of diagonal elements in any given matrix ($tr \mathbf{A} = \sum_i a_{ii}$).

The partitioning scheme outlined above assumes that the complementary trait set does not represent an actual hypothesis; however, we may choose to consider that both sets ($\mathbf{A}_h$ and $\mathbf{A}_c$) represent two distinct hypothesis. The estimation of RV coefficients remains the same; however, MHI values are estimated considering that $\bar{\rho}_+$ is the average correlation in both $\mathbf{C}_h$ and $\mathbf{C}_c$, while $\bar{\rho}_-$ represents the average correlation only in $\mathbf{C}_b$. In the case of the distinction between Facial and Neurocranial traits, we estimated MHI values in this manner, reporting values for this estimate under the denomination ‘Neuroface’, following Marroig & Cheverud (2001), along independent MHI estimates for each region. Furthermore, since both Face and Neurocranium are two disjoint trait sets when any morphometric variable type is considered, RV coefficient values for either set are equal; therefore, a single RV value

is reported for both regions, for each variable type.

In order to test the hypothesis that a trait set represents a variational module, we used a randomization procedure generating 1000 random trait sets with the same number of traits as the original set, calculating MHI and RV values for each iteration. For each trait set and covariance/correlation matrix, we used these values to construct distributions for both statistics representing the null hypothesis that a given trait set is a random arrangement without meaningful relationships; we then compare this null distribution to the real value obtained. For MHIs we consider this null hypothesis rejected when the real value is higher than the upper bound for the distribution, considering the significance level established; for RV coefficients, the null hypothesis is rejected when the real RV value is lower than the lower bound for the distribution, also considering significance level. For Procrustes residuals the randomization procedure maintains coordinates within the same landmark together in each randomly generated trait set, following Klingenberg & Leamy (2001).

While the procedure for estimating significance for MHIs is derived from Mantel's (1967) approach (as outlined by Cheverud *et al.*, 1989), we chose to generate null distributions for MHI directly, instead of estimating matrix correlation values for both real and randomized matrices. Estimated p -values in both cases remain the same, and the additional step of calculating matrix correlations would produce an unnecessary difference between the estimation of significance for MHI and RV.

Estimation of Error Rates

To evaluate the statistical properties of metrics, types of morphometric variables, and the randomization procedure, we built theoretical correlation matrices

$$C_s = \begin{bmatrix} W_1 & B \\ B^t & W_2 \end{bmatrix}$$

where W_1 and W_2 are correlation blocks associated with two distinct trait sets, and B represents the correlation block between sets.

For each correlation matrix in our empirical dataset, we estimated average correlations within and between all trait sets we considered, obtaining a distribution of within and between sets correlations associated with each type of morphometric variable (Procrustes Residuals, Interlandmark Distances and Local Shape Variables), also considering whether size was retained or removed (Figure 2.1). We constructed corresponding sets of C_s matrices sampling each distribution obtained; for each matrix, we sampled two within-set correlations

and one between-set correlation, filling the corresponding blocks ($\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{B}$) in each theoretical matrix with the sampled correlations. For example, with only four traits divided into two blocks of two traits, sampling the values 0.5 and 0.3 from the within-set distribution and -0.1 from the between-set distribution produces

$$\mathbf{C}_s = \begin{bmatrix} 1 & 0.5 & -0.1 & -0.1 \\ 0.5 & 1 & -0.1 & -0.1 \\ -0.1 & -0.1 & 1 & 0.3 \\ -0.1 & -0.1 & 0.3 & 1 \end{bmatrix}$$

filling all cells in each block with the associated sampled correlation. Thus, we represent each type of morphometric variable as two correlation distributions (Figure 2.1), building six sets of theoretical correlation matrices that are representative of each type (considering the presence or absence of size variation) which retain their statistical properties.

For each pair of correlation distributions, we built 10000 correlation matrices for 40 traits; previous tests indicate that matrix dimensionality does not qualitatively affect our results. We also sampled variances from each type of morphometric variable, thus constructing an associated covariance matrix, since the estimation of RV values uses covariance matrices. We considered only positive-definite matrices; if any given matrix did not fit this criterion, we discarded that matrix and sampled new correlations; thus, we can sample observations from a multivariate normal distribution using each of these 60000 matrices as the Σ parameter. For each matrix, we also randomly determine the number of traits in both sets, establishing a minimum value of five for set size.

We used this set of 60000 covariance matrices of known structure ($\mathbf{C}_s$) to build another set of covariance matrices of unknown, random structure ($\mathbf{C}_r$) by simply shuffling both lines and columns of each matrix. For all matrices ($\mathbf{C}_r$ and $\mathbf{C}_s$) we obtained samples of increasing size (20, 40, 60, 80, 100 individuals) and re-estimated a covariance matrix for each sample, thus considering in our tests the uncertainty derived from sampling. For each matrix estimated, both $\mathbf{C}_r$ and $\mathbf{C}_s$, we test the hypothesis that the two sets used to generate each $\mathbf{C}_s$ matrix represent two variational modules, using both MHI and RV coefficients, as described in the section ‘Empirical Tests’.

If samples were generated from a $\mathbf{C}_r$ matrix, this represents a situation of a true null hypothesis for each test, since the correlation matrix used to produce the sample was generated by a permutation of the hypothesis being tested. Therefore, testing hypotheses over $\mathbf{C}_r$ matrices allows the estimation of Type I error rates, or the proportion of cases in which a true null hypothesis is rejected, given a significance level. In an adequate test, we

Figure 2.1: Distribution of within-set and between-set mean correlations. Such correlations are derived from the six types of empirical correlation matrices.

would expect that both quantities, significance level and Type I error rate, will be identical.

The opposite case, when we sampled C_s matrices, represents a situation in which we know that the null hypothesis of either test is false, since we are testing the hypothesis that the partitioning scheme used to generate that particular matrix actually represents two variational modules. Thus, we estimated Type II error rates, or the probability that a false null hypothesis is not rejected, given a significance level; here, we represent Type II error using the power for each test, by simply calculating the complementar probability to Type II error rate. In an adequate test, we expect that power would rapidly reach a plateau when significance level is still close to zero, and further increasing $P(\alpha)$ will not produce a great increase in power.

Our estimates of power for both statistics should also be controlled for effect size, since sampled correlations may generate a correlation structure that is not detected due to small differences among within-set and between-set correlations. For each correlation matrix sampled, we estimate squared between-set correlations (b^2), in order to use it as an estimate of effect size that is not directly associated with either MHI and RV metrics. We expect that power for either tests decreases with increasing b^2 values, as effect size would also decrease.

Software

All analysis were performed under R 3.2.2 (R Core Team, 2015). Source code for all analyses can be found at <http://github.com/wgar84>. Previous tests indicated no differences between our estimation of empirical RV coefficients, based upon our own code, and estimates provided by MorphoJ (Klingenberg, 2011). In order to obtain symmetrical landmarks configurations, we used code provided by Annat Haber, available at <http://life.bio.sunysb.edu/morph/soft-R.html>.

Results

Empirical Tests

Tests performed using MHI for localized trait sets (Oral, Nasal, Zygomatic, Orbit, Base, and Vault; Figure 2.2a) detected a consistent pattern among OTUs for interlandmark distances and local shape variables. In the first set the Oral subregion was detected as a modular partition, and, with size removed, the Vault subregion was also detected; both Orbit and Base region were not detected in any of these tests. With local shape variables, Oral, Vault

Nasal, and Zygomatic sets were detected consistently across OTUs; the removal of allometric variation affected only the detection of the Vault in some groups. Furthermore, the Base sub-region was detected only in 3 of 42 tests performed over local shape variables, pooling together size and size-free correlation patterns.

- 5 For Procrustes residuals, the pattern of detection among sub-regions and OTUs was more inconsistent; for instance, the Base sub-region was detected in several OTUs, which contrasts this type of morphometric variable with the other two types. Observing the actual MHI values, Figure 2.2a also indicates that Procrustes residuals had a low variance of this metric within each OTU, while interlandmark distances and local shape variables have a consistent
10 pattern of variation, with lower values for the Base and, for interlandmark distances, Orbit trait sets, while the Oral, Nasal and Vault regions displayed higher values consistently.

Tests performed using RV coefficients (Figure 2.2b) show a more irregular pattern for each variable type. When interlandmark distances were considered, most tests detected the Vault sub-region with size variation retained, and the Base sub-region when size variation
15 was removed. For Procrustes residuals, few tests were able to reject the null hypotheses, detecting only a handful of valid modular partitions. Tests performed on local shape variables displayed the opposite behavior: almost all partitions were detected, regardless of whether allometric variation had been retained or removed. Moreover, RV values displayed a pattern of marked variation among OTUs, more so than between values within each
20 OTU; notably, *Macaca fascicularis* and *Papio anubis* showed RV values much higher than those estimated on remaining species. Such pattern can be observed both on interlandmark distances with size retained and in Procrustes residuals.

With respect to regional trait sets (Face and Neurocranium), tests performed using MHI (Figure 2.3a) indicate a pattern consistent with those found on localized sets (Figure 2.2a).
25 Considering interlandmark distances, Facial traits were detected as a valid modular partition both with size variation retained or removed, while Neurocranial traits were detected as a valid partition only when size variation was removed. This pattern mirrors the contrast between Oral and Vault traits found in the localized sets regarding interlandmark distances. For Procrustes residuals, Neurocranial traits were detected as a valid partition with both
30 size retained and removed; once again, this pattern mirrors the detection of the Basicranial partition found in the localized sets. Finally, in local shape variables, both Face and Neurocranium were detected as valid with size retained; with size removed, only the Face was recognized consistently. The same can be observed for localized sets, where the removal of allometric variation affected the detection of the Vault set in some OTUs. The tests for
35 the distinction of within-set and between-set correlations for these two sets (designated

Figure 2.2: MHI (a) and RV (b) values for localized trait sets. Circles indicate whether a trait set is recognized as a variational module in a given OTU, with $P(\alpha)$ indicated by the legend. Notice that values represented in blue are either high MHI values and low RV values, as the alternate hypothesis for each statistic is formulated accordingly; see ‘Methods’ for details.

‘Neuroface’) show a pattern that is consistent with tests for the individual sets: if one of the sets was previously detected, this distinction was also detected as valid.

Figure 2.3: MHI (a) and RV (b) values for regional trait sets. Circles indicate whether a trait set is recognized in a given OTU, with $P(\alpha)$ indicated by the legend. Notice that values represented in blue are either high MHI values and low RV values, as the alternate hypothesis for each statistic is formulated accordingly; see ‘Methods’ for details.

Testing the distinction between Face and Neurocranium using RV coefficients (Figure 2.3b) showed that in most cases both regions are considered distinct and valid variational modules for interlandmark distances and local shape variables; for Procrustes residuals, only a handful of taxa indicated the same result. In this case, the correspondence with localized trait sets (Figure 2.2b) is more difficult due to the lack of independent tests for each region.

Error Rates

Comparing the distributions of MHI and RV values from theoretical matrices with respect to their structure (Figure 2.4) indicated marked differences between metrics and morphometric variables from which correlations were sampled. In general, the distribution of MHI

values obtained from C_r matrices was the same regardless of type of variable, while the distributions for C_s matrices for this metric were heterogeneous; for either interlandmark distances and local shape variables, the distribution of MHI values for these two types of matrices did not overlap, while for Procrustes residuals, these distributions overlapped to some degree. For RV values, all distributions overlap to some degree; for local shape variables and interlandmark distances with size retained, either distributions (C_r and C_s) overlapped to a lesser extent.

The relationship between significance levels and Type I error rates estimated over C_r matrices approached an identity relationship (Figure 2.5), regardless of whether variational modularity was estimated using MHI or RV; even at low sample sizes Type I error rates were very close to significance levels. Furthermore, the effect of sampling correlations from size-free distributions did not change Type I error rates.

With respect to the relationship between power and significance levels (estimated over C_s matrices), there were substantial differences with respect to the chosen metric (MHI or RV) and to the type of variable that provides sampled correlations. Considering local shape variables (Figure 2.6), tests using either MHI and RV implied high power, even at low sample or effect sizes; increasing these quantities further increased power. However, for lower effect sizes (represented by high average squared correlation between sets, b^2) power for tests using MHI was higher than for those using RV; as effect size increased (lower b^2 values), the difference in power between the two statistics decreased. For local shape variables, sampling from its associated size-free correlation distribution implied minor differences in power for both statistics.

For interlandmark distances (Figure 2.7) there were substantial differences on the relationship between power and significance level. In general, power for tests using MHI are always higher than for tests using RV; this effect is more pronounced on C_s matrices derived from size-free interlandmark distances, although tests on these matrices have a substantial decrease in power for either tests. However, this decrease is more pronounced for tests based on the RV statistic, since for lower effect sizes (high b^2 values) power approaches an identity relationship with significance level. Sample size also interferes with this relationship since increasing this quantity also increases power when higher effect sizes (low b^2 values) are considered.

With respect to Procrustes residuals (Figure 2.8), tests using either MHI or RV have reduced power regardless of effect or sample size. Sampling from size-free correlation distributions to build C_s matrices also has little effect. In this case, power for tests performed using RV values approaches an identity relationship with significance level; increasing sample size

Figure 2.4: Distribution of Modularity Hypothesis Index and RV Coefficient for theoretical correlation matrices.

Figure 2.5: Type I error rates. Rates are represented as a function of the chosen significance level regarding tests for variational modularity applied on C_r correlation matrices. The solid black line represents the identity relationship.

Figure 2.6: Power for MHI and RV statistics for local shape variables. Power is represented as a function of the chosen significance levels with respect to tests for variational modularity applied on sampled C_s correlation matrices. Lines are colored with respect to quantiles of the b^2 distribution.

Figure 2.7: Power for MHI and RV statistics for interlandmark distances. Power is represented as a function of the chosen significance levels with respect to tests for variational modularity applied on sampled C_s correlation matrices. Lines are colored with respect to quantiles of the b^2 distribution. The solid black line represents the identity relationship.

has some effect, but it does not increase power above 50% in any case.

Discussion

Covariance matrices derived from morphological traits are supposed to represent the pattern of codependence that stems from interactions among developing morphological elements (Olson & Miller, 1958; Cheverud, 1996a). Such interactions are the expression of local developmental factors, as they interact with the gene expression profiles of surrounding cell types, producing coordinated changes in cellular growth and tissue interaction, thus integrating these elements in the adult population. Although these events of local integration overlap, and the composed effect over adult covariance patterns may be confusing (Hallgrímsson & Lieberman, 2008; Hallgrímsson *et al.*, 2009), we believe that a careful comparison of different yet equally proper ways of measuring and representing form may be informative of the underlying processes that produce covariances.

Due to the minimization of quadratic distances among homologous landmarks during GPA, covariance matrices derived from Procrustes estimators lose the signal of localized effects on covariance patterns (Linde & Houle, 2009). Therefore, the use of Procrustes estimators to investigate morphological integration or modularity implicitly implies a dissociation between the phenomenon under investigation and the chosen representation. This disconnection between theory and measurement may have serious consequences for the hypotheses we wish to test (Houle *et al.*, 2011); such consequences are observable in both our empirical tests and those tests performed on theoretical matrices, as we explore below.

The mammalian Basicranium originates from thirteen precursor tissues derived from both paraxial mesoderm and neural crest, and they may merge to form single bones, such as the sphenoid (Jiang *et al.*, 2002; Lieberman, 2011). Furthermore, these precursors display a mosaic of endochondral and intramembranous ossification early in development, and, as the brain grows, it induces a pattern of internal resorption and exterior deposition on the underlying posterior Basicranium (Lieberman *et al.*, 2000); meanwhile, the anterior portion suffers influence from the development of Facial elements (Bastir & Rosas, 2005). Thus, since the Basicranium ossifies early in development, the composed effect of all posterior steps of cranial development will overshadow any pattern of integration this region might have, as predicted by the palimpsest model of development (Hallgrímsson *et al.*, 2009). Moreover, the angulation between anterior and posterior elements of the Basicranium has significantly changed during primate evolution, and such property appears to have evolved in coordination with Facial growth relative to the cranial Vault, accomodating both

Figure 2.8: Power for MHI and RV statistics for Procrustes residuals. Power is represented as a function of the chosen significance levels with respect to tests for variational modularity applied on sampled C_s correlation matrices. Lines are colored with respect to quantiles of the b^2 distribution. The solid black line represents the identity relationship.

structures on each other (Scott, 1958; Lieberman *et al.*, 2000, 2008).

Due to this heterogeneity of developmental processes acting on the Basicranium, it is not expected that this region would be detected as a variational module; thus, a test of this property over skull covariance patterns should fail to reject the null hypothesis of random association. However, considering the 42 tests performed over covariance matrices derived from Procrustes residuals regarding localized hypotheses (midpanels of Figure 2.2a), the Basicranium is detected as a valid variational module in 27 cases, distributed through matrices with size either retained or removed; in some cases (e.g. *Alouatta*, *Cercopithecus*), only the Basicranium is detected. Hence, Procrustes residuals show a pattern of detection of variational modules opposed to the expectation based on developmental dynamics. In covariances matrices derived from interlandmark distances (upper panels of Figure 2.2a), the Basicranium was never detected; with local shape variables (lower panels of Figure 2.2a), the Basicranium was detected only three out of 42 times.

These two remaining types of representations (i.e. interlandmark distances and local shape variables) show patterns of detection of variational modules that are consistent with the expectations derived from developmental and functional interactions, especially with respect to how these patterns articulate with the retention or removal of size variation. Considering that interlandmark distances are on a ratio scale (Houle *et al.*, 2011), isometric variation will be represented to a greater extent when compared to subtle allometric relationships, due to the multiplicative nature of biological growth (Huxley, 1932). Therefore, the Oral trait set is detected as a valid variational module, considering covariance matrices among interlandmark distances (upper left panel of Figure 2.2a), since this region is strongly affected by the induction of bone growth due to muscular activity beginning in the pre-weaning period (Zelditch & Carmichael, 1989; Hallgrímsson *et al.*, 2009). Furthermore, in these matrices, allometric interactions are associated with the strength of association between traits in Oral, Nasal, Zygomatic and Vault regions in an evolutionary scale (Chapter 3); thus, allometric relationships also play some role in the determination of covariance/correlation patterns for other subregions, although such pattern may be masked in interlandmark distances by the effect of isometric variation. On the other hand, the patterns expressed in covariance/correlation matrices for local shape variables are influenced by allometric relationships defined on a proper scale (Jolicœur, 1963; Houle *et al.*, 2011; Márquez *et al.*, 2012) and local developmental processes. Thus, they reflect this association between integration and allometry.

By removing allometric effects from local shape variables (lower right panel of Figure 2.2a), variational modularity can still be detected in both Oral and Nasal regions, while in a

number of species, the Vault region is no longer detected as a variational module. Vault integration may be achieved through both allometric relationships and the effect of relative brain growth, since Vault elements arise mostly through intramembranous ossification, induced by the secretion of signaling factors from the outer brain tissues, with a modest but necessary contribution of mesoderm-derived tissue that undergoes endochondral ossification (Jiang *et al.*, 2002; Rice *et al.*, 2003; Franz-Odenaal, 2011; Lieberman, 2011). In humans, the growth of Vault osteological elements occur, from a topological standpoint, without deviations from an isometric growth model, forming regular connections among bones (Esteve-Altava & Rasskin-Gutman, 2014), since the boundary of interactions between brain and Vault bones is also regular. Thus, as this effect dominates the latter stages of pre-natal development (Hallgrímsson & Lieberman, 2008; Lieberman, 2011), the overall effect of brain growth over skull growth patterns at this stage may mirror an affine transformation, rendering it undetectable over covariance patterns derived from local shape variables, since their estimation excludes shape variation associated with affine transformations (Márquez *et al.*, 2012). In contrast, Oral and Nasal elements have more complex patterns of connectivity arising from their tight integration with the soft tissues that compose the remaining elements of the Face (Lieberman, 2011; Esteve-Altava *et al.*, 2013; Esteve-Altava & Rasskin-Gutman, 2014), thus producing an intricate pattern of associations that may be responsible for the variational modularity we are able to detect in all species, regardless of whether size variation is retained or removed.

The differences between the pattern of detection in all representations for local trait sets for RV values (Figure 2.2b) makes a similar interpretation of the results for this metric difficult, as opposed to the results regarding MHI values. The variance of RV values among species seems to indicate that RV values are sensitive to the magnitude of morphological integration, since species with high RV values, such as *Papio anubis* and *Macaca fascicularis*, are within genera that also exhibit higher magnitudes among Catarrhines (Oliveira *et al.*, 2009b), at least when we consider interlandmark distances. Interestingly enough, this effect of magnitude of integration (which is thought to emerge as a consequence of size variation; Wagner, 1984; Marroig & Cheverud, 2001, 2005; Porto *et al.*, 2013) seems to also affect Procrustes residuals regardless of whether covariance structure arising from allometric relationships is removed or not.

For the set of hypotheses concerning differences between Facial and Neurocranial traits (Figure 2.3), there is substantial agreement between tests performed using MHI and RV values. For MHI values (Figure 2.3a), the overall pattern of detection is similar to the pattern detected in local trait sets, for all variable types; for RV values (Figure 2.3b), there is

strong support for the hypotheses that both Face and Neurocranium represent variational modules, in both local shape variables and interlandmark distances. These regions have marked differences in developmental timing and pattern formation, as observed from the behavior of their composing units (Zelditch & Carmichael, 1989; Hallgrímsson *et al.*, 2009; 5 Lieberman, 2011; Esteve-Altava & Rasskin-Gutman, 2014); therefore, the more general pattern of distinction between Face and Neurocranium is detected regardless of the metric chosen to represent modularity.

Theoretical Matrices and Error Rates

The distribution of MHI and RV values obtained from the theoretical matrices (Figure 2.4) is 10 a starting point for understanding the differences in power for tests using these two metrics (Figures 2.6–2.8). For MHI values, the distribution obtained from random (C_r) matrices is consistently the same, regardless of what representation we used to sample correlations, or whether size was retained or removed. On the other hand, the distribution of RV values for random matrices change depending on the representation sampled or whether size variation 15 has been removed or retained. Moreover, this change in behavior for the distribution of RV values for C_r matrices implies different levels of overlap between these null distributions and the distribution of values obtained for structured (C_s) matrices.

For Procrustes residuals, a substantial overlap occurs regardless of whether size variation was removed or retained; unsurprisingly, our estimates of power for RV in this type of 20 representation are very low (Figure 2.8). Nonetheless, power estimated for tests based on MHI is lower than in other types of representation, since the difference between within-set and between-set correlations in Procrustes residuals (Figure 2.1) is the lowest of all representations. A substantial overlap in RV distributions for C_r and C_s matrices also occurs with interlandmark distances when size variation is removed, and it results in low 25 power for tests using RV values in this type of representation (right column of Figure 2.7). However, in this case there is a substantial difference between within-set and between-set correlations (Figure 2.1), and tests using MHI to represent modularity are still able to detect such difference (albeit with reduced power) when compared to tests over C_s matrices derived from interlandmark distances with size retained (Figure 2.7). In those cases where both 30 distributions for C_r and C_s matrices do not overlap substantially — for example, when local shape variables are considered (Figure 2.6) — power for tests performed using MHI values is always higher than for tests using RV except when sample sizes are very high; in this case power for both metrics are similar. The same behavior can also be observed in interlandmark distances when size is retained (left column of Figure 2.7).

These results indicate that RV coefficients are more sensitive to the absolute value of both within-set and between-set correlation distributions than MHI values. For interlandmark distances (Figure 2.7), removing size variation reduced the average value of both correlation distributions by a similar amount (Figure 2.1); the difference between average correlations in these two sets actually increases, going from 0.042 to 0.061 when size is removed. However, since the actual average correlations for these two distributions approach zero, tests based on RV lose power more rapidly than tests based on MHI. This sensitivity might be associated to the use of squared covariances, as shown by Equation 2.4, while Modularity Hypothesis Indexes use correlations directly (Equation 2.3). Furthermore, as observed by Fruciano et al. (2013), sample sizes influence the estimation of RV coefficients, and we demonstrate here that such sensitivity also extends to estimates of power for tests using this metric.

Our estimates of power for tests using MHI indicate that it is more robust to differences in absolute correlation values or sample sizes, thus allowing comparisons across more heterogeneous settings, such as our comparison between different representations of form, with substantial variation in sample sizes. The detection of variational modularity is akin to Student's t test, since we try to determine whether two groups of observations (correlations between traits in the same subset *versus* correlations between traits in different subsets) have a significant difference in average values; we use resampling procedures to estimate significance in this case only due to the interdependency between pairwise correlations. Thus, as the estimation of MHI values resembles the estimation of t -values (difference between location parameters for two groups, divided by a scale parameter — ICV; Equation 2.3), we believe that this statistic is appropriate to represent variational modularity, and that is reinforced by the robustness of the tests which employed MHI.

Our approach for constructing theoretical matrices attempts to simulate the simplest situation: only two subsets of traits, akin to the distinction between Face and Neurocranium in our empirical dataset (Figure 2.3). In this setting, both statistics are capable of detecting this distinction, except when both are used on covariance/correlation patterns derived from Procrustes residuals. However, even though we built theoretical matrices using correlations sampled from these estimators, to simulate the interference in covariance structure that such estimators produce in our theoretical matrices is quite difficult. Furthermore, constructing such matrices with more complicated patterns (with three modules, for instance) while maintaining their connection to the correlation distributions of each morphometric type is also difficult, due to the restriction on positive-definiteness we enforce on them. Thus, the lack of differences in type I error rates for all cases may be a limitation of our scheme for building theoretical matrices.

The issues we found with the use of Procrustes estimators for covariance matrices and the use of RV coefficients to estimate and detect variational modularity may explain results found by other authors. For instance, Martínez-Abadías *et al.* (2011) has found no evidence that genetic and phenotypic covariance structure for human skulls conforms to a modular structure, since all tests performed by these authors failed to reject the null hypothesis of random association. These authors use Procrustes estimators to represent covariance structure, and test their hypothesis of partitioning (Face, Vault and Base) using RV as the statistic representing variational modularity. Since this combination yields very low power (Figure 2.8), not rejecting the null hypothesis in their case might be a consequence of the choice of estimates of both covariance structure and variational modularity; thus, these authors' conclusion of pervasive genetic integration in the human skull might be misleading, considering that skull covariance patterns in humans are one of the most modular examples of such patterns when compared to other mammals (Porto *et al.*, 2009) or catarrhine primates (Oliveira *et al.*, 2009b).

The approach we explore in the present work is only one of the different alternatives to investigate the association between genetic, functional and developmental interactions and correlation structure (Mitteroecker & Bookstein, 2007). For example, Perez *et al.* (2009) relies on abstracting correlation matrices into networks, then using community-detection algorithms to search for modular patterns without *a priori* hypotheses, associating their results with knockout experiments that support the communities they found among traits; however, it is not clear how much relevant biological information is retained in these network representations. Furthermore, the authors use Procrustes estimators, which may bias the detection of modularity patterns in this setting in the same manner as we demonstrated here.

Monteiro *et al.* (2005) assumes that the underlying morphogenetic components of the rodent mandible behave as modules, further investigating the patterns of correlation between these units in both within-species variation and between-species variation among Echimids; Monteiro & Nogueira (2010), relying on the correspondence of these units through mammalian diversification did the same to phylostomid bats. Although using a landmark-based approach to represent morphological variation, the authors do not use Procrustes estimators to represent covariance structure among these units, and the pattern of reorganization of correlation structure among these units associated with niche diversification in phylostomids seems robust, considering that this radiation may have been associated with a very heterogeneous adaptive landscape, and such heterogeneity may lead to a reorganization of correlation patterns (Jones, 2007; Jones *et al.*, 2012; Melo & Marroig, 2015).

Another valid approach is to model certain aspects of development as null hypotheses; Esteve-Altava & Rasskin-Gutman (2014), investigating the pattern of connections among human cranial bones, formulates the null hypothesis that unconstrained bone growth is sufficient to explain the observed patterns, further constructing a null network based on this expectation which is compared to the actual network. Such approach could be extended to investigate covariance structure; considering the geometric properties of the features we measure (Mitteroecker & Bookstein, 2007), one could formulate the null hypothesis that topological proximity is a sufficient explanation for the observed covariance structure, against the alternative hypothesis that local developmental processes coupled with functional interactions produce stronger relationships among close elements that surpass topological interactions. Alternatively, one could actively look for variational boundaries between regions, as boundary formation is a phenomenon that has been well studied under a dynamical perspective on development (e.g. Turing, 1952; Meinhardt, 1983; Tiedemann *et al.*, 2012).

The approach of partitioning covariance matrices into blocks that correspond to inferred modular associations has the advantage that it is operationally simple; however, modularity patterns are almost certainly not expressed in phenotypic data as the binary hypotheses we used here (Hallgrímsson *et al.*, 2009). Thus, hypotheses and inferences derived from them have to be contextualized in the light of developmental dynamics, since the measurements we make and the parameters we estimate have to be properly connected to the models we are considering; otherwise, inferences made from such models may be devoid of meaning (Wagner, 2010; Houle *et al.*, 2011).

Conclusion

Here we show that Procrustes estimators for covariance matrices fail to capture the modularity patterns embedded in phenotypic data, regardless of which metric is chosen to represent such patterns, although the combination of this type of variable with RV coefficients for investigating modularity has even more problems than either has alone. Both interlandmark distances and local shape variables seem valid options to represent morphological variation, if their limitations are taken into consideration. We wish to stress this point: any representation of morphological variation has limitations since they are themselves models — at the very least of what it is important to represent — not fully capturing the phenomena we may be interested in.

Chapter 3

The Evolution of Skull Allometry in Anthropoid Primates

Abstract

5 Anthropoid Primates display an impressive body size variation, which implies many morphological, ecological and physiological differences among lineages. These allometric relationships may impose constraints on phenotypic evolution; for instance, most of the morphological diversity in skull shape in Anthropoids can be explained as an indirect consequence of body size evolution, except within Hominids. In the present work, we tested
10 whether allometric relationships for skull shape in Anthropoids have remained stable during the diversification of this group and also examined the association between these relationships and integration patterns. We show that allometric intercepts have changed a number of times, mostly within New World Monkeys, and that slopes have changed significantly only within Hominidae, in two independent instances, *Homo* and *Gorilla*. Furthermore, the
15 variation in allometric slopes contributes to Facial and Neurocranial integration in opposing directions. Thus, the constancy of allometric slopes seems to be related to the maintenance of functional interactions during development imposed by stabilizing selection, so that the few changes in slopes may be associated with shifts in the adaptive landscape in the course of Hominid diversification.

Introduction

Anthropoid primates display an approximately thousand-fold variation on body size, from pygmy marmosets ($\sim 110\text{g}$) to lowland gorillas ($\sim 120\text{kg}$). This impressive diversity implies substantial morphological, physiological and ecological differences among these organisms, since changes in body size fundamentally affect how living organisms interact with their environments (Gould, 1966; West *et al.*, 1997). These differential scaling relationships between organismal traits and body size are referred to as allometry (Huxley, 1932; Gould, 1966; Cheverud, 1982b; Lande, 1985) and may be represented as power laws, which become linear on a log-log scale (Huxley, 1932; but see Nijhout & German, 2012). Allometry may thus be represented by two parameters: (allometric) slopes and intercepts, and these parameters are associated with the timing and rate of developmental milestones (Houle, 2001; Voje *et al.*, 2013; Pélabon *et al.*, 2014).

Allometry may act as a constraint, diverting evolutionary trajectories in natural populations (Gould & Lewontin, 1979; Lande, 1979; Schluter, 1996). For example, selection for body size driven by shifts in feeding habits has dominated diversification in New World monkeys, and differences in skull morphology among species in this group appear to be an indirect consequence of the allometric relationships between skull traits and body size (Marroig & Cheverud, 2005, 2010; Marroig *et al.*, 2012); within Old World Monkeys, the same scenario could explain skull morphological diversity in Cercopithecids (Cardini *et al.*, 2007; Cardini & Elton, 2008). Regarding Hominids, selection for reduction of facial elements and increased neurocranial capacity has played a limited role in the divergence between *Homo* and *Australopithecus* (Ackermann & Cheverud, 2004; Schroeder *et al.*, 2014).

Regarding morphological traits, allometric slopes may be constrained throughout development by stabilizing selection (Lande, 1980; Cheverud, 1984, 1996a; Hansen, 2011; Zelditch & Swiderski, 2011), while intercepts may be more labile on an evolutionary scale (Gould, 1974; Egset *et al.*, 2012; Bolstad *et al.*, 2015). However, these parameters are dependent on the mean individual growth rate and the covariance between these growth rates and body size achieved at adulthood; thus, selection on body size may also change either allometric slopes or intercepts (Pélabon *et al.*, 2013; Voje *et al.*, 2013). Experimental evidence on the evolution of allometric relationships indicates that growth parameters harbor genetic variation and thus may be available to selection, an effect that will then cascade to populational allometric relationships (e.g. Frankino *et al.*, 2005; Tobler & Nijhout, 2010; Egset *et al.*, 2012; Bolstad *et al.*, 2015), although methodological issues may hamper some of this evidence (Houle *et al.*, 2011; Voje *et al.*, 2013). Therefore, whether and to which extent allometric slopes can

evolve remain open questions.

In the present work, we investigate the relationship between skull size and shape in a phylogenetic framework using Anthropoid Primates as model organisms. We focus on testing whether and to which proportion intercepts and slopes have changed during the diversification of this group. Moreover, integration patterns in skull form are quite stable in both New World (Marroig & Cheverud, 2001) and Old World Monkeys, with the exception of some Hominids (Oliveira *et al.*, 2009b). Several authors (e.g. Porto *et al.*, 2013; Armbruster *et al.*, 2014) have argued that integration in this group is a consequence of the joint effect of both size variation and allometric relationships within populations; thus, we test the hypothesis that variation in allometric parameters among Anthropoid species, if present, will contribute to variation in the strength of association between skull traits. Considering both the spatial and temporal dynamics of mammalian skull development (Zelditch & Carmichael, 1989; Hallgrímsson *et al.*, 2009; Cardini & Polly, 2013), we expect that such changes will affect Facial and Neurocranial traits in opposing ways, due to the distinction between early- and late-developmental factors in skull development.

Using a Bayesian phylogenetic random regression model of allometric shape against size, we estimate allometric parameters along the phylogeny of Anthroidea. We show that allometric slopes have remained quite stable throughout Anthropoid diversification, while intercepts have changed in several groups, and that the little variation associated with allometric slopes produces opposite changes in Facial and Neurocranial integration, providing a clear link between allometry and integration. Despite the general stability in allometric slopes, we find remarkable changes occurring in *Homo* and *Gorilla*, demonstrating that constraints imposed by allometry are in no way absolute and, in some situations, may be overcome.

Methods

Sample

Our database consists of 5108 individuals from 109 species, distributed across all major Anthropoid clades above the genus level, comprising all Platyrrhini genera and a substantial portion of Catarrhini genera. We associated this database with an ultrametric phylogenetic hypothesis for Anthroidea (Figure 1.1), derived from Springer *et al.* (2012).

Individuals in our sample are represented by 36 registered landmarks, using a Polhemus 3Draw and a Microscribe 3DS for Platyrrhini and Catarrhini, respectively. Twenty-two

unique landmarks represent each individual (Figure 1.2), since fourteen of the 36 registered landmarks are bilaterally symmetrical. For more details on landmark registration, see Marroig & Cheverud (2001) and Oliveira et al. (2009b). Databases from both previous studies were merged into a single database, retaining only individuals in which all landmarks, from both sides, were present. 5

This database of landmark registration data was used to estimate both interlandmark distances and shape variables. For interlandmark distances, those measurements that involve bilaterally symmetrical landmarks were averaged after computing distances; for shape variables, a symmetrical landmark configuration was obtained by taking the mean shape between each individual configuration and its reflection along the sagittal plane 10 (Klingenberg *et al.*, 2002).

Allometric Relationships

We estimated allometric parameters using local shape variables (Márquez *et al.*, 2012), which are measurements of infinitesimal log volume transformations, calculated as the natural logarithm determinants of derivatives of the TPS function between each individual in our sample and a reference shape (in our case, the mean shape for the entire sample, estimated from a Generalized Procrustes algorithm). Such derivatives were evaluated at 38 locations, the midpoints between pairs of landmarks connected in Figure 1.2. After estimating local shape variables, we controlled for sources of variation of little interest in the present context, such as sexual dimorphism, and variation between subspecies or populations within each OTU, according to Figure 1.1, using generalized linear models. 15 20

Summarizing Shape Variation

After controlling for fixed effects within each species, we obtained a pooled within-species covariance matrix estimate for all 5108 individuals in our sample. We consider this estimate our best surrogate for phenotypic covariance patterns for log Centroid Size (logCS) and local shape variables in the ancestral Anthropoid population. Using this estimate, we obtained the Common Allometric Component (CAC: regression coefficients of shape onto logCS; Mitteroecker *et al.*, 2004) by dividing the covariances between local shape variables and logCS by the standard deviation of logCS and normalizing this vector of regression coefficients. Hence, the CAC represents the pooled within-species overall effect of size 25 30 onto shape. We projected all individuals in our sample along this direction, thus obtaining individual scores over the CAC for each sample unit. With the model considering the

relationship between CAC and logCS, we are able to test whether the strength of association between this ancestral allometric component and size has remained stable or changed during Anthropoid diversification, while a full multivariate model of local shape variables over logCS would allow testing whether the direction of within-species allometric components has changed. However, this reduction in dimensionality is necessary because a full multivariate random regression model using all 38 local shape variables would be computationally untractable, considering the actual state of MCMC samplers available.

Random Regression Model

In order to estimate whether allometric parameters for each of the 109 species have changed during Anthropoid diversification, we used a random regression model. Thus, for individual i in species j , the relationship between size and shape can be written as

$$y_{ij} = a_{sj} + b_{sj}x_{ij} \quad (3.1)$$

where x_{ij} and y_{ij} represent logCS values and projections of local shape variables onto the CAC respectively, and a_{sj} and b_{sj} represent static (within populations; Cheverud, 1982b; Pélabon *et al.*, 2013) allometric intercepts and slopes for species j . Phylogenetic random regression models assume that the vector of a_s and b_s for all $2p-2$ nodes on the phylogenetic hypothesis considered (p terminal nodes plus the $p-2$ intermediate nodes, excluding the basal node) are realizations of a multivariate normal distribution of the form

$$(a_s, b_s) \sim \mathcal{N}_{2(2p-2)}\left(\mathbf{1}_{2p-2} \otimes (\bar{a}_s, \bar{b}_s), \mathbf{S} \otimes \mathbf{C}\right) \quad (3.2)$$

where $\bar{a}_s$ and $\bar{b}_s$ represent the allometric intercept and slope estimated at the basal (root) node, $\mathbf{S}$ represents the complete phylogenetic covariance matrix among all $2p-2$ nodes, and $\mathbf{C}$ represents the covariance matrix between slopes and intercepts among species; $\mathbf{1}_{2p-2}$ is a vector of ones of size $2p-2$ and $\otimes$ represents the outer (Kronecker) product. The phylogenetic covariance matrix $\mathbf{S}$ encapsulates the relationships between terminal branches and internal nodes incorporating branch lengths between them (Hadfield & Nakagawa, 2010); thus, it is implicitly assumed that both intercept and slopes evolve in a correlated manner under Brownian motion. In our specific case, previous tests using model comparison indicated that the evolutionary covariance between slopes and intercepts ($\mathbf{C}$) can be suppressed from the model without loss of explanatory power; the multivariate gaussian distributions of a_s and b_s are then independent and can be written separately as

$$a_s \sim \mathcal{N}_{2p-2}(\bar{a}_s \mathbf{1}_{2p-2}, \mathbf{S}) \quad (3.3)$$

and

$$b_s \sim \mathcal{N}_{2p-2}(\bar{b}_s \mathbf{1}_{2p-2}, \mathbf{S}). \quad (3.4)$$

Therefore, this model produces estimates for the posterior distribution of the parameters that describe the relationship between logCS and CAC projection for each terminal OTU and also for ancestral nodes simultaneously, enabling us to track changes in both parameters along the phylogeny. In order to test the robustness of such ancestral node reconstructions for a_s and b_s , we also produced point estimates under a linear parsimony model, implemented in Mesquite (Maddison & Maddison, 2011); estimates for all parameters (slopes and intercepts) using both models yield similar results (not shown), indicating that the Bayesian estimates we used here are robust regardless of the model chosen to reconstruct ancestral allometric parameters.

We used uniform prior distributions for all a_s and b_s . In order to sample the posterior distribution for our model, we used a MCMC sampler with 250000 iterations, comprising a burnin period of 50000 iterations and a thinning interval of 200 iterations after burnin to avoid autocorrelations in the posterior sample, thus generating 1000 posterior samples for all parameters we estimate. We performed a handful of runs with different starting values and pseudo-random number generator seeds to ensure convergence; with these values for iteration steps, we achieved convergence in all MCMC runs.

Using these posterior samples for static allometric parameters, we tested whether a given intercept or slope in any node of the tree (terminal or ancestral) deviates from the values estimated for the root of the tree by computing 95% credible intervals for the difference between the parameter estimated at both points. If this interval excludes the null value, we considered that as evidence that the parameter (a_s or b_s) has changed in that particular node.

Morphological Integration

Using our landmark configuration database, we calculated 38 interlandmark distances (Figure 1.2), based on previous works on mammalian covariance patterns (Cheverud, 1995, 1996b; Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b). Within each OTU, we estimated

and removed fixed effects of little interest within the present context using generalized linear models (Figure 1.1). Using residuals from each model, we estimated covariance and correlation **P**-matrices for all OTUs.

In order to represent modularity patterns embedded in these **P**-matrices, we used Modularity Hypothesis Indexes (MHI; Porto *et al.*, 2013), which are scale-free measurements of
5 variational modularity. We estimated MHIs for each trait subset depicted in Table 1.2 for all 109 **P**-matrices, using the equation

$$\text{MHI} = \frac{\bar{\rho}_+ - \bar{\rho}_-}{\text{ICV}}$$

where $\bar{\rho}_+$ represents the average correlation within a given subset, $\bar{\rho}_-$ represents the average correlation of the complementary set, and ICV is the coefficient of variation of eigenvalues
10 of the associated covariance matrix, which is a measurement of the overall integration between all traits (Shirai & Marroig, 2010).

We estimated the relationship between MHIs and static allometric parameters using phylogenetic mixed linear models; we adjusted such models separately for each trait subset, using MHI values as response variables and allometric parameters as predictors. We also use a
15 Bayesian framework to estimate such models. In order to evaluate which parameters are sufficient to explain variation in MHI values, we adjusted four different models: one for static intercepts alone (a_s), one for static slopes (b_s), and a third model that considers the joint effect of both parameters, without interactions ($a_s + b_s$). We compared these three models using the Deviance Information Criterion (DIC; Gelman *et al.*, 2004), which decreases as a
20 function of the average posterior log likelihood and increases as a function of the number of parameters considered; in a equivalent manner to Akaike's (1974) Information Criterion, the model with the smallest DIC is considered the best fit; models whose difference in DICs is lower than two are considered equivalent.

Software

25 All analysis were performed under R 3.2.2 (R Core Team, 2015). Our code for the estimation of local shape variables can be found at <http://github.com/wgar84>. We performed MCMC sampling for all models using the MCMCglmm package in R (Hadfield, 2010). In order to obtain symmetrical landmarks configurations, we used code provided by Annat Haber, available at <http://life.bio.sunysb.edu/morph/soft-R.html>.

Results

We obtained local shape variables (Márquez *et al.*, 2012) to represent skull shape variation, and projected them over the Common Allometric Component (CAC; Mitteroecker *et al.*, 2004), which represents the pooled within-species allometric shape variation. We used such component as a proxy for the ancestral allometric relationships for Anthropoid Primates. This axis is associated with positive loadings for Facial traits and negative loadings for Neurocranial traits, thus representing a contrast between these regional sets (Figure 3.1a).

To investigate whether species deviate from such ancestral relationships, we carried out a Bayesian mixed model, considering that slopes and intercepts for the regression between individuals projected over the CAC and log Centroid Size (logCS) within each species (Figure 3.1b) are codependent upon their phylogenetic relationships. The posterior distribution of mean-centered intercepts and slopes (Figure 3.2) indicates that intercepts deviate from the mean in at least seven lineages: increases occur independently in *Pithecia* and *Callicebus* within Pitheciids, in the clade composed of Callithrichinae and *Aotus*, and in *Alouatta* within Atelids, while lower intercepts are found in *Homo* and *Pan* within Hominidae, in *Hylobates* within Hylobatids, and also in *Ateles* within Atelids (Figure 3.2a). It is noteworthy that we mean-centered logCS values for the entire sample, and intercepts then refer to the point at which allometric trajectories for all species cross this mean value (Figure 3.1b); more generally, as intercepts are defined in the same scale as the response variable (CAC projections), lower intercepts represent increased neurocranial capacity and shorter faces, while the converse is valid for higher intercepts (Figure 3.1a). Allometric slopes deviate from the mean in only three species: *Homo sapiens* and both representatives of *Gorilla*, all with slopes below the mean (Figure 3.2b).

The association between allometric parameters and modularity was measured using regression models of Modularity Hypothesis Indexes (MHI; Porto *et al.*, 2013) for two regions (Face and Neurocranium) and three sub-regions within each region (Oral, Nasal and Zygomatic in the Face; Orbit, Vault and Basicranium in the Neurocranium) on the estimated allometric parameters. Model selection using Deviance Information Criteria (DIC; Gelman *et al.*, 2004) showed that models considering only intercept variation only always have the worst fit. In most cases, models including only the slope were better than or equal ($\Delta\text{DIC} < 2$) to models with both parameters (intercept and slope), suggesting a negligible effect of the intercept. For the Neurocranium region and the Vault sub-region, the model including intercept and slope yielded a better fit than the slope-only model; in both cases, however, the effects of the intercept were not significant when one takes into account the posterior distribution of the

Figure 3.1: Allometric shape variation. (a) Shape variation associated with the Common Allometric Component (CAC). (b) Random regression of projections on the CAC over log Centroid Size. A regression line is adjusted for each species, and dots represent their mean values.

Figure 3.2: Deviations from ancestral allometry. (a) Intercepts; (b) Slopes. Asterisks indicate whether a given deviation is distinct from zero, considering 95% credible intervals.

regression coefficients (results not shown). Thus, the models including only allometric slopes provided the best representation of the association between allometric parameters and MHI values (Table 3.1). It should be noted that none of the models tested for the Orbit and Basicranium sub-regions showed any association between allometric parameters and MHI values, therefore these sub-regions were not included in the model selection. Allometric slopes showed opposite effects on MHI values for Face and Neurocranium (Figure 3.3), and, with the exception of the Orbit and Basicranium, sub-regions followed the same pattern as the more inclusive partition in which they are contained (Figure 3.4).

Table 3.1: Model selection for phylogenetic linear models of AVG indexes over allometric parameters. Values are shown as deviations from the model with the lowest DIC.

	ΔDIC		
	b_s	a_s	$a_s + b_s$
Face	0.00	41.92	1.08
Neuro	8.42	30.27	0.00
Oral	0.88	15.01	0.00
Nasal	0.00	5.86	0.93
Zygomatic	0.00	9.48	2.59
Vault	3.04	14.95	0.00

Discussion

The hypothesis that allometry may act as a constraint, deflecting changes in morphological, physiological and other organismal aspects along a size-trait gradient (e.g. Gould, 1974) has historically been confronted with the opposing view that allometry may be in fact much more dynamic and evolvable (e.g. Kodric-Brown *et al.*, 2006; Frankino *et al.*, 2009). This paradox may exist due to different meanings for allometry (Houle *et al.*, 2011): one related to any relationship of a given trait with size measured on any scale, while the other associates allometry with the multiplicative nature of biological growth and local differences in growth driven by development (Huxley, 1932), thus defining allometry strictly in a log-log scale. Huxley's original definition allows linking allometric relationships to the population- and individual-level processes that control both the timing and rate of life-history events, such as birth, weaning, and sexual maturation (Houle, 2001; Voje *et al.*, 2013; Pélabon *et al.*, 2014) and their physiological consequences (West *et al.*, 1997, 2001). Considering the few studies estimating evolvability of narrow-sense allometry, only two demonstrate changes in allometric slopes (Tobler & Nijhout, 2010; Bolstad *et al.*, 2015).

Figure 3.3: Allometry and Integration. Modularity Hypothesis Index (MHI) regressed over allometric slope deviations (b_s). For each global trait set (Face and Neurocranium), the shaded region around regression lines indicate 95% credible intervals around estimated regression parameters.

The results from the phylogenetic random regression indicate that substantial changes have occurred on intercepts, especially in New World Monkeys (Figure 3.2a). As we expected, the role of natural selection on the evolution of body size in this group (Marroig & Cheverud, 2005, 2010; Marroig *et al.*, 2012) can be related to some of these changes. For example, in the lineage composed of marmosets, tamarins and owl monkeys (Callithrichinae and Aotinae), selection for reduced body size is pervasive, and associated with shifts in prenatal growth rates, considering a maintenance in relative gestation lengths with respect to their sizes (Marroig & Cheverud, 2009). The increased intercepts in this lineage may be related to a reduction in endocranial volume relative to body size in comparison to other Platyrrhines (Hartwig *et al.*, 2011), since brain growth occurs mostly during the prenatal phase in primates (Lieberman, 2011). Increased intercepts has also been observed in howler monkeys (*Alouatta*), in which reduced brain capacity both in terms of mass and volume has been associated with a trade-off between a herbivorous diet and the energetic requirements of a large brain (Aiello & Wheeler, 1995). It is also noteworthy that, in some lineages of New World Monkeys, selection for changes in body size did not affect intercepts or slopes; for instance, selection for increased body size in *Cebus* (Marroig *et al.*, 2012) produced only a correlated response on skull shape, along the same allometric relationships observed in its sister genus *Saimiri*

Figure 3.4: Modularity Hypothesis Indexes (MHI) *vs.* allometric slope deviations (b_s). For each local trait set, the shaded region around regression lines indicate 95% credible intervals around estimated regression parameters.

(Marroig, 2007). The converse is also true: while increased intercepts were observed in saki monkeys (*Pithecia*), diversification in this group can be explained solely as a consequence of drift, at least with respect to skull morphology (Marroig & Cheverud, 2004).

We have detected significant divergence in allometric slopes in only two lineages, *Homo* and *Gorilla*, although some changes might also have occurred in other lineages such as *Alouatta* and *Callithrix* (Figure 3.2b). Even though *Homo* and *Gorilla* are closely related and their estimated slopes are lower than average, these lineages exhibit two distinct patterns of changes in ontogenetic trajectories (Mitteroecker & Bookstein, 2008). For *Gorilla*, this change mostly represents an increase in neurocranial capacity near the end of their trajectory, while for *Homo* it represents deviations occurring since the beginning of post-natal growth affecting mostly Oral and Vault traits, which results in an overall dissociation between skull shape and size in this lineage, when compared to other Hominoids. These changes in ontogenetic trajectories thus have effects on the allometric relationships in the adult populations of *Homo* and *Gorilla*, which can be observed as shallow slopes in our model (Figure 3.2b) but also can be observed in the multivariate allometric relationships in both *Homo* and *Gorilla*, when compared to the Common Allometric Component (Figure 3.5).

Considering that the early evolution of *Homo* has involved directional selection for both increased neurocranial capacity and reduced facial elements at least in its initial stages of diversification (Ackermann & Cheverud, 2004; Schroeder *et al.*, 2014) and given the ancestral allometric relationships for Anthropoids (CAC; Figure 3.1a), this pattern of selection would produce a correlated response for reduction in skull size (Lande, 1979). Thus, since skull size for *Homo* is similar to its sister lineages (Figure 3.1b), the shift in allometric relationships on the former may be indicative of a joint effect of directional selection for reduction of the average CAC value and stabilizing selection on skull size, given that the changes in allometric slopes detected in *Homo* might be the result of shifts in the relationship between size and shape in the adaptive landscape for this lineage (Jones, 2007; Arnold *et al.*, 2008; Jones *et al.*, 2012; Melo & Marroig, 2015), which cascades into changes in pleiotropic effects mediated by epistatic interactions (Pavlicev *et al.*, 2008, 2011).

Concurrently, these changes in *Homo* may also be the result of changes in environmental effects associated with food processing, especially for Facial traits, since part of the integration observed in this region can be attributed to the effect of post-natal muscular activity on bone growth, induced by masticatory loadings (Hallgrímsson & Lieberman, 2008; Zelditch *et al.*, 2009; Herring, 2011). Furthermore, there is evidence for differences in integration for Facial traits associated with differences in diet among human populations (Paschetta *et al.*, 2015). The act of processing food and cooking meals may already have produced

Figure 3.5: Direction of Multivariate Allometry in *Homo* and *Gorilla*. Colors represent loadings of the allometric component for local shape variables in both representatives of *Gorilla* and in *Homo sapiens*, depicted over the mean shape for each species. The direction of the Common Allometric Component (Figure 3.1a) is also represented for comparison.

changes in physiological aspects of food intake in humans (Wrangham & Conklin-Brittain, 2003), as the oldest paleoanthropological evidence on the use of fire by Hominins dates back to 1.6 Million Years ago (Roebroeks & Villa, 2011), which coincides with the emergence of modern forms *Homo* such as *H. erectus*, with reduced Facial morphology (Wrangham *et al.*, 1999). Hence, this cultural aspect of our ecology may also contribute to shallower allometric relationships in skull traits, since food processing and cooking reduces masticatory loadings associated with its intake.

Shifts in allometric slopes throughout Anthropoid diversification imply changes in the strength of association between Facial and Neurocranial traits, in opposing directions (Figure 3.3); furthermore, allometric slopes explain this effect on integration to a greater extent than intercepts (Table 3.1). Thus, although the distinction between these two regions is the by-product of developmental interactions since the initial steps of skull ontogeny (Hallgrímsson & Lieberman, 2008), allometric relationships reinforce the effect of such interactions due to the induction of facial growth by muscular activity from weaning to adulthood (Zelditch *et al.*, 1992, 2009), since post-natal growth represents a great portion of body size variation observed in this group. Hence, groups with steeper slopes, such as *Ateles* and *Alouatta*, exhibit strong Facial integration and lower Neurocranial integration, while groups with shallower slopes, such as Callithrichines, *Homo*, and *Gorilla*, exhibit similar integration values for both regions, indicating that constraints over allometric relationships are associated with the maintenance of the functional interactions among skull traits. Although the bulk of this effect over phenotypic variation may have an environmental origin (Cheverud, 1982b), muscular activity is necessary for the proper development of osteological elements (Herring, 2011), and even though such functional interactions depend on the environment, they are canalized in a predictable manner given the existence of epigenetic effects through skull development (Lieberman, 2011). These features of developmental systems may be a consequence of the selective pressures these systems suffer for both robustness and replicability (Wagner & Altenberg, 1996; Hansen, 2011; Pavlicev & Hansen, 2011), as they exhibit behaviors similar to machine learning algorithms (Watson *et al.*, 2013).

Here, we demonstrated under a comparative framework that allometric intercepts are more labile than slopes in a macroevolutionary scale and also provided a clear link between allometric relationships and morphological integration in Anthropoid Primates. The consistence of allometric slopes through Anthropoid diversification and the remarkable changes in *Homo* and *Gorilla* therefore provide novel evidence showing that allometric slopes may represent an important constraint for mammals; however, such constraint is not absolute and may be overcome in particular situations. The approach we used here

to estimate allometric parameters benefits from properly separating variation within and among populations, considering the interdependence arising from phylogenetic relatedness simultaneously. For instance, the same approach could be used to evaluate variation in allometric parameters between individuals in a population whose genealogy is available,
5 enabling us to quantify the influence of both genetic and environmental factors over these parameters. In fact, given the availability of a strong theoretical background relating allometries defined at different levels of biological organization (Pélabon *et al.*, 2013), further work on the genetic basis of variation in allometry and its interactions with developmental systems may be critical to fully understand the conditions in which this feature evolves.

Chapter 4

A Phylogenetic Analysis of Shape Covariance Structure

Abstract

Phenotypic traits evolve in a concerted manner due to developmental and functional interactions, mediated by the dynamics of adaptive landscapes; the codependence among traits arising from these three factors is captured by genetic (G) and phenotypic (P) covariance matrices. Mammalian skull development produces an intricate pattern of tissue organization and mutual signaling that integrates this structure, although the set of functions it performs is quite disparate. The interplay between these interactions, and their relationships with the adaptive landscape may thus influence divergence in covariance structure among sister lineages. Here, we evaluate the stability of phenotypic covariance structure in skull size and shape along the diversification of Anthropoid Primates under a phylogenetic framework. We estimate diversity in covariance structure, testing hypotheses concerning the phylogenetic distribution of P-matrix variation; a second approach allows us to pinpoint which traits are associated with such variation. Tests for the divergence in matrix structure indicate that most changes occurred in the basal split between Platyrrhini and Catarrhini; such changes occur within both Orbital and Basicranial trait sets, while Oral, Nasal and Vault trait sets present stable associations along the Anthropoid phylogeny. Therefore, changes in P-matrix structure among Anthropoids are restricted to trait sets whose functional significance is associated with the accommodation of the two precursor tissues that compose the skull, while the stability in the remaining regions hints at the stability of the underlying functional relationships imposed by the adaptive landscape.

Introduction

Phenotypic traits evolve in a concerted manner either because of shared genetic and developmental processes or joint effects on fitness (Gould & Lewontin, 1979; Lande, 1979; Felsenstein, 1988; Zeng, 1988), and response to natural selection is maximized when both
5 effects coincide, while discordance between them may hinder a direct response (Schluter, 1996). The additive genetic covariances among traits (G) and the partial regression coefficients of fitness on traits (β) provide linear approximations for both these effects in the characterization of phenotypic change across generations (Lande, 1979; Rice, 2002). Therefore, the additive covariances in G represent the codependency among traits due
10 to pleiotropy and linkage disequilibrium, characterizing a linear genotype/phenotype map centered on the mean phenotype (Wagner, 1984, 1996; Cheverud, 1996a).

However, the structure of pleiotropic interactions depends on the local curvature of the genotype/phenotype map, traditionally represented in quantitative genetics as either dominance or epistasis (Rice, 1998, 2004; Wolf *et al.*, 2001). The effect of epistatic *loci* on
15 the modulation of pleiotropic interactions has been identified in experimental settings (e.g. Cheverud *et al.*, 2004; Wolf *et al.*, 2005; Pavlicev *et al.*, 2008), thus indicating that populations may harbor genetic variation in the linear components of genotype/phenotype maps. Genetic covariances among phenotypic traits then evolve as a consequence of shifts in allele frequencies in these *loci*, for example in response to genetic drift and founder effects
20 (Goodnight, 2000; Brito *et al.*, 2005). local features of the adaptive landscape may also impact genetic covariances among traits, as G is thought to match the patterns imposed by stabilizing selection and mutational effects (Lande, 1980; Cheverud, 1984; Jones, 2007). It is noteworthy that changes in genetic covariances due to the non-linear structure of adaptive landscapes can be explained just by considering shifts in linkage disequilibrium
25 among *loci* (Turelli, 1988), without any consideration to epistatic pleiotropy. However, recent experimental data (Careau *et al.*, 2015) and simulation-based models (Jones *et al.*, 2012; Melo & Marroig, 2015) have demonstrated the effects of directional selection on the structure described by G , thus indicating that the linear component of adaptive landscape can also have an impact on genetic covariances due to genetic variation in pleiotropic interactions.

30 The relationship between adaptive landscapes and intrapopulation covariance structure is mediated through performance (Arnold, 1983), which may be thought as a dynamical property of phenotypes; thus, the separation between developmental and functional interactions as two distinct factors shaping phenotypic covariance structure is blurry at least (Cheverud, 1996a; Zelditch & Swiderski, 2011). For example, in the mammalian skull, precursor tissues

(paraxial mesoderm and neural crest) originate two distinct regions, Face and Neurocranium, which exhibit marked contrasts in terms of functions which they perform, interactions with soft tissues, and response to developmental milestones, such as birth or weaning (Hallgrímsson & Lieberman, 2008; Lieberman, 2011). Thus, the Oral, Nasal, and Zygomatic regions are associated with the Face and are responsible for mastication, respiration, and the attachment of the muscular apparatus involved in mandibular articulation. The Vault, Orbit and Base regions are associated with the Neurocranium and are responsible for encasing and protecting both brain and eye, and for supporting and connecting the skull with the remaining organismal systems; the Vault also houses muscle attachment sites associated with mandibular articulation, indicating that, to some degree, regions may be involved in more than one function and that some functions may be shared between them. The contrast between these regions is the result of distinct gene expression profiles, which are further changed by the diffusion of signaling factors, thus generating a feedback loop of cell and tissue differentiation (Turing, 1952; Marcucio *et al.*, 2005; Meinhardt, 2008; Franz-Odenaal, 2011; Xu *et al.*, 2015).

These signaling factors may target specific cell lineages, but the contact between adjoining tissues may produce correlated changes between them due to mechanical interactions and through mutual signaling cascades that induce specific behaviors in neighbor tissues, and such interactions are necessary for the proper development of both tissues (Cheverud *et al.*, 1992; Ravosa *et al.*, 2000; Jiang *et al.*, 2002; Marcucio *et al.*, 2011). For instance, cranial Vault growth, which occurs on the final stages of pre-natal development, is a result of the tension exerted by the growing brain on its encircling membranes, inducing them to secrete signaling factors which promote bone growth (Opperman, 2000; Rice *et al.*, 2003), and such mechanism promotes a tight association between the Vault and the brain. Furthermore, post-natal facial growth is induced by muscular activity related to masticatory function, and this effect dominates post-natal skull growth (Zelditch *et al.*, 1992; Herring, 2011), since muscle activity mainly affects the focal Oral region in which muscular forces are exerted, but other skull regions to which these muscles are attached are also affected to a lesser extent, such as the Zygomatic and Vault. Thus, development is composed of a series of such events, and it may be difficult to isolate the effect each individual process has on covariance structure, given their spatial overlap (Hallgrímsson *et al.*, 2009); hence, both the temporal hierarchy and the spatial organization of developmental and functional interactions influence the patterns embedded in genetic covariance structure. Moreover, these molecular aspects of developmental regulation motivate distinct definitions for pleiotropy (Zelditch & Swiderski, 2011): one in which the focal *locus* affects a series of developmental processes within a single cell lineage (*intrinsic, genetic effect*), and another in which processes in one cell lineage

depends on the expression of the focal *locus* on another (*extrinsic, epigenetic effect*); thus, such definitions impart spatial structure to pleiotropic and epistatic interactions.

Empirical evidence on long-term changes on genetic covariance structure (comparative quantitative genetics; Steppan *et al.*, 2002) rely on the correspondence between **G** and its phenotypic counterpart **P**, since estimating **G** demands high sample sizes and nonetheless such estimate is prone to substantial error, since sample units are families rather than individuals (Meyer, 1991; Houle & Meyer, 2015). The similarity in covariance structure between **P** and **G** (“Cheverud’s Conjecture”; Cheverud, 1988; Roff, 1995), has been supported in different trait systems and organisms (e.g.: Waitt & Levin, 1998; Dochtermann, 2011; Garcia *et al.*, 2014), indicating that environmental effects (**E**) are either uncorrelated or exhibit patterns similar to **G**, since these effects exert their influence over phenotypes through the same developmental processes in which genetic variation is structured (Rice, 2002, 2004).

The comparative analysis of phenotypic covariance structure shows that **P** can remain stable in macroevolutionary scales (e.g. Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b; Kolbe *et al.*, 2011), and such stability may be a consequence of the alignment between phenotypic covariance structure and the local features of the adaptive landscape acting over different lineages. In New World Monkeys (Platyrrhini), divergence in body size among lineages is closely associated with shifts in diet composition (Rosenberger, 1992). Size represents the main feature of both genetic (Cheverud, 1995, 1996b) and phenotypic (Marroig & Cheverud, 2005) covariance structure, and for some groups within New World Monkeys, such as Atelids and Callithrichines, divergence in body size is a direct consequence of directional selection (Marroig & Cheverud, 2010; Marroig *et al.*, 2012). On the other hand, dietary or locomotory divergence between sister lineages can rearrange the patterns expressed in **P** in some situations (Young & Hallgrímsson, 2005; Young *et al.*, 2010; Monteiro & Nogueira, 2010; Haber, 2015), and such reorganization hints at changes in the underlying architecture of the genotype/phenotype map, indicating that the patterns embedded in the adaptive landscape can overcome constraints imposed by development (Jamniczky & Hallgrímsson, 2009).

The adaptive radiation of phyllostomid bats, for instance, has involved rearrangements of mandibular phenotypic covariance structure (Monteiro & Nogueira, 2010), as the dietary divergence in this group from a probable insectivore ancestor towards more specialized diets, such as frugivory or sanguivory, imply distinct functional relationships among mandibular components; thus, such reorganization may indicate heterogeneity in the non-linear aspects of the adaptive landscape for different phyllostomid lineages. For bats in general, their specialized locomotory behavior is associated with the decoupling between fore- and

hindlimb covariance structure when compared to other mammalian lineages (Young & Hallgrímsson, 2005); to a lesser extent, a similar pattern can be observed in hominoids when compared to remaining Anthropoids (Young *et al.*, 2010). Another example of divergence in covariance structure among lineages due to changes in functional demands are those observed in ruminants (Haber, 2015), which may related to increased metabolic requirements for occupying open habitats, as the divergence between Bovidae and Cervidae is associated with increased Nasal integration in the former. Thus, divergence among sister lineages may impart a reorganization of both genetic and phenotypic covariance structure of phenotypic traits if such divergence is associated with distinct functional relationships, represented by the non-linear components of adaptive landscape.

Currently, there is an abundance of methods dedicated to the comparison of covariance structures (e.g. Krzanowsky, 1979; Phillips & Arnold, 1999; Cheverud & Marroig, 2007), and such methods are often focused on constructing hypothesis of similarity or dissimilarity; however, in some situations, either hypotheses can be rejected in the same comparison (Ackermann & Cheverud, 2000; Marroig & Cheverud, 2001; Haber, 2015). By expressing covariance matrix (dis)similarities with a single metric, these methods lack an explicit way of describing the differences in covariance structure between pairs of covariance matrices; furthermore, these methods also lack a direct manner to incorporate phylogenetic relatedness in pairwise comparisons. While extensions for these methods have been proposed to deal with the first limitation (e.g. Hansen & Houle, 2008; Hine *et al.*, 2009; Marroig *et al.*, 2011), the second issue is usually resolved by comparing the set of pairwise matrix comparisons with the set of phylogenetic distances among lineages (e.g. Marroig & Cheverud, 2001; Oliveira *et al.*, 2009b). Although such comparison provides a first approximation to this problem, it has the same problems as comparing covariance matrices themselves, that is, summarising the relationship between patterns expressed in $\mathbf{P}$ or $\mathbf{G}$ in a set of lineages and their phylogenetic relatedness to a single value.

Objectives

The interplay between developmental and functional interactions, and their relationships with the topology of adaptive landscapes may influence the divergence between covariance structure in sister lineages. In the present work, we evaluate the stability of phenotypic covariance structure in skull size and shape along the diversification of Anthropoid Primates. We build upon approaches proposed by other authors (Marroig *et al.*, 2011; Aguirre *et al.*, 2013; Haber, 2015) in order to explicitly incorporate phylogenetic relationships into the comparative analysis of covariance structure, under the hypothesis that different cranial

regions will exhibit different degrees of stability among sister lineages, thus producing a non-random pattern of changes in covariance structure.

Methods

Sample

5 Our sample consists of 5108 individuals in 109 species, distributed throughout all major Anthropoid clades above the genus level, comprising all Platyrrhini genera and a substantial portion of Catarrhini genera. We associate this database with a ultrametric phylogenetic hypothesis for Anthroidea (Figure 1.1), derived from Springer *et al.* (2012).

Individuals in our sample are represented by 36 landmarks; these landmarks were registered
10 using a Polhemus 3Draw and a Microscribe 3DS for Platyrrhini and Catarrhini, respectively. Twenty-two unique landmarks represent each individual (Figure 1.2), since fourteen of the 36 registered landmarks are bilaterally symmetrical. For more details on landmark registration, see Marroig & Cheverud (2001) and Oliveira *et al.* (2009b). Databases from both previous studies were merged into a single database, retaining only those individuals in which all
15 landmarks from both sides were present. In the present work, we considered only covariance structure for the symmetrical component of variation; therefore, prior to any analysis, we controlled the effects of variation in asymmetry. We followed the procedure outlined in Klingenberg *et al.* (2002) for bilateral structures by obtaining for each individual a symmetrical landmark configuration, averaging each actual shape with its reflection along
20 the sagittal plane.

We used this database to obtain local shape variables (Márquez *et al.*, 2012), which represent infinitesimal volumetric expansions or retractions, calculated as the natural logarithm determinants of derivatives of the TPS function between each individual in our sample and a reference shape (in our case, the mean shape for the entire sample, estimated from
25 a Generalized Procrustes algorithm). Such derivatives were evaluated at the midpoints between pairs of landmarks represented in Figure 1.2, for a total of 38 local shape variables.

After obtaining these values, we estimated covariance **P**-matrices for size (represented by the natural logarithm of Centroid Size) and local shape variables after removing fixed effects of little interest in the present context, such as sexual dimorphism, for example.

30 These effects were removed through a multivariate linear model adjusted for each species, according to Figure 1.1. We adjusted such models under a Bayesian framework, sampling

100 residual covariance matrices from the posterior distribution of each model. These distributions allow us to estimate uncertainty for any parameters derived from these matrices as credibility intervals; furthermore, since posterior distributions are conditional upon the prior distribution we used — a uniform Wishart distribution — every matrix sampled from these posterior distributions is also a realization of a Wishart distribution, therefore positive-
 definite regardless of sample size (Gelman *et al.*, 2004); in this framework, lower sample
 sizes imply in broader and less informative credibility intervals. For each posterior sample,
 we estimated geometric mean covariance matrices, since such mean respects the underlying
 structure of the Riemannian manifold in which positive-definite symmetric matrices lie
 (Moakher, 2005, 2006); thus, these mean **P**-matrices are also positive-definite, regardless of
 sample size. For each species, we ran independent models, with 13000 iterations of MCMC
 sampling, discarding the 3000 initial iterations as a burnin period and further sampling
 one covariance matrix per 100 iterations to avoid autocorrelations induced by sequential
 sampling.

Phylogenetic Decomposition of Matrix Diversity

In order to evaluate the distribution of covariance structure diversity during Anthropoid diversification, we estimated Riemannian distances among all pairs of mean **P**-matrices, according to the definition given by Mitteroecker & Bookstein (2009); for any pair of positive-definite covariance matrices $\mathbf{C}_i$ and $\mathbf{C}_j$ of size $p \times p$, the distance $d(\mathbf{C}_i, \mathbf{C}_j)$ is given by

$$d(\mathbf{C}_i, \mathbf{C}_j) = \sqrt{\sum_{k=1}^p \ln^2 \lambda_k(\mathbf{C}_i \mathbf{C}_j^{-1})} \quad (4.1)$$

where $\lambda_k(\cdot)$ refers to the k -th eigenvalue obtained from the spectral decomposition of a given matrix, in this case the product $\mathbf{C}_i \mathbf{C}_j^{-1}$. This distance among pairs of **P**-matrices is negatively correlated with Random Skewers comparisons (Figure 4.1), a measurement of matrix similarity explored elsewhere (Cheverud & Marroig, 2007); such correlation indicates that our exploration of Riemmanian distances among **P**-matrices could be extended to
 other types of measurement of similarity or dissimilarity among these objects without
 hindrances.

Using these distances among **P**-matrices, we estimate matrix diversity at each node of the phylogenetic tree of Anthropoidea using a measurement of the weighted distance among the distributions of matrix distances for the two descending edges, based on Pavoine et al.

Figure 4.1: The relationship between Riemannian Distances and Random Skewers among pairs of P-matrices. Each observation is the posterior mean value for each statistic, estimated from 100 posterior samples obtained for each OTU. Colors indicate the lower sample size in each pair, according to the legend.

(2010). For a fully resolved tree, diversity w_i on node i is given by

$$w_i = \frac{1}{2} \frac{n_\alpha n_\beta}{n_i n_T} D_\Delta^2(P_\alpha, P_\beta) \quad (4.2)$$

where α and β represent the subsets of descendants from node i , and n refers to the number of OTUs on each set (n_i for the total descendants of node i ; n_T for the total number of OTUs considered; n_α and n_β for the size of descending subsets). $D_\Delta(P_\alpha, P_\beta)$ represents the actual distance between the two distributions P_α and P_β for descending nodes, as formulated by Rao (1982):

$$D_\Delta(P_\alpha, P_\beta) = \sqrt{2 \left(2H_\Delta \left(\frac{P_\alpha + P_\beta}{2} \right) - H_\Delta(P_\alpha) - H_\Delta(P_\beta) \right)} \quad (4.3)$$

where

$$H_\Delta(P) = \sum_{i,j \in P} \frac{d^2(\mathbf{C}_i, \mathbf{C}_j)}{2} \quad (4.4)$$

represents Rao's quadratic entropy among Riemannian distances $d(\mathbf{C}_i, \mathbf{C}_j)$ as defined in Equation 4.1.

Following the framework established by Pavoine et al. (2010), diversity w_i can be normalized as $v_i = w_i / \sum_i w_i$ to represent the percentage of diversity with respect to the total diversity on the phylogenetic tree. We test three different hypothesis regarding the distribution of v_i values through Anthropoid diversification: (1) that P-matrix diversity is concentrated in a single node; (2) that P-matrix diversity is concentrated in a reduced number of nodes; (3) that P-matrix diversity is skewed towards either the root or tips of the phylogeny, in a two-tailed test. We test each hypothesis against the null hypothesis that the distribution of matrix diversity is randomly arranged over the phylogeny; such null hypothesis is represented by randomizing the association between terminal branches and covariance matrices, constructing 9999 distributions of v_i values that represent this scenario. Each test is carried out using a different parameter derived from the distribution of v_i values (described in detail by Pavoine *et al.*, 2010), comparing the actual value obtained with a null distribution constructed using permutations. The third hypothesis can be tested either by considering only the topology of the tree and by also considering branch lengths; both tests are similar to Blomberg's (2003) K test, as they search for a phylogenetic signal in covariance structure diversity.

Characterizing Covariance Matrix Variation

The tests described in the previous section allow us to pinpoint which nodes contribute mostly to divergence in covariance structure; however, these tests are not designed to properly describe the actual changes in **P**-matrix structure that are responsible for such divergence. To actually represent these changes in a comprehensible manner, we combine a number of ordination techniques to reduce the dimensionality of the manifold that contains covariance matrices of size $p \times p$ (Figure 4.2).

For a Riemannian manifold, there exists at least one bijective function defined in the neighbourhood of a given covariance matrix **M** that maps the manifold to an Euclidean space — a hyperplane with $p(p-1)/2$ dimensions also contained in $\mathbb{R}^{p \times p}$ — and equips the manifold with a notion of inner product, thus allowing the construction of an orthonormal basis that can be used to describe variation in **P**-matrix structure. For a covariance matrix **X** in the neighbourhood of **M**,

$$f(\mathbf{X}) = \log(\mathbf{M}^{-\frac{1}{2}} \mathbf{X} \mathbf{M}^{-\frac{1}{2}}) \quad (4.5)$$

represents one possible function. Here, the logarithm operator refers to matrix logarithm; for symmetric positive-definite matrices, this transformation is equivalent to applying the usual logarithm function to the eigenvalues of such matrix and reverting the spectral decomposition. The function defined in Equation 4.5 also transforms the Riemannian distance among covariance matrices defined in Equation 4.1 into Euclidean distances between transformed matrices (Moakher, 2005).

We defined the average matrix among all sampled **P**-matrices as the location parameter **M** to map the entire set of posterior **P**-matrices into an Euclidean space. We then used these **P**-matrices to produce axes of matrix variation using an eigentensor decomposition (Basser & Pajevic, 2007; Hine *et al.*, 2009) obtaining a set of eigentensors and eigenvalues that summarise matrix variation (Figure 4.2a). As a consequence of using the mean covariance matrix for the entire sample over Equation 4.5, the projections over eigentensors we obtained are naturally centered on **M**.

We used the projections of **P**-matrices over these eigentensors as traits in a phylogenetic Principal Component Analysis (pPCA; Jombart *et al.*, 2010), which produces a new set of axes of matrix variation that considers both trait dispersal and phylogenetic relationships among OTUs simultaneously. If **Z** represents a matrix with projections of n **P**-matrices over each eigentensor on its columns, phylogenetic PCs are the eigenvectors obtained from

Figure 4.2: Representation of the steps used to characterize covariance matrix variation. In (a), the set of covariance matrices A, B and C in the neighbourhood of M are projected into an Euclidean space and eigentensors are estimated (PM1 and PM2); in (b), these eigentensors are rotated to incorporate phylogenetic relatedness; in (c), covariance matrices at the upper and lower bounds of the confidence intervals for each axis are returned back to the original manifold. See text for more details.

$$\frac{1}{2n} \mathbf{Z}^t (\mathbf{W} + \mathbf{W}^t) \mathbf{Z} \quad (4.6)$$

where $\mathbf{W}$ represents the matrix of phylogenetic distances between OTUs; here, the distance w_{ij} between tips i and j is the sum of branch lengths from their last common ancestor to both tips. Other definition of phylogenetic distances may be used (see Jombart *et al.*, 2010); the results we show here were not changed by considering different measures of phylogenetic distance, regardless of whether these distances consider branch lengths among OTUs or not.

Such analysis produces both positive and negative eigenvalues, which are respectively associated with variation close to the root of the tree ('Global') and variation close to the tips ('Local') in matrix structure (pPC1 and pPC2 in Figure 4.2b, respectively). Pavoine *et al.* (2010) argues that this contrast between Global and Local components in phylogenetic PCs reflects phylogenetic signal and convergence in trait values, respectively, as observed in the distribution of Moran's (1948, 1950) autocorrelation Indexes for each axis constructed in this manner. This index can be understood as the degree onto which an observed value in a given OTU is determined by the values on its phylogenetic "neighborhood" (as expressed by $\mathbf{W}$), in a similar manner to autoregressive models (Cheverud & Dow, 1985; Cheverud *et al.*, 1985).

For each pPC obtained in this manner, we obtained two covariance matrices by estimating the upper and lower limits of the 95% confidence interval for each axis and mapping these values back to the manifold of symmetric positive-definite matrices (Figure 4.2c), defining the inverse operation associated with Equation 4.5 as

$$f^{-1}(\mathbf{X}) = \mathbf{M}^{\frac{1}{2}} \exp(\mathbf{X}) \mathbf{M}^{\frac{1}{2}} \quad (4.7)$$

where the exponential operator refers to matrix exponential. We used these covariance matrices to describe matrix variation associated with each axis comparing each pair of matrices with the Selection Response Decomposition tool (Marroig *et al.*, 2011), in order to pinpoint which traits are associated with the divergence in covariance structure associated with each pPC.

In order to characterize such divergence in covariance structure with respect to the uncertainty in $\mathbf{P}$ -matrix estimation, we carried out the analyses described in this section with both mean $\mathbf{P}$ -matrices obtained from posterior samples, and with posterior samples themselves, obtaining 100 sets of phylogenetic PCs and 100 sets of paired covariance matrices for each

axis, thus allowing us to estimate a posterior distribution of mean SRD scores for each trait in all pPCs. We use the phylogenetic PCA estimated over mean P-matrices in order to represent the phylogenetic patterns described by each pPC.

We use the posterior distribution of mean SRD scores over traits and pPC to investigate whether these changes in trait-specific covariance structure along Anthropoid diversification are randomly distributed with respect to the skull regions delimited in Table 1.2 by comparing SRD scores estimated within each region for all pPCs. The association between cranial traits and such regions reflect their functional significance and developmental origins.

Software

We performed all analyses under R 3.2.2 (R Core Team, 2015). We fitted Bayesian linear models for estimating posterior P-matrix samples using the *MCMCglmm* package (Hadfield, 2010). Source code for the eigentensor decomposition can be found at <http://github.com/wgar84>. The phylogenetic decomposition of diversity was provided by Pavoine et al. (2010) in their Supplemental Material, while pPCA is implemented in the *adephylo* package (Jombart & Dray, 2010); the SRD method is provided by the *evolgg* package (Melo *et al.*, 2015). In order to obtain symmetrical landmarks configurations, we used code provided by Annat Haber, available at <http://life.bio.sunysb.edu/morph/soft-R.html>.

Results

The distribution of covariance matrix diversity along the Anthropoid phylogeny (Figure 4.3a) indicates that the divergence between Catarrhini and Platyrrhini contributes to approximately 10% of all covariance matrix diversity; within New World Monkeys, the divergence between Atelidae and Cebidae contributes with 4% to covariance structure diversity, while for Catarrhines, the divergence between Hominoidea and Cercopithecoidea contributes 3% to overall covariance matrix diversity. The remaining P-matrix diversity is distributed along the tree, with a consistent decay of explained diversity the closer any given node is from terminal branches. With respect to the hierarchy of tests exploring the phylogenetic distribution of matrix diversity (Table 4.1), all tests reject their null hypotheses of random arrangements of diversity along the Anthropoid phylogeny, thus indicating that covariance matrix diversity exhibits some degree of phylogenetic structure, with these few basal nodes contributing to a greater extent to such diversity.

Figure 4.3: Phylogenetic decomposition of P-matrix variation. (a) Decomposition of matrix diversity over the phylogenetic hypothesis for Anthropoidea; the size of each circle indicates the percentage of diversity on each node, according to the legend. (b) Mean P-matrices of each species projected over the first three and the last phylogenetic Principal Components (G1-3 and L1, respectively); cell colors represent projection values, according to the legend.

Table 4.1: Phylogenetic decomposition of covariance matrix structure diversity.

	Value	Expected ^a	Distance ^b	<i>p</i> -value
Single Node	0.106	0.029	13.455	$< 10^{-4}$
Few Nodes	0.248	0.139	13.545	$< 10^{-4}$
Root/Tip Skewness ^c	0.632	0.505	12.197	$< 10^{-4}$
Root/Tip Skewness ^d	0.381	0.505	-11.067	$< 10^{-4}$

^a refers to the distribution of permuted values;

^b difference between empirical and expected values in standard deviations of the permuted values distribution;

^c considering only topology;

^d including branch lengths.

The eigenvalue distribution of phylogenetic Principal Components that describe **P**-matrix structure (Figure 4.4) shows that individual Global components surpass their Local counterparts in terms of explained variance, such that the first Global component has a larger contribution than the first Local component to interspecific **P**-matrix variation. While there are less positive than negative eigenvalues, Global components explain more than half (57%) of the total **P**-matrix variation, considering mean absolute eigenvalues obtained from their posterior distribution. The posterior distribution of Moran's Index for phylogenetic Principal Components (Figure 4.5) is assymmetric towards positive values, also indicating that the similarity produced by phylogenetic conservatism is greater than the similarity produced by convergence in **P**-matrix structure.

Considering the distribution of matrix projections for each OTU on these pPCs (Figure 4.3b), we observe that the first Global pPC separates New World and Old Monkeys, while the second Global pPC consists of a contrast between Atelids and Cebids, and the third Global pPC generally contrasts Hominids with the remaining Anthropoids, thus indicating a pattern of **P**-matrix variation consistent with those observed in the diversity decomposition summarized in Figure 4.3a. The first Local pPC consists of localized contrasts between sister species, such as the two representatives of *Saimiri*, for instance; however, matrix projections over this axis can be explained by the effect of sample sizes (Figure 4.7). These localized contrasts can thus be explained on the account of substantial sample size differences between sister species.

The variation in trait-specific covariance structure described by these four phylogenetic Principal Components, as captured by comparing the posterior distribution of confidence intervals for each axis using Selection Response Decomposition (Figure 4.6) indicates that, regardless of which axis is considered, traits with lower posterior SRD scores are usually

Figure 4.4: Posterior distribution of eigenvalues obtained for pPCs. Positive eigenvalues are associated with phylogenetic signal; negative eigenvalues are associated with convergence.

Figure 4.5: Posterior bivariate distribution of variances explained by each phylogenetic PC *versus* Moran's Indexes estimated for each axis.

Figure 4.6: Covariance structure variation associated with the first three and last pPCs (Global 1-3 and Local 1, respectively), represented using posterior mean SRD scores. Dotted lines represent average SRD scores for each comparison. Traits are colored according to their association with each skull region, according to the legend.

Figure 4.7: Relationship between sample sizes and the first Local phylogenetic Principal Component.

localized in either Orbit or Basicranium, along with log Centroid Size, which represents the covariance structure associated with allometric relationships. Traits in remaining skull regions (Oral, Nasal, Zygomatic, Vault) consistently exhibit higher SRD scores, thus indicating a more stable covariance structure associated with these regions throughout Anthropoid diversification.

The overall distribution of average posterior SRD scores along the entire set of phylogenetic PCs (Figure 4.8) indicates that this behavior detailed on Figure 4.6 for G1-3 and L1 is the norm also for the remaining pPCs, that is, Orbit and Basicranial traits along with allometric relationships consistently have lower average SRD scores than other skull regions. Notice that the explained amount of variance for intermediate pPCs (roughly from pPC 31 to 61) is very close to zero (Figure 4.4); thus, although matrices describing confidence intervals in these axes can be recovered, the pattern expressed by them should not be taken into account.

Discussion

Since its conception, the hypothesis that functional interactions among morphological traits shape their phenotypic covariance structure (Olson & Miller, 1958) has been complemented with the notion that developmental interactions mediate the functional relationships among traits dynamically; thus, it is difficult to separate the relative contribution of either development or function to phenotypic integration (Cheverud, 1996a), especially considering that the structure of developmental interactions is thought to match the pattern of optimal functional interactions (Lande, 1980; Cheverud, 1984; Wagner, 1996; Jones, 2007), which further entangles both phenomena. Thus, changes in covariance structure between sister lineages should be associated with the interplay between functional and developmental interactions.

Our results indicate that the changes in phenotypic covariance structure associated with all phylogenetic principal components follow a similar pattern with respect to changes in trait-specific covariances (Figures 4.6 and 4.8). Such changes are mostly associated with either Basicranial and Orbital trait sets, along with allometric relationships, represented by the covariances between log Centroid Size and local shape variables. It is noteworthy that while the comparison between covariance matrices in the limits of each phylogenetic PC represents the overall pattern of dissimilarity in trait-specific covariance structure associated with each axis, the actual values obtained represent only a fraction of the overall divergence between Anthropoid lineages. Given any two P-matrices associated with a pair

Figure 4.8: Distribution of mean SRD scores for each trait along the entire ordered set of phylogenetic PCs.

of species, the actual SRD comparison between then will be some linear combination of their divergences along each pPC axis.

Orbital traits are mostly associated with the development of the postorbital bar in Eu-
primates, as opposed to Plesiadapiforms, and in Anthropoids it fully develops into the
5 orbital cavity and postorbital wall (Ravosa & Savakova, 2004). The origins of such structure
have been linked to the distribution of masticatory loadings around the comparatively
large primate eye for the maintenance of a stable, forward-facing visual field even during
feeding behavior (Ravosa *et al.*, 2000). Although most of the cranial Vault originates
from intramembranous ossification induced by the growing brain and thus derived from
10 the neural crest, the influence of mesoderm-derived condensations and its pattern of en-
dochondral ossification is necessary for the proper development of the fronto-nasal and
fronto-zygomatic sutures, which affect the brow ridge and both medial and lateral orbital
walls (Jiang *et al.*, 2002).

In the same manner, the Basicranium originates from a set of thirteen condensations
15 derived from both precursors, which exhibit a mosaic pattern of both endochondral and
intramembranous ossification (Lieberman, 2011). Such processes occurs early during skull
development, and the spatial overlap of developmental processes (Hallgrímsson *et al.*, 2009)
may also explain Basicranium variation in covariance structure. Moreover, the angulation
between anterior and posterior elements of the Basicranium has significantly changed during
20 primate evolution, and such property appears to have evolved in coordination with Facial
growth relative to the cranial Vault, accomodating both structures on each other (Scott,
1958; Lieberman *et al.*, 2000, 2008). Therefore, both Orbit and Basicranium are located in
the boundary between domains of two precursor tissues that originate osteological elements
in the skull, and their proper development is affected by both of them; furthermore, their
25 functional significance is also associated with the accomodation of remaining structures.
These characteristic properties of their development may thus be sufficient to explain their
divergence among lineages in terms of covariance structure.

On the other hand, the structure of phenotypic covariances for Oral, Nasal and Vault
traits remained stable through the course of Anthropoid diversification. These regions
30 exhibit more consistent patterns of developmental processes, when compared to the Orbit
or Basicranium. Both Oral and Nasal regions originate from mesoderm-derived tissue,
and exhibit a great degree of interactions with soft tissue during prenatal development;
the Oral region further suffers the influence of musculoskeletal interactions associated with
Facial growth in postnatal development, which also contributes to its pattern of stability
35 in covariance structure (Zelditch *et al.*, 1992; Herring, 2011; Lieberman, 2011). The cranial

Vault originates mostly from precursors derived from the neural crest and exhibits a more regular pattern of growth, induced by the underlying brain (Lieberman, 2011; Esteve-Altava & Rasskin-Gutman, 2014). While there are prenatal developmental processes associated with the integration of both structures (Marcucio *et al.*, 2005, 2011) and postnatal muscle-bone interactions may be understood as a overall integrating factor, since the skull as whole is affected by such interactions, each of these regions is located within the bounds of precursor components, thus exhibiting a stable association between developmental processes that originate these regions and their functional aspects.

Finally, allometric relationship exhibit changes in covariance structure across phylogenetic principal components with magnitudes equivalent to Basicranial traits (Figures 4.6 and 4.8). While such relationships are explored elsewhere (Chapter 3), it is noteworthy that these changes may be associated with the role of directional selection for body size, which shaped the extant diversity of New World (Marroig & Cheverud, 2005, 2010) and Old World Monkeys (Cardini *et al.*, 2007; Cardini & Elton, 2008), as well as the more complex relationships between size and shape observed within Hominidae (Ackermann & Cheverud, 2004; Mitteroecker *et al.*, 2004; Mitteroecker & Bookstein, 2008; Schroeder *et al.*, 2014). To some extent, selection for body size produces mean shape differences among lineages in a manner consistent with the ancestral allometric relationships (Lande, 1979; Schluter, 1996), but selective pressures for size may alter such relationships depending on the structure of interactions between size, shape and ontogeny (Pélabon *et al.*, 2013, 2014).

The phylogenetic distribution of changes in covariance structure (Figure 4.3) reveals an even distribution of such changes throughout Anthropoid diversification, with three different instances in which more substantial changes have occurred, in order of the estimated P-matrix diversity at each point (Figure 4.3a): the divergence between New World and Old World Monkeys, between Atelidae and Cebidae, and between Hominoidea and the remaining Anthropoids. Results from the phylogenetic principal component analysis (Figure 4.3b) recover a similar pattern, considering the contribution of these lineages to each axis of matrix variation, and the posterior distributions of the associated eigenvalues for the first, second and third phylogenetic principal components (Figure 4.4) indicate that these cladogenetic events can be set apart in terms of their relative contributions to P-matrix diversity.

Although the phenotypic covariance structure for Hominids has been regarded as the most disparate when compared to remaining Catarrhines (Oliveira *et al.*, 2009b) and to mammals in general (Porto *et al.*, 2009), these findings may differ qualitatively from our results due to the use of traditional morphometric measurements, as opposed to the geometric

morphometrics approach we use here. Without any means of controlling the contribution of size variation, interlandmark distances will overestimate the influence of this source of variation over subtle shape relationships; thus, as the divergence in Hominid covariance structure is mostly related to differences in allometric relationships in this lineage (Figure 3.5),
5 interlandmark distances will put more emphasis on these differences. On the other hand, the actual pattern of divergence in the third phylogenetic component (Figure 4.3b) contrasts the Hominoid lineage apart not only from its sister group, Cercopithecoidea, but also from remaining lineages. Thus, the apparent greater contribution from the divergence between Atelids and Cebids to P-matrix diversity may be a consequence of the phylogenetic structure
10 imposed by either analyses, as the contrast produced by the third pPC is not a proper contrast between sister lineages, but rather between Hominoidea and the paraphyletic grouping obtained from removing this group.

However, both events are overshadowed by the separation between New World and Old World Monkeys, since this divergence is associated with an unequivocally higher eigenvalue,
15 as indicated by their posterior distribution (Figure 4.4). The subtle differences in P-matrix diversity such separation represents may indicate that changes in the genetic architecture of cranial traits due to founder effects (Goodnight, 2000) may be a sufficient explanation for this divergence, considering that the effective population sizes in the ancestral population for all Platyrrhines probably plummeted during the transition between Africa and South
20 America. Such transition probably occurred through the colonization of successive island environments that may have occurred in the South Atlantic Ocean during the Eocene (Oliveira *et al.*, 2009a).

It is noteworthy that the pattern recovered by the first Local component (lower right panel of Figure 4.6) also describes a pattern similar to that of Global components, considering
25 that the distribution of lineages in this component scales with sample sizes on a logarithmic scale (Figure 4.7). Basicranial and Orbital traits exhibit overall lower covariances, both among themselves and between other regions, which contributes to more uncertainty in their estimation, and to a stronger sampling effect over their values. However, it is also important to consider that these directions depicted as changes in trait-specific covariance structure
30 using the SRD method are orthogonal in the Euclidean image of the exponential-logarithm map, and as such, they represent different aspects of covariance divergence among lineages, which here are concentrated in the same traits regardless of the phylogenetic structure considered. The association between the first Local component with sample sizes hinders its interpretation as biologically meaningful, in the terms of Jombart's (2010) notion that Local
35 components should be related to convergence among phylogenetically disparate lineages;

rather, this component seems to be associated with divergence in the statistical properties of estimated covariance matrices.

Regardless of such considerations with respect Local components, the prevalence of variation associated with positive, Global components in relation to variation in negative, Local components (Figures 4.4 and 4.5) implies that the contribution of historical constraints is greater than the contribution of convergence among disparate lineages. Considering that either process is associated with a stable structure of changes in covariance structure (Figures 4.6 and 4.8) related to some of the properties of skull development, these results stress out the overall stability of covariance structure in Anthropoids.

Conclusions

The stability of shape covariance structure in Anthropoids may be a consequence of either constraints on mammalian skull development or the prevalence of a constant pattern of functional relationships imposed by stabilizing selection; such dichotomy has already been pointed out by other authors (e.g. Marroig & Cheverud, 2001; Porto *et al.*, 2009; Oliveira *et al.*, 2009b). In the present work, we tend to favor the second point of view, in light of the evidence for the evolution of genetic or phenotypic covariances under directional selection (Jones *et al.*, 2012; Melo & Marroig, 2015; Careau *et al.*, 2015) or under relaxation of stabilizing selection (Jamniczky & Hallgrímsson, 2009). Furthermore, some available comparative data (Monteiro & Nogueira, 2010; Haber, 2015) indicate that some mammal lineages, such as phyllostomid bats and ruminants, have diverged in phenotypic covariance structure probably due to an heterogeneity in the shape of adaptive landscapes resulting from differences in ecological processes acting on different lineages. On the other hand, for Anthropoids, their stability in P-matrix structure may thus be possible as a consequence of stability of the functional relationships imposed over by the adaptive landscape in this lineage, as it seems that in this group, dietary shifts imposed changes on body size only, at least for Platyrrhines (Marroig & Cheverud, 2010).

From an information theory point of view (Brooks *et al.*, 1989), natural selection will increase the correlation between information encoded in a population and the information that represents its environment; the recent suggestion that developmental systems share properties with machine learning algorithms (Watson *et al.*, 2013) only reinforces such view; furthermore, the regulation of developmental systems, through both genetic and epigenetic effects, may also be targeted by selection for both robustness and replicability (Hansen, 2011). These different ways of thinking about developmental and functional interactions mean that the probabilistic distribution of phenotypes, as a consequence of either genetic

information or epigenetic effects, will tend to match the properties of the distribution of fitness in a particular environment. Thus, the maintenance of the same distribution over macroevolutionary timescales indicates that the fitness distribution itself may be stable as well.

Capítulo 5

Considerações Finais

Embora os três capítulos anteriores sejam suficientemente independentes, os resultados obtidos e as discussões desenvolvidas com base nestes resultados, quando tomadas em conjunto, auxiliam no entendimento da influência da escolha de representações para caracterizar variação morfológica e o seu impacto na avaliação de propriedades variacionais em escala macroevolutiva. 5

No Capítulo 2, observamos que grande parte da contenda a respeito das diferenças entre a detecção de padrões de modularidade dentre as diferentes formas de representar variação morfológica são uma consequência dos problemas que resíduos de Procrustes apresentam em representar processos locais de desenvolvimento responsáveis por gerar covariâncias fenotípicas, o que torna difícil a associação entre os padrões obtidos através de tal representação e o conhecimento atual a respeito de desenvolvimento craniano. Tais problemas são resultado dos vieses introduzidos pela superposição, em que uma estrutura de variação isotrópica acaba por ser imposta sobre as estimativas de matrizes de covariância; neste sentido, variáveis locais de forma são uma alternativa válida para a representação destes processos, considerando, em primeiro lugar, sua correspondência com o desenvolvimento craniano, conforme discutido naquele capítulo. Em segundo lugar, esta representação se associa de forma explícita à representação de processos locais de crescimento e diferenciação, considerando sua definição, enquanto resíduos de Procrustes dificultam uma interpretação focada neste aspecto, levando-se em conta a natureza da superposição necessária ao tratamento estatístico de configurações geométricas. Finalmente, a flexibilidade em definir a dimensionalidade da amostra e a posição relativa dos pontos amostrados propicia uma vantagem interessante deste tipo de representação, ainda que a informação contida em toda a configuração registrada seja levada em consideração pelo modelo de interpolação. Ainda que o modelo de comportamento de tecidos adjacentes 25

utilizado aqui seja também questionável, visto que equiparar o comportamento de tecidos biológicos em crescimento e diferenciação ao comportamento de uma placa metálica sob tensão parece distanciar esta representação de qualquer realidade biológica, tal aproximação pode ser substituída por um modelo mais realista, considerando que este modelo é explicitamente incorporado nesta representação.

No contexto desta tese, o Capítulo 3 mostra a associação entre os padrões de modularidade observados sobre distâncias euclidianas e as relações alométricas obtidas por uma representação geométrica da variação morfológica, em que tamanho e forma são separados por definição. Observa-se que grande parte da variação em modularidade das diferentes regiões cranianas quando o primeiro tipo de representação é considerado são uma consequência da variação nas relações alométricas no segundo tipo de representação, dentre as diferentes linhagens de Anthropeidea. Considerando-se quanto a variação de forma pode ser de fato separada da variação de tamanho do ponto de vista dos processos de desenvolvimento, visto que crescimento de tecidos e diferenciação celular são fenômenos intrinsecamente ligados aos processos metabólicos subjacentes a estes níveis de organização, um tratamento matemático que separa tamanho e forma, especialmente se esta última for representada através das variáveis locais de forma, evidencia um contraste entre fenômenos globais de desenvolvimento, responsáveis pelo crescimento, e a diferenciação produzida por fenômenos locais. Dessa forma, ainda que seja possível separar estas duas fontes de variação do ponto de vista analítico, a influência do tamanho sobre a forma nunca deveria deixar de ser considerada quando propriedades variacionais são avaliadas, em qualquer representação.

O Capítulo 4 coloca a questão da similaridade em padrões de covariância em tamanho e forma craniana sob uma perspectiva explicitamente filogenética, e observa-se a estabilidade geral destes padrões ao longo da diversidade de Anthropeidea, reforçando os resultados obtidos anteriormente para este grupo em contribuições anteriores. Aqui, cabem três colocações importantes. Em primeiro lugar, dado o avanço da literatura a respeito de interações epistáticas e pleiotrópicas e do impacto destas interações sobre a variação fenotípica, um cenário onde restrições de desenvolvimento são responsáveis pela manutenção dos padrões de covariância perde força frente ao cenário onde tais padrões são mantidos através da estabilidade dos padrões de seleção estabilizadora impostos pela paisagem adaptativa. Em segundo lugar, favorece-se neste capítulo uma abordagem baseada na estimativa e subsequente interpretação de intervalos de incerteza referentes aos parâmetros descritores de propriedades variacionais, frente a uma abordagem baseada em distribuições nulas, conforme o exposto no Capítulo 2. Apesar dos resultados destes dois capítulos serem compatíveis, uma abordagem baseada nestas estimativa de incerteza permite maior flexibilidade

no tipo de hipóteses que podem ser formuladas e subsequentemente testadas a respeito destas propriedades, principalmente considerando-se de fato o quanto podemos considerar a organização aleatória de processos de desenvolvimento um cenário nulo realmente válido.

Finalmente, durante o desenvolvimento do terceiro capítulo, o conceito de modularidade não é evocado em nenhum momento, e esta foi uma escolha consciente. Ainda que seja possível interpretar aquelas regiões do crânio que se mostram estáveis ao longo da diversificação de Anthroidea como sendo efetivamente módulos, a aplicação do conceito de modularidade no contexto da integração morfológica talvez tenha introduzido distorções na literatura deste campo. O conceito de modularidade é útil no sentido de demonstrar certos paralelos entre níveis distintos de organização biológica; no entanto, o mapa genótipo-fenótipo exposto por Wagner (1996) e Wagner & Altenberg (1996) se mostra distante de uma perspectiva mais aprofundada do desenvolvimento, como por exemplo aquela exposta por Hallgrímsson et al. (2009), que considera a dinâmica inerente ao desenvolvimento como fundamental para o entendimento da estrutura de covariância entre caracteres fenotípicos. É necessário ressaltar que a estrutura do mapa apresentado nestes trabalhos é apenas uma simplificação, derivada da aproximação linear do mapa, conforme Wagner (1984). Neste sentido, a presença de modularidade em um mapa genótipo/fenótipo linear também é uma aproximação à estrutura inerente ao desenvolvimento que desconsidera sua dinâmica, e caso esta simplificação seja encarada de forma literal, o entendimento a respeito dos processos de desenvolvimento e consequentemente da história de vida das populações pode ser ignorado em prol de uma busca desprovida de significado por padrões de modularidade. Ademais, esta desconexão entre os padrões apresentados pela estrutura de covariância e as hipóteses a respeito das interações que produzem tais padrões se reflete e se reforça pela desconexão inerente à morfometria baseada em configurações geométricas, considerando-se os vieses introduzidos pela superposição de Procrustes.

Referências Bibliográficas

- Ackermann, R.R. & Cheverud, J.M. 2004. Detecting genetic drift versus selection in human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 17946–17951.
- 5 Ackermann, R.R. & Cheverud, J.M. 2000. Phenotypic covariance structure in tamarins (genus *Saguinus*): A comparison of variation patterns using matrix correlation and common principal component analysis. *American Journal of Physical Anthropology* **111**: 489–501.
- Adams, D.C. & Otárola-Castillo, E. 2013. geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* **4**:
10 393–399.
- Adams, D.C., Rohlf, F.J. & Slice, D.E. 2004. Geometric morphometrics: Ten years of progress following the “revolution”. *Italian Journal of Zoology* **71**: 5–16.
- Aguirre, J.D., Hine, E., McGuigan, K. & Blows, M.W. 2013. Comparing G : multivariate analysis of genetic variation in multiple populations. *Heredity* **112**: 21–29.
- 15 Aiello, L.C. & Wheeler, P. 1995. The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Current anthropology* 199–221.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716–723.
- Andrade, R.F.S., Rocha-Neto, I.C., Santos, L.B.L., Santana, C.N. de, Diniz, M.V.C. & Lobão,
20 T.P.*et al.* 2011. Detecting Network Communities: An Application to Phylogenetic Analysis. *PLoS Computational Biology* **7**: e1001131.
- Armbruster, W.S., Pélabon, C., Bolstad, G.H. & Hansen, T.F. 2014. Integrated phenotypes: understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369**: 20130245.
- 25 Arnold, S.J. 1983. Morphology, Performance and Fitness. *American Zoologist* **23**: 347–361.

- Arnold, S.J., Bürger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. 2008. Understanding the evolution and stability of the G-matrix. *Evolution* **62**: 2451–2461.
- Basser, P.J. & Pajevic, S. 2007. Spectral decomposition of a 4th-order covariance tensor: Applications to diffusion tensor MRI. *Signal Processing* **87**: 220–236.
- Bastir, M. & Rosas, A. 2005. Hierarchical nature of morphological integration and modularity in the human posterior face. *American Journal of Physical Anthropology* **128**: 26–34. 5
- Blomberg, S.P., Garland Jr., T. & Ives, A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**: 717–745.
- Bolstad, G.H., Cassara, J.A., Márquez, E., Hansen, T.F., Linde, K. van der & Houle, D.*et al.* 2015. Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 201505357. 10
- Bond, M., Tejedor, M.F., Campbell Jr, K.E., Chornogubsky, L., Novo, N. & Goin, F. 2015. Eocene primates of South America and the African origins of New World monkeys. *Nature* **520**: 538–541.
- Bookstein, F.L. 1982. Foundations of Morphometrics. *Annual Review of Ecology and Systematics* **13**: 451–470. 15
- Bookstein, F.L. 1991. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge.
- Bookstein, F.L. 1989. Principal warps: Thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**: 567–585. 20
- Bookstein, F.L., Chernoff, B., Elder, R., Humphries, Smith, G. & Strauss, R. 1985. *Morphometrics in Evolutionary Biology*. The Academy of Natural Sciences of Philadelphia, Philadelphia.
- Brito, R.A., Pletscher, L.S. & Cheverud, J.M. 2005. The evolution of genetic architecture. I. Diversification of genetic backgrounds by genetic drift. *Evolution* **59**: 2333–2342.
- Brooks, D.R., Collier, J., Maurer, B.A., Smith, J.D.H. & Wiley, E.O. 1989. Entropy and information in evolving biological systems. *Biology and Philosophy* **4**: 407–432. 25
- Cardini, A. & Elton, S. 2008. Variation in guenon skulls (I): species divergence, ecological and genetic differences. *Journal of Human Evolution* **54**: 615–637.
- Cardini, A. & Polly, P.D. 2013. Larger mammals have longer faces because of size-related constraints on skull form. *Nature Communications* **4**. 30
- Cardini, A., Jansson, A.-U. & Elton, S. 2007. A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. *Journal of Biogeography* **34**:

1663–1678.

Careau, V., Wolak, M.E., Carter, P.A. & Garland, T. 2015. Evolution of the additive genetic variance–covariance matrix under continuous directional selection on a complex behavioural phenotype. *Proc. R. Soc. B* **282**: 20151119.

5 Cheverud, J.M. 1988. A comparison of genetic and phenotypic correlations. *Evolution* **42**: 958–968.

Cheverud, J.M. 1996a. Developmental integration and the evolution of pleiotropy. *American Zoology* **36**: 44–50.

10 Cheverud, J.M. 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *American Naturalist* **145**: 63–89.

Cheverud, J.M. 1982a. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution* **36**: 499–516.

15 Cheverud, J.M. 1996b. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *Journal of Evolutionary Biology* **9**: 5–42.

Cheverud, J.M. 1984. Quantitative genetics and developmental constraints on evolution by selection. *Journal of Theoretical Biology* **110**: 155–172.

Cheverud, J.M. 1982b. Relationships among ontogenetic, static, and evolutionary allometry. *American Journal of Physical Anthropology* **59**: 139–149.

20 Cheverud, J.M. & Dow, M.M. 1985. An autocorrelation analysis of genetic variation due to lineal fission in social groups of rhesus macaques. *American Journal of Physical Anthropology* **67**: 113–121.

25 Cheverud, J.M. & Marroig, G. 2007. Comparing covariance matrices: Random skewers method compared to the common principal components model. *Genetics and Molecular Biology* **30**: 461–469.

Cheverud, J.M. & Richtsmeier, J.T. 1986. Finite-Element Scaling Applied to Sexual Dimorphism in Rhesus Macaque (*Macaca Mulatta*) Facial Growth. *Systematic Biology* **35**: 381–399.

30 Cheverud, J.M., Dow, M.M. & Leutenegger, W. 1985. The Quantitative Assessment of Phylogenetic Constraints in Comparative Analyses: Sexual Dimorphism in Body Weight Among Primates. *Evolution* **39**: 1335–1351.

- Cheverud, J.M., Ehrich, T.H., Vaughn, T.T., Koreishi, S.F., Linsey, R.B. & Pletscher, L.S. 2004. Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation in morphological integration. *Journal of experimental zoology. Part B, Molecular and developmental evolution* **302**: 424–435.
- Cheverud, J.M., Kohn, L.A.P., Konigsberg, L.W. & Leigh, S.R. 1992. Effects of fronto-occipital artificial cranial vault modification on the cranial base and face. *American Journal of Physical Anthropology* **88**: 323–345. 5
- Cheverud, J.M., Routman, E.J. & Irschick, D.J. 1997. Pleiotropic Effects of Individual Gene Loci on Mandibular Morphology. *Evolution* **51**: 2006–2016.
- Cheverud, J.M., Wagner, G.P. & Dow, M.M. 1989. Methods for the comparative analysis of variation patterns. *Evolution* **38**: 201–213. 10
- Dochtermann, N.A. 2011. Testing Cheverud’s Conjecture for Behavioral Correlations and Behavioral Syndromes. *Evolution* **65**: 1814–1820.
- Drake, A.G. & Klingenberg, C.P. 2010. Large Scale Diversification of Skull Shape in Domestic Dogs: Disparity and Modularity. *The American Naturalist* **175**: 289–301. 15
- Egset, C.K., Hansen, T.F., Le Rouzic, A., Bolstad, G.H., Rosenqvist, G. & Pélabon, C. 2012. Artificial selection on allometry: change in elevation but not slope. *Journal of Evolutionary Biology* **25**: 938–948.
- Escoufier, Y. 1973. Le Traitement des Variables Vectorielles. *Biometrics* **29**: 751–760.
- Espinosa-Soto, C. & Wagner, A. 2010. Specialization can drive the evolution of modularity. *PLoS Comput. Biol.* **6**: e1000719. 20
- Esteve-Altava, B. & Rasskin-Gutman, D. 2014. Beyond the functional matrix hypothesis: a network null model of human skull growth for the formation of bone articulations. *Journal of Anatomy* **225**: 306–316.
- Esteve-Altava, B., Marugán-Lobón, J., Botella, H., Bastir, M. & Rasskin-Gutman, D. 2013. Grist for Riedl’s mill: A network model perspective on the integration and modularity of the human skull. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **320**: 489–500. 25
- Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th ed. Addison Wesley Longman, Harlow, Essex. 30
- Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* **19**: 445–471.

Fleagle, J.G. 2013. *Primate Adaptation and Evolution: 3rd Edition*. Academic Press.

Fortuna, M.A., García, C., Guimarães Jr., P.R. & Bascompte, J. 2008. Spatial mating networks in insect-pollinated plants. *Ecology Letters* **11**: 490–498.

Frankino, W.A., Emlen, D.J. & Shingleton, A.W. 2009. Experimental approaches to studying the evolution of animal form. In: *Experimental evolution: concepts, methods, and applications of selection experiments* (T. Garland Jr. & M. R. Rose, eds), pp. 419–478.

Frankino, W.A., Zwaan, B.J., Stern, D.L. & Brakefield, P.M. 2005. Natural Selection and Developmental Constraints in the Evolution of Allometries. *Science* **307**: 718–720.

Franz-Odenaal, T.A. 2011. Epigenetics in Bone and Cartilage Development. In: *Epigenetics: Linking Genotype and Phenotype in Development and Evolution* (B. Hallgrímsson & B. K. Hall, eds), pp. 195–220. University of California Press.

Fruciano, C., Franchini, P. & Meyer, A. 2013. Resampling-Based Approaches to Study Variation in Morphological Modularity. *PLoS ONE* **8**: e69376.

Garcia, G., Hingst-Zaher, E., Cerqueira, R. & Marroig, G. 2014. Quantitative Genetics and Modularity in Cranial and Mandibular Morphology of *Calomys expulsus*. *Evolutionary Biology* **41**: 619–636.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2004. *Bayesian data analysis*, 2^a ed. CRC Press, New York.

Genini, J., Morellato, L.P.C., Guimarães Jr., P.R. & Olesen, J.M. 2010. Cheaters in mutualism networks. *Biology Letters* **6**: 494–497.

Goodall, C. 1991. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)* **53**: 285–339.

Goodnight, C.J. 2000. Quantitative trait loci and gene interaction: the quantitative genetics of metapopulations. *Heredity* **84**: 587–598.

Goswami, A. & Polly, P.D. 2010. The influence of modularity on cranial morphological disparity in Carnivora and Primates (Mammalia). *PLoS ONE* **5**: e9517.

Gould, S.J. 1966. Allometry and Size in Ontogeny and Phylogeny. *Biological Reviews* **41**: 587–638.

Gould, S.J. 1974. Allometry in primates, with emphasis on scaling and the evolution of the brain. *Contributions to primatology* **5**: 244–292.

Gould, S.J. & Lewontin, R.C. 1979. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of*

London B: Biological Sciences **205**: 581–598.

Grabowski, M.W., Polk, J.D. & Roseman, C.C. 2011. Divergent patterns of integration and reduced constraint in the human hip and the origins of bipedalism. *Evolution* **65**: 1336–1356.

Haber, A. 2015. The Evolution of Morphological Integration in the Ruminant Skull. *Evolutionary Biology* **42**: 99–114.

5

Hadfield, J.D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The {MCMCglmm} {R} Package. *Journal of Statistical Software* **33**: 1–22.

Hadfield, J.D. & Nakagawa, S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology* **23**: 494–508.

10

Hallgrímsson, B. & Lieberman, D.E. 2008. Mouse models and the evolutionary developmental biology of the skull. *Integrative and Comparative Biology* **48**: 373–384.

Hallgrímsson, B., Jamniczky, H., Young, N.M., Rolian, C., Parsons, T.E. & Boughner, J.C. *et al.* 2009. Deciphering the Palimpsest: Studying the Relationship Between Morphological Integration and Phenotypic Covariation. *Evolutionary Biology* **36**: 355–376.

15

Hansen, T.F. 2011. Epigenetics: adaptation or contingency. In: *Epigenetics: linking genotype and phenotype in development and evolution* (B. Hallgrímsson & B. K. Hall, eds), pp. 357–376.

Hansen, T.F. & Houle, D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21**: 1201–1219.

Hartwig, W., Rosenberger, A.L., Norconk, M.A. & Owl, M.Y. 2011. Relative Brain Size, Gut Size, and Evolution in New World Monkeys. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* **294**: 2207–2221.

20

Herring, S.W. 2011. Muscle-Bone Interactions and the Development of Skeletal Phenotype. In: *Epigenetics: Linking Genotype and Phenotype in Development and Evolution* (B. Hallgrímsson & B. K. Hall, eds), pp. 221–237. University of California Press.

25

Hine, E., Chenoweth, S.F., Rundle, H.D. & Blows, M.W. 2009. Characterizing the evolution of genetic variance using genetic covariance tensors. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**: 1567–78.

Houle, D. 2001. Characters as the units of evolutionary change. In: *The character concept in evolutionary biology*, pp. 109–140.

30

Houle, D. & Meyer, K. 2015. Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *Journal of Evolutionary Biology* **28**:

1542–1549.

Houle, D., Pélabon, C., Wagner, G.P. & Hansen, T.F. 2011. Measurement and Meaning In Biology. *The Quartely Review of Biology* **86**: 3–34.

Huckemann, S. 2011. Inference on 3D Procrustes Means: Tree Bole Growth, Rank Deficient
5 Diffusion Tensors and Perturbation Models: Inference on 3D Procrustes means. *Scandinavian Journal of Statistics* no–no.

Huckemann, S.F. 2012. On the meaning of mean shape: manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics* **64**: 1227–1259.

Huxley, J.S. 1932. *Problems of relative growth*.

10 Jamniczky, H.A. & Hallgrímsson, B. 2009. A comparison of covariance structure in wild and laboratory muroid crania. *Evolution* **63**: 1540–1556.

Jiang, X., Iseki, S., Maxson, R.E., Sucov, H.M. & Morriss-Kay, G.M. 2002. Tissue Origins and Interactions in the Mammalian Skull Vault. *Developmental Biology* **241**: 106–116.

Jolicoeur, P. 1963. The Multivariate Generalization of the Allometry Equation. *Biometrics*.

15 Jombart, T. & Dray, S. 2010. adephylo: exploratory analyses for the phylogenetic comparative method. *Bioinformatics* **26**: 1907–1909.

Jombart, T., Pavoine, S., Devillard, S. & Pontier, D. 2010. Putting phylogeny into the analysis of biological traits: A methodological approach. *Journal of Theoretical Biology* **264**: 693–701.

Jones, A.G. 2007. The mutation matrix and the evolution of evolvability. *Evolution* **61**:
20 727–745.

Jones, A.G., Bürger, R., Arnold, S.J., Hohenlohe, P.A. & Uyeda, J.C. 2012. The effects of stochastic and episodic movement of the optimum on the evolution of the G-matrix and the response of the trait mean to selection. *Journal of evolutionary biology* 1–22.

Kamilar, J.M. & Baden, A.L. 2014. What drives flexibility in primate social organization?
25 *Behavioral Ecology and Sociobiology* **68**: 1677–1692.

Kappeler, P.M. & Schaik, C.P. van. 2002. Evolution of Primate Social Systems. *International Journal of Primatology* **23**: 707–740.

Kendall, D.G. 1984. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* **16**: 81–121.

30 Kent, J.T. & Mardia, K.V. 1997. Consistency of Procrustes Estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**: 281–290.

- Klingenberg, C.P. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* **11**: 353–357.
- Klingenberg, C.P. 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. *Evolution & Development* **11**: 405–421.
- Klingenberg, C.P. & Leamy, L.J. 2001. Quantitative genetics of geometric shape in the mouse mandible. *Evolution* **55**: 2342–2352. 5
- Klingenberg, C.P., Barluenga, M. & Meyer, A. 2002. Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution* **56**: 1909–1920.
- Klingenberg, C.P., Leamy, L.J. & Cheverud, J.M. 2004. Integration and Modularity of Quantitative Trait Locus Effects on Geometric Shape in the Mouse Mandible. *Genetics* **166**: 1909–1921. 10
- Kodric-Brown, A., Sibly, R.M. & Brown, J.H. 2006. The allometry of ornaments and weapons. *Proceedings of the National Academy of Sciences* **103**: 8733–8738.
- Kolbe, J.J., Revell, L.J., Székely, B., Brodie III, E.D. & Losos, J.B. 2011. Convergent Evolution of Phenotypic Integration and Its Alignment with Morphological Diversification in Caribbean Anolis Ecomorphs. *Evolution* **65**: 3608–3624. 15
- Krzanowsky, W.J. 1979. Between-Groups Comparison of Principal Components. *Journal of the American Statistical Association* **74**: 703–707.
- Lande, R. 1985. Genetic and Evolutionary Aspects of Allometry. In: *Size and Scaling in Primate Biology* (W. L. Jungers, ed), pp. 21–32. Springer US. 20
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. *Evolution* **33**: 402–416.
- Lande, R. 1980. The Genetic Covariance Between Characters Maintained by Pleiotropic Mutations. *Genetics* **94**: 203–215.
- Lele, S. 1993. Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Mathematical Geology* **25**: 573–602. 25
- Lele, S.R. & McCulloch, C.E. 2002. Invariance, Identifiability, and Morphometrics. *Journal of the American Statistical Association* **97**: 796–806.
- Lieberman, D.E. 2011. Epigenetic Integration, Complexity and Evolvability of the Head: Rethinking the Functional Matrix Hypothesis. In: *Epigenetics: Linking Genotype and Phenotype in Development and Evolution* (B. Hallgrímsson & B. K. Hall, eds), pp. 271–289. University of California Press. 30

Lieberman, D.E., Hallgrímsson, B., Liu, W., Parsons, T.E. & Jamniczky, H.A. 2008. Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. *Journal of Anatomy* **212**: 720–735.

Lieberman, D.E., Ross, C.F. & Ravosa, M.J. 2000. The primate cranial base: Ontogeny,
5 function, and integration. *American Journal of Physical Anthropology* **113**: 117–169.

Linde, K. van der & Houle, D. 2009. Inferring the Nature of Allometry from Geometric Data. *Evolutionary Biology* **36**: 311–322.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland.

10 Maddison, W.P. & Maddison, D.R. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75. 2011. URL <http://mesquiteproject.org>.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.

Marcucio, R.S., Cordero, D.R., Hu, D. & Helms, J.A. 2005. Molecular interactions coordina-
15 ting the development of the forebrain and face. *Developmental Biology* **284**: 48–61.

Marcucio, R.S., Young, N.M., Hu, D. & Hallgrímsson, B. 2011. Mechanisms that underlie co-variation of the brain and face. *genesis* **49**: 177–189.

Marivaux, L., Essid, E.M., Marzougui, W., Khayati Ammar, H., Adnet, S. & Marandat, B.
20 *et al.* 2014. A morphological intermediate between eosimiiform and simiiform primates from the late middle Eocene of Tunisia: Macroevolutionary and paleobiogeographic implications of early anthropoids. *American Journal of Physical Anthropology* n/a–n/a.

Marroig, G. 2007. When size makes a difference: allometry, life-history and morphological evolution of capuchins (*Cebus*) and squirrels (*Saimiri*) monkeys (Cebinae, Platyrrhini). *BMC Evolutionary Biology* **7**: 20.

25 Marroig, G. & Cheverud, J.M. 2001. A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys. *Evolution* **55**: 2576–2600.

Marroig, G. & Cheverud, J.M. 2004. Cranial evolution in sakis (*Pithecia*, Platyrrhini) I: Interspecific differentiation and allometric patterns. *American Journal of Physical Anthropology*
30 **125**: 266–278.

Marroig, G. & Cheverud, J.M. 2009. Size and Shape in Callimico and Marmoset Skulls: Allometry and Heterochrony in the Morphological Evolution of Small Anthropoids. In: *The*

Smallest Anthropoids, Developments in Primatology: Progress and Prospects (S. M. Ford, L. M. Porter, & L. C. Davis, eds), pp. 331–353. Springer US, Boston.

Marroig, G. & Cheverud, J.M. 2005. Size as a line of least evolutionary resistance: Diet and adaptive morphological radiation in new world monkeys. *Evolution* **59**: 1128–1142.

Marroig, G. & Cheverud, J.M. 2010. Size as a line of least resistance II: direct selection on size or correlated response due to constraints? *Evolution* **64**: 1470–1488. 5

Marroig, G., Melo, D., Porto, A., Sebastião, H. & Garcia, G. 2011. Selection Response Decomposition (SRD): A New Tool for Dissecting Differences and Similarities Between Matrices. *Evolutionary Biology* **38**: 225–241.

Marroig, G., Melo, D.A.R. & Garcia, G. 2012. Modularity, Noise and Natural Selection. *Evolution* **66**: 1506–1524. 10

Martínez-Abadías, N., Esparza, M., vold, T. Sjø, González-José, R., Hernández, M. & Klingenberg, C.P. 2011. Pervasive genetic integration directs the evolution of human skull shape. *Evolution* **66**: 1010–1023.

Márquez, E.J., Cabeen, R., Woods, R.P. & Houle, D. 2012. The Measurement of Local Variation in Shape. *Evolutionary Biology* **39**: 419–439. 15

Meinhardt, H. 1983. A boundary model for pattern formation in vertebrate limbs. *Journal of Embryology and Experimental Morphology* **76**: 115–137.

Meinhardt, H. 2008. Models of biological pattern formation: from elementary steps to the organization of embryonic axes. *Current topics in developmental biology* **81**: 1–63. 20

Melo, D. & Marroig, G. 2015. Directional selection can drive the evolution of modularity in complex traits. *Proceedings of the National Academy of Sciences* **112**: 470–475.

Melo, D., Garcia, G., Hubbe, A., Assis, A.P. & Marroig, G. 2015. EvolQG - An R package for evolutionary quantitative genetics. *F1000Research*, doi: 10.12688/f1000research.7082.1.

Meyer, K. 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genet. Sel. Evol.* **23**: 67–83. 25

Minelli, A. 2011. A principle of developmental inertia. *Epigenetics: Linking Genotype and Phenotype in Development and Evolution* 116–133.

Mitani, J.C., Call, J., Kappeler, P.M., Palombit, R.A. & Silk, J.B. 2012. *The Evolution of Primate Societies*. University of Chicago Press. 30

Mitteroecker, P. & Bookstein, F.L. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Systematic Biology* **56**: 818–836.

Mitteroecker, P. & Bookstein, F.L. 2008. The evolutionary role of modularity and integration in the hominoid cranium. *Evolution* **62**: 943–958.

Mitteroecker, P. & Bookstein, F.L. 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution* **63**: 727–737.

Mitteroecker, P., Gunz, P., Bernhard, M., Bookstein, F.L. & Schaefer, K. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. *Journal of Human Evolution* **46**: 679–697.

Moakher, M. 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications* **26**: 735–747.

Moakher, M. 2006. On the Averaging of Symmetric Positive-Definite Tensors. *Journal of Elasticity* **82**: 273–296.

Monteiro, L.R. & Nogueira, M.R. 2010. Adaptive radiations, ecological specialization, and the evolutionary integration of complex morphological structures. *Evolution* **64**: 724–744.

Monteiro, L.R., Bonato, V. & Reis, S.F. 2005. Evolutionary integration and morphological diversification in complex morphological structures: mandible shape divergence in spiny rats (Rodentia, Echimyidae). *Evolution & Development* **7**: 429–439.

Moran, P.A.P. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika* **37**: 17–23.

Moran, P.A.P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)* **10**: 243–251.

Newman, M.E.J. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103**: 8577–8582.

Neyman, J. & Scott, E.L. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16**: 1–32.

Nijhout, H.F. & German, R.Z. 2012. Developmental Causes of Allometry: New Models and Implications for Phenotypic Plasticity and Evolution. *Integrative and Comparative Biology* **52**: 43–52.

Oliveira, F.B. de, Molina, E.C. & Marroig, G. 2009a. Paleogeography of the South Atlantic: a Route for Primates and Rodents into the New World? In: *South American Primates* (P. A. Garber, A. Estrada, J. C. Bicca-Marques, E. W. Heymann, & K. B. Strier, eds), pp. 55–68. Springer New York.

- Oliveira, F.B., Porto, A. & Marroig, G. 2009b. Covariance structure in the skull of Catarrhini: a case of pattern stasis and magnitude evolution. *Journal of Human Evolution* **56**: 417–430.
- Olson, E. & Miller, R. 1958. *Morphological integration*. University of Chicago Press, Chicago.
- Opperman, L.A. 2000. Cranial sutures as intramembranous bone growth sites. *Developmental Dynamics* **219**: 472–485.
- Paschetta, C., De Azevedo, S., González, M., Quinto-Sánchez, M., Cintas, C. & Varela, H.*et al.* 2015. Shifts in subsistence type and its impact on the human skull's morphological integration. *American Journal of Human Biology* n/a–n/a.
- Pavlicev, M. & Hansen, T.F. 2011. Genotype-Phenotype Maps Maximizing Evolvability: Modularity Revisited. *Evolutionary Biology* **38**: 371–389.
- Pavlicev, M., Kenney-Hunt, J.P., Norgard, E.A., Roseman, C.C., Wolf, J.B. & 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**: 199–213.
- Pavlicev, M., Norgard, E.A., Fawcett, G.L. & Cheverud, J.M. 2011. Evolution of pleiotropy: epistatic interaction pattern supports a mechanistic model underlying variation in genotype-phenotype map. *Journal of Experimental Zoology B* **316**: 371–85.
- Pavoine, S., Baguette, M. & Bonsall, M.B. 2010. Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecological Monographs* **80**: 485–507.
- Pearson, K. & Davin, A.G. 1924. On the Biometric Constants of the Human Skull. *Biometrika* **16**: 328–363.
- Perez, S.I., Aguiar, M.A.M., Guimarães Jr., P.R. & Reis, S.F. dos. 2009. Searching for Modular Structure in Complex Phenotypes: Inferences from Network Theory. *Evolutionary Biology*, doi: 10.1007/s11692-009-9074-7.
- Pélabon, C., Bolstad, G.H., Egset, C.K., Cheverud, J.M., Pavlicev, M. & Rosenqvist, G. 2013. On the Relationship between Ontogenetic and Static Allometry. *The American Naturalist* **181**: 195–212.
- Pélabon, C., Firmat, C., Bolstad, G.H., Voje, K.L., Houle, D. & Cassara, J.*et al.* 2014. Evolution of morphological allometry. *Annals of the New York Academy of Sciences* n/a–n/a.
- Phillips, P.C. & Arnold, S.J. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* **53**: 1506–1515.
- Polly, P.D. 2008. Developmental Dynamics and G-Matrices: Can Morphometric Spaces be Used to Model Phenotypic Evolution? *Evolutionary Biology* **35**: 83–96.

- Porto, A., Oliveira, F.B., Shirai, L.T., Conto, V. de & Marroig, G. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. *Evolutionary Biology* **36**: 118–135.
- Porto, A., Shirai, L.T., Oliveira, F.B. de & Marroig, G. 2013. Size Variation, Growth Strategies,
5 and the Evolution of Modularity in the Mammalian Skull. *Evolution* **67**: 3305–3322.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C.R. 1982. Diversity and dissimilarity coefficients: a unified approach. *Theoretical population biology* **21**: 24–43.
- 10 Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. & Barabási, A.L. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**: 1551–1555.
- Ravosa, M.J. & Savakova, D.G. 2004. Euprimate origins: the eyes have it. *Journal of Human Evolution* **46**: 355–362.
- Ravosa, M.J., Noble, V.E., Hylander, W.L., Johnson, K.R. & Kowalski, E.M. 2000. Masticatory
15 stress, orbital orientation and the evolution of the primate postorbital bar. *Journal of Human Evolution* **38**: 667–693.
- Rice, D.P.C., Rice, R. & Thesleff, I. 2003. Molecular mechanisms in calvarial bone and suture development, and their relation to craniosynostosis. *The European Journal of Orthodontics* **25**: 139–148.
- 20 Rice, S.H. 2002. A general population genetic theory for the evolution of developmental interactions. *Proceedings of the National Academy of Sciences* **99**: 15518–15523.
- Rice, S.H. 2004. Developmental Associations Between Traits: Covariance and Beyond. *Genetics* **166**: 513–526.
- Rice, S.H. 1998. The Evolution of Canalization and the Breaking of Von Baer’s Laws:
25 Modeling the Evolution of Development with Epistasis. *Evolution* **52**: 647–656.
- Roebroeks, W. & Villa, P. 2011. On the earliest evidence for habitual use of fire in Europe. *Proceedings of the National Academy of Sciences* **108**: 5209–5214.
- Roff, D.A. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud’s conjecture. *Heredity* **74**: 481–490.
- 30 Rohlf, F.J. & Slice, D. 1990. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Systematic Biology* **39**: 40–59.

- Rosenberger, A.L. 1992. Evolution of feeding niches in New World Monkeys. *American Journal of Physical Anthropology* **88**: 525–562.
- Rueffler, C., Hermisson, J. & Wagner, G.P. 2012. Evolution of functional specialization and division of labor. *Proceedings of the National Academy of Sciences* **109**: E326–E335.
- Sanger, T.J., Mahler, D.L., Abzhanov, A. & Losos, J.B. 2012. Roles for modularity and 5 constraint in the evolution of cranial diversity among Anolis lizards. *Evolution* **66**: 1525–42.
- Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**: 1766–1774.
- Schroeder, L., Roseman, C.C., Cheverud, J.M. & Ackermann, R.R. 2014. Characterizing the Evolutionary Path(s) to Early Homo. *PLoS ONE* **9**: e114307. 10
- Scott, J.H. 1958. The cranial base. *American Journal of Physical Anthropology* **16**: 319–348.
- Shirai, L.T. & Marroig, G. 2010. Skull modularity in neotropical marsupials and monkeys: size variation and evolutionary constraint and flexibility. *Journal of experimental zoology. Part B, Molecular and developmental evolution* **314B**: 663–683.
- Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J. & Rabosky, D.L.*et al.* 15 2012. Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix. *PLoS ONE* **7**: e49521.
- Steppan, S.J., Phillips, P.C. & Houle, D. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends in Ecology and Evolution* **17**: 320–327.
- Theobald, D.L. & Wuttke, D.S. 2006. Empirical Bayes hierarchical models for regularizing 20 maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proceedings of the National Academy of Sciences* **103**: 18521–18527.
- Tiedemann, H.B., Schneltzer, E., Zeiser, S., Hoesel, B., Beckers, J. & Przemeck, G.K.H.*et al.* 2012. From dynamic expression patterns to boundary formation in the presomitic mesoderm. *PLoS computational biology* **8**: e1002586. 25
- Tobler, A. & Nijhout, H.F. 2010. Developmental constraints on the evolution of wing-body allometry in *Manduca sexta*. *Evolution & Development* **12**: 592–600.
- Turelli, M. 1988. Phenotypic Evolution, Constant Covariances, and the Maintenance of Additive Variance. *Evolution* **42**: 1342–1347.
- Turing, A.M. 1952. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the 30 Royal Society of London* **237**: 37–72.

Voje, K.L., Hansen, T.F., Egset, C.K., Bolstad, G.H. & Pélabon, C. 2013. Allometric Constraints and the Evolution of Allometry. *Evolution* **68**: 866–885.

Wagner, G.P. 1996. Homologues, natural kinds and the evolution of modularity. *The American Zoologist* **36**: 36–43.

5 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**: 77–95.

Wagner, G.P. 2010. The Measurement Theory of Fitness. *Evolution* **64**: 1358–1376.

10 Wagner, G.P. & Altenberg, L. 1996. Perspective: complex adaptations and the evolution of evolvability. *Evolution* **50**: 967–976.

Wagner, G.P., Pavlicev, M. & Cheverud, J.M. 2007. The road to modularity. *Nature reviews. Genetics* **8**: 921–931.

Waitt, D.E. & Levin, D.A. 1998. Genetic and phenotypic correlations in plants: a botanical test of Cheverud’s conjecture. *Heredity* **80**: 310–319.

15 Walker, J.A. 2000. Ability of Geometric Morphometric Methods to Estimate a Known Covariance Matrix. *Systematic Biology* **49**: 686–696.

Watson, R.A., Wagner, G.P., Pavlicev, M., Weinreich, D.M. & Mills, R. 2013. The Evolution of Phenotypic Correlations and “developmental Memory”. *Evolution* **67**: 1124–1138.

20 West, G.B., Brown, J.H. & Enquist, B.J. 2001. A general model for ontogenetic growth. *Nature* **413**: 628–631.

West, G.B., Brown, J.H. & Enquist, B.J. 1997. A General Model for the Origin of Allometric Scaling Laws in Biology. *Science* **276**: 122–126.

25 Willmore, K.E., Roseman, C.C., Rogers, J., Cheverud, J.M. & Richtsmeier, J.T. 2009. Comparison of Mandibular Phenotypic and Genetic Integration between Baboon and Mouse. *Evolutionary Biology* **36**: 19–36.

Wolf, J.B., Frankino, W.A., Agrawal, A.F., Brodie III, E.D. & Moore, A.J. 2001. Developmental Interactions and the Constituents of Quantitative Variation. *Evolution* **55**: 232–245.

30 Wolf, J.B., Leamy, L.J., Routman, E.J. & Cheverud, J.M. 2005. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. *Genetics* **171**: 683–94.

Woods, R.P. 2003. Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. *NeuroImage* **18**: 769–788.

Wrangham, null, Jones, null, Laden, null, Pilbeam, null & Conklin-Brittain, null. 1999. The Raw and the Stolen. Cooking and the Ecology of Human Origins. *Current Anthropology* **40**: 567–594.

Wrangham, R. & Conklin-Brittain, N. 2003. “Cooking as a biological trait”. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **136**: 35–46. 5

Xu, Q., Jamniczky, H., Hu, D., Green, R.M., Marcucio, R.S. & Hallgrímsson, B.*et al.* 2015. Correlations Between the Morphology of Sonic Hedgehog Expression Domains and Embryonic Craniofacial Shape. *Evolutionary Biology* **42**: 379–386.

Young, N.M. & Hallgrímsson, B. 2005. Serial homology and the evolution of mammalian limb covariation structure. *Evolution* **59**: 2691–2704. 10

Young, N.M., Wagner, G.P. & Hallgrímsson, B. 2010. Development and the evolvability of human limbs. *Proceedings of the National Academy of Sciences* **107**: 3400–3405.

Zelditch, M.L. & Carmichael, A.C. 1989. Ontogenetic variation in patterns of developmental and functional integration in skulls of *Sigmodon fulviventer*. *Evolution* **43**: 814–824.

Zelditch, M.L. & Swiderski, D.L. 2011. Epigenetic interactions: the developmental route to functional integration. In: *Epigenetics: linking genotype and phenotype in development and evolution*, pp. 290–316. 15

Zelditch, M.L., Bookstein, F.L. & Lundrigan, B.L. 1992. Ontogeny of integrated skull growth in the cotton rat *Sigmodon fulviventer*. *Evolution* **46**: 1164–1180.

Zelditch, M.L., Swiderski, D.L., Sheets, H.D. & Fink, W.L. 2004. *Geometric Morphometrics for Biologists: A Primer*, 1st ed. Elsevier. 20

Zelditch, M.L., Wood, A.R. & Swiderski, D.L. 2009. Building Developmental Integration into Functional Systems: Function-Induced Integration of Mandibular Shape. *Evolutionary Biology* **36**: 71–87.

Zeng, Z.-B. 1988. Long-Term Correlated Response, Interpopulation Covariation, and Interspecific Allometry. *Evolution* **42**: 363–374. 25

Apêndice: Artigos Publicados

2012

Modularity, Noise and Natural Selection.

Gabriel Marroig, Diogo A. R. Melo, & **Guilherme Garcia**.

5 Evolution 66: 1506–1524.

2014

Quantitative Genetics and Modularity in Cranial and Mandibular Morphology of Calomys expulsus.

Guilherme Garcia, Erika Hingst-Zaher, Rui Cerqueira & Gabriel Marroig.

10 Evolutionary Biology 41: 619–636.

2015

EvolQG - An R package for evolutionary quantitative genetics.

Diogo Melo, **Guilherme Garcia**, Alex Hubbe, Ana Paula Assis & Gabriel Marroig.

F1000Research, doi: 10.12688/f1000research.7082.1.

MODULARITY, NOISE, AND NATURAL SELECTION

Gabriel Marroig,^{1,2} Diogo A R Melo,¹ and Guilherme Garcia¹

¹Laboratório de Evolução de Mamíferos, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências Universidade de São Paulo, Caixa Postal 11.461, CEP 05422 970, São Paulo, SP, Brasil

²E-mail: gmarroig@usp.br

Received April 14, 2011

Accepted November 24, 2011

Most biological systems are formed by component parts that are to some degree interrelated. Groups of parts that are more associated among themselves and are relatively autonomous from others are called modules. One of the consequences of modularity is that biological systems usually present an unequal distribution of the genetic variation among traits. Estimating the covariance matrix that describes these systems is a difficult problem due to a number of factors such as poor sample sizes and measurement errors. We show that this problem will be exacerbated whenever matrix inversion is required, as in directional selection reconstruction analysis. We explore the consequences of varying degrees of modularity and signal-to-noise ratio on selection reconstruction. We then present and test the efficiency of available methods for controlling noise in matrix estimates. In our simulations, controlling matrices for noise vastly improves the reconstruction of selection gradients. We also perform an analysis of selection gradients reconstruction over a New World Monkeys skull database to illustrate the impact of noise on such analyses. Noise-controlled estimates render far more plausible interpretations that are in full agreement with previous results.

KEY WORDS: Adaptation, models/simulations, morphological evolution, pleiotropy, quantitative genetics, selection—natural.

The study of biological systems and its component parts, whether molecules, cells, tissues, organisms and its forming parts, and even species and their interactions, is rapidly converging to the central theme of modularity. This refers to the connections among some of the component parts of a biological system (genes or morphological traits, for example) and the lack of such associations among other parts of the same system. (Olson and Miller 1958; Berg 1960; Wagner et al. 2007). The notion that interacting parts are not independent is intuitive and appears early in the history of Biology (see Mayr 1982). Therefore, modularity is quickly becoming one of the central questions in modern biology (Wagner et al. 2007; Klingenberg 2008) and a point of convergence of various specialties and areas (Mathematics, Statistics, Genetics and Genomics, Evolutionary Biology, Ecology, Biochemistry, and Physiology).

In biology, several types of modules have been recognized, including (1) functional, consisting of characters or features that

act together on performing a task or function and are quasi-autonomous in relation to other functional sets; (2) developmental, which corresponds to parts of an embryo that are relatively autonomous with respect to pattern formation and differentiation, or an autonomous signaling cascade; (3) variational, composed of characters that vary together and are relatively independent of other such sets (Wagner et al. 2007).

The study of modularity is centered on statistical estimation of association among traits (Olson and Miller 1958; Berg 1960). Whether such association is measured by correlation, covariance, or distance/similarity measures, it is usually represented by matrices. Even if a particular system or network does not present a modular structure or is not being interpreted under this theory, associations among traits, parts, genes, or lineages will still be quantified by statistical association or dissociation matrices among these elements. We will focus here on correlation or covariance matrices (from now on **C-matrix**) among variables, although the

same problem appears in any statistics of association/dissociation among the component parts of any system. As biologists, what we usually do is to sample nature and infer properties from natural systems using measures and statistics obtained from such samples. We should be aware of the fact that by sampling a population we do not have the true population parameter values but only estimates of these quantities. These estimates will be approximations that should converge to the true population value depending on a number of things, such as sample size, number of parameters considered, precision of the measuring device, and the quality of parameter estimators themselves, measured by their precision and accuracy (Sokal and Rohlf 1995). The general trend is that, as the ratio between number of parameters and sample size decreases, signal-to-noise ratio will increase. For C-matrices, this effect is summarized by their sampling distribution (the Wishart distribution) and can be expressed in terms of their eigenvalues (see next section and Meyer and Kirkpatrick 2008).

In this article, we illustrate the problem of noise in matrix estimation using modularity as our framework and addressing effects of matrix estimates in the context of natural selection. The evolutionary response of a set of quantitative traits is described by $\Delta\bar{z} = G\beta$, where $\Delta\bar{z}$ is the vector of differences in means between generations, β is the selection gradient vector, and G is the additive genetic covariance matrix (Lande 1979). Rearranging the evolutionary response equation, the pattern of selection responsible for populations' divergence can be reconstructed from observed mean differences using the relationship:

$$\beta = G^{-1}\Delta\bar{z}, \quad (1)$$

where β is the cumulative selection gradient summed over generations (net- β sensu Lande 1979), G^{-1} is the inverse of G , and $\Delta\bar{z} = (\bar{z}_i - \bar{z}_j)$ is the difference in means between populations i and j (Lande 1979; Lofsvold 1988; Cheverud 1996). Selection reconstruction can be extremely useful in understanding patterns of multivariate selection within a microevolutionary context (Boag 1983; Lande and Arnold 1983; Grant and Grant 1995), and if certain assumptions hold (see Marroig and Cheverud 2001, 2004, 2005, 2010) it can be extended to a macroevolutionary context.

We illustrate here how noise associated with sampling in matrix estimation can lead to error in β reconstruction using a simulation approach. We also illustrate how changing both number of modules and magnitude of association among elements in a system will affect matrix estimation and β reconstruction. We present possible solutions that could help ameliorate the noise problem in matrix estimation and test the performance of such solutions. Finally, we illustrate the analyses of selection reconstruction in a well-studied case, New World Monkeys (NWMs) skull evolution, and show how previously estimated β 's (Marroig and Cheverud 2005) are most likely dominated by noise and how estimates of

these gradients, after controlling for noise, lend an interpretation in full agreement with results obtained using a different approach on the same dataset (Marroig and Cheverud 2010).

Materials and Methods

BACKGROUND—EXTENSION APPROACH

The study of quantitative traits is largely based on C-matrices. Estimation of these matrices is usually done in a straightforward manner, simply sampling a population of interest and estimating the population covariance matrix for a set of p traits using the sample covariance matrix obtained from a group of n individuals. This method has its appeal derived from the fact that this estimator is the maximum likelihood (ML) estimator under multivariate normality.

However, when dealing with small datasets, in which the number of parameters being estimated is close to the number of individuals or even an order of magnitude smaller, one can expect substantial error to be present in the estimated matrix. This problem is even more severe when sample size n is smaller than the number of traits p , when the sample matrix becomes singular. This is clearly a problem, which stems from the fact that the ML estimator is only asymptotically convergent to the true value, and for small datasets, it is possible to improve greatly on it (Stein 1956). The effect can become even more critical in the genomic era where large matrices are estimated for hundreds or thousands of genes with very few observations. Usually in quantitative genetics, the ML estimator is good enough, so this convergence problem is overlooked in most cases. But when theory demands matrix inversion, this "noise" associated with undersampling becomes very apparent. In particular, it greatly affects selection reconstruction analysis and must be taken into account in such studies, as we show later on.

One simple example of the noise involved in matrix estimation can be visualized in the complete absence of modularity (40 independent traits), so even when dealing with uncorrelated traits of equal variance (or with the correlation matrix for these traits) where one would expect eigenvalues of the C-matrix to be equal, in practice only with very large sample sizes these eigenvalues tend to approach the expected value (see Fig. 1A and Horn 1965). This phenomenon occurs because although the ML estimator for C-matrices provides an unbiased estimator for the mean of its eigenvalues, as expected by the sampling distribution of C-matrices, their variance is overestimated (Meyer and Kirkpatrick 2008). This "spreads" individual eigenvalue estimates, increasing the larger and decreasing the smaller ones.

We can visualize the effect of noise by plotting the distribution of estimated correlations for different sample sizes, as presented in Figure 1B for this uniform matrix. Notice that, by chance alone, with small sample sizes, although the average of

Figure 1. Bias introduced in C-matrix estimates by sample size from a population originated with a multivariate normal distribution with uniform covariance matrix: (A) eigenvalue distribution for a series of matrices estimated for a series of increasing sample sizes; (B) distributions of correlations observed in sample C-matrices estimated with increasing sample size.

all 780 pairwise correlations is zero as expected, the range of observed values is between -0.6 and 0.6 for $n = 20$, and even with $n = 200$ we still observe correlations ranging from -0.25 to 0.25 . All those correlations would indicate moderate association between traits despite the fact that they are independent. The situation is dramatically improved by using the extension method detailed below.

The extension approach was recently introduced by Hayden and Twede (2002) (also Twede and Hayden 2004) as way to improve performance of spectral filtering algorithms based on covariance matrix inverses. The idea is based on the argument that matrices are estimated with some degree of error and this noise is actually amplified and dominates inverted matrices. A C-matrix (C) can be represented by the following equation:

$$C = V^T \Lambda V, \quad (2)$$

where V is a square matrix of normalized eigenvectors, Λ is a square diagonal matrix of eigenvalues λ_i ; the superscript T denotes matrix transpose. The inverse of C is given by:

$$C^{-1} = V^T \Lambda^{-1} V, \quad (3)$$

where the diagonal elements of Λ^{-1} are the inverse of each eigenvalue ($\frac{1}{\lambda_i}$). When λ_i becomes small, the estimate C^{-1} becomes a poor estimate of the true inverse. This is due to a number of factors, but particularly in biological systems because of sampling error in the estimation and the fact that modularity reduces the dimensionality of C . The smallest eigenvalues are the ones that concentrate the most noise in the estimate of C , and will become the dominant figure in the inverted matrix C^{-1} , which is obvious if we think in terms of the inverted eigenvalues. For example, a 0.1 eigenvalue in Λ will become 10 in Λ^{-1} , and a smaller one of 0.0001 will become $10,000$. Because those last eigenvalues are precisely those with larger error in their estimates (Fig. 1), this unduly contribution to C^{-1} will affect the performance of analyses based on this matrix and consequently the inferences draw from it.

Here we treat the sequence of sorted eigenvalues as a discrete function defined on (say) the natural numbers from 1 to p , where p is the number of traits. The second derivative of this function is defined by $\lambda_{i-1} - 2\lambda_i + \lambda_{i+1}$ for the i th point. Using this definition, it is possible to find the region of the eigenvalues distribution corresponding to noise by inspecting these values and finding the minimum value of the second derivative. Subsequent eigenvalues should then be replaced by the last reliable eigenvalue (the extended eigenvalue). This was the original criterion from Hayden and Twede (2002) to find the noise-floor region. We here modified this approach by taking the absolute values of this second derivative and then calculating the variance of those values in groups of 3–5 consecutive second derivatives. When

this variance approaches zero (0.001 or 0.0001), the noise-floor region is reached and subsequent eigenvalues should be replaced by the last reliable eigenvalue (see Fig. 2 for an example of both approaches). The point when the variance is sufficiently close to zero is somewhat variable, depending on the scale of the organism (the total amount of variation in the matrix) and the amount of noise in the estimated matrix, and should be analyzed graphically on a case-by-case basis. In an attempt to minimize this effect, instead of working directly with the eigenvalues, we use the amount of variation that each one explains, or, more clearly, we divide each eigenvalue by their sum. We compare both criteria in the following sections.

Notice that C (and C^{-1}) are not truncated and still maintain their full rank, and that eigenvectors (V) are not changed but only adjusted in their variances (eigenvalues) in dimensions with very little variation. This means that matrix structure is essentially preserved by using this approach.

SIMULATION AND BOOTSTRAP

We used a simulation strategy with resampling via bootstrap to explore the impact of sampling error, modularity, and noise on reconstructing selection gradients. This strategy was designed to represent the variation in modularity patterns and, most notably, magnitude that is observable across skull variation in mammalian taxa (see Marroig and Cheverud 2001; Oliveira et al. 2009; Porto et al. 2009). The first step was to create a series of C -matrices with varying degrees of modularity and overall integration. All matrices are 40×40 in rank and simulate morphological data. Notice that all these matrices represent biological systems with low dimensionality; in other such systems, like gene coexpression databases, for example, dimensionality can easily reach the order of thousands (e.g., Nowick et al. 2009; Oldham et al. 2006). We start with uniform C -matrices, with equal eigenvalues, then move to matrices with two separate modules (A matrices), then four modules (B), eight modules (C), and finally matrices with eight modules and two submodules within each main module (D). In all these matrices, all 40 traits were divided equally among modules. In each of these modular structures, we also increase both between and within module correlation, which alters the influence of the first principal component (PC1) (size) on the total amount of variation (matrices A through D of types 1, 2, and 3 and two uniform matrices—see Supporting information). Using these C -matrices, we created 14 multivariate normal distributions with null mean, and drew from each one populations with 10,000 individuals. These are meant to represent natural populations of interest.

Next, 1000 bootstrap samples of 50, 100, 200, and 500 individuals were taken from the populations, representing experiments using different sample sizes. Each of these individual samples was used to calculate a C -matrix estimate.

Figure 2. Eigenvalues for a C3 class matrix (see section Simulation and Bootstrap). Sample eigenvalue estimation mean and coefficient of variation taken from 1000 bootstrap matrices with 50 individuals taken from a multivariate normal distribution with C3 as the covariance matrix. Arrows show cutoff points for the minimum (top right) and variance (bottom right) criteria. See text for details.

We are interested in how well these sample C -matrices can be used to estimate selection gradients. To determine this we generate a random known selection gradient (β_T), which is used in Lande's equation along with the population C -matrix to determine the observed $\Delta\bar{z}$. Using this $\Delta\bar{z}$ and sampled C -matrices, we use equation (1) to calculate an estimate of the selection gradient (β_s). The similarity between β_T and β_s is then a measure of the quality of C -matrix estimation and of its inverse. This procedure is summarized in Figure 3. We use vector correlation to compare all vectors in this study. Using random vectors drawn from a multivariate normal distribution with 40 elements, we determined that any vector correlation with absolute value above 0.5 is significant at $P < 0.001$. The same sample C -matrices were then submitted to the extension approach and a number of comparisons were made to check if controlling noise improves the results. We present below three criteria for selecting the numbers of eigenvalues that should be retained: an optimal criterion (OC), a variance criterion (VC), and a minimum criterion.

The OC is the best possible fit between β_T and the reconstructed selection gradient based on the noise-controlled sample C -matrix (β_{opt}). For every C -matrix sampled from the population, we determine how many retained eigenvalues give the highest correlation between these selection gradients. The rationale here is to have a benchmark against which to compare the impact of

Figure 3. Simulation scheme for testing noise control techniques in selection reconstruction analysis.

the extension procedure, since such approach is obviously not applicable to real world problems where we do not know the true population $\mathbf{C}$ -matrix or the true selection gradient.

We introduce here the VC, in which we calculate the cut-off estimate described in the background section without any consideration to true population values. It is entirely based on the decay of variance of the second derivative of sorted eigenvalues. The comparison between these reconstructed selection gradients based on the VC (β_{var}) with those obtained from the OC will be useful to benchmark whether we are still able to find a reasonable cut-off point without any a priori knowledge of the population $\mathbf{C}$ -matrix.

The minimum criterion is the original cut-off estimative of Twede and Hayden (2004) for the extension method, which takes the minimum of the second derivative of sorted eigenvalues as the cut-off point.

We also apply another noise control technique described for covariance matrices estimation. The shrinkage method is based on a theorem by Stein (1956) that shows that it is often possible to improve the ML estimator for the $\mathbf{C}$ -matrix of a given high-dimensionality dataset. The idea is to use a target matrix and find the linear combination between the ML estimator and an arbitrary target matrix that best describes the true covariance structure of the population. A good choice for the target matrix in our case is a diagonal matrix of variances with null covariances. This particular target only modifies covariances, leaving variances untouched. Using a positive definite target matrix, this procedure guarantees a resulting matrix that is always positive definite and well conditioned. The optimal linear combination between the target and original matrix, called shrinkage intensity, is determined analytically as to minimize the mean squared error of the resulting shrinkage matrix estimate (Ledoit and Wolf 2003). This analytical calculation of shrinkage intensity is very robust, and a poor choice of target matrix will result in a linear combination that privileges the original matrix estimated directly from the dataset. For an ML estimated $\mathbf{C}$ -matrix $\mathbf{C}$, the shrinkage estimate $\mathbf{C}^*$ is given by:

$$\mathbf{C}^* = \alpha \mathbf{T} + (1 - \alpha) \mathbf{C}, \quad (4)$$

where $\mathbf{T}$ is the target matrix and α is the shrinkage intensity. For an extensive description of the method, different shrinkage targets and how to calculate shrinkage intensity for such targets see Schafer and Strimmer (2005). The shrinkage method results were then compared to our cut-off estimative for the extension method in their efficiency in estimating the “true” $\mathbf{C}$ -matrix of the whole sample.

In all methods, we also measure the amount of variation explained by the first eigenvalue, average between and within module correlation, and mean squared correlation for all traits (an

overall measure of integration, see Porto et al. 2009). All these measures are related to the amount of integration in the system that in turn is inversely proportional to evolutionary flexibility (how close to the direction selection is pushing the population is able to respond or, in other words, the correlation between the selection gradient and evolutionary response, sensu Marroig et al. 2009). Accordingly, we can study the effects of these modularity/integration measures on selection reconstruction analysis.

To measure the impact of the extension approach over the correlation distribution in a sample matrix with respect to its “real” counterpart, we registered differences between estimated and “real” correlations, before and after applying the extension method for a single set of $\mathbf{A}$ -matrices generated from a sample of 50 individuals, thus illustrating what the extension method does to correlation estimates within each matrix.

A STUDY CASE—SELECTION RECONSTRUCTION IN NWMS SKULL EVOLUTION

To illustrate the application of the extension approach to evolutionary biology problems, we carry through a selection reconstruction analysis on our NWMS skull database. The data collection, measurements, sample sizes, and procedures to obtain within-population phenotypic covariance matrices are described in Marroig and Cheverud (2001). All matrices used from now on were pooled for each node on the tree from within-population matrices observed in the terminal taxa (genus) using the phylogenetic tree of Platyrrhini (Wildman et al. 2009) as described in Marroig and Cheverud (2005, 2010). Average values for the 16 genera were used along with the phylogeny (Wildman et al. 2009) to obtain estimates of the direction of evolution ($\Delta \bar{z}$) based on estimates of ancestral data (Marroig and Cheverud 2010). Using the equation (1), we obtain the net selection gradients. Here, $\Delta \bar{z}$ is the difference vector between two nodes (or between a living genus and its ancestor) and we use the inverse of the phenotypic pooled within-group covariance matrix ($\mathbf{P}^{-1}$) as a proxy for $\mathbf{G}^{-1}$.

For this study case, our criterion for determining the cut-off point for the extension noise control was expanded to maintain the $\Delta \bar{z}$ calculated using the noise-controlled β as close as possible to the original $\Delta \bar{z}$ with respect to its magnitude and direction. Therefore, the cut-off point was chosen so that the percentile difference between the norms of both responses was not greater than 3%, and that their vector correlation stayed in the range of 0.97–1.00. After reconstructing selection gradients, we calculate their correlations with an isometric size vector to test whether these β 's represent selection in the direction of P_{max} or not. Also, by calculating the correlations between P_{max} and the response to selection, we can discern if this response was affected by the attractor effect of the direction of least genetic resistance or if the $\Delta \bar{z}$ is a direct response to its corresponding β . We also report the correlation between these reconstructed β 's and the

corresponding $\Delta\bar{z}$ s (flexibility sensu Marroig et al. 2009), before and after controlling $\mathbf{P}$ for noise, to compare these flexibility values with those obtained by simulating 6000 random β 's, using Lande's equation to calculate $\Delta\bar{z}$ and creating a distribution of flexibilities, akin to the approach taken by Marroig and Cheverud (2010).

Simulating $\mathbf{G}$ -Matrices

We are also interested in the impact of noise control techniques in estimating $\mathbf{G}$ -matrices. Therefore, we would like to test if noise control techniques improve the estimation of $\mathbf{G}$ -matrices, comparing these estimates with "real" $\mathbf{G}$ -matrices before and after applying noise control techniques. Because "real" $\mathbf{G}$ -matrices are impossible to obtain from actual data, we address this question via a simulation-based approach.

Since $\mathbf{G}$ -matrices represent patterns of covariation between additive effects of genes over phenotypic traits, we are able to relate the features of these *loci* (like the frequencies and average effects of different alleles within each *locus*, for example) with the genetic variances and covariances of phenotypic traits (Falconer and Mackay 1996; Lynch and Walsh 1998; Kelly 2009). If these features are known, we can compare the real $\mathbf{G}$ -matrix with its estimate before and after applying noise control techniques with respect to reconstructing selection gradients.

We simulate three diploid populations, each comprised of 20,000 individuals. Each individual belonging to one of these populations is initially represented by a variable number of *loci* with purely additive effects (800 for the first population, and 1150 for the second) affecting 40 traits. We assume initially that each locus has two possible alleles and for each individual we draw two alleles with equal probabilities; genotypic values for the three possible genotypes are -0.5 and 0.5 for both homozygotes, and zero for the heterozygote, as there is no dominance between each pair of alleles.

The 800 genes in the first population affect the 40 traits in a modular way, as they are divided into eight groups of 100 genes that affect groups of five traits. Each gene affects two traits within its given module; hence, the resulting $\mathbf{G}$ -matrix for this population has zero covariances between traits in different modules. The two other populations have genetic architectures that are modified from this basic framework. The second population has 350 additional genes that randomly affect two traits in each of the five modules simultaneously, so that its $\mathbf{G}$ -matrix had an average positive covariance between traits in different modules; therefore, this second $\mathbf{G}$ -matrix has a size factor (Bookstein et al. 1985) integrating all its modules. The third population has 800 genes controlling its traits, as the first, but on each module, two traits are affected by the same genes, so that their genetic correlation is

one; hence, the resulting $\mathbf{G}$ -matrix has five dimensions with zero variance and does not have full rank.

After sampling all individuals belonging to both populations, we calculate average effects for each allele and breeding values for each individual (Falconer and Mackay 1996). To create phenotypes from these breeding values, we assign each trait to a score of heritability, randomly sampled from a F -distribution with 100 degrees of freedom for both numerator and denominator, centered to mean 0.4; we use this distribution here just to ensure that our random sample of heritabilities is nonnegative and has a mean value compatible with those observed in morphological traits for endotherms (Mousseau and Roff 1987). Starting with these heritabilities, we calculate environmental variances for traits and use these variances to generate independent Gaussian noise to the breeding values of all individuals. Therefore, environmental effects for all traits are uncorrelated, so that the $\mathbf{P}$ - and $\mathbf{G}$ -matrices for each population have strong similarities in their eigenstructure, as empirically found for morphological traits in mammals (see Cheverud 1995, 1996; Porto et al. 2009).

The individuals belonging to each population are paired randomly in 10,000 couples. These couples are used to produce a F1 generation, by sampling one of the two alleles for all *loci* independently, without linkage disequilibrium; genetic covariances between characters therefore arise only from pleiotropy. Each couple is used to produce two siblings, and the same procedure used to generate phenotypes in the parental generation was used in F1. Hence, each population comprised 10,000 families, each composed of sire, dam, and two full-siblings.

Because genealogical data for all populations are known, we are able to estimate $\mathbf{G}$ -matrices with restricted maximum likelihood (REML) methods (Shaw 1987, 1991; Lynch and Walsh 1998) by randomly sampling subsets of families within each population. Starting with a subset composed of 100 families, we increased sample size by 100 for each subsequent sample, until reaching 1000 families. For each of these sample sizes, 100 $\mathbf{G}$ -matrices were calculated from resampling the total set of families.

After obtaining estimates for the $\mathbf{G}$ -matrix from phenotypic values, we are able to use these estimates to reconstruct selection gradients, applying 1000 random β 's over the population $\mathbf{G}$ -matrix, and then using the $\Delta\bar{z}$ s generated to reconstruct the original β s with the estimated $\mathbf{G}$ -matrices. Therefore, we test the quality of the reconstructions with the same simulation scheme used on the theoretical matrices in the previous section (Fig. 3). The shrinkage approach cannot be used in the REML estimated matrices because it demands actual measurements being correlated (in this case, estimated breeding values of each individual, which are not calculated in a simple REML estimation), so we restrict ourselves to the extension approach. All codes, matrices, software, and simulated data can be obtained at <http://dreyfus.ib.usp.br/gmarroig/extension/>

Figure 4. Correlation between true and estimated β 's for the matrices type C1, C2, and C3. See Supporting information and text for detailed description of the matrices.

Results

A typical plot of observed eigenvalues of the **C**-matrix obtained for a population created from a uniform **C**-matrix is presented in Figure 1A, for different sample sizes. In the original matrix, eigenvalues are all equal. This sort of negative exponential eigenvalue plot is characteristic of morphological systems (Wagner 1984; Pavlicev et al. 2009a). In Figure 1B, notice the large 95% confidence interval and observed range in correlation estimates (which should all be zero) in the original matrices compared to those where noise was controlled using the extended procedure. As expected by sampling theory, larger sample sizes are associated with smaller error (noise) in the estimates, but in all sample sizes the extended controlled estimates are substantially improved in comparison to the original ones.

Figure 4 shows the correlation distributions between true β values and those estimated directly or using the various noise control methods. We show results for a single type of matrix (C1, C2, and C3) and for different sample sizes, but the same pattern was found for other matrix types (see Supporting information). Usually, both extension (VC) and shrinkage estimates of β are dramatically improved in regard to noise estimates.

Figure 5 shows how close to the best possible cut-off point (OC) noise control methods get in terms of the correlation between the estimated β and β_T . Our VC is the one with the least difference to the OC (more points near zero, which indicate agreement with the OC cut-off point). The shrinkage method usually performs slightly worse, even though in a few samples this method is slightly better than the extended method using the OC in

Figure 5. Comparison of each method with the OC. More points near zero indicate complete agreement in the correlation with using the OC and each method.

approaching β_T . In practice, both methods should be equivalent in their ability to improve our estimations of C-matrices. The original cut-off point of Twede and Hayden (2004) is inferior to other methods, estimating β 's that are significantly different from the OC estimates in many cases. The original estimation without any noise control technique is also shown, and it can clearly be observed that this estimate is poor when compared to β_T .

Figures 6 and 7 show how size variation (integration) and modularity affects the reconstructions. Within each matrix type (A, B, C, and D), numbers 1, 2, and 3 indicate an increasing amount of variation associated with the PC1. In general, matrices 1 do not have size variation with off-module correlations equal to zero, whereas matrices 2 and 3 do have variation in size (see Supporting information). Conversely, the number of modules is higher in matrices D (eight modules with two modules nested within each one to create a substructure— $8 \times 2 = 16$ modules), followed by matrices C (eight modules), B (four modules), and A (only two modules). Therefore, within each matrix number (1, 2, and 3), matrices $A > B > C > D$ have a decreasing amount of variation associated with their PC1 (Fig. 6). The order of the matrices in terms of increasing integration is $D1 > C1 > B1 > A1 > D2 > C2 > D3 > B2 > A2 > C3 > B3 > A3$ (see

Fig. 6). In our matrices, where the total amount of variation is similar, more variation in PC1 means that the last eigenvalues are even smaller, and thus even more difficult to estimate (Fig. 2) which in turn affects our ability to obtain a reasonable inverse. Accordingly, matrices with high integration, low modularity, and low flexibility (sensu Marroig et al. 2009) lead to estimates of

Figure 6. Correlation between true and estimated β 's without controlling for noise with $n = 500$ as a function of total integration.

β without noise control that are quite poor (see Table S1). Even controlling for noise, there is an inverse relationship between magnitude of integration and efficiency of the selection reconstruction (Fig. 6).

Regarding the effect of the extension method in estimating correlation values in each matrix, Figure 8 shows the distribution of deviations from population values for a selected set of three A-matrices. We can observe that the difference between population and estimated correlations increases in absolute value when we apply the extension method, especially in the A3-matrix, which has a more dominant size factor (higher correlations). However, the same figure also shows that overall matrix structure does not change, so that applying the extension method does not affect the observation of modularity present in the population matrix. In fact, matrix structure is less patchy in those matrices controlled for noise because the general effect of noise control methods is to reduce correlation values and thus control for spurious correlations.

G-MATRIX

None of our REML G-matrix estimates for the first (size-free) population are positive-definite (which occurs when all eigenvalues of a given square matrix are strictly positive) even at our maximum sample size of 1000 families. The occurrence of negative eigenvalues was common in our REML estimated matrices. For the third population, which is not positive-definite by definition, even some phenotypic matrices have negative eigenvalues. However, in the second population, affected by a size factor, 3% of our estimated G-matrices at sample size 700 are positive-definite; this percentage increases with larger samples so that, at our maximum sample size, 81% of these estimates are positive-definite.

Regarding selection reconstruction, raw estimates of **G** for both populations yielded poor estimates of β with respect to the correlation between those β 's and the ones applied over the population G-matrix. In the size-free population (Fig. 9), these estimates are always poor, regardless of sample size; however, in the population with size effect, there is a slight improvement of reconstructions with respect to sample size (Fig. 10). After applying either optimal (OC) or variance (VC) criteria, estimates of **G** for both populations improve the reconstructed β 's, so that their correlation with the corresponding β_T has an asymptotic growth toward a maximum possible value of correlation with respect to sample size, as it can be observed in Figures 9 and 10, regarding the size-free and size-factor populations, respectively. Furthermore, the variance of the reconstructed β correlations is much lower in the noise-controlled matrices (VC and OC) compared to the raw matrices without any form of noise control.

With respect to the phenotypic counterparts to estimates of **G**, these **P**-matrices are much better behaved. Even without any noise control and at low sample sizes, **P**-matrices yield β reconstructions

close to the true β 's considered. However, it appears that there is an upper limit to correlations between estimated and true β 's (around 0.9 for the size population and 0.75 for the size-free population). It is also noteworthy that, with respect to the size population, applying the extension method over phenotypic matrices slightly decreases the average correlation between real and estimated β 's. The **P**-matrices in these simulations are so well estimated that the VC is unable to produce an improvement in β estimation, whereas the OC produces marginal improvement. This level of precision is of course highly unlikely in real studies, and the difference in β estimation is so slight that it would probably go unnoticed. Table S2 summarizes all results obtained for this section.

Results for the third population are presented in Supporting information (Fig. 2); however, it is noteworthy that such results are by no means different for those presented for the size-free population.

Study Case

The noise and noise-controlled estimates of reconstructed β 's for our study case are presented in Table S3. Table I presents results regarding analyses comparing the directions of β 's, $\Delta\bar{z}$ s, and $P_{\max}$. These results show that most responses to selection are significantly aligned with $P_{\max}$, in concordance with Marroig and Cheverud (2005). Also, in a dozen cases across the phylogeny of NWM, the direction of β after controlling for noise is also aligned with the direction of size. The bulk of such cases are concentrated on the Callithrinae clade (Fig. 11).

Regarding estimates for flexibility before and after applying the VC (Table I), the original β 's (with noise) presents a range of values for flexibility between 0.11 and 0.45, usually outside the range of expected flexibilities for these matrices, obtained from the simulated distribution with known selection gradients; only three values of flexibility calculated with β 's without noise control along the phylogeny fall within the 95% confidence interval of the simulated distribution. When we use gradients reconstructed with noise control, the range of observed flexibility values becomes 0.45–0.76, which is compatible with the distribution of simulated flexibilities; all observed values along the phylogeny fall either within the 95 percentile empirical distribution of simulated values or slightly above.

Discussion

The difficulty in estimating eigenvalues of a uniform matrix, as shown in Figure 1, is a clear sign that noise control methods are a necessity when working with **C**-matrices. In fact, Figure 1 illustrates that with limited sample sizes, in a system where all traits are independent and equally variable, noise alone will lead

Figure 7. Correlation between true and estimated β 's using the OC. Here, we see how having a first principal component (size) accounting for most of the variability in the sample affects the reconstruction analysis. The larger the first PC, the harder it is to obtain a reasonable reconstruction of the selection gradient.

to a negative exponential distribution of eigenvalues, a fact well known (Anderson 2003; Meyer and Kirkpatrick 2008). Furthermore, false-positive correlation values with moderate to strong intensity tend to appear, an effect which is more pronounced in small sample sizes, as can be seen in Figure 1B.

Figures 4 and 5 illustrate the effect of noise and compares the shrinkage and extended methods, showing the influence of sample size on matrix estimation and consequently on selection reconstruction analyses. All methods perform better than C-matrix estimates without noise control by our criterion of how closely the reconstructed β is aligned to the true gradient. Although the effects of noise naturally diminish with increasing sample sizes, all noise control methods applied here perform better than the simple sam-

ple estimates. In fact, with sample sizes around 50 individuals, β 's reconstructed without controlling for noise are no more similar to the true selection gradient than expected by chance alone (see Table S1). After controlling for noise, however, selection gradient estimates are not only substantially improved but also present high and statistically significant similarity to the true gradients. Yet, it should be noted that the VC developed here for the extended approach generally performs better than other methods, as it can be seen in Figure 5. The difference between the criteria suggested by Hayden and Twede (2002) and our VC can be observed in Figure 2. In general, the VC will select a smaller number of eigenvalues to be retained, whereas the original (Hayden and Twede 2002) based on the minimum second derivative of the

Figure 8. The effect of the extension method over correlation values. In the first line of graphics, we show the distribution of correlation values for A1-, A2-, and A3-matrices before and after applying the extension method, with correlations between and within each module set apart. Lines on each graph indicate the true correlation values for each population; dotted lines indicate extramodular correlations and dashed lines indicate the intramodular ones. In the three subsequent lines, we show graphic representations of these matrices: the population matrix, the estimated matrix without noise control, and the estimated matrix after applying the extension method. The grayscale used in these matrices represent the maximum correlation level as white and the minimum as black, with intermediate values as shades of gray.

Figure 9. Reconstruction of selection gradients using the estimated G- and P-matrices for the size-free population, with and without noise controlling methods. Correlations between true and estimated β 's using both optimal (OC) and variance (VC) criteria.

eigenvalues will usually identify the noise floor at a later point over the ordered eigenvalues.

Matrices with enhanced modularity (larger number of modules and less integration) are less affected by noise and have both a higher similarity and a smaller variance in the correlations between the reconstructed β 's and true gradients (Fig. 6). We can see the effect of modularity and integration in Figures 6 and 7, where increasing eigenvalue variance, either by raising between module correlations (thus increasing the first eigenvalue, matrices type 1, 2, and 3, respectively, from lower to higher) or by having fewer relevant directions in the morphospace (with increasing degree of modularity, matrices type A, B, C, and D) greatly affects the reconstruction analysis. In general, matrices that are less modular and have a strong integration (a substantial portion of the variabil-

ity is concentrated on the first eigenvalue) will be more affected by noise than those matrices where dimensionality (number of modules) is higher and overall integration is lower (Figs. 4, 6, and 7).

The connection between variational modularity and the difficulty in estimating selection gradients can be interpreted by noting that in less modular highly integrated systems a larger portion of the trait variation is shared with other traits and very little of the trait variation is not shared (common and trait-specific variation—see McGuigan and Blows 2010) making it harder to detect any signal in that specific part of the trait variation against noise. We show here that this type of noise can be readily identified with small eigenvalues of the C-matrix, and modularity aggravates this problem (Figs. 2 and 6). This occurs because some

Figure 10. Reconstruction of selection gradients using the estimated G- and P-matrices for the size-factor population, with and without noise controlling methods. Correlations between true and estimated β 's using both optimal (OC) and variance (VC) criteria.

directions on the morphological space contain less variation, due to the high correlation of traits within modules and consequently the concentration of the total variation in few dimensions. Thus, the last eigenvalues are correspondingly smaller and carry only a tiny fraction of the total variation in the system (Pavlicev et al. 2009a). Because smallest quantities had relatively more error in its estimates than larger quantities (just because they are small and approach the error inherent to both observer and measuring device), it is easy to see why the last eigenvalues have a low signal-to-noise ratio. Just to give an idea of the noise control effect on the millimeter scale of real measures taken from a sample of marmosets skulls, we compared the observed values in a sample of 502 skulls with the values after controlling for the noise in the C-matrix (adjusting eigenvalues via extended approach and reprojecting the data back to the original basis). The average dif-

ference in millimeter between observed data and noise-controlled data is around 0.12 mm that is well within observer plus measuring device (0.05-mm accuracy of the three-dimensional digitizer) error.

When we consider the effect of extension on the individual correlation estimates (Fig. 8), it is clear that by controlling noise we are increasing to some extent the variation in the system (the small eigenvalues are increased), and so estimated correlation values tend to become smaller than true values, an effect more pronounced in size-dominated matrices. It is noteworthy, however, that this bias represents a small one when compared to the absolute value of correlations in these matrices, within and between modules. In the example provided in Figure 8, intramodular correlations in the estimated A3-matrix reduce to an average of 0.6 when the extension method is applied. This is a small change

Figure 11. Association between reconstructed β 's and the direction of size across the phylogeny of Platyrrhini monkeys. The strength of correlation between these vectors is indicated by the color scale and by circles and squares, whether such correlation is negative or positive, respectively.

when compared to the real value of 0.8, and the reduced value is still statistically different from zero at a sample size of 50 individuals. At same time, small spurious correlations that reflect noise in the system are controlled by extension and consequently the structure of the matrix is a much better representation of the true pattern structure than the raw matrix (see Fig. 8, especially the left panel, matrix A1 where off-modules correlation should be zero). This illustrates the effects of the extension approach over the eigenvalue distribution and its impacts on C-matrix structure. As we extend the lower eigenvalues, we reduce the variance of their distribution, effectively controlling for its bias, while introducing a small bias in the eigenvalue mean. Our argument here is that this bias in eigenvalue mean does not affect C-matrix structure as much as a bias in eigenvalue variance, as shown above.

Regarding results with simulated G-matrices for the three populations, it is clear that the extension approach improves dramatically estimation of selection gradients, regardless of sample size. Often, estimated G-matrices will have negative eigenvalues, representing linear combinations of traits that have negative additive genetic variances (Hill and Thompson 1978); hence, these matrices cannot be considered to be true C-matrices. These nega-

tive eigenvalues appear in estimated G-matrices due to impossible values of estimated pairwise correlations, either because they fall outside the range of -1 to 1 , or because there are combinations of partial correlations for certain subsets of traits that are impossible in terms of variance partitioning. Therefore, G-matrices estimated through separate REML models for each of its variances and covariances need to be corrected for bias in its eigenvalues; our argument here is that the extension method is the best correction because it preserves modularity structure in any C-matrix estimate, be it phenotypic or genetic; the extension method is also a simple method in terms of implementation when compared to the shrinkage method.

In our simulated population, virtually all G-matrix estimates in the size-free population have negative eigenvalues where there are positive eigenvalues in the population G-matrix. Because the eigenvalue distribution of a G-matrix estimate is used for estimating the dimensionality of the population G-matrix (e.g., Mezey and Houle 2005; Hine and Blows 2006; Pavlicev et al. 2009b), we suggest that one should proceed with caution when interpreting negative eigenvalues as dimensions without genetic variation (see also Pavlicev et al. 2009b).

Table 1. Reconstruction of noise-controlled β 's across the Platyrrhini phylogeny and their relationship with Δz s and $P_{\max}$. Bold font indicate vector correlation significantly different ($P < 0.001$) from a random distribution.

Flexibility											
Node	PC's retained	% dif.	$\Delta z \times \Delta z_{\text{observed}}$	$\beta \times \text{size}$	Constraints	Noise	Control	Simulation			
								Lower bound	Average	Upper bound	SD
9-5	7	1.554	0.993	-0.589	-0.901	0.257	0.718	0.374	0.511	0.665	0.075
15-13	12	0.808	0.997	-0.594	-0.934	0.220	0.634	0.416	0.547	0.684	0.069
13-11	8	1.266	0.994	-0.701	-0.964	0.283	0.760	0.415	0.548	0.683	0.070
11-9	8	1.533	0.993	-0.690	-0.921	0.347	0.764	0.439	0.590	0.732	0.078
6-3	10	1.580	0.995	0.125	0.729	0.249	0.604	0.374	0.511	0.665	0.075
5-3	10	2.845	0.988	-0.409	-0.865	0.279	0.593	0.374	0.511	0.665	0.075
19-2	24	2.990	0.992	0.110	0.548	0.448	0.547	0.372	0.527	0.691	0.084
19-3	9	1.162	0.995	-0.474	-0.927	0.278	0.613	0.366	0.520	0.685	0.082
18-19	9	1.429	0.993	0.453	0.904	0.273	0.612	0.366	0.520	0.685	0.082
20-18	10	1.768	0.992	0.375	0.739	0.320	0.708	0.343	0.505	0.685	0.091
21-20	18	2.522	0.991	0.180	0.795	0.296	0.501	0.387	0.512	0.659	0.070
26-2	24	2.990	0.992	-0.110	-0.557	<u>0.448</u>	0.547	0.372	0.527	0.691	0.084
28-27	18	1.213	0.994	0.258	0.698	0.281	0.627	0.383	0.546	0.703	0.086
27-26	13	2.895	0.984	0.275	0.787	0.315	0.618	0.384	0.549	0.710	0.088
Pithecia-27	18	2.771	0.990	-0.212	-0.284	0.258	0.759	0.383	0.546	0.703	0.086
Callicebus-26	8	2.269	0.991	-0.512	-0.843	0.298	0.703	0.384	0.549	0.710	0.088
Ateles-20	17	2.865	0.990	0.200	0.640	0.363	0.659	0.387	0.512	0.659	0.070
Alouatta-18	16	1.530	0.995	-0.046	0.727	<u>0.370</u>	0.592	0.343	0.505	0.685	0.091
Aotus-5	15	2.563	0.991	-0.071	-0.736	0.274	0.580	0.374	0.511	0.665	0.075
Saguinus-9	8	3.000	0.987	-0.572	-0.868	0.331	0.750	0.439	0.590	0.732	0.078
Leontopithecus-11	21	2.485	0.991	-0.117	0.719	<u>0.426</u>	0.607	0.415	0.548	0.683	0.070
Callimico-13	24	2.969	0.988	0.032	-0.478	0.327	0.566	0.416	0.547	0.684	0.069
Callithrix-15	12	1.293	0.995	-0.479	-0.886	0.225	0.661	0.422	0.553	0.690	0.069
Cebuella-15	12	0.783	0.997	-0.623	-0.925	0.245	0.469	0.422	0.553	0.690	0.069
Cebus-6	5	2.077	0.990	0.507	0.864	0.110	0.738	0.302	0.427	0.577	0.071
Saimiri-6	12	1.708	0.994	-0.197	-0.844	0.203	0.480	0.302	0.427	0.577	0.071
Brachyteles-21	18	2.866	0.988	0.213	0.744	0.219	0.570	0.405	0.548	0.702	0.077
Lagothrix-21	18	2.774	0.988	-0.053	0.263	0.226	0.680	0.405	0.548	0.702	0.077
Cacajao-28	13	2.504	0.989	0.333	0.746	0.276	0.643	0.386	0.542	0.695	0.081
Chiropotes-28	24	2.961	0.988	-0.072	-0.017	0.380	0.690	0.386	0.542	0.695	0.081

In the size-factor population, some of the $\mathbf{G}$ -matrix estimates at higher sample sizes become free of negative eigenvalues, which makes these estimates more reliable even without any correction for their eigenvalue distribution. This is by no means a property of biological systems that have size variation; it is only a byproduct of our simulation design. Because heritabilities are fixed, genetic variation in size increases the overall genetic and phenotypic variances in the system, reducing environmental variances. Therefore, estimates for genetic variances and covariances are easier to obtain in the size-factor population, simply because they are higher than those in the size-free population (Young et al. 2010).

A corollary of our simulations is that, when environmental effects are low and normally distributed and interactions between genetic and environmental effects are nonexistent, as it is the case

here, $\mathbf{P}$ -matrices produce better estimates of selection gradients than $\mathbf{G}$ -matrices without using any noise control techniques, as can be seen in the leftmost panels of Figures 9 and 10. This is consistent with Cheverud's conjecture (Cheverud 1988; Roff 1995; Reusch and Blanckenhorn 1998; Waitt and Levin 1998; Porto et al. 2009; Dochtermann 2011), which suggests that $\mathbf{P}$ -matrices are good approximations of their genetic counterparts for evolutionary analyses, given that environmental effects are low or have similar structure to $\mathbf{G}$. This result reflects the fact that, under the conditions of our simulation, $\mathbf{P}$ -matrices are estimated with less noise due to their large sample sizes than their corresponding $\mathbf{G}$ -matrices due to the familial structure (Roff 1995).

In all the examples based on simulations, sample sizes considered as adequate for studies based on museum or laboratory

collections ($n = 100$ or $n = 200$) still present a bias, with the first eigenvalues being overestimated and the last ones underestimated. Even when working with matrices with moderate rank, only huge, realistically unobtainable samples lead to ML estimates with reasonable error in eigenvalues. This is an important message for researchers about an effect that can be exacerbated in situations where the number of estimated parameters is larger than or close to sampled specimens, as is common in geometric morphometrics nowadays. The situation is even more dramatic in gene coexpression studies where thousands of dimensions (genes) might be correlated with only a few individuals. This has already been noticed in other fields in which the error in estimation is more evident, such as genomics or even economics, where covariance matrices that are not positive definite are obtained directly from data, and thus easily perceived as in need of improvement (Ledoit and Wolf 2003; Schafer and Strimmer 2005; see also Efron 1982 for a discussion on the limitations of ML estimators). Although the problem has been noticed, noise control is not yet a common practice in genomic studies (e.g., Oldham et al. 2006; Nowick et al. 2009). Both shrinkage and extended approaches can be used in estimates of variance/covariance or correlation as a way of controlling noise and improving representations of population patterns. We illustrate this in the context of selection reconstruction analyses exploring both modularity and sampling in signal-to-noise ratio in simulated datasets as well as estimating selection gradients in NWMs skull evolution. We shall now consider these results.

Previously, Marroig and Cheverud (2004, 2005) showed that skull evolution of NWMs was predominantly under natural selection and that the adaptive radiation of this group was dominated by size changes (and allometrically correlated shape changes) associated with dietary differences. However, net selection gradients reconstructed were not similar to size, suggesting that size evolution was a byproduct of the attractor effect of the line of least resistance and its long-term conservation (Marroig and Cheverud 2005). Yet, net selection gradients reconstructed previously (Fig. 4 in Marroig and Cheverud 2005 and Table S3) present vector correlations (between 0 and 0.44), with the observed evolutionary responses ($\Delta\bar{z}_{\text{observed}}$) outside the range normally observed in our simulation (0.4–0.85) of evolutionary responses produced by random selection vectors (Table 1, see Marroig and Cheverud 2010; Marroig et al. 2009), suggesting that they were poor estimates of the true net- β due to noise in those C-matrices. Furthermore, Marroig and Cheverud (2010) developed a method based on random selection simulations that allows to test whether a particular observed case of evolutionary change along a size dimension can be explained by the line of least resistance attractor effect. In the NWMs skull evolution study, they show that at least in a third of the total cases the observed size/allometry evolutionary change probably result from selection to some extent aligned with size.

We confirm this finding here by removing noise from matrices estimates prior to reconstructing net selection gradients (Table 1). Five of the six cases identified by Marroig and Cheverud (2010) as being under size selection (*Callicebus*, *Cebus*, *Saguinus*, *Callicebus*, and *Cebuella*) also present here a significant correlation of the net- β with a size vector (Table 1 and Fig. 11). Furthermore, there is a clear signal of size selection in branches corresponding to marmosets and tamarins where a long-term trend for reducing overall body size is observed (Fig. 11).

Conclusion

It is important that we realize that the consequences of not controlling noise from our estimated matrices would not only impact selection reconstruction but any statistical inference based on matrix estimates and, particularly, in a very pronounced way, inverted matrices. Thus, although we illustrate the problem here within the context of selection reconstruction analysis, the problem (noise) and solutions to control it (extension and shrinkage) have far broader implications. Both traditional Euclidean and Geometric morphometrics make extensive use of covariance matrices and invert them routinely. In ecology, genomics, and genetics, many procedures involve matrix inversion. In these contexts, using ML estimates lead to results that may be extremely different from reality. For example, in phylogeography and phylogenetic analyses, we can think about historical relationships among units (whether populations or species) in terms of a covariance matrix of ancestry (Cavalli-Sforza and Piazza 1975). These matrices will be estimated with some degree of noise and the problem so far has been largely ignored. Noise control methods discussed here can be useful in these applications and we plan to explore that in a subsequent contribution. In evolutionary ecology, methods dealing with demography and population growth are usually represented in form of covariance or correlation matrices (Caswell, 2001, 2008) and the same problem will be present in such estimates. In fact, given the central importance of correlation and variance/covariance matrices in quantitative biology, we suggest that researchers should incorporate noise control techniques in matrix estimation as a routine procedure and to fully incorporate the extension approach in the housekeeping statistical toolkit.

Acknowledgments

We thank C. Roseman, D. Meyer, N. Barton, R. Bürger, and three anonymous reviewers for comments on an earlier draft of the manuscript. Support was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

LITERATURE CITED

- Anderson, T. W. 2003. An introduction to multivariate statistical analysis. 3rd ed. John Wiley & Sons, Hoboken, NJ.
- Berg, R. L. 1960. The ecological significance of correlation pleiades. *Evolution* 14:171–180.
- Boag, P. T. 1983. The heritability of external morphology in Darwin's ground finches (*Geospiza*) on Isla Daphne Major, Galapagos. *Evolution* 37:877–894.
- Bookstein, F. L., B. Chernoff, R. Elder, G. Humphries Smith, and R. Strauss. 1985. Morphometrics in evolutionary biology. The Academy of Natural Sciences of Philadelphia, Philadelphia, PA.
- Caswell, H. 2001. Matrix Population Models: construction, analyses and interpretation. Sinauer Associates Inc. Publishers, Sunderland, MA.
- Caswell, H. 2008. Perturbation analysis of nonlinear matrix population models. *Demogr. Res.* 18:59–116.
- Cavalli-Sforza, L. L., and A. Piazza. 1975. Analysis of evolution: evolutionary rates, independence, and treeness. *Theor. Pop. Biol.* 8:127–165.
- Cheverud, J. M. 1988. A comparison of genetic and phenotypic correlations. *Evolution* 42:958–968.
- . 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *Am. Nat.* 145:63–89.
- . 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *J. Evol. Biol.* 9:5–42.
- Dochtermann, N. A. 2011. Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evol.* 65:1814–1820.
- Efron, B. 1982. Maximum likelihood and decision theory. *Ann. Stat.* 10:340–356.
- Falconer, D. S., and Mackay, T. F. C. 1996. Introduction to quantitative genetics. Addison Wesley Longman, Harlow, Essex.
- Grant, P. R., and Grant, B. R. 1995. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution* 49:241–251.
- Hayden, A. F., and E. R. Twede. 2002. Observations on the relationship between eigenvalues, instrument noise and detection. *Proc. SPIE* 4816:355–362.
- Hill, W. G., and R. Thompson. 1978. Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* 34:429–439.
- Hine, E., and M. W. Blows. 2006. Determining the effective dimensionality of the genetic variance-covariance matrix. *Genetics* 173:1135–1144.
- Horn, J. L. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30:179–185.
- Kelly, J. K. 2009. Connecting QTLs to the G-matrix of evolutionary quantitative genetics. *Evolution* 63:813–825.
- Klingenberg, C. P. 2008. Morphological integration and developmental modularity. *Annu. Rev. Ecol. Evol. Syst.* 39:115–132.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. *Evolution* 33:402–416.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Ledoit, O., and M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Finance* 10:603–621.
- Lofsvold, D. 1988. Quantitative genetics of morphological differentiation in *Peromyscus*. II. Analysis of selection and drift. *Evolution* 42:54–67.
- Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.
- Marroig, G., and J. M. Cheverud. 2001. A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys. *Evolution* 55:2576–2600.
- . 2004. Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys? *Am. Nat.* 163:417–428.
- . 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in new world monkeys. *Evolution* 59:1128–1142.
- . 2010. Size as a line of least resistance II: direct selection on size or correlated response due to constraints? *Evolution* 64:1470–1488.
- Marroig, G., L. Shirai, A. Porto, F. de. Oliveira, and V. de. Conto. 2009. The evolution of modularity in the mammalian skull II: evolutionary consequences. *Evol. Biol.* 36:136–148.
- Mayr, E. 1982. The growth of biological thought: diversity, evolution and inheritance. Harvard Univ. Press, Cambridge, MA.
- McGuigan, K., and M. W. Blows. 2010. Evolvability of individual traits in a multivariate context: partitioning the additive genetic variance into common and specific components. *Evolution* 64:1899–1911.
- Meyer, K., and Kirkpatrick, M. 2008. Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices. *Genetics* 180:1153–1166.
- Mezey, J. G., and D. Houle. 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* 59:1027–1038.
- Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. *Heredity* 59:181–197.
- Nowick, K., T. Gernat, E. Almaas, and L. Stubbs. 2009. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl. Acad. Sci.* 106:22358–22363.
- Oldham, M. C., S. Horvath, and D. H. Geschwind. 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl. Acad. Sci.* 103:17973–17978.
- Oliveira, F. B., A. Porto, and G. Marroig. 2009. Covariance structure in the skull of Catarrhini: a case of pattern stasis and magnitude evolution. *J. Hum. Evol.* 56:417–430.
- Olson, E., and R. Miller. 1958. Morphological integration. Univ. of Chicago Press, Chicago, IL.
- Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2009a. Measuring morphological integration using eigenvalue variance. *Evol. Biol.* 36:157–170.
- Pavlicev, M., G. P. Wagner, and J. M. Cheverud. 2009b. Measuring evolutionary constraints through the dimensionality of the phenotype: adjusted bootstrap method to estimate rank of phenotypic covariance matrices. *Evol. Biol.* 36:339–353.
- Porto, A., F. B. Oliveira, L. T. Shirai, V. de Conto, and G. Marroig. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. *Evol. Biol.* 36:118–135.
- Reusch, T., and W. U. Blanckenhorn. 1998. Quantitative genetics of the dung fly *Sepsis cynipsea*: Cheverud's conjecture revisited. *Heredity* 81:111–119.
- Roff, D. A. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. *Heredity* 74:481–490.
- Schafer, J., and K. Strimmer. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4:1–32.
- Shaw, R. G. 1987. Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* 41:812–826.
- . 1991. The comparison of quantitative genetic parameters between populations. *Evolution* 45:143–151.
- Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. W. H. Freeman, New York.
- Stein, C. 1956. Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Berk. Symp. Math. Stat. Prob.* 1:197–206.

- Twede, E. R., and A. F. Hayden 2004. Refinement and generalization of the extension method of covariance matrix inversion by regularization for spectral filtering optimization imaging. *Proc. SPIE* 5159:299–306.
- Wagner, G. P. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. *J. Math. Biol.* 21:77–95.
- Wagner, G. P., M. Pavlicev, and J. M. Cheverud, 2007. The road to modularity. *Nature reviews. Genetics* 8:921–931.
- Waitt, D. E., and D. A. Levin, 1998. Genetic and phenotypic correlations in plants: a botanical test of Cheverud's conjecture. *Heredity* 80: 310–319.
- Wildman, D. E., N. M. Jameson, J. C. Opazo, and S. V. Yi 2009. A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). *Mol. Phylogenet. Evol.* 53:694–702.
- Young, N. M., G. P. Wagner, and B. Hallgrímsson. 2010. Development and the evolvability of human limbs. *Proc. Natl. Acad. Sci.* 107:3400–3405.

Associate Editor: R. Bürger

Supporting Information

The following supporting information is available for this article:

Table S1. Table with all means and standard deviations for bootstrap simulations.

Table S2. Table with all means and standard deviations for **G**-matrix simulations.

Table S3. Selection gradients with noise and controlled for noise in the **P**-matrices along with the $\Delta\bar{z}$ s. Node numbers correspond to Figure 11.

Table S4. All 14 population matrices from which samples were drawn from a multivariate normal distribution.

Figure S1. Graphical representation of the matrices used in section “Simulation and Bootstrap.”

Figure S2. Reconstruction of selection gradients using the estimated **G**- and **P**-matrices for the reduced-rank population, with and without noise controlling methods. Correlations between true and estimated β s using both optimal (OC) and variance (VC) criteria.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Quantitative Genetics and Modularity in Cranial and Mandibular Morphology of *Calomys expulsus*

Guilherme Garcia · Erika Hingst-Zaher ·
Rui Cerqueira · Gabriel Marroig

Received: 10 April 2014 / Accepted: 11 August 2014 / Published online: 27 August 2014
© Springer Science+Business Media New York 2014

Abstract Patterns of genetic covariance between characters (represented by the covariance matrix G) play an important role in morphological evolution, since they interact with the evolutionary forces acting over populations. They are also expected to influence the patterns expressed in their phenotypic counterparts (P), because of limits imposed by multiple developmental and functional restrictions on the genotype/phenotype map. We have investigated genetic covariances in the skull and mandible of the vesper mouse (*Calomys expulsus*) in order to estimate the degree of similarity between genetic and phenotypic covariances and its potential roots on developmental and functional factors shaping those integration patterns. We use a classic ad hoc analysis of morphological integration based on current state of art of developmental/functional factors during mammalian ontogeny and also applied a novel methodology that makes use of simulated evolutionary responses. We have obtained P and G that are strongly similar, for both skull and mandible; their similarity is achieved through the spatial and temporal

organization of developmental and functional interactions, which are consistently recognized as hypothesis of trait associations in both matrices.

Keywords Morphological integration · Modularity · G-matrix · Cheverud's Conjecture · Genotype/phenotype map

Introduction

Morphological integration refers to the interconnection among morphological elements due to genetic, functional and developmental relationships between such elements, as expressed during the course of development (Olson and Miller 1958; Cheverud 1996). From a classical quantitative genetics perspective (Falconer and Mackay 1996), these interconnections among morphological elements are represented as phenotypic and genetic covariance or correlation matrices (P and G , respectively). The structure represented by G is the product of additive effects of multiple *loci*, affecting multiple traits through pleiotropy and linkage disequilibrium. Phenotypic covariance or correlation structure is then defined as the result of the interactions between these genetic effects and environmental perturbations. However, genes act upon phenotypes through development, which can be thought as a function that maps genotypes into phenotypes (Wagner 1984; Polly 2008). In this context, integration refers to the structure of gene-trait mapping. Theoretical considerations made by Wagner (1996) suggest that this genotype/phenotype map would display a modular organization; i.e., that traits are grouped in subsets, and each trait subset is affected by a subset of genes with pleiotropic effects mostly confined to each group.

Electronic supplementary material The online version of this article (doi:10.1007/s11692-014-9293-4) contains supplementary material, which is available to authorized users.

G. Garcia (✉) · G. Marroig
Laboratório de Evolução de Mamíferos, Departamento de
Genética e Biologia Evolutiva, Instituto de Biociências,
Universidade de São Paulo, São Paulo, SP, Brazil
e-mail: wgar@usp.br

E. Hingst-Zaher
Museu Biológico, Instituto Butantan, São Paulo, SP, Brazil

R. Cerqueira
Laboratório de Vertebrados, Departamento de Ecologia, Instituto
de Biologia, Universidade Federal do Rio de Janeiro,
Rio de Janeiro, RJ, Brazil

While G is associated with the additive portion of genetic variation, which can be understood as the linear approximation of this developmental function centered at the mean phenotype (Wagner 1984), its structural aspects are also dependent upon non-linear developmental factors, traditionally viewed in quantitative genetics as dominance and epistasis (Wolf et al. 2001). Developmental dynamics involves not only the interaction among many genes, but the interactions among the developing cells and tissues, both in terms of differential expression and signaling, and mechanical interactions; these complex interactions often lead to non-linear effects within the genotype/phenotype map (Turing 1952; Polly 2008; Krupinski et al. 2011; Tiedemann et al. 2012; Watson et al. 2013). In the mammalian skull, for instance, there is a series of overlapping steps that define the adult phenotype, such as neural crest cell migration, the formation of cell condensations of osteoblasts and the subsequent differentiation of these cells into osteocytes, brain growth and muscle-bone interactions (Hallgrímsson and Lieberman 2008; Franz-Odenaal 2011). Therefore, while development has a tendency to produce covariance among morphological elements, the pattern described by phenotypic or genetic covariance matrices is often difficult to categorize, due to the superposition of these effects in shaping covariance patterns (Hallgrímsson et al. 2009; Roseman et al. 2009). Furthermore, developmental dynamics may limit the full expression of genetic variation (Hallgrímsson et al. 2009), since mutational effects may destabilize the system as a whole; hence, developmental dynamics may act as internal stabilizing selection (Cheverud 1984) and, in equilibrium, the genetic covariance structure will match those pattern arising both from stabilizing selection, drift, and mutational effects (Lande 1980).

Empirical evidence on the association between morphological traits and their underlying genetic architecture focused on the identification of pleiotropic quantitative trait *loci* supports the hypothesis of modular organization in gene-trait associations. For instance, the partitioning of the skull into Face and Neurocranium (Leamy et al. 1999) and partitioning of the mandible into Alveolar Process and Ascending Ramus (Cheverud et al. 1997; Mezey et al. 2000; Cheverud 2006); however, Klingenberg et al. (2004) has found evidence contrary to these mandibular partitions, although the disagreement between these results might arise due to methodological differences between traditional and geometric morphometrics. Without a strict adherence to the view of a modular genotype/phenotype map, Roseman et al. (2009) found a positive association between correlation among pairs of traits and their amount of shared pleiotropic effects.

The structural aspects expressed in G are central to all discussions regarding the evolution of complex phenotypes

Fig. 1 Theoretical expectations regarding the relationship between selection gradients and response to selection with respect to covariance structure. In this system composed of three traits, the covariance between x_1 and x_2 is positive, while both these traits are independent from y_1 . In (a) and (b), the β applied over the covariance matrix between x_1 , x_2 and y_1 matches the covariance matrix structure; therefore, Δz follows β closely, yielding a high f value. In (c) and (d), β does not match the structure embedded in the covariance matrix; hence, Δz is deflected due to the covariance between x_1 and x_2 (c); in this case, the associated f value is lower

since G mediates the response to directional selection (Lande 1979), imposing several properties of such response in a microevolutionary scale (Hansen and Houle 2008), with possible consequences on a macroevolutionary scale (Marroig and Cheverud 2005). For example, the cosine of the angle between a given selection gradient (β ; the vector of fitness slopes over phenotypes; Lande 1979; Falconer and Mackay 1996) and the response it produces (Δz), called flexibility (f) by Marroig et al. (2009), is affected both by the imposed selection direction and the covariance structure of the trait system considered. For example, in a system composed of three traits (Fig. 1), if directional selection is aligned with the genetic covariance structure in a population (which describes morphological integration), response to such selection will be direct (Fig. 1a, b), yielding a high flexibility value for this selection gradient. However, if the selection gradient is unaligned to population covariance structure, response to selection will be deflected from the optimal direction (Fig. 1c, d), lowering the associated flexibility value.

The relationship between selection gradients and response to selection may also be affected by the line of least genetic resistance (LLR; Schluter 1996), the first

Fig. 2 Expectations regarding the relationship between flexibility values and variation along the LLR. In (a) and (b), the β applied over the covariance matrix between x_1 and x_2 only selects for average increase in x_1 . In both situations, response to selection is deflected due to the covariance between the two traits. However, the response is more strongly biased in (a) than in (b), leading to a lower associated f value for the first situation

eigenvector of G , which represent the major axis of multivariate genetic variation on a population for a given trait system. If genetic variation is concentrated in this direction in a given population (Fig. 2a), response to selection will be biased towards it; only selection vectors nearly orthogonal to the LLR would escape such bias. Hence, the population would exhibit overall lower flexibility values, compared to another population with less variation concentrated along the LLR (Fig. 2b). With respect to mammalian morphological systems, the LLR is, in most cases, associated with size variation (Marroig et al. 2004; Porto et al. 2009, 2013), and can be thought as a global integrating factor, since it impacts morphological variation as a whole. Therefore, the effect of this source of variation over morphological integration has to be taken into account when dealing with such systems (Marroig et al. 2004; Mitteroecker and Bookstein 2007; Porto et al. 2013).

Due to the importance of the genetic covariance structure in multivariate trait systems, there has been a consistent effort in estimating G (Steppan et al. 2002). However, such matrices are often difficult to estimate, since this estimation depends on the availability of genealogical relationship information within a population (Falconer and Mackay 1996; Lynch and Walsh 1998). Furthermore, even if such information exists, the estimation of G is severely prone to error (Hill and Thompson 1978; Meyer and Kirkpatrick 2008; Marroig et al. 2012), since their estimation is usually based on sample units (families) that are an order of magnitude lower than sample sizes for P (Roff 1997).

A solution to this limitation was proposed by Cheverud (1988), who compared a wide range of P s and G s estimated mostly from morphological traits, finding similarities in covariance patterns, especially for morphological traits. Therefore, the so called ‘Cheverud’s Conjecture’ (Roff

1995) states that the patterns expressed by G will be mirrored by their phenotypic counterparts, given that the covariance between environmental effects (E) are either low, uncorrelated or similar to G in their structure. While it may seem odd that environmental and genetic covariance patterns would have similar structures, both sources of variation exert their effects over phenotypic covariance structure through common developmental pathways (Klingenberg 2008). Several authors (e.g.: Roff 1995, 1997; Reusch and Blanckenhorn 1998; Roff and Fairbairn 2011; Dochtermann 2011) have formally tested this conjecture for different traits systems across a wide range of taxa, finding supporting evidence on its favor.

In the present article, we estimated phenotypic and genetic covariance and correlation matrices for both cranial and mandibular traits for a population of the vesper mouse *Calomys expulsus*, a sigmodontine rodent, comparing these matrices with respect to the correspondence of patterns predicted by Cheverud’s Conjecture (Cheverud 1988; Roff 1995). We also investigated the association between covariance/correlation structure and hypotheses of trait associations based upon functional and developmental relationships, using the traditional methods established by several authors (Cheverud 1995; Marroig and Cheverud 2001; Marroig et al. 2004; Porto et al. 2009) and a novel methodology, based on the expectations regarding the relationship between selection gradients that represent hypotheses of trait association and response to selection to such gradients, as suggested by Hansen and Houle (2008).

Materials and Methods

Sample

The genus *Calomys* (Muroidea, Cricetidae) consists of six species of small sigmodontine rodents that occurs in open and forested areas across Central South America (Herskovitz 1962; Bonvicino and Almeida 2000; Bonvicino et al. 2003). Molecular phylogenetic analysis indicates that *Calomys* is a basal clade of the Phyllotini tribe (Steppan et al. 2004). The vesper mouse, *Calomys expulsus*, occurs over the dry biomes of Central Brazil in sympatry with the delicate vesper mouse (*C. tener*) at the southern limit of its distribution. Recently, both morphological, karyotypic and molecular data confirm the validity of the species (Bonvicino and Almeida 2000; Almeida et al. 2007).

In order to estimate covariance matrices for cranial and mandibular traits in *C. expulsus*, we used a series of specimens deposited at the mammal collection of the Museu Nacional do Rio de Janeiro (MNRJ). This series constitute a captive-bred colony originated from 37 sexually immature individuals captured in a single locality in

Fig. 3 The 20 registered landmarks represented over an adult skull of *C. expulsus* in dorsal (a), lateral (b) and ventral (c) views. Landmark names are highlighted in the particular view at which they are best defined. The set of 35 inter-landmark distances are also represented; different line types represent distinct modularity hypotheses to which traits are associated. Some distances (ISPNS, NSLZI, PTTSP) are associated with more than one hypothesis, indicated by double lines

the State of Minas Gerais, Brazil (Fazenda Canoas, 16°50'S, 43°35'W, 800 m). Twenty-one males and 16 females were placed under controlled conditions and randomly paired, producing over 400 sibs in a 1-year period; each litter was maintained with its dam until weaning. During this period, couples were rearranged at random; therefore, the colony design is unbalanced, containing both paternal and maternal half-sibs. Our sample is constituted of 365 skulls and 228 mandibles of individuals from this colony whose genealogical information is available, comprising individuals of both sexes and of different age classes.

Landmarks and Measurements

We registered three-dimensional coordinates for 20 cranial landmarks (Fig. 3) using a Microscribe MX digitizer (Microscribe, IL, USA). We registered bilaterally symmetrical landmarks in both sides, when available, for a total of 32

Fig. 4 The 10 registered landmarks represented over an adult left hemimandible of *C. expulsus*. Lines between landmarks represent the set of 20 Euclidean distances calculated between landmarks; the dashed line represents the two distinct morphological integration hypotheses tested over these traits, the (anterior) Alveolar Processes and the (posterior) Ascending Ramus. Landmark definitions follow: DA dorsal incisor alveolus, MA anterior limit of m1, MP posterior limit of m1, COR coronoid process dorsal limit, MCA anterior limit of mandibular condyle, MCP posterior limit of mandibular condyle, ANG angular process posterior limit, IDM maximum dorsal inflection between angular and Alveolar Processes, VC ventral limit of chin, VA ventral incisor alveolus

landmarks. Details of the digitizing procedure and landmark definition can be found in Cheverud (1995). We calculated a set of 35 inter-landmark distances (Fig. 3); trait names follow the landmarks of which they are composed. We measured each individual twice, in order to access measurement error through the estimation of repeatability (Lessels and Boag 1987). After this estimation, we averaged repeated measures, thus reducing our measurement error (Falconer and Mackay 1996); we also averaged distances that are present on both sides of the skull. Therefore, our set of 35 cranial traits is comprised of these averaged inter-landmark distances. Our inter-landmarks distances are designed to measure individual bones, thus capturing localized aspects of the skull development and avoiding the shortcomings of dealing with full length measures capturing several bones at same time (skull length for example) or principal components of shape.

We registered bi-dimensional coordinates for 10 mandibular landmarks (Fig. 4) over pictures we took of both hemimandibles (when available) using tpsDig2 (Rohlf 2006), and calculated a set of 20 inter-landmark distances (Fig. 4). We averaged measures from both hemimandibles to compose our set of 20 mandibular traits; as with the skull, trait names follow the landmarks of which they are composed. Since the mandible is a single bone, both landmark placements and inter-landmark distances suffer from an intrinsic uncertainty, because most landmarks are defined as the limits of mandibular processes (Bookstein 1991); therefore, our mandibular traits may not be as biologically meaningful as our cranial traits, since they only

describe mandibular morphology, not being associated with particular developmental processes.

We took a second set of pictures of a subset of 30 individuals; we used this subset to estimate repeatability taking into account both measurement error and the error associated with picture registration. To this purpose, we collected four sets of data for these individuals, two from each set of pictures. Repeatability in this case is the percentage of variance explained by the individual factor alone, excluding variation from both measurement and photo acquisition errors.

Estimation of Covariance Matrices

Since sexual dimorphism and ontogenetic variation are of little interest within the present context, we explored such sources of variation with respect to our sets of 35 cranial traits and 20 mandibular traits through analysis of multivariate covariance (MANCOVA), and evaluated the significance of these factors (and their interaction term) using Wilks' λ test. In order to remove the impact these effects may have on covariance structure, we used the residual covariance matrices of the two separate linear models (one for cranial traits and a second one for mandibular traits) as estimates of phenotypic covariance matrices for the *C. expulsus* population. We also estimated $\mathbf{P}$ s for different ages (20, 30, 50, 100, 200, 300 and 400 days) and for males and females. These matrices were compared in order to test whether these matrices are similar and can be grouped into a single $\mathbf{P}$ for the entire sample.

G-Matrix Estimation

For the estimation of genetic covariance matrices of these two sets of traits, we use the approach proposed by Runcie and Mukherjee (2013) that uses a Bayesian mixed-effects model in order to produce $\mathbf{G}$ estimates. This model builds upon the simple animal model (Lynch and Walsh 1998) by defining both $\mathbf{G}$ and $\mathbf{P}$ as depending on a set of latent traits, i.e., linear combination of the original set of traits. The prior distributions that define these factors impose two important constraints on their structure: first, that there is a declining influence of subsequent factors in terms of amount of explained phenotypic and genetic variance; second, the prior distributions impose sparsity over each factor; therefore, each latent trait is associated to a limited number of traits in the original set. Therefore, based on this factorization, $\mathbf{G}$ can be represented as

$$\mathbf{G} = \mathbf{A}\mathbf{\Sigma}_{h^2}\mathbf{A}^t + \mathbf{\Psi}_a \quad (1)$$

where $\mathbf{A}$ is the matrix of trait loadings on each factor, $\mathbf{\Sigma}_{h^2}$ is the diagonal matrix of factor heritabilities, and $\mathbf{\Psi}_a$ is the

diagonal matrix of trait-specific genetic variances. Due to its characteristic decomposition of $\mathbf{G}$ -matrices, this model was named Bayesian sparse factor analysis of genetic covariances (BSFG) by Runcie and Mukherjee (2013).

In order to obtain posterior distribution samples for $\mathbf{A}$, $\mathbf{\Sigma}_{h^2}$, and $\mathbf{\Psi}_a$ in both cranial and mandibular trait sets, we used a MCMC algorithm, obtaining 1,000 samples from 10,000 iterations, with a burnin period of 1,000 iterations. We initialized the MCMC run with 17 latent traits for both cranial and mandibular trait sets. We did a handful of MCMC runs with different initial prior distribution parameters, but posterior distributions were not affected by these differences. For each sample taken from the posterior distribution for those parameters described above, we constructed a different $\mathbf{G}$, obtaining a posterior distribution for this matrix; we also estimate a posterior mean matrix, which is our best estimate for $\mathbf{G}$.

Using the posterior distribution of heritabilities associated with each factor ($\mathbf{\Sigma}_{h^2}$), we test whether any given factor has a genetic variance different from zero, by computing the highest posterior density (HPD) intervals for all diagonal elements of $\mathbf{\Sigma}_{h^2}$. Using the posterior sample of $\mathbf{G}$ -matrices, we also estimated posterior intervals for trait heritabilities.

We also estimated $\mathbf{G}$ using a classical REML algorithm (Shaw 1987), treating genetic variances for each trait and pairwise genetic covariances as independent estimates; we grouped these estimates in matrix form to compose a $\mathbf{G}$ estimate. We did such proceeding using both a REML algorithm we wrote using R (R Core Team 2013) and also in WOMBAT (Meyer 2007), estimating additive variances and covariances using the simple animal model (Lynch and Walsh 1998). We compared these estimates with those obtained from the Bayesian sparse factor analysis outlined above; however, we chose to use the Bayesian model for four reasons. First, it produces estimates for the entire $\mathbf{G}$ structure simultaneously; when using REML algorithms, this structure has to be divided into its components (genetic variances and covariances) to be computationally tractable. Second, since $\mathbf{G}$ estimates from the BSFG model are samples from the posterior distribution, they are constrained to have only positive eigenvalues, even if only a reduced number of latent traits is included in the model. Furthermore, the latent traits ($\mathbf{A}$) obtained from the BSFG model might be informative in terms of the underlying processes influencing the genotype/phenotype map (*sensu* Wagner and Altenberg 1996). Finally, the posterior distribution for $\mathbf{G}$ we obtained from the model can be used to obtain posterior intervals for all matrix parameters we estimated (see the following subsections for details).

Although we prefer the BSFG model estimates for $\mathbf{G}$, the conventional REML estimates can be used to estimate

the effective sample size (N_{eff}) of each h^2 estimate; this represents the approximate number of independent additive values used to estimate that particular h^2 value (Cheverud 1995). The relationship

$$N_{eff} = \frac{2h^4}{V(h^2)} + 1 \quad (2)$$

can be used to estimate effective sample sizes, where h^4 is the squared h^2 estimate, and $V(h^2)$ is the estimated variance of the estimate, considering a normal approximation to the likelihood profile for that estimate. The geometric mean of individual N_{eff} estimates for each trait can then be used as a proxy for the experiment-wise effective sample size for $\mathbf{G}$ as a whole (Cheverud 1995).

Size Variation In order to investigate the influence of size variation over morphological integration, we also obtained matrices whose variation associated with the first principal component has been removed, as this component is usually associated with size in mammalian morphological systems (Wagner 1984; Marroig et al. 2004; Porto et al. 2013). As suggested by Bookstein et al. (1985), a residual matrix whose size variation has been removed can be obtained using the following relationship:

$$\mathbf{R} = \mathbf{C} - \mathbf{v}\mathbf{v}' \quad (3)$$

where $\mathbf{C}$ represents the raw (size retained) matrix and $\mathbf{v}$ denotes the unstandardized first eigenvector of $\mathbf{C}$; its norm is equivalent to the square root of the associated eigenvalue.

Matrix Comparisons

We compared our estimated $\mathbf{P}$ with the mean $\mathbf{G}$ estimated from the BSFG model, by using the Random Skewers method (Cheverud and Marroig 2007) for covariance matrices and matrix correlation followed by Mantel's test for significance (Cheverud et al. 1989) for correlation matrices. We also compared $\mathbf{P}$ with all $\mathbf{G}$ s from the posterior distribution of the BSFG model in order to estimate uncertainty in matrix similarity, obtaining a distribution of average Random Skewers and matrix correlation values. These comparisons directly tests Cheverud's Conjecture (Cheverud 1988; Roff 1995) of similarity between genetic and phenotypic covariance/correlation patterns.

We also compared covariance matrices using the selection response decomposition method (Marroig et al. 2011), which pinpoints differences between matrices with respect to trait covariance structure. Using this method, we are able to observe the correspondences between phenotypic and genetic covariance patterns for both mandibular and cranial traits. We compared both matrices with size variation retained and removed using the SRD method.

Morphological Integration Analysis

In order to investigate whether patterns described by $\mathbf{P}$ and $\mathbf{G}$ in our sets of cranial and mandibular traits conform to hypothesis of trait association due to functional and developmental interactions, we measured both magnitude and pattern of integration in phenotypic and genetic covariance/correlation matrices. We represented magnitude of morphological integration using the ICV (Shirai and Marroig 2010), which is the coefficient of variation of eigenvalues in a given covariance matrix.

We took two different approaches to investigate patterns of integration. In the classical approach (Olson and Miller 1958; Cheverud 1995), we constructed theoretical matrices to represent hypotheses of association between traits due to functional and/or developmental relationships expected a priori (Cheverud 1995). The hypotheses for cranial traits followed those proposed by Porto et al. (2009) for the Order Rodentia (Fig. 3). We constructed a total of nine theoretical matrices for cranial traits: five associated with localized hypotheses (Oral, Nasal, Zygomatic, Vault and Base), two associated with more global hypotheses that contrast early and late developmental patterns in mammals (Neurocranium and Face, respectively), and two composite hypotheses associated with these two groups (Total and Neuroface, respectively). Notice that Total integration correspond to the sum of all five individual hypothesis into one composite hypothesis and the Neuroface corresponds to the conjoint test of late and early developmental influence (or the sum of both individual hypothesis, Face and Neurocranium). With respect to mandibular traits, we considered the distinction between the Alveolar Processes and the Ascending Ramus, according to Fig. 4, as proposed by several authors (e.g.: Cheverud et al. 1997; Klingenberg et al. 2004; Willmore et al. 2009). We also tested a third, composite hypothesis (Total), corresponding to the sum of Alveolar and Ascending hypotheses. We calculated matrix correlations using Pearson-product moment correlation between each constructed theoretical matrix and the correlation matrices estimated from data in order to estimate the association between them (Cheverud et al. 1989; Cheverud 1995). We estimated significance for each correlation calculated this way by Mantel's test. Notice that, in the context of testing for hypotheses of association between traits, this classical approach is equivalent to a simple test of differences between the averages of two groups, integrated versus non-integrated traits, as a Student's t test, albeit with the significance test modified (using Mantel) to account for the non-independence of the observations (in this case, correlations among traits).

We investigated pattern and magnitude of morphological integration in matrices with size variation retained and removed; in order to compare estimates between these two

types of matrices, we used the modularity index proposed by Porto et al. (2013), calculated as the difference between the mean correlation taken from the set of correlations bounded by a given hypothesis and the mean correlation taken from the complementary set of correlations, divided by the ICV calculated for the corresponding covariance matrix. This index is, therefore, comparable across matrices with different magnitudes of integration, which is the case when comparing matrices with and without size variation.

By using the $\mathbf{G}$ posterior distribution obtained from the BSFG model, we estimated posterior intervals for both ICV and modularity indexes for each hypothesis. By estimating these intervals, we intend to represent the error associated with these parameters estimated over our mean $\mathbf{G}$. It also allows us to compare the estimated parameter values for both $\mathbf{P}$ and $\mathbf{G}$, under the null hypothesis that differences between these parameters for the two types of matrices are only due to error in estimating the mean $\mathbf{G}$. This comparison addresses the possibility of local differences between $\mathbf{P}$ and $\mathbf{G}$, in a manner similar to the SRD method.

Simulated Evolutionary Responses

The second and novel approach used to test hypotheses of trait associations is based on simulated evolutionary responses to selection, as suggested *en passant* by Hansen and Houle (2008). We used all covariance matrices available, including those whose size variation has been removed ($\mathbf{P}$ and $\mathbf{G}$ for both skull and mandible). For each matrix, we obtained an empirical distribution of flexibility (Marroig et al. 2009) without any a priori assumptions. Using 10,000 random normalized vectors drawn from a multivariate normal distribution without correlation structure, we estimated mean value ($\bar{f}$) and the 95 % confidence interval for f .

Using this f distribution, we explored our hypotheses of trait associations (as suggested by Hansen and Houle 2008) by constructing theoretical selection gradients that represent our hypothetical modules. Each β associated with a given hypothesis has a value of one for traits within the hypothetical module and zero otherwise; afterwards, each vector is also normalized. For each β constructed in this fashion (Oral, Nasal, Zygomatic, Vault, Base, Face and Neurocranium for the skull; Alveolar and Ascending for the mandible), we estimated the associated f values; the Neuroface (for the skull) and Total (for both skull and mandible) hypotheses cannot be properly represented as selection gradients, and are therefore excluded from this analysis. In order to test whether each of these vectors corresponds to a set of integrated traits in a given covariance matrix, we compare these f values with the critical

95 % interval around $\bar{f}$ for that matrix, obtained from selection gradients that are random with respect to that matrix covariance structure. If that particular f is higher than the critical value from the distribution, we considered this as evidence that the associated β represents trait associations embedded in that covariance matrix. In this case, the corresponding Δz follows β more closely than expected by chance alone; therefore, covariance structure interferes to a lesser extent in the response to selection in this case, demonstrating that the population has, to some degree, independent variation in the direction of that particular β .

In a manner similar to the ICV and modularity index, the $\mathbf{G}$ posterior distribution allows us to estimate posterior distributions for both $\bar{f}$ and f values associated with hypotheses of morphological integration. These distributions intend to represent the error associated with estimation of flexibilities in $\mathbf{G}$ -matrices; hence, they allow us to compare f values between $\mathbf{P}$ and $\mathbf{G}$, under the null hypotheses that differences in these values are only due to error in estimating $\mathbf{G}$.

BSFG Factors and Morphological Integration

In order to investigate the relationship between the factors estimated by the BSFG model and our hypotheses of morphological integration, we calculated vector correlations between these factors and the vectors constructed to represent our hypotheses, as outlined in the previous section. We estimated significance for these correlations by using a distribution of correlations obtained from random vectors, in a manner similar to the estimation of significance in the Random Skewers method (Cheverud and Marroig 2007). If vector correlations between any given combination of latent traits and hypothesis vectors is higher than the critical value estimated from random vector correlations, we consider that correlation as evidence of non-random association between latent traits and hypothesis.

Results

Measurement Error Assessment

Repeatabilities for cranial traits (Table S1) ranged from 0.654 to 0.998 with an average value of 0.956 and standard deviation 0.068. For mandibular traits (Table S2), values ranged from 0.869 to 0.992, with mean 0.961 and SD 0.037. Traits with low repeatabilities were those with low variances, such as MTPNS. For cranial traits, these results for repeatability should not impact other results, since repeatabilities for averaged traits that have been measured

Table 1 Unstandardized first principal components obtained from covariance matrices and factors retrieved from the BSFG model for cranial traits

Trait	Hypothesis	Region	p_1	g_1	λ_1	λ_2	λ_3	λ_4
ISPM	Oral	Face	−0.868	−0.268	−0.347	0.013	−0.003	0.003
PMZS	Oral	Face	−1.196	−0.37	−0.464	0.132	0.012	−0.009
PMZI	Oral	Face	−1.47	−0.469	−0.582	0.167	0.031	−0.001
PMMT	Oral	Face	−0.764	−0.23	−0.302	0.001	−0.008	−0.053
MTPNS	Oral	Face	−0.085	−0.023	−0.031	0.002	0.011	−0.003
ISPNS	Oral/Nasal	Face	−1.804	−0.551	−0.708	0.095	0.005	−0.14
NSLZI	Oral/Nasal	Face	−2.607	−0.815	−1.029	0.19	0.01	0.009
ISNSL	Nasal	Face	−0.563	−0.181	−0.235	0.017	−0.005	0
NSLNA	Nasal	Face	−1.822	−0.55	−0.699	0.09	−0.044	−0.021
NSLZS	Nasal	Face	−2.22	−0.68	−0.858	0.194	−0.001	0.006
NAPNS	Nasal	Face	−0.879	−0.269	−0.356	−0.012	−0.006	−0.095
PTZYGO	Zygomatic	Face	−0.951	−0.296	−0.4	−0.202	0	0.016
ZSZI	Zygomatic	Face	−0.454	−0.154	−0.2	−0.027	0.008	0.002
ZIMT	Zygomatic	Face	−0.854	−0.274	−0.349	0.102	−0.011	0.023
ZIZYGO	Zygomatic	Face	−0.474	−0.137	−0.2	0.005	0.003	−0.009
ZITSP	Zygomatic	Face	−0.851	−0.262	−0.342	0.03	−0.022	−0.001
EAMZYGO	Zygomatic	Face	−0.504	−0.144	−0.191	0.001	−0.021	0.005
ZYGOTSP	Zygomatic/ Vault	Face/Neuro	−0.68	−0.221	−0.287	0.019	0.004	0.007
PTTSP	Vault	Neurocranium	−0.382	−0.116	−0.155	−0.036	0.003	−0.011
NABR	Vault	Neurocranium	−0.797	−0.25	−0.321	0.146	−0.058	0.017
BRPT	Vault	Neurocranium	0.001	0.016	0.014	−0.06	−0.017	0.002
BRAPET	Vault	Neurocranium	−0.556	−0.16	−0.219	−0.109	0.047	−0.032
PTAPET	Vault	Neurocranium	−0.906	−0.272	−0.371	−0.177	−0.002	0.009
PTBA	Vault	Neurocranium	−1.414	−0.43	−0.572	−0.16	0.016	0.042
PTEAM	Vault	Neurocranium	−1.13	−0.344	−0.468	−0.243	0.002	0.025
LDAS	Vault	Neurocranium	−0.148	−0.04	−0.052	−0.001	0.081	0.002
BRLD	Vault	Neurocranium	−0.437	−0.122	−0.152	−0.017	0.369	−0.001
OPILD	Vault	Neurocranium	−0.69	−0.212	−0.284	−0.094	−0.065	−0.014
PTAS	Vault	Neurocranium	−0.741	−0.22	−0.291	−0.173	0.085	0.016
JPAS	Vault	Neurocranium	−0.627	−0.192	−0.262	−0.107	−0.006	0.004
PNSAPET	Base	Neurocranium	−1.015	−0.319	−0.415	0.069	0	0.169
APETBA	Base	Neurocranium	−0.688	−0.211	−0.271	0.048	0.002	0.013
APETTS	Base	Neurocranium	−0.168	−0.052	−0.069	−0.006	−0.001	−0.019
BAEAM	Base	Neurocranium	−0.496	−0.155	−0.2	0.039	0.016	0.008
BAOPI	Base	Neurocranium	0.139	0.048	0.06	−0.011	0.003	−0.012

Bold values indicate those factor loadings that differ from zero, according to the 95 % posterior interval of factor loadings

twice (as we have done here) are higher than traits with single measurements (Falconer and Mackay 1996). For mandibular traits, the lowest repeatability value is above 0.85; therefore, we expect that further analysis should not be impacted by the error arising from both photo acquisition and measurement procedure.

Matrix Estimation

Regarding estimation of phenotypic matrices, both linear models adjusted for removal of age and sex fixed effects (and their interaction) were significant for all factors used,

for both sets of traits; separate estimates for P at different ages and in both males and females are also similar (all comparisons above 0.68; results not shown). Therefore, we used the residual covariance matrices from these models as estimates of P .

Heritabilities for latent traits recovered from the BSFG model (Table S3) indicate that, for both cranial and mandibular trait sets, four of the 17 latent traits have posterior distributions of h^2 that do not include the null value. For both cranial and mandibular trait sets, the first latent trait recovered is a direction associated with size variation, as indicated by the negative sign associated with almost all

Table 2 Unstandardized first principal components obtained from covariance matrices and factors retrieved from the BSFG model for mandibular traits

Trait	Hypothesis	p_1	g_1	λ_1	λ_2	λ_3	λ_4
DAMA	Alveolar	−0.127	−0.078	0.170	0.006	0.015	0.053
DAMP	Alveolar	−0.162	−0.096	0.196	0.007	0.016	−0.047
MAMP	Alveolar	−0.036	−0.018	0.024	0	0	−0.105
DAVC	Alveolar	−0.142	−0.067	0.115	0.003	0.128	−0.058
MAVC	Alveolar	−0.174	−0.085	0.156	0.003	0.034	0.040
MPVC	Alveolar	−0.194	−0.103	0.186	−0.007	−0.031	−0.034
DAVA	Alveolar	−0.076	−0.043	0.080	0.003	0.141	−0.026
MAVA	Alveolar	−0.141	−0.080	0.162	0.006	−0.001	0.072
MPVA	Alveolar	−0.168	−0.094	0.178	0	−0.04	−0.044
VCVA	Alveolar	−0.063	−0.021	0.029	0.001	−0.009	−0.036
CORMCA	Ascending	−0.160	−0.090	0.070	−0.263	−0.002	−0.019
CORMCP	Ascending	−0.212	−0.108	0.093	−0.284	0.001	0.004
MCAMCP	Ascending	−0.050	−0.013	0.019	−0.001	0.004	0.028
CORANG	Ascending	−0.387	−0.180	0.238	−0.167	−0.011	−0.026
MCAANG	Ascending	−0.234	−0.109	0.168	−0.013	−0.003	−0.021
MCPANG	Ascending	−0.183	−0.083	0.120	0.004	−0.009	−0.027
CORIDM	Ascending	−0.364	−0.152	0.257	0.033	0.011	−0.021
MCAIDM	Ascending	−0.360	−0.169	0.258	−0.068	0.013	−0.036
MCPIDM	Ascending	−0.390	−0.169	0.250	−0.058	0.013	−0.017
ANGIDM	Ascending	−0.413	−0.182	0.278	−0.062	0.010	−0.020

Bold values indicate those factor loadings that differ from zero, according to the 95 % posterior interval of factor loadings

traits in λ_1 for cranial traits (Table 1, with the exception of BRPT and BAOPI) and the positive sign associated with λ_1 for mandibular traits (Table 2). Considering the posterior distribution of factor loadings in λ_1 for both skull and mandible, almost all traits have factor loadings different from zero, except for the cranial trait BRPT and mandibular traits VCVA and MCAMCP.

The second cranial latent trait (λ_2 in Table 1) has positive loadings associated with most facial traits, and negative loadings related to neurocranial traits. Therefore, this factor can be understood as a contrast between these two regions. The third factor, λ_3 , is only associated with Vault characters, while the fourth, λ_4 , recovers a pattern similar to λ_2 , involving contrasts between facial and neurocranial traits located in the ventral side of the skull.

The three remaining mandibular latent traits (λ_2 , λ_3 , λ_4 in Table 2) have only a handful of traits whose factor loadings are different than zero, and such sparse factor loadings have localized effects with respect to mandibular partitioning. The second factor affect only traits belonging to the Ascending Ramus, while the other two factors are associated with Alveolar traits.

Following Eq. 1, we constructed our estimated mean $\mathbf{G}$ and the posterior distribution for $\mathbf{G}$ using all factors we estimate. Comparing the mean $\mathbf{G}$ estimated by the BSFG model with $\mathbf{G}$ s estimated using REML (both our own R-based algorithm and Wombat) using Random Skewers (for covariance matrices) and matrix correlation followed

by Mantel's test (for correlation matrices) indicate that matrices estimated from the three methods are fairly similar. Random Skewers correlations ranged from 0.84 to 0.91, and matrix correlations ranged from 0.72 to 0.84; all correlations are indicative of a lack of structural dissimilarities under their respective tests of significance (for $P(\alpha) < 10^{-4}$). Estimated trait heritabilities are also similar among these three different types of estimation methods (Figures S1 and S2). Therefore, the BSFG estimates of $\mathbf{G}$ we use here, for both cranial and mandibular trait sets, are very similar to the traditional REML estimates.

Using the REML estimates for h^2 , we were able to estimate the effective sample sizes for both cranial (Table S1) and mandibular (Table S2) traits. For cranial traits, we obtained an average N_{eff} of 20 individuals; for mandibular traits, the estimated average N_{eff} is 32 individuals.

For both cranial and mandibular matrices, the first eigenvector of both $\mathbf{P}$ and $\mathbf{G}$ are indeed size vectors (represented as p_1 and g_1 in Tables 1 and 2). Hence, by removing the first eigenvector of these matrices following Eq. 3, we obtain matrices without size variation, removing both variation from scaling and allometry.

Matrix Comparisons

All comparisons between $\mathbf{P}$ and the mean $\mathbf{G}$, using both Random Skewers correlations for covariance matrices and matrix correlation followed by Mantel's test, reject the null

Table 3 Comparison among *P* and *G* for both cranial and mandibular traits

		Value	Mean	Median	PI–	PI+
Skull	Cov	0.988	0.955	0.963	0.902	0.987
	Cor	0.975	0.927	0.936	0.856	0.969
Mandible	Cov	0.960	0.919	0.923	0.869	0.963
	Cor	0.942	0.832	0.839	0.731	0.920

“Value” represents the average Random Skewers correlation or matrix correlation Γ -values between *P* and the mean *G* for covariance and correlation matrices, respectively. Other statistics (mean, median, 95 % posterior interval) refer to the distribution of the same matrix correlation statistics between *P* and the posterior distribution for *G*. All comparisons reject the null hypothesis of no structural similarity between compared matrices at $P(\alpha) = 0.001$

hypothesis of structural dissimilarity between these matrices (Table 3), with very high correlation values (all above 0.9). Comparing *P* with any *G* derived from the posterior distribution of the BSFG model results in matrix correlations which reject this null hypothesis for all comparisons. Correlation values for covariance matrices are systematically higher than those derived from correlation matrices, considering both the comparison between *P* with the mean *G* and the posterior distribution of matrix correlations (Table 3).

The comparison between *P* and *G* using the SRD method (Fig. 5) indicates that, in matrices whose size variation has been retained, few traits have marked differences in covariance structure. In the cranial set (Fig. 5a; with an average SRD score of 0.99), only trait BRPT shows a substantial difference in trait covariance structure between *P* and *G*; in the mandibular set (Fig. 5c; with an average SRD score of 0.98), trait MCAMCP has the lowest SRD score. Comparing matrices whose size variation has been removed (Fig. 5b, d) yields lower SRD scores (although still quite high: 0.92 for the skull, and 0.89 for the mandible); differences between the two matrices are more distributed through each set.

Morphological Integration

When considering matrices whose size variation has been retained, ICV's estimated for both cranial and mandibular trait sets do not differ between *P* and *G*, as both phenotypic ICV's are within the 95 % posterior interval constructed using sampled *G*s (Table 4). For size-free matrices, for cranial traits, ICV values between *P* and *G* also do not differ; however, the phenotypic value for the mandibular *P* ICV (1.332) is below the lower bound of the posterior interval (1.576). It is also noteworthy that, when size is

removed, there is a substantial reduction in ICV values in both trait sets.

Regarding pattern of morphological integration in cranial traits (Table 4), both Nasal and Facial trait subsets are identified by Mantel's test; the composite Neuroface hypothesis is also identified. When size variation is removed, the Oral, Vault and Neurocranial regions are identified; the Neuroface region is identified again. There is agreement in highlighted regions between *P* and *G*; modularity indexes estimated for phenotypic matrices are within the range of their respective 95 % posterior intervals, indicating that patterns of morphological integration are not different between *P* and *G*, when considering the uncertainty associated with estimating *G*.

Considering the mandible, only the total composite hypothesis is identified by Mantel's test in matrices with size variation (Table 4). When removing size variation, all three hypotheses (Alveolar, Ascending and Total) are identified. As with the cranial trait set, there is agreement between the hypothesis identified for both phenotypic and genetic matrices. The posterior distribution of modularity indexes indicate, for those matrices with size variation retained, that patterns of morphological integration are not different between matrices compared; however, when considering the comparison between matrices without size variation, modularity indexes estimated for *P* for both Alveolar (0.055) and Total (0.08) hypotheses are slightly above the upper limits of their respective posterior intervals (0.054 and 0.079).

Flexibility

Regarding our analysis of morphological integration based on simulated evolutionary responses (Table 5), for both cranial matrices, the β s associated with Facial, Oral, and Nasal traits had flexibility values significantly higher than expected by chance alone; when size variation is removed, the β s associated with the Vault and Base were also higher than average f values. For mandibular traits, the β associated with the Ascending Ramus had f values higher than $\bar{f}$ for both *P* and *G*; when size variation is removed, the Alveolar Process β was also associated with a high f value in *P*, but not in *G*.

Considering the differences in flexibility values between *P* and *G*, only the f -value estimated for the Alveolar region in *P* with size removed (0.838) is above the upper bound of the posterior distribution of f -values derived from sampled *G*s (0.806); the $\bar{f}$ value calculated for *P* (0.626) is also above the limit drawn by the posterior distribution (0.561). Remaining f -values, in all other phenotypic matrices, are within the range of their respective posterior intervals.

Fig. 5 Selection response decomposition plots for the comparison between phenotypic and genetic matrices for both cranial (a, b) and mandibular (c, d) traits; size variation has either been retained (a, c) and removed (b, d) for each matrix in these comparisons. In all

comparisons, *dashed lines* represent average SRD scores. Cranial or mandibular traits indicated with *triangles* are those which differ significantly between each *P/G* set, with $P(x) = 0.025$.

Factors and Morphological Integration

Correlations between factors recovered from the BSFG model and our hypotheses (Table 6) indicate that the first factor for cranial traits has high correlation values with the Oral, Nasal and Facial β s, while the second factor is mildly correlated with the Vault β . Regarding mandibular latent traits, λ_1 has high correlations with both Alveolar and

Ascending β s, while λ_2 has a high correlation with only the β associated with the Ascending Ramus.

Discussion

The results we obtained here point out that phenotypic and genetic covariance structure are very similar in the vesper

Table 4 Magnitude and pattern of phenotypic and genetic integration, as measured by ICV and modularity index, respectively

	Size	Retained				Removed			
		<i>P</i>	<i>G</i>	<i>PI</i> −	<i>PI</i> +	<i>P</i>	<i>G</i>	<i>PI</i> −	<i>PI</i> +
Skull	<i>ICV</i>	4.091	3.794	2.615	4.791	1.502	1.448	1.299	2.042
	Oral	<i>0.049</i>	<i>0.056</i>	0.031	0.087	0.018	0.025	0.008	0.037
	Nasal	0.083	0.092	0.065	0.127	0.016	<i>0.024</i>	0.007	0.037
	Zygomatic	−0.005	−0.005	−0.031	0.025	−0.013	−0.015	−0.026	0.004
	Vault	−0.011	−0.009	−0.024	0.017	0.035	0.045	0.015	0.067
	Base	−0.058	−0.056	−0.073	−0.039	−0.005	0.001	−0.016	0.018
	Total	0.007	0.011	0.002	0.029	0.021	0.029	0.010	0.046
	Face	0.039	0.045	0.032	0.059	0.002	0.004	−0.001	0.009
	Neuro	−0.031	−0.032	−0.043	−0.02	0.017	0.019	0.006	0.031
	Neuroface	0.010	0.014	0.007	0.025	0.014	0.017	0.005	0.027
Mandible	<i>ICV</i>	2.737	2.656	2.283	3.120	1.332	1.822	1.576	2.355
	Alveolar	0.003	0.011	−0.019	0.049	0.055	0.027	0.002	0.054
	Ascending	<i>0.050</i>	0.033	−0.010	0.078	0.056	0.040	0.009	0.060
	Total	0.039	0.031	0.006	0.060	0.080	0.048	0.011	0.079

Both cranial and mandibular matrices are represented, with size variation either retained or removed. Posterior intervals ($PI_{\gamma=0.95}$) are associated with the posterior distribution for *G*, representing parameter uncertainty in *G* parameters. Bold values are associated with morphological integration hypotheses that are recognized in a given matrix by Mantel's test ($P(\alpha) = 0.05$); italic values are marginally significant ($P(\alpha) = 0.1$)

Table 5 Flexibilities of phenotypic and genetic matrices, Both cranial and mandibular matrices are represented, with size variation either retained and removed

	Size	Retained				Removed			
		<i>P</i>	<i>G</i>	<i>PI</i> −	<i>PI</i> +	<i>P</i>	<i>G</i>	<i>PI</i> −	<i>PI</i> +
Skull	$\bar{f}$	0.299	0.322	0.244	0.427	0.583	0.322	0.488	0.628
	Critical	0.536	0.556	–	–	0.690	0.713	–	–
	Oral	0.567	0.576	0.555	0.612	0.449	0.423	0.330	0.499
	Nasal	0.684	0.690	0.668	0.722	0.464	0.490	0.385	0.547
	Zygomatic	0.348	0.352	0.321	0.414	0.606	0.640	0.437	0.710
	Vault	<i>0.464</i>	<i>0.486</i>	0.407	0.624	0.734	0.717	0.634	0.757
	Base	0.199	0.226	0.186	0.300	0.739	0.720	0.549	0.767
	Face	0.746	0.749	0.739	0.760	0.428	0.392	0.292	0.454
	Neuro	<i>0.467</i>	0.479	0.429	0.574	0.654	0.642	0.553	0.709
Mandible	$\bar{f}$	0.418	0.410	0.346	0.447	0.626	0.410	0.444	0.561
	Critical	0.630	0.594	–	–	0.778	0.719	–	–
	Alveolar	0.554	<i>0.568</i>	0.514	0.638	0.838	0.628	0.438	0.806
	Ascending	0.860	0.843	0.785	0.887	0.437	0.384	0.245	0.494

Posterior intervals ($PI_{\gamma=0.95}$) are associated with the posterior distribution for *G*, representing parameter uncertainty in *G* parameters. Bold values are associated with morphological integration hypotheses that are recognized in a given matrix by comparison with the null distribution of flexibilities for any given matrix ($P(\alpha) = 0.05$); italic values are marginally significant ($P(\alpha) = 0.1$). ‘Critical’ refers to the 95 % upper bound value for these comparisons

mouse, for both skull and mandible; while differences in trait-specific covariances exist, as shown by the SRD method (Fig. 5), these differences are local and do not affect the overall covariance structure, as shown by matrix correlation between *P* and *G* (Table 3), and also by average SRD scores, even when the effects of size variation is controlled (Fig. 5). Hence, these comparisons corroborate

Cheverud's Conjecture (Cheverud 1988; Roff 1995) of similarity between patterns of correlation/covariance structure in both *P* and *G*. We will now focus on exploring the potential underlying causes of this similarity.

The observed differences in trait-specific covariances, pinpointed by the SRD method (Fig. 5) do not affect our hypotheses of trait associations attributed to developmental

Table 6 Correlation between factors recovered from the BSFG model and morphological integration hypotheses

		λ_1	λ_2	λ_3	λ_4
Skull	λ_2	0.204			
	λ_3	0.060	0.100		
	λ_4	0.040	0.065	0.009	
	Oral	0.546	0.357	0.055	0.284
	Nasal	0.662	0.369	0.041	0.380
	Zygomatic	0.313	0.061	0.032	0.043
	Vault	0.377	0.468	0.323	0.067
	Base	0.167	0.097	0.022	0.275
	Face	0.740	0.281	0.019	0.242
Mandible	Neuro	0.408	0.341	0.284	0.205
	λ_2	0.369			
	λ_3	0.204	0.015		
	λ_4	0.322	0.154	0.180	
	Alveolar	0.533	0.016	0.394	0.308
	Ascending	0.721	0.636	0.044	0.258

Values highlighted represent vector correlation values higher than those expected by chance, with $P(x) = 0.05$ for italic values and $P(x) = 0.01$ for bold values

and functional interactions. Overall, the same subsets of traits are integrated in both phenotypic and genetic correlation (Table 4) and covariance (Table 5) patterns. Although these results agree between each P/G set, there are minor differences between the traditional method employed on correlation matrices and the method we propose here, used on covariance matrices. For instance, while the Neurocranium is identified as a valid partition in correlation matrices without size variation, the same partition is not identified by our evolutionary simulations approach.

While these differences might occur due to differences in type II error rates for both tests (which can be subject to future investigations), our simulation-based approach has the advantage of being directly related to the use of covariance matrices in evolutionary quantitative genetics theory. The classical test of morphological integration (Cheverud et al. 1989) has its roots on the comparison of distance matrices (Mantel 1967), while our simulation approach is based on the analysis of evolutionary response to selection as suggested by Hansen and Houle (2008). Therefore, this approach is more appropriate for dealing with systems (and representations of such systems, i.e., covariance matrices) that could be subjected to evolutionary change. On the other hand, the classical approach has the advantage of being able to handle more complex hierarchical structures that cannot be properly represented as a vector. For example, the contrast between early and late developmental factors, represented by a composite

hypothesis that compares correlations within facial and neurocranial traits with correlations between these two groups, i.e. the Neuroface hypothesis, cannot be tested in the selection gradient framework, since this test considers the correlation structure of all traits simultaneously and cannot be represented as a binary vector. Therefore, both approaches complement each other, being under different premises and expectations.

The distribution of posterior G statistics that represent pattern and magnitude of morphological integration (ICV, modularity index, flexibility values) indicate that almost all differences in these values between P and G may be attributed only to error in estimating G , with the exception of features associated with the mandibular matrices when size is removed. Considering together results in Tables 4 and 5 indicate that the mandibular P is less integrated and more modular than its genetic counterpart, as indicated by the lower ICV value and higher f -values and modularity indexes associated with Alveolar traits. However, these differences are found only when size variation is removed; such source of variance represents a substantial portion of overall magnitude of integration. Therefore, these differences between P and G in the mandible are minor compared to the overall similarity in matrix structure.

Size Variation and Integration Patterns

Removing size from both phenotypic and genetic matrices produces substantial changes in covariance structure. First and foremost, differences between trait-specific covariances increase (Fig. 5). Magnitude of integration, as measured by ICV, drops considerably, in all matrices whose size variation has been removed. Patterns of integration also change; different hypotheses of trait association are recognized as valid.

The variation in size within populations has a pervasive effect throughout development, from zygote to adult. However, growth will have an heterogeneous effect over traits, producing allometric variation. Overall, regarding the mammalian skull, facial traits are more influenced by growth than neural traits during the pre-weaning period and afterwards (Zelditch and Carmichael 1989; Hallgrímsson et al. 2009), due to overall growth and muscle-bone interactions associated with both breastfeeding and mastication. Thus, the influence of size variation is more pervasive in the integration of facial traits in the skull (Tables 4, 5).

Size variation also has an influence on integration patterns of the mandible; the Ascending Ramus is recognized as a valid hypothesis of trait association by our evolutionary simulation approach when we consider matrices whose size variation has been retained (Table 5). The masseter muscle, which affects the growth of the associated

Ascending Ramus through its function, has a marked influence on pre-weaning development (Zelditch and Carmichael 1989). The interactions between the masseter and the mandible also produce an extensive reworking of the condylar cartilage (Herring 2011), which might explain the difference in MCAMCP (condyle length) covariance structure between P and G (Fig. 5). Therefore, both functional and developmental interactions shape the observed integration patterning through growth in both skull and mandible.

When size variation is removed, different features of the developmental system are observed. In the skull, we observe integration in neurocranial traits, in both analysis (Tables 4, 5). For the mandible, the traditional analysis (Table 4) recognizes both Alveolar and Ascending Ramus hypothesis as valid. These other aspects of integration reflect aspects of prenatal development. In the skull, they are associated with prenatal brain growth (Hallgrímsson and Lieberman 2008; Hallgrímsson et al. 2009); in the mandible, the contrast between the two regions may reflect the distinct cell condensation from which each structure arises (Atchley and Hall 1991; Ramaesh and Bard 2003). While trait-specific covariance structure diverges between P and G when size is removed (Fig. 5b, d), these differences do not impact the relationships between these traits, as the same hypotheses of trait associations are recognized in both matrices (Tables 4, 5). Therefore, these differences in trait-specific covariance structure depicted by the SRD method may reflect differences in covariance structure outside of those regions that are recognized as integrated when size is removed.

Integration patterns produced by size variation and developmental patterning are temporally organized, since their main influence occur at postnatal and prenatal development, respectively. Hence, these developmental factors affect covariance patterns in the skull and mandible with different strengths, with size integration partially masking the effect of other patterns produced by prenatal growth (Hallgrímsson and Lieberman 2008; Hallgrímsson et al. 2009). The relative contribution from both these effects to response to selection will reflect this hierarchy, with size variation contributing to a major extent relative to prenatal patterning, acting as a line of least genetic resistance, with evolutionary consequences (Schluter 1996; Marroig and Cheverud 2005).

Our treatment of size variation contrasts with the approach advocated by landmark-based geometric morphometrics (Bookstein 1991; Zelditch et al. 2004), in which size is treated as a separate variable, centroid size. Certainly, this approach has advantages; for instance, it allows for a clear representation of allometric relationships, by regressing shape variables on centroid size. With our approach, we are not able to differentiate between

isometric and allometric variation; our size factors are associated with both variation components. However, quantitative genetics analyses using landmark-based geometric morphometrics (e.g. Klingenberg and Leamy 2001; Klingenberg et al. 2004; Martínez-Abadías et al. 2011) have not fully appreciated this advantage, since size variation is treated as a nuisance with respect to shape and its effects are removed or simply ignored. Therefore, the effects of size variation over shape genetic covariance structure remain unexplored under the purview of geometric morphometrics.

However, the main issue regarding the use of landmark-based data to investigate covariance structure lies at the assumption of isotropic variation around each landmark necessary to perform the Generalized Procrustes Superimposition (GPS; Dryden and Mardia 1998; Linde and Houle 2009). Under a morphological integration perspective, we expect that local function and/or developmental factors will produce differences in both magnitude and direction of landmark variation; therefore, the assumption of isotropic variation is incompatible with analyses of covariance structure. If a particular dataset breaks this assumption and GPS is performed nonetheless, the result is that landmark variation will be forced to conform to the assumption by distributing total variation across all landmarks (Linde and Houle 2009) and consequently erasing most of the actual integration patterns.

This issue might explain the contrast between the results from Cheverud et al. (1997) and Klingenberg et al. (2004) regarding the organization of mandibular pleiotropic QTL effects in the same *Mus* population using traditional and geometric morphometrics, respectively. While Cheverud et al. (1997) found that most of these effect are confined within both Alveolar Process and Ascending Ramus (see also Mezey et al. 2000; Cheverud 2006), Klingenberg et al. (2004) found evidence that most pleiotropic effects are shared by the two regions. Some authors have proposed solutions that circumvent this problem (e.g. Theobald and Wuttke 2006; Linde and Houle 2009; Márquez et al. 2012), but these works have not yet been appreciated outside the literature of morphometric theory.

Latent Traits and Developmental Factors

The latent traits recovered from the BSFG model (Tables 1, 2) may be understood as a local linear approximation of the developmental function, centered at the mean phenotype, as they are derived from a particular hyper-parametrization of G we use here (Runcie and Mukherjee 2013). When considering a set of linear approximations to developmental factors, other sets obtained through rotations of the original set may also be considered as a valid approximation (Wagner 1984; Wolf

et al. 2001; Mitteroecker 2009; Runcie and Mukherjee 2013). However, the set of latent traits we recovered here are constrained by sparsity, and thus are not arbitrary with respect to rotation.

The first factor recovered for both cranial and mandibular traits is associated with size variation (Tables 1, 2), and reflect the influence of this source of variation over morphological integration. Interestingly, the size factor recovered for both skull and mandible breaks the a priori assumption of sparsity made by the BSFG model, which is indicative of strong evidence favoring the existence of such factor. Remaining latent traits reflect our hypotheses of morphological integration (Table 6); in the skull, the second trait recovered is associated with facial and neurocranial traits simultaneously, albeit with opposing signs (Table 1). This factor is usually recovered from mammalian skull variation from our previous studies (e.g.: Marroig et al. 2004); in the mandible, each latent trait recovered aside from the first has localized effects over either Alveolar or Ascending traits (Table 2).

The structure of latent traits we recovered here are an abstraction over more complex developmental dynamics; individually, they may or may not reflect actual aspects of the organization of pleiotropic effects. The first factor in both skull and mandible recovers a biologically realistic factor, that is, size variation; in fact, any gene involved in controlling the amount of cell metabolic output or cellular growth and division will contribute to size variation. The second factor in the skull essentially contrasts the face with the neurocranium. While this factor certainly captures one essential aspect of mammalian skull development (one that contrasts early- and late-developmental growth as recognized here in our integration hypotheses) on the other hand it does not correspond to the genetic basis of skull covariation. In fact, most QTLs found in previous studies (Cheverud et al. 1997; Leamy et al. 1999) affect either the Face (31 %) or Neurocranium (31 %) separately and only 38 % affected the whole skull. More importantly, of those 38 % with general effects on the skull only 20 % (7.6 % out of the total QTLs) had effects with opposite signals on the Face and Neurocranium which would indicate antagonist pleiotropy. For the mandible essentially the same pattern was found with 26 % of the 41 QTLs found did affect the whole mandible and the remaining 74 % have only localized effects with no evidence of antagonistic pleiotropy. Therefore, latent traits for the skull and mandible display different structures, although there are only minor differences in their underlying genetic architectures with respect to pleiotropic effects, as demonstrated by QTL studies. These latent traits should then be interpreted as a set, and are not to be taken individually as representations of true developmental factors, in the same manner that individual principal components also cannot be regarded as

biologically meaningful (Zelditch et al. 2004; Adams et al. 2011; Berner et al. 2011; Berner 2012).

These results suggest that similarity between P and G is the result of limits imposed by the developmental system to the action of additive effects over morphological traits (Cheverud 1984; Wolf et al. 2001). These limits are the result of both developmental and functional interactions, and are reflected in our hypothesis of trait associations, which are consistently recognized in both matrices and in the latent traits identified by the BSFG model. While these latent traits may be understood as developmental modules (Runcie and Mukherjee 2013), or more likely, as combinations or contrasts between interacting modules, they are temporally organized throughout development (Zelditch and Carmichael 1989; Hallgrímsson and Lieberman 2008), and their effect over both P and G will reflect this organization.

It is paramount to understand modularity not as a static feature of morphological systems, but as a feature embedded within a dynamical process, that is, development, which will produce an integrated phenotype (Cheverud 1996; Wagner 1996; Hallgrímsson et al. 2009). While our hypothesis of trait associations might be quite simple (as they are essentially binary descriptors of subsets of traits), our results are interpreted in the light of developmental dynamics. Therefore, rather than address only to the identification of modules, one should consider the dynamics of the underlying development when trying to understand both the structure and evolution of complex morphological elements subject to morphological integration.

Conclusions

Integration and modularity are paramount features of morphological systems, and the evolution of such systems is strongly affected by them. The patterns embedded within genetic and phenotypic covariance matrices capture both these phenomena simultaneously; hence, by estimating covariance patterns within populations, we are able to quantify the relative influence of both magnitude and pattern of integration, and the relationship between G and P , established by the genotype/phenotype map. Here, we investigate both these levels of organization in the skull and mandible of a *C. expulsus* population, finding remarkable similarities in their overall structure and in the patterns they describe, as predicted by Cheverud's Conjecture (Cheverud 1988; Roff 1995). This similarity between covariance patterns is achieved through the constraints imposed by the developmental system, through the association of morphological elements due to developmental and functional interactions, formulated as

hypotheses of trait associations that are recognizable in both genetic and phenotypic covariance structure.

Acknowledgments We would like to thank N. P. Barros, A. M. Marcondes, F. Almeida, L. Araripe, J. M. Freschi for help with lab work; F. A. Machado, D. Melo, and two anonymous reviewers for comments on early drafts. We also thank D. Runcie and S. Murkherjee for help with their BSFG model codes. This work has been supported by grants from CNPq (Conselho Nacional de Pesquisa e Desenvolvimento), FAPERJ (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), MMA (Ministério do Meio Ambiente) and MCT (Ministério de Ciência e Tecnologia).

References

- Adams, D. C., Cardini, A., Monteiro, L. R., O'Higgins, P., & Rohlf, F. J. (2011). Morphometrics and phylogenetics: Principal components of shape from cranial modules are neither appropriate nor effective cladistic characters. *Journal of Human Evolution*, 60(2), 240–243. doi:10.1016/j.jhevol.2010.02.003.
- Almeida, F. C., Bonvicino, C. R., & Cordeiro-Estrela, P. (2007). Phylogeny and temporal diversification of calomys (rodentia, sigmodontinae): Implications for the biogeography of an endemic genus of the open/dry biomes of South America. *Molecular Phylogenetics and Evolution*, 42(2), 449–466. doi:10.1016/j.ympev.2006.07.005.
- Atchley, W. R., & Hall, B. K. (1991). A model for development and evolution of complex morphological structures. *Biological Reviews*, 66, 101–157.
- Berner, D. (2012). How much can the orientation of g's eigenvectors tell us about genetic constraints? *Ecology and Evolution*, 2(8), 1834–1842. doi:10.1002/ece3.306.
- Berner, D., Kaeuffer, R., Grandchamp, A. C., Raeymaekers, J. A. M., Räsänen, K., & Hendry, A. P. (2011). Quantitative genetic inheritance of morphological divergence in a lake-stream stickleback ecotype pair: Implications for reproductive isolation. *Journal of Evolutionary Biology*, 1–9. doi:10.1111/j.1420-9101.2011.02330.x.
- Bonvicino, C., Lima, J., & Almeida, F. (2003). A new species of calomys waterhouse (rodentia, Sigmodontinae) from the cerrado of central Brazil. *Revista Brasileira de Zoologia*, 20(2), 301–307.
- Bonvicino, C. R., & Almeida, F. C. (2000). Karyotype, morphology and taxonomic status of calomys expulsus (rodentia: Sigmodontinae). *Mammalia*, 64, 339–351.
- Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. Cambridge: Cambridge University Press.
- Bookstein, F. L., Chernoff, B., Elder, R., & Strauss, R. (1985). *Morphometrics in evolutionary biology*. Philadelphia: The Academy of Natural Sciences of Philadelphia.
- Cheverud, J. M. (1984). Quantitative genetics and developmental constraints on evolution by selection. *Journal of Theoretical Biology*, 110, 155–172.
- Cheverud, J. M. (1988). A comparison of genetic and phenotypic correlations. *Evolution*, 42(5), 958–968.
- Cheverud, J. M. (1995). Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *American Naturalist*, 145(1), 63–89.
- Cheverud, J. M. (1996). Developmental integration and the evolution of pleiotropy. *American Zoology*, 36, 44–50.
- Cheverud, J. M. (2006). Modular pleiotropic effects of quantitative trait loci on morphological traits. In G. Schlosser & G. P. Wagner (Eds.), *Modularity in development and evolution* (1st ed., pp. 132–153). Chicago: The University of Chicago Press.
- Cheverud, J. M., & Marroig, G. (2007). Comparing covariance matrices: Random skewers method compared to the common principal components model. *Genetics and Molecular Biology*, 30, 461–469. doi:10.1590/S1415-47572007000300027.
- Cheverud, J. M., Routman, E. J., & Irschick, D. J. (1997). Pleiotropic effects of individual gene loci on mandibular morphology. *Evolution*, 51(6), 2006–2016.
- Cheverud, J. M., Wagner, G. P., & Dow, M. M. (1989). Methods for the comparative analysis of variation patterns. *Evolution*, 38(3), 201–213.
- Dochtermann, N. A. (2011). Testing cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution*, 65(6), 1814–1820. doi:10.1111/j.1558-5646.2011.01264.x.
- Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. New York: Wiley.
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow, Essex: Addison Wesley Longman.
- Franz-Odenaal, T. A. (2011). Epigenetics in bone and cartilage development. In B. Hallgrímsson & B. K. Hall (Eds.), *Epigenetics: Linking genotype and phenotype in development and evolution* (1st ed., pp. 195–220). Oakland, CA: University of California Press.
- Hallgrímsson, B., Jerniczy, H., Young, N. M., Rolian, C., Parsons, T. E., Boughner, J. C., et al. (2009). Deciphering the palimpsest: Studying the relationship between morphological integration and phenotypic covariation. *Evolutionary Biology*, 36(4), 355–376. doi:10.1007/s11692-009-9076-5.
- Hallgrímsson, B., & Lieberman, D. E. (2008). Mouse models and the evolutionary developmental biology of the skull. *Integrative and Comparative Biology*, 48(3), 373–384. doi:10.1093/icb/icn076.
- Hansen, T. F., & Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, 21(5), 1201–1219. doi:10.1111/j.1420-9101.2008.01573.x.
- Herring, S. W. (2011). Muscle-bone interactions and the development of skeletal phenotype. In B. Hallgrímsson & B. K. Hall (Eds.), *Epigenetics: Linking genotype and phenotype in development and evolution* (1st ed., pp. 221–237). Oakland, CA: University of California Press.
- Hershkovitz, P. (1962). Evolution of neotropical cricetine rodents (muridae) with special reference to the phyllotine group. *Fieldiana: Zoology*, 46, 1–524.
- Hill, W. G., & Thompson, R. (1978). Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics*, 34(3), 429–439.
- Klingenberg, C. P. (2008). Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics*, 39(1), 115–132. doi:10.1146/annurev.ecolsys.37.091305.110054.
- Klingenberg, C. P., & Leamy, L. J. (2001). Quantitative genetics of geometric shape in the mouse mandible. *Evolution*, 55, 2342–2352.
- Klingenberg, C. P., Leamy, L. J., & Cheverud, J. M. (2004). Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics*, 166, 1909–1921.
- Krupinski, P., Chickarmane, V., & Peterson, C. (2011). Simulating the mammalian blastocyst-molecular and mechanical interactions pattern the embryo. *PLoS Computational Biology*, 7(5), e1001128. doi:10.1371/journal.pcbi.1001128.
- Lande, R. (1979). Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. *Evolution*, 33(1), 402–416.

- Lande, R. (1980). The genetic covariance between characters maintained by pleiotropic mutations. *Genetics*, 94, 203–215.
- Leamy, L. J., Routman, E. J., & Cheverud, J. M. (1999). Quantitative trait loci for early and late developing skull characters in mice: A test of the genetic independence model of morphological integration. *The American Naturalist*, 153, 201–214. doi:10.1086/303165.
- Lessels, C. M., & Boag, P. T. (1987). Unrepeatable repeatabilities: A common mistake. *The Auk*, 2, 116–121.
- Linde, Kvd, & Houle, D. (2009). Inferring the nature of allometry from geometric data. *Evolutionary Biology*, 36(3), 311–322. doi:10.1007/s11692-009-9061-z.
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland: Sinauer Associates.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Marroig, G., & Cheverud, J. M. (2001). A comparison of phenotypic 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., & Cheverud, J. M. (2005). Size as a line of least evolutionary resistance: Diet and adaptive morphological radiation in new world monkeys. *Evolution*, 59(5), 1128–1142.
- Marroig, G., de Vivo, M., & Cheverud, J. M. (2004). Cranial evolution in sakis (Pithecia, Platyrrhini) II: Evolutionary processes and morphological integration. *Journal of Evolutionary Biology*, 17(1), 144–155. doi:10.1046/j.1420-9101.2003.00653.x.
- Marroig, G., Melo, D. A. R., & Garcia, G. (2012). Modularity, noise and natural selection. *Evolution*, 66(5), 1506–1524. doi:10.1111/j.1558-5646.2011.01555.x.
- Marroig, G., Shirai, L. T., Porto, A., de Oliveira, F., & de Conto, V. (2009). The evolution of modularity in the mammalian skull II: Evolutionary consequences. *Evolutionary Biology*, 36(1), 136–148. doi:10.1007/s11692-009-9051-1.
- Marroig, G., Melo, D., Porto, A., Sebastião, H., & Garcia, G. (2011). Selection response decomposition (SRD): A new tool for dissecting differences and similarities between matrices. *Evolutionary Biology*, 38(2), 225–241. doi:10.1007/s11692-010-9107-2.
- Martínez-Abadías, N., Esparza, M., Sjøvold, T., González-José, R., Hernández, M., & Klingenberg, C.P. (2011). Pervasive genetic integration directs the evolution of human skull shape. *Evolution*, 66(4), 1010–1023. doi:10.5061/dryad.12g3c7ht.
- Márquez, E. J., Cabeen, R., Woods, R. P., & Houle, D. (2012). The measurement of local variation in shape. *Evolutionary Biology*, 39(3), 419–439. doi:10.1007/s11692-012-9159-6.
- Meyer, K. (2007). WOMBAT: A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science B*, 8(11), 815–821. doi:10.1631/jzus.2007.B0815.
- Meyer, K., & Kirkpatrick, M. (2008). Perils of parsimony: Properties of reduced-rank estimates of genetic covariance matrices. *Genetics*, 180(2), 1153–1166. doi:10.1534/genetics.108.090159.
- Mezey, J. G., Cheverud, J. M., & Wagner, G. P. (2000). Is the genotype/phenotype map modular? A statistical approach using mouse quantitative trait loci data. *Genetics*, 156, 305–311.
- Mitteroecker, P. (2009). The developmental basis of variational modularity: Insights from quantitative genetics, morphometrics, and developmental biology. *Evolutionary Biology*, 36(4), 377–385. doi:10.1007/s11692-009-9075-6.
- Mitteroecker, P., & Bookstein, F. L. (2007). The conceptual and statistical relationship between modularity and morphological integration. *Systematic Biology*, 56(5), 818–836. doi:10.1080/10635150701648029.
- Olson, E., & Miller, R. (1958). *Morphological integration*. Chicago: University of Chicago Press.
- Polly, P. D. (2008). Developmental dynamics and g-matrices: Can morphometric spaces be used to model phenotypic evolution? *Evolutionary Biology*, 35(2), 83–96. doi:10.1007/s11692-008-9020-0.
- Porto, A., Oliveira, F. B., Shirai, L. T., de Conto, V., & Marroig, G. (2009). The evolution of modularity in the mammalian skull I: Morphological integration patterns and magnitudes. *Evolutionary Biology*, 36(1), 118–135.
- Porto, A., Shirai, L. T., de Oliveira, F. B., & Marroig, G. (2013). Size variation, growth strategies, and the evolution of modularity in the mammalian skull. *Evolution*, 67(11), 3305–3322. doi:10.1111/evo.12177.
- R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramaesh, T., & Bard, J. B. L. (2003). The growth and morphogenesis of the early mouse mandible: A quantitative analysis. *Journal of Anatomy*, 203, 213–222. doi:10.1046/j.1469-7580.2003.00210.x.
- Reusch, T., & Blanckenhorn, W. U. (1998). Quantitative genetics of the dung fly sepsis cynipsea: Cheverud's conjecture revisited. *Heredity*, 81, 111–119.
- Roff, D. A. (1995). The estimation of genetic correlations from phenotypic correlations: A test of cheverud's conjecture. *Heredity*, 74, 481–490.
- Roff, D. A. (1997). *Evolutionary quantitative genetics*. New York: Chapman & Hall.
- Roff, D. A., & Fairbairn, D. J. (2011). Path analysis of the genetic integration of traits in the sand cricket: a novel use of BLUPs. *Journal of Evolutionary Biology*, 1–13. doi:10.1111/j.1420-9101.2011.02315.x.
- Rohlf, F. J. (2006). *tpsDig2, version 2.6. Department of Ecology and Evolution*. Stony Brook, New York: SUNY.
- Roseman, C. C., Kenny-Hunt, J. P., & Cheverud, J. M. (2009). Phenotypic integration without modularity: Testing hypotheses about the distribution of pleiotropic quantitative trait loci in a continuous space. *Evolutionary Biology*, 36(3), 282–291. doi:10.1007/s11692-009-9067-6.
- Runcie, D. E., & Mukherjee, S. (2013). Dissecting high-dimensional phenotypes with bayesian sparse factor analysis of genetic covariance matrices. *Genetics*, 194(3), 753–767. doi:10.1534/genetics.113.151217.
- Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution*, 50(5), 1766–1774.
- Shaw, R. G. (1987). Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution*, 41, 812–826.
- Shirai, L. T., & Marroig, G. (2010). Skull modularity in neotropical marsupials and monkeys: Size variation and evolutionary constraint and flexibility. *Journal of experimental zoology Part B, Molecular and developmental evolution*, 314B(June), 663–683. doi:10.1002/jez.b.21367.
- Stephan, S. J., Adkins, R., & Anderson, J. (2004). Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Systematic Biology*, 53, 533–553.
- Stephan, S. J., Phillips, P. C., & Houle, D. (2002). Comparative quantitative genetics: Evolution of the g matrix. *Trends in Ecology and Evolution*, 17, 320–327.
- Theobald, D. L., & Wuttke, D. S. (2006). Empirical bayes hierarchical models for regularizing maximum likelihood estimation in the matrix gaussian procrustes problem. *Proceedings of the National Academy of Sciences*, 103(49), 18,521–18,527. doi:10.1073/pnas.0508445103.
- Tiedemann, H. B., Schneltzer, E., Zeiser, S., Hoesel, B., Beckers, J., Przemeck, G. K. H., et al. (2012). From dynamic expression

- patterns to boundary formation in the presomitic mesoderm. *PLoS Computational Biology*, 8(6), e1002586. doi:[10.1371/journal.pcbi.1002586](https://doi.org/10.1371/journal.pcbi.1002586).
- Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London*, 237(641), 37–72.
- 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, 77–95.
- Wagner, G. P. (1996). Homologues, natural kinds and the evolution of modularity. *The American Zoologist*, 36, 36–43.
- Wagner, G. P., & Altenberg, L. (1996). Perspective: Complex adaptations and the evolution of evolvability. *Evolution*, 50, 967–976.
- Watson, R. A., Wagner, G. P., Pavlicev, M., Weinreich, D. M., & Mills, R. (2013). The evolution of phenotypic correlations and 'developmental memory'. *Evolution*, 67(4), 1124–1138. doi:[10.1111/evo.12337](https://doi.org/10.1111/evo.12337).
- Willmore, K. E., Roseman, C. C., Rogers, J., Cheverud, J. M., & Richtsmeier, J. T. (2009). Comparison of mandibular phenotypic and genetic integration between baboon and mouse. *Evolutionary Biology*, 36(1), 19–36. doi:[10.1007/s11692-009-9056-9](https://doi.org/10.1007/s11692-009-9056-9).
- Wolf, J. B., Frankino, W. A., Agrawal, A. F., Iii, E. D. B., & Moore, A. J. (2001). Developmental interactions and the constituents of quantitative variation. *Evolution*, 55(2), 232–245. doi:[10.1111/j.0014-3820.2001.tb01289.x](https://doi.org/10.1111/j.0014-3820.2001.tb01289.x).
- Zelditch, M. L., & Carmichael, A. C. (1989). Ontogenetic variation in patterns of developmental and functional integration in skulls of *sigmodon fulviventer*. *Evolution*, 43, 814–824.
- Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2004). *Geometric morphometrics for biologists: A primer* (1st ed.). San Diego: Elsevier.

SOFTWARE TOOL ARTICLE

EvolQG - An R package for evolutionary quantitative genetics

[version 1; referees: 1 approved, 1 approved with reservations]

Diogo Melo¹, Guilherme Garcia¹, Alex Hubbe², Ana Paula Assis¹, Gabriel Marroig¹

¹Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

²Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, Salvador, Brazil

v1 First published: 30 Sep 2015, 4:925 (doi: [10.12688/f1000research.7082.1](https://doi.org/10.12688/f1000research.7082.1))
 Latest published: 30 Sep 2015, 4:925 (doi: [10.12688/f1000research.7082.1](https://doi.org/10.12688/f1000research.7082.1))

Abstract

We present an open source package for performing evolutionary quantitative genetics analyses in the R environment for statistical computing. Evolutionary theory shows that evolution depends critically on the available variation in a given population. When dealing with many quantitative traits this variation is expressed in the form of a covariance matrix, particularly the additive genetic covariance matrix or sometimes the phenotypic matrix, when the genetic matrix is unavailable. Given this mathematical representation of available variation, the EvolQG package provides functions for calculation of relevant evolutionary statistics, estimation of sampling error, corrections for this error, matrix comparison via correlations and distances, and functions for testing evolutionary hypotheses on taxa diversification.

This article is included in the **RPackage** channel.

Open Peer Review

Referee Status:

Invited Referees		
	1	2
version 1		
published 30 Sep 2015	report	report
1 Mark Grabowsky , University of Oslo Norway		
2 David Houle , Florida State University USA		
Discuss this article		
Comments (0)		

Corresponding author: Diogo Melo (diogro@usp.br)

How to cite this article: Melo D, Garcia G, Hubbe A *et al.* EvolQG - An R package for evolutionary quantitative genetics [version 1; referees: 1 approved, 1 approved with reservations] *F1000Research* 2015, 4:925 (doi: [10.12688/f1000research.7082.1](https://doi.org/10.12688/f1000research.7082.1))

Copyright: © 2015 Melo D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was supported by funding from FAPESP. D.M. was funded by grants 2012/20180-0 and 2014/01694-9; G.G. was funded by grant 2011/52469-4; A.H. was funded by grants 2008/56886-6 and 2012/24937-9; A.P.A. was funded by grant 2010/52369-0; G.M. was funded by grant 2011/14295-7.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 30 Sep 2015, 4:925 (doi: [10.12688/f1000research.7082.1](https://doi.org/10.12688/f1000research.7082.1))

Introduction

Quantitative genetics deals with the evolution and inheritance of continuous traits, like body size, bone lengths, gene expressions or any other inheritable characteristic that can be measured on a continuous scale, or which can be transformed to a continuous scale. This framework has been used in selective breeding and in describing the different sources of variation in natural populations, as well as understanding the interaction of evolutionary processes with this variation²⁵. Quantitative genetics has been successful in describing short term evolution, and is also useful in understanding diversification at a macroevolutionary level. The core development of modern evolutionary quantitative genetics started with the generalization of the univariate breeders equation to the multivariate response to selection equation, derived by Lande and also referred to as the Lande equation^{22,46}.

The Lande equation relates the evolutionary change in trait means of a given population (Δz) to the interaction between the additive genetic variation (**G**-matrix) of this population and the directional selection (β) acting on this population. The additive genetic variation of a population is represented by a symmetric square matrix called the **G**-matrix, which contains the additive genetic variance of each trait on the diagonal and the additive genetic covariance between traits on the off-diagonal elements. From the Lande equation, $\Delta z = G\beta$, we can see that different populations may present markedly different responses (Δz) to the same directional selection (β) simply because these populations have distinct **G**-matrices. Other evolutionary forces affecting populations are also influenced by available variation, e.g., based on the **G**-matrix it is possible to test if morphological differentiation of extant taxa is compatible with genetic drift or stabilizing selection (e.g., 2,28). Thus, describing and understanding changes in standing variation among populations^{5,24,27} as well as understanding constraints imposed by populations standing variation (e.g., 15,26,42,49) are major elements in evolutionary quantitative genetics.

In this article we describe the **EvoIQG** package, developed to deal with the evolutionary quantitative genetics questions addressed above in the R environment for statistical computing⁴⁴. Our goal was to provide a suite of tools in a single consistent source, and to standardize and facilitate the adoption of these tools.

Measurement error estimation

Before estimating a **G**-matrix, it is important to evaluate the influence of measurement error in data collection, since measurement error can drastically bias further analyses⁹. Measurement error can be estimated by measuring each individual at least twice and estimating the amount of variation associated with each individual, which is the measurement error, in relation to total variation (i.e., the sum of within and between individuals variation) using an analysis of variance. The proportion of variance not associated with the individuals is called the repeatability²³. A repeatability of 1 means that no variation is associated with measurement error. The function **CalcRepeatability()** performs the calculation described in 23 for a set of multivariate traits measured at least twice for each individual.

Matrix estimation

In the rest of this article we assume that the covariance matrix of interest has already been estimated by some procedure. This can be a simple covariance of all the observed traits, or an estimated parameter from a more complicated linear model. The simplest case of a linear model approach would be using a multivariate analysis of covariance (MANCOVA) to control for differences in trait means that are not of immediate interest in the analyses (e.g., sexual dimorphism, geographic variation, etc.). The residual pooled within-group covariance matrix can be used in subsequent analysis²⁷. The **EvoIQG** package function **CalculateMatrix()** uses R's **lm()** model object to calculate variance-covariance matrices adjusting for the proper degrees of freedom in a simple fixed-effects MANCOVA. Of course more complicated methods may be used to obtain **G**-matrices, such as an animal model or a mixed model^{25,48}.

Accurate **G**-matrix estimation can be hard to achieve, requiring large sample sizes, many families and known genealogies⁵⁰. One alternative that is usually more feasible is to use the phenotypic covariance matrix (the **P**-matrix) as a proxy of the population's **G**-matrix^{6,47}. Conditions on where this approximation is reasonable depend on the structure of developmental and environmental effects, and testing for similarity is an empirical question that should be undertaken before using the **P**-matrix as a proxy for the **G**-matrix, ideally by direct comparison (e.g., 11). As a general rule, high similarity between populations' **P**-matrices is a good indicator of high similarity between **P** and **G**, and of a stable shared **G**-matrix pattern, since the similarity between populations must come from either a common genetic structure, or the unlikely scenario of a different genetic structure buffered by an exactly compensating environmental structure in each population that leads to high similarity between phenotypic covariation.

Some of the methods described below are not applicable to covariance matrices, only to correlation matrices. Correlations are standardized measures of association that are bounded between [-1, 1], and, unlike covariances, can be directly compared for pairs of traits with different scales. In most instances, correlation matrices can be obtained directly from covariance matrices by using the R function **cov2cor()**.

Matrix error and repeatabilities

A **G**-matrix will always be estimated with error^{16,30,33}, and it is important to take this error into account in further analyses. In some circumstances we want to compare two or more **G**-matrices, calculating the matrices correlations (see section **Matrix Comparison**). However, due to error in estimating these matrices, their correlations will never be one, even if the actual population parameter values are identical⁶. Thus, matrix repeatabilities are used to correct matrix correlations by taking sampling error into account. The basic premise of all the methods is that taking repeated samples from the same population and comparing the resulting matrices would still give correlations that are lower than 1. We estimate the maximum correlation between matrices taken from the same population and correct the observed correlation by this maximum value. The corrected correlation between two observed matrices will be

given by the original correlation divided by the geometric mean of their repeatabilities. If the repeatability of both matrices is one, the observed correlation does not change under the correction, and lower repeatabilities yield larger corrections. A number of methods for repeatability estimation are provided, and their results can be passed on to the functions that calculate matrix correlations (section **Matrix Comparison**):

AlphaRep(): Cheverud⁶ describes an analytical expression for the repeatability of a correlation matrix. This expression is asymptotically convergent, so it should be used only when sample sizes are large, at least larger than the number of traits.

BootstrapRep(): We may estimate the repeatability of the covariance (or correlation) structure of a given data set using a bootstrap procedure, sampling individuals with replacement from the data set and calculating a covariance (or correlation) matrix from each sample. The mean value of the correlation between the random sample matrix and the original estimated matrix is an estimate of the repeatability. This method has the advantage of not assuming any distribution on the data, but does provide inflated estimates of repeatabilities for small data sets. Even so, upwardly biased matrix repeatabilities are not so problematic, since they lead to conservative corrections of matrix correlations. However, users should be aware of this bias and should not interpret a high repeatability obtained from a small data set as indicating that the parameter is well estimated.

MonteCarloRep(): We can use the observed covariance (or correlation) matrix as the Σ parameter in a multivariate normal distribution, and produce samples from this distribution, using a fixed sample size. The covariance (or correlation) matrix for each sample is compared to the observed matrix, and the mean of these comparisons is an estimate of the repeatability²⁷. This method has the advantage of being easy to apply to matrices coming from linear models with many controlled effects, and not requiring the original data.

Sometimes the question we are trying to answer does not involve matrix comparisons, so other methods of assessing and correcting for error are needed.

Rarefaction(): Rarefaction consists of taking progressively smaller samples with replacement from the original data set, calculating some statistic on each data set and comparing this with the full data set. This gives a general idea of how the inferences would change if we had smaller sample sizes, and how robust our data set is with respect to sampling. The default operation is to calculate the covariance or correlation matrices and compare them using any of the matrix comparison methods (see section **Matrix Comparison**).

ExtendMatrix(): Marroig *et al.*³⁰ showed that sampling error on covariance matrix estimation can have a dramatic effect on the reconstruction of net selection gradients using the multivariate response to selection equation²². One way to improve estimates is the simple procedure of “extending” the eigenvalues of the covariance matrix, where all the eigenvalues lower than a certain threshold are substituted by the smallest eigenvalue above the threshold.

This causes minimal changes in the distribution of phenotypes, but improves dramatically the estimates of net selection gradients. See 30 for a thorough examination of the performance and consequences of the extension method on simulated and real data sets.

Evolutionary statistics

Hansen and Houle¹⁵ provide a suite of statistics that have fairly good biological interpretations for a given **G**- or **P**-matrix. Marroig *et al.*³¹ is a comprehensive example of how these statistics may be used for interpreting morphological data.

The function **MeanMatrixStatistics()** calculates most of these statistics and their distributions, as shown below.

In the following, $E[\cdot]_\beta$ represents the expected value over many random β vectors with unit norm, $\langle \cdot, \cdot \rangle$ represents the dot product between two vectors, $\cos(\cdot, \cdot)$ is the cosine between two vectors, **G** is an arbitrary covariance matrix, **G**⁻¹ is the inverse **G**, $\text{tr}(\mathbf{G})$ is the trace of **G**, and $\|\cdot\|$ the Euclidean norm of a vector. **MeanMatrixStatistics()** calculates:

- Mean squared correlation (r^2): Given a correlation matrix, the elements below the diagonal are squared and averaged resulting in a measure of integration, that is, overall association between traits (also see the section **Modularity and Integration** and 40).
- Coefficient of variation of eigenvalues (ICV): A measure of integration that is suitable for covariance matrices, as it takes the amount of variation into account. Notice that at least for mammals, mean squared correlations and ICV generally have very high correlation, but can lead to different conclusions if the traits included in the analysis have very different variances (due to scale, for example). If σ_λ is the standard deviation of the eigenvalues of a covariance matrix, and $\bar{\lambda}$ is the mean of the eigenvalues, the ICV is:

$$ICV = \frac{\sigma_\lambda}{\bar{\lambda}}$$

- Percent of variation in first principal component: If λ_1^G is the leading eigenvalue of **G**, we calculate this percentage as:

$$PC1\% = \frac{\lambda_1^G}{\text{tr}(\mathbf{G})}$$

- Evolvability (Figure 1): The mean projection of the response to random selection gradients with unit norm onto the selection gradient. This is a measure of a population's available variation in the direction of a particular selection gradient, averaged across all directions¹⁵.

$$\bar{e} = E[\langle \mathbf{G}\beta, \beta \rangle_\beta]$$

- Flexibility (Figure 1): The mean cosine of the angle between random selection gradients and the corresponding

responses. Flexibility measures on average how the response to selection aligns with the selection gradient³¹.

$$\bar{f} = E[\cos(\mathbf{G}\beta, \beta)]_{\beta}$$

- **Responsability (Figure 1):** Mean norm of the response to random selection gradients with unit norm. It also estimates how fast the population mean will change under directional selection^{15,31}.

$$\bar{r} = E[\|\mathbf{G}\beta\|]_{\beta}$$

- **Conditional Evolvability:** Measures the mean response to selection in the direction of a given β when other directions are under stabilizing selection¹⁵.

$$\bar{c} = E\left[\left(\langle \mathbf{G}^{-1}\beta, \beta \rangle\right)^{-1}\right]_{\beta}$$

- **Autonomy:** Measures the proportion of variance in the direction of a given β that is independent from variation in other directions. Therefore, mean Autonomy can also be calculated as the mean ratio between Conditional Evolvability ($\bar{c}$) and Evolvability ($\bar{e}$)¹⁵.

$$\bar{a} = E\left[\left(\langle \mathbf{G}^{-1}\beta, \beta \rangle\right)^{-1} \left(\langle \mathbf{G}\beta, \beta \rangle\right)^{-1}\right]_{\beta}$$

- **Constraints:** The mean correlation between the response vector to random selection gradients and the matrix's first principal component³¹. If Λ_1^G is the first principal component of $\mathbf{G}$, constraints are measured as:

$$E[\cos(\mathbf{G}\beta, \Lambda_1^G)]_{\beta}$$

Figure 1. Graphical representation of evolvability ($\bar{e}$), responsibility ($\bar{r}$) and flexibility ($\bar{f}$) for a single selection gradient (β) and the corresponding response (Δz) in the two dimensions defined by traits x and y .

Matrix comparison

A $\mathbf{G}$ -matrix describes how the variation in particular populations is structured, but frequently the relevant question is how similar or dissimilar two populations are with respect to this standing variation. Since no two populations are identical, different patterns of variation are the norm. Depending on the evolutionary question at hand, different methods of comparing variation may be required. One possible application of matrix comparisons is when we wish to apply the Lande equation to micro and macroevolution, since this requires some additional assumptions, such as a relative stability of the $\mathbf{G}$ -matrix over generations. Comparing extant covariance matrices is a test of this required stability (e.g. 27). For a thoughtful discussion on the biological relevance of statistical significance in matrix comparisons, see the discussion in 13.

Matrix correlations

One approach to estimate the similarity or dissimilarity between matrices is to calculate the correlation between these matrices. **EvolQG** provides several functions for pairwise matrix correlation.

RandomSkewers(): The Random Skewers (RS) method makes use of the Lande equation²², $\Delta z = \mathbf{G}\beta$, where Δz represents the vector of response to selection, $\mathbf{G}$ the $\mathbf{G}$ -matrix and β the directional selection vector, or selection gradient. In the RS method, the two matrices being compared are multiplied by a large number of normalized random selection vectors, and the resulting response vectors to the same selection vector are compared via a vector correlation (the cosine between the two vectors). The mean value of the correlation between the responses to the same selective pressure is used as a simple statistic of how often two populations respond similarly (in the same direction) to the same selective pressure:

$$RS(A, B) = E[\cos(A\beta, B\beta)]_{\beta} \quad (1)$$

Where $E[\cdot]_{\beta}$ is the expected value over random selection vectors β . Significance in the random skewers comparison can be determined using a null expectation of correlation between random vectors. If the observed correlation between two matrices is above the 95% percentile of the distribution of correlations between random vectors, we consider the correlation significant and infer that there is evidence the two populations behave similarly under directional selection. Other implementations of the RS method sometimes resort to other forms of calculating significance, such as generating random matrices and creating a random distribution of correlations between matrices. This is difficult to do because generating random matrices with the properties of biological covariance structures is hard, see the **RandomMatrix()** function for a quick discussion on this. The RS values range between -1 (the matrices have the opposite structure) and 1 (the matrices share the same structure), and zero means the matrices have distinct structures.

MantelCor(): Correlation matrices can be compared using a simple Pearson correlation between the corresponding elements. Significance of this comparison must take the structure into account, so it is calculated by a permutation scheme, in which a null

distribution is generated by permutation of rows and columns in one of the matrices and repeating the element-by-element correlation. The observed correlation is significant when it is larger than the 95% quantile of the permuted distribution. This method can not be used on covariance matrices because the variances might be very different, leading to large differences in the scale of the covariances. This scale difference can lead to a massive inflation in the correlation between matrices. The correlation between matrices range between -1 (the matrices have the opposite structure) and 1 (the matrices share the same structure), and zero means the matrices have distinct structures.

KrzCor(): The Krzanowski shared space, or Krzanowski correlation, measures the degree to which the first principal components (eigenvectors) span the same subspace^{3,20}, and is suitable for covariance or correlation matrices. If two $n \times n$ matrices are being compared, the first $k = \frac{n}{2} - 1$ principal components from one matrix are compared to the first k principal components of the other matrix using the square of the vector correlations, and the sum of the correlations is a measure of how congruent the spanned subspaces are. We can write the Krzanowski correlation in terms of the matrices' principal components (Λ_i^A being the i th principal component of matrix A):

$$\text{KrzCor}(A, B) = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k \cos^2(\Lambda_i^A, \Lambda_j^B) \quad (2)$$

The Krzanowski correlation values range between 0 (two subspaces are dissimilar) and 1 (two subspaces are identical).

PCAsimilarity(): The Krzanowski correlation compares only the subspace shared by roughly the first half of the principal components, but does not consider the amount of variation each population has in these directions of the morphological space⁵². In order to take the variation into account, we can add the eigenvalue associated with each principal component into the calculation, effectively pondering each correlation by the variance in the associated directions. If Λ_i^A is the i th eigenvalue of matrix A , we have:

$$\text{PCAsimilarity}(A, B) = \frac{\sum_{i=1}^n \sum_{j=1}^n \lambda_i^A \lambda_j^B \cos^2(\Lambda_i^A, \Lambda_j^B)}{\sum_{i=1}^n \lambda_i^A \lambda_j^B} \quad (3)$$

Note the sum spans all the principal components, not just the first k as in the Krzanowski correlation method. This method gives correlations that are very similar to the RS method, but is much faster. The PCA similarity values range between 0 (the shared subspaces have no common variation) and 1 (the shared subspaces have identical variation).

SRD(): The RS method can be extended to give information into which traits contribute to differences in terms of the pattern of correlated selection due to covariation between traits in two populations²⁹. The Selection Response Decomposition does this by treating the terms of correlated response in the Lande equation as separate entities. Writing out the terms in the multivariate response to selection equation:

$$\begin{aligned} A\beta &= \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \\ &= \begin{pmatrix} A_{11}\beta_1 + A_{12}\beta_2 + \cdots + A_{1n}\beta_n \\ A_{21}\beta_1 + A_{22}\beta_2 + \cdots + A_{2n}\beta_n \\ \vdots \\ A_{n1}\beta_1 + A_{n2}\beta_2 + \cdots + A_{nn}\beta_n \end{pmatrix} = \Delta z \end{aligned} \quad (4)$$

Separating the terms in the sums of the right hand side:

$$\begin{pmatrix} A_{11}\beta_1 & A_{12}\beta_2 & \cdots & A_{1n}\beta_n \\ A_{21}\beta_1 & A_{22}\beta_2 & \cdots & A_{2n}\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}\beta_1 & A_{n2}\beta_2 & \cdots & A_{nn}\beta_n \end{pmatrix} \quad (5)$$

Each of these row vectors $r_i^{AB} = (A_{ij}\beta_j)_{j=1..n}$ are the components of the response to the selection gradient β on trait i . The term $A_{ii}\beta_i$ represents the response to direct selection on trait i , and the terms $(A_{ij}\beta_j)_{i \neq j}$ represent the response to indirect selection due to correlation with the other traits. Given two matrices, A and B , we can measure how similar they are in their pattern of correlated selection on each trait by calculating the correlation between the vectors r_i for each trait for random selection vectors of unit norm. The mean SRD score for trait i is then:

$$\mu\text{SRD}(A, B)_i = E[\text{cor}(r_i^{AB}, r_i^{BB})]_\beta \quad (6)$$

And the standard deviation of the correlations gives the variation in SRD scores:

$$\sigma\text{SRD}(A, B)_i = \sqrt{E[(\text{cor}(r_i^{AB}, r_i^{BB}) - \mu\text{SRD}_i)^2]_\beta} \quad (7)$$

When the same trait in different matrices share a correlated response pattern, μSRD is high and σSRD is low; if the correlated response pattern is different, μSRD is low and σSRD is high. See 29 for details and examples.

Matrix distances

Another approach to estimate the similarity or dissimilarity between matrices is to calculate the distance between a pair of matrices. Matrices distances are different from correlations in that correlations are limited to [-1,+1], while distances must only be positive. Also, smaller values of distances mean more similarity. Two distances are in use in the current evolutionary literature, and are implemented in the function **MatrixDistance()**.

- **Overlap distance:** Ovaskainen *et al.*³⁹ proposed a distance based on probability distributions, where two covariance matrices would have a distance proportional to how distinguishable they are. This distance is natural if we think of covariance matrices as describing the probability

distribution of phenotypes or additive values in the population. The higher the probability of a random draw coming from the distribution defined by one of the matrices being misclassified as coming from the distribution defined by the other, the lower the distance. For two probability distributions f and g , the probability of misclassifying a draw from f as coming from g is:

$$q(f, g) = \int_{\mathbb{R}^n} \frac{g(x)}{f(x) + g(x)} f(x) dx \quad (8)$$

where n is the dimensionality of the space in which the distributions are defined. If the distributions are indistinguishable, $q(f, g) = 1/2$, if they are completely distinguishable $q(f, g) = 0$. We can then define the distance as:

$$d(f, g) = \sqrt{1 - 2q(f, g)} \quad (9)$$

Since $q(f, g)$ is symmetrical, $d(f, g)$ is also symmetrical, and the square root guarantees that $d(f, g)$ satisfies the triangle inequality³⁹. Calculation is straight forward and can be done with a simple sampling Monte Carlo scheme, see 39 for details.

- **Riemann distance:** Mitteroecker and Bookstein³⁴ use a Riemannian metric in the space of positive definite matrices (either covariance or correlation matrices), based on exponential mapping³⁵ to quantify transition in the ontogenetic trajectory of phenotypic covariance matrices. This metric is based on the eigenvalues of the product of one matrix to the inverse of the other. If λ_i are the eigenvalues of $A^{-1}B$ (or AB^{-1}), we have:

$$\|A, B\|_{cov} = \|B, A\|_{cov} = \sqrt{\sum_{i=1}^p [\log(\lambda_i)]^2} \quad (10)$$

This distance has the advantage of being invariable under changes in the base used to represent the matrices. See 34 for a discussion on the biological relevance of this distance.

Phylogenetic comparisons

AncestralStates(): Ancestral state reconstruction of continuous traits using maximum likelihood can be performed in R (for example, using *ape* or *phytools*), but most packages don't provide a simple interface for multivariate data. **AncestralStates()** is a wrapper for **fastAnc()** in *phytools*⁴⁵ that works on multivariate data, reconstructing each character independently.

PhyloW(): Given a set of covariance matrices for the terminal taxa in a phylogeny, we can estimate the covariance matrix for internal nodes by taking means over sister taxa, weighted by sample size. The mean matrix at the root node is the within-group covariance matrix in a MANCOVA with the terminal clades as the fixed effects. **PhyloW()** does this by taking a tree and a set of measurements (covariance matrices) and returns means for internal nodes. The implementation is generic, so this function can also be used to

calculate weighted means for any numerical measurement with an addition operation implemented in R.

While using the within-group covariance matrix is a reasonable alternative as the estimator of an ancestral covariance matrix, this ignores branch lengths, and so should be used carefully when matrix differences are correlated to phylogenetic distance. An alternative when matrix evolution depends of branch lengths is to reconstruct every position of the covariance matrix independently via maximum likelihood, but this method can result in non positive-definite estimates.

PhyloCompare(): Sometimes it is not practical to pairwise compare every single population in a study, since for a large number of populations these results can be difficult to interpret. In these cases, comparing populations in a phylogenetically structured way can be helpful in detecting major transitions or differences between clades. **PhyloCompare()** takes estimates for all the nodes in a tree and compares sister groups by any comparison method, providing comparison values for every inner node.

Hypothesis testing

Modularity and integration

Modularity is a general concept in biology, and refers to a pattern of organization that is widespread in many biological systems. In modular systems, we find that some components of a given structure are more related or interact more between themselves than with other components. These highly related groups are termed modules. The nature of this interaction will depend on the components being considered, but may be of any kind, like physical contact between proteins, joint participation of enzymes in given biochemical pathways, or high correlation between quantitative traits in a population. This last kind of modularity is called variational modularity, and is characterized by high correlations between traits belonging to the same module and low correlation between traits in different modules⁵¹. In the context of morphological traits, variational modularity is associated with the concept of integration³⁸, that is, the tendency of morphological systems to exhibit correlations due to common developmental factors and functional demands^{7,14}.

Both modularity and integration may have important evolutionary consequences, since sets of integrated traits will tend to respond to directional selection in an orchestrated fashion due to genetic correlations between them; if these sets are organized in a modular fashion, they will also respond to selection independently of one another³¹. At the same time, selection can alter existing patterns of integration and modularity, leading to traits becoming more or less correlated^{18,32}. The pattern of correlation between traits in a **G**-matrix then carries important information on the expected response to selection and on the history of evolutionary change of a given population.

TestModularity(): Variational modularity can be assessed by comparing a modularity hypothesis (derived from development and functional considerations) with the observed correlation matrix. If two traits are in the same variational module, we expect the correlation between them to be higher than between traits belonging to

different modules. We test this by creating a modularity hypothesis matrix and comparing it via Mantel correlation with the observed correlation matrix. The modularity hypothesis matrix consists of a binary matrix where each row and column corresponds to a trait. If the trait in row i is in the same module of the trait in column j , position (i, j) in the modularity hypothesis matrix is set to one, if these traits are not in the same module, position (i, j) is set to zero. Significant correlation between the hypothetical matrix representing a modularity hypothesis and the observed correlation matrix represents evidence of the existence of this variational module in the population. We also measure the ratio between correlations within a module (AVG+) and outside the module (AVG-). This ratio (AVG+/AVG-) is called the AVG Ratio, and measures the strength of the within-module association compared to the overall association for traits outside the module. The higher the AVG Ratio, the bigger the correlations within a module in relation to all other traits associations in the matrix (e.g., 41). **TestModularity()** also provides the Modularity Hypothesis Index, which is the difference between AVG+ and AVG- divided by the coefficient of variation of eigenvalues. Although the AVG Ratio is easier to interpret (how many times greater the within-module correlation is compared to the between-module correlation) than the Modularity Hypothesis Index, the AVG Ratio cannot be used when the observed correlation matrix presents correlations that differ in sign, and this is usually the case for residual matrices after size removal (for example with **RemoveSize()**, but see 19 for other alternatives). In these cases, the Modularity Hypothesis Index is useful and allows comparing results between raw and residual matrices⁴².

LModularity(): If no empirical or theoretical information is available for creating a modularity hypothesis, such as functional or developmental data, we can try to infer the modular partition of a given population by looking only at the correlation matrix and searching for the trait partition that minimizes some indicator of modularity. Borrowing from network theory, we can treat a correlation matrix as a fully connected weighted graph, and define a Newman-like modularity index³⁶. If A is a correlation matrix we define L modularity as:

$$L = \sum_{i \neq j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(g_i, g_j) \quad (11)$$

The terms g_i and g_j represent the partition of traits, that is, in what modules the traits i and j belong to. The function $\delta(\cdot, \cdot)$ is the Kronecker delta, where:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \quad (12)$$

This means only traits in the same module contribute to the value of L . The term k_i represent the total amount of correlation attributed to trait i , or the sum of the correlation with trait i :

$$k_i = \sum_{j \neq i} A_{ij} \quad (13)$$

And m is the sum of all k ($m = \sum_i k_i$). The term $\frac{k_i k_j}{2m}$ plays the role of a null expectation for the correlation between the traits i and j . This choice for the null expectation is natural when we impose that it must depend on the values of k_i and k_j and must be symmetrical³⁶. So, traits in the same module with correlations higher than the null expectation will contribute to increase the value of L , while traits in the same module with correlation less than the null expectation will contribute to decrease L . With this definition of L , we use a Markov Chain Monte Carlo annealing method to find the partition of traits (values of g_i) that maximizes L . This partition corresponds to the modularity hypothesis inferred from the correlation matrix, and the value of L is a measure of modularity comparable to the AVG Ratio. The igraph package⁹ provides a number of similar community detection algorithms that can also be used in correlation matrices.

RemoveSize(): If the first principal component of a covariance or correlation matrix corresponds to a very large portion of its variation, and all (or most) of the entries of the first principal component are of the same sign (a *size* principal component, see 26), it is useful to look at the structure of modularity after removing this dominant integrating factor. This is done using the method described in 4. Porto *et al.*⁴² show that modularity is frequently more easily detected in matrices where the first principal component variation was removed and provide biological interpretations for these results.

Drift

Selection is frequently invoked to explain morphological diversification, but the null hypothesis of drift being sufficient to explain current observed patterns must always be entertained. We can test the plausibility of drift for explaining multivariate diversification by using the regression method described in 1, or the correlation of principal component scores³⁸. Since both these tests use drift as a null hypothesis, failure to reject the null hypothesis is not evidence that selection was not involved in the observed pattern of diversification, only that the observed pattern is compatible with drift.

DriftTest(): Under drift, we expect that the current between group variance for many populations will be proportional to the ancestral population's covariance structure, which is approximated by the pooled within-group covariance matrix. Conditions for the validity of these assumptions are reviewed in 43. Under these conditions, if B is the between group covariance matrix, and W is the within group covariance matrix, t is the time in number of generations and N_e is the effective population size, we have:

$$B \propto (t / N_e) W \quad (14)$$

If we express all these matrices in terms of the eigenvectors of W , so that W is diagonal, we can write B as the variance of the scores of the means on these eigenvectors. The relationship between B and W can be expressed as a log regression, where B_i is the variance between groups in the projected means and λ_i^W are the eigenvalues of W :

$$\log(B_i) = \log(t / N_e) + \beta \log(\lambda_i^W) \quad (15)$$

where β is the regression coefficient. Under drift we expect β to be one. If β is significantly different from one, we have evidence that drift is not sufficient to explain currently observed diversification.

MultivDriftTest(): This drift test verifies the plausibility of drift in a multivariate context when only two populations are available, one ancestral (or reference) and one derived. Let z_0 represent a vector of means from m traits in an ancestral population. After t generations, the expected traits mean for n populations under drift would correspond to z_0 with variance given by $B = (t/N_e)W$, where B represents the expected between group covariance matrix, W is the genetic covariance matrix from the ancestral (or reference) population, and N_e is the effective population size^{17,21,22}. So, given the ancestral population mean and G -matrix, we can use this model to estimate the B -matrix expected under drift. We can then use this B -matrix as the Σ parameter in a multivariate normal distribution and sample n populations from this distribution. Using this sample of random populations, we can assess the amount of divergence expected by drift, estimated as the norm of the difference vectors between ancestral (or reference) and simulated population means. Then, we can compare the observed amount of divergence between the ancestral and derived populations, calculated as the norm of the difference vector between them, taking into account the standard error of traits means. An observed divergence higher than the expectations under drift indicates that genetic drift is not sufficient to explain currently observed divergence, suggesting a selective scenario.

PCScoreCorrelation(): This test of drift relies on the correlation between principal component scores of different populations. Under drift, we expect the mean scores of different populations in the principal components of the within-group covariance matrix to be uncorrelated²⁸. Significant correlations between the scores of the means on any two principal components is an indication of correlated directional selection¹⁰.

Random matrices

RandomMatrix(): Generating realistic random covariance matrices for null hypothesis testing is a challenging task, since random matrices must adequately sample the space of biologically plausible evolutionary parameters, like integration and flexibility. Most common covariance and correlation matrix sampling schemes fail at this, producing matrices with unrealistically low levels of integration, unless the level of integration is supplied *a priori* (as in 12).

The method described in 37 provides correlation matrices with a reasonable range of evolutionary characteristics. However, the adequacy of the generated matrices in hypothesis testing has not been well established, and we recommend these random matrices be used only for informal tests requiring an arbitrary covariance or correlation matrix.

Summary

We have described a suite of functions dedicated to analyzing multivariate data sets within an evolutionary quantitative genetics framework. These functions focus on the central role that covariance and correlation matrices play in this framework; therefore, we provide functions that perform both descriptive statistics and hypothesis testing related to such matrices within an evolutionary context.

We have intentionally neglected to include techniques like phylogenetic regression or more extensive linear model functionality. We also lack Bayesian implementations that would be possible for some functions (i.e. 3). Our reasons for this are twofold: the difficulty in transposing these methods efficiently to multiple traits, especially with respect to Bayesian implementations of existing functions, and the many different robust packages for performing some of these analyses, such as phytools, phylolm, pgl, nlme, MCMCglmm and others.

Some of the material implemented here is available in other sources or through custom implementations. We have attempted to create a single consistent source for these techniques. This is by no means an exhaustive effort, and we hope to expand it given demand from the community and further developments in the field. We hope to contribute to standardization and wide adoption of these tools, and, since we opted for an open source implementation under R, this also allows the involvement of the R community in using, debugging and complementing these tools, in an effort to contribute to an open scientific environment in which, for example, truly reproducible results are the norm rather than the exception.

Software availability

The most recent version of the **EvolQG** package can be installed from GitHub using the package **devtools**:

```
> library(devtools)
> install_github("lem-usp/evolqg")
```

A less up-to-date version is also available from CRAN:

```
> install.packages("evolqg")
```

1. Software available from: <https://cran.r-project.org/web/packages/evolqg/>

2. Latest source code: <https://github.com/lem-usp/EvolQG>
3. Archived source code as at time of publication: <http://dx.doi.org/10.5281/zenodo.3091053>
4. License: The MIT License (<https://opensource.org/licenses/MIT>)

Author contributions

D.M. compiled existing implementations, re-factored existing functions and contributed new code, documentation and unit tests. G.G. provided the initial set of functions and first implementations. A.H. tested and revised code and documentation. A.P.A. contributed code and documentation, G.M. developed methods and devised necessary elements for package. All authors wrote the paper.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by funding from FAPESP. D.M. was funded by grants 2012/20180-0 and 2014/01694-9; G.G. was funded by grants 2008/56886-6 and 2011/52469-4; A.H. was funded by grant 2012/24937-9; A.P.A. was funded by grant 2010/52369-0; G.M. was funded by grant 2011/14295-7.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to thank Barbara Costa, Daniela Rossoni, Edgar Zanella and Fabio Machado for contributing code and revising documentation and results.

References

1. Ackermann RR, Cheverud JM: **Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation.** *Am J Phys Anthropol.* 2002; **117**(3): 260–271.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Ackermann RR, Cheverud JM: **Detecting genetic drift versus selection in human evolution.** *Proc Natl Acad Sci U S A.* 2004; **101**(52): 17946–17951.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Aguirre JD, Hine E, McGuigan K, et al.: **Comparing G: multivariate analysis of genetic variation in multiple populations.** *Heredity (Edinb).* 2014; **112**(1): 21–29.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Bookstein FL, Chernoff B, Elder RL, et al.: **Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes.** 1985.
[Reference Source](#)
5. Cheverud JM: **Phenotypic, Genetic, and Environmental Morphological Integration in the Cranium.** *Evolution.* 1982; **36**(3): 499–516.
[Publisher Full Text](#)
6. Cheverud JM: **A Comparison of Genetic and Phenotypic Correlations.** *Evolution.* 1988; **42**(5): 958–968.
[Publisher Full Text](#)
7. Cheverud JM: **Developmental Integration and the Evolution of Pleiotropy.** *Integr Comp Biol.* 1996; **36**(1): 44–50.
[Publisher Full Text](#)
8. Csardi G, Nepusz T: **The igraph software package for complex network research.** *Inter Journal. Complex Systems:* 1695, 2006.
[Reference Source](#)
9. Falconer DS, Mackay TF: **Introduction to Quantitative Genetics.** Benjamin Cummings, London, 4th edition, 1996.
[Reference Source](#)
10. Felsenstein J: **Phylogenies And Quantitative Characters.** *Annu Rev Ecol Syst.* 1988; **19**(1): 445–471.
[Publisher Full Text](#)
11. Garcia G, Hingst-Zaher E, Cerqueira R, et al.: **Quantitative Genetics and Modularity in Cranial and Mandibular Morphology of *Calomys expulsus*.** *Evol Biol.* 2014; **41**(4): 619–636.
[Publisher Full Text](#)
12. Haber A: **A Comparative Analysis of Integration Indices.** *Evol Biol.* 2011; **38**(4): 476–488.
[Publisher Full Text](#)
13. Haber A: **The Evolution of Morphological Integration in the Ruminant Skull.** *Evol Biol.* 2015; **42**(1): 99–114.
[Publisher Full Text](#)
14. Hallgrímsson B, Jerniczky H, Young NM, et al.: **Deciphering the Palimpsest: Studying the Relationship Between Morphological Integration and Phenotypic Covariation.** *Evol Biol.* 2009; **36**(4): 355–376.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Hansen TF, Houle D: **Measuring and comparing evolvability and constraint in multivariate characters.** *J Evol Biol.* 2008; **21**(5): 1201–1219.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Hill WG, Thompson R: **Probabilities of Non-Positive Definite between-Group or Genetic Covariance Matrices.** *Biometrics.* 1978; **34**(3): 429–439.
[Publisher Full Text](#)
17. Hohenlohe PA, Arnold SJ: **MIPoD: a hypothesis-testing framework for microevolutionary inference from patterns of divergence.** *Am Nat.* 2008; **171**(3): 366–385.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Jones AG, Arnold SJ, Bürger R: **Evolution and stability of the G-matrix on a landscape with a moving optimum.** *Evolution.* 2004; **58**(8): 1639–1654.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Jungers WL, Falsetti AB, Wall CE: **Shape, relative size, and size-adjustments in morphometrics.** *Am J Phys Anthropol.* 1995; **38**(S2): 137–161.
[Publisher Full Text](#)
20. Krzanowski WJ: **Between-Groups Comparison of Principal Components.** *J Am Stat Assoc.* 1979; **74**(367): 703–707.
[Publisher Full Text](#)
21. Lande R: **Natural Selection and Random Genetic Drift in Phenotypic Evolution.** *Evolution.* 1976; **30**(2): 314–334.
[Publisher Full Text](#)
22. Lande R: **Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry.** *Evolution.* 1979; **33**(1): 402–416.
[Publisher Full Text](#)
23. Lessells CM, Boag PT: **Unrepeatable repeatabilities: a common mistake.** *Auk.* 1987; **104**(1): 116–121.
[Publisher Full Text](#)
24. Lofsvold D: **Quantitative Genetics of Morphological Differentiation in *Peromyscus*. I. Tests of the Homogeneity of Genetic Covariance Structure Among Species and Subspecies.** *Evolution.* 1986; **40**(3): 559–573.
[Publisher Full Text](#)
25. Lynch M, Walsh B: **Genetics and analysis of quantitative traits.** Sinauer, Sunderland, MA., 1998.
[Reference Source](#)
26. Marroig G, Cheverud J: **Size as a line of least resistance II: Direct selection on size or correlated response due to constraints?** *Evolution.* 2010; **64**(5): 1470–1488.
[PubMed Abstract](#) | [Publisher Full Text](#)

27. Marroig G, Cheverud JM: **A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys.** *Evolution*. 2001; **55**(12): 2576–2600.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Marroig G, Cheverud JM: **Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys?** *Am Nat*. 2004; **163**(3): 417–428.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Marroig G, Melo D, Porto A, *et al.*: **Selection Response Decomposition (SRD): A New Tool for Dissecting Differences and Similarities Between Matrices.** *Evol Biol*. 2011; **38**(2): 225–241.
[Publisher Full Text](#)
30. Marroig G, Melo DA, Garcia G: **Modularity, noise, and natural selection.** *Evolution*. 2012; **66**(5): 1506–1524.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Marroig G, Shirai LT, Porto A, *et al.*: **The Evolution of Modularity in the Mammalian Skull II: Evolutionary Consequences.** *Evol Biol*. 2009; **36**(1): 136–148.
[Publisher Full Text](#)
32. Melo D, Marroig G: **Directional selection can drive the evolution of modularity in complex traits.** *Proc Natl Acad Sci U S A*. 2015; **112**(2): 470–475.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Meyer K, Kirkpatrick M: **Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices.** *Genetics*. 2008; **180**(2): 1153–66.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Mitteroecker P, Bookstein F: **The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans.** *Evolution*. 2009; **63**(3): 727–737.
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Moakher M: **On the Averaging of Symmetric Positive-Definite Tensors.** *J Elast*. 2006; **82**(3): 273–296.
[Publisher Full Text](#)
36. Newman ME: **Modularity and community structure in networks.** *Proc Natl Acad Sci U S A*. 2006; **103**(23): 8577–8582.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Numpacharoen K, Atsawarungrangkit A: **Generating correlation matrices based on the boundaries of their coefficients.** *PLoS One*. 2012; **7**(11): e48902.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Olson R, Miller E: **Morphological integration.** University of Chicago Press, Chicago, 1958.
[Reference Source](#)
39. Ovaskainen O, Cano JM, Merilä J: **A Bayesian framework for comparative quantitative genetics.** *Proc Biol Sci*. 2008; **275**(1635): 669–678.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Pavlicev M, Cheverud JM, Wagner GP: **Measuring Morphological Integration Using Eigenvalue Variance.** *Evol Biol*. 2009; **36**(1): 157–170.
[Publisher Full Text](#)
41. Porto A, de Oliveira FB, Shirai LT, *et al.*: **The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes.** *Evol Biol*. 2009; **36**(1): 118–135.
[Publisher Full Text](#)
42. Porto A, Shirai LT, de Oliveira FB, *et al.*: **Size variation, growth strategies, and the evolution of modularity in the mammalian skull.** *Evolution*. 2013; **67**(11): 3305–3322.
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Prôa M, O'Higgins P, Monteiro LR: **Type I error rates for testing genetic drift with phenotypic covariance matrices: a simulation study.** *Evolution*. 2013; **67**(1): 185–195.
[PubMed Abstract](#) | [Publisher Full Text](#)
44. R Core Team. **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2014.
[Reference Source](#)
45. Revell LJ: **phytools: An R package for phylogenetic comparative biology (and other things).** *Methods Ecol Evol*. 2012; **3**(2): 217–223.
[Publisher Full Text](#)
46. Roff D: **Evolutionary quantitative genetics: Are we in danger of throwing out the baby with the bathwater?** *Ann Zool Fennici*. 2003; **40**(4): 315–320.
[Reference Source](#)
47. Roff DA: **The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture.** *Heredity*. 1995; **74**(5): 481–490.
[Publisher Full Text](#)
48. Runcie DE, Mukherjee S: **Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices.** *Genetics*. 2013; **194**(3): 753–767.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Schluter D: **Adaptive Radiation Along Genetic Lines of Least Resistance.** *Evolution*. 1996; **50**(5): 1766–1774.
[Publisher Full Text](#)
50. Steppan SJ, Phillips PC, Houle D: **Comparative quantitative genetics: evolution of the G matrix.** *Trends Ecol Evol*. 2002; **17**(7): 320–327.
[Publisher Full Text](#)
51. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity.** *Nat Rev Genet*. 2007; **8**(12): 921–931.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. Yang K, Shahabi C: **A PCA-based similarity measure for multivariate time series.** In *Proceedings of the 2nd ACM international workshop on Multimedia databases*. ACM, 2004; 65–74.
[Publisher Full Text](#)
53. Melo D, Alvarenga EZ, Hubbe A, *et al.*: **EvoIQG: F1000Research.** *Zenodo*. 2015.
[Data Source](#)

Open Peer Review

Current Referee Status:

Version 1

Referee Report 16 November 2015

doi:10.5256/f1000research.7623.r10610

David Houle

Department of Biological Science, Florida State University, Tallahassee, FL, USA

This paper describes the functions implemented in the R package “evolqg.” This package implements a large set of methods for characterizing and comparing variance-covariance matrices. A compendium of such methods is useful, but the user should be aware of shortcomings in this package.

First, the authors justify grouping these methods together under the umbrella of “evolutionary quantitative genetics.” Under this label, the very diversity of methods is a bit misleading - the majority of implemented methods have no explicit relationship to evolution, and this is not clear in the paper. To a large degree this confusion reflects the state of the field. A great many ad hoc methods are proposed and widely applied, such as the popular Mantel test. For example, this package implements methods for quantifying repeatability of the phenotypic covariance and correlation matrices, which is a fine thing to do, but has little relevance to the genetic matrices that underlie evolution. For those matrices, any procedure for estimating them will partition out the non-genetic factors as part of the hierarchical model fitting. Repeatability is usually the least of our estimation problems. Other methods that are in this not-really-evolutionary category include, matrix correlations, the ‘modularity’ analyses.

A related issue is that this package does not include state-of-the-art techniques. In particular, the advent of MCMC methods (plus other methods for getting sampling variation of matrices: Houle and Meyer 2015) provides proper measures of uncertainty for the evolutionarily relevant G matrices. The previously available R package “evolability” uses the posterior distributions for G matrices as well as selection gradients, when available, to place confidence intervals on the Hansen and Houle measures of evolvabilities. The “evolqg” package under review here primarily implements methods developed and used by the Marroig group, lacking some of the most promising approaches and metrics developed by others (e.g. Hine *et al.* 2009, Houle and Fierst 2013; Aguirre *et al.* 2014). Programming time is of course a limitation, and no one is under obligation to implement everything. However, it is deficient scholarship that this paper makes no mention of these approaches, or of the availability of software that does these analyses.

Based on their descriptions here, the phylogenetic comparison methods implemented in this package seem to be quite deficient. AncestralStates is contradictorily described as dealing with multivariate data, but reconstructs each character independently. It appears to implement an interface to a univariate method, and is not actually a multivariate approach. PhyloW and PhloCompare compute “weighted” estimates of matrices at internal nodes, but the proper weight to apply to a G matrix is not clear at all. This appears to only deal with sampling variance at the individual level, neglecting the more important sources of matrix variation. This disconnect appears to rest on the author’s assumption that “As a general rule,

high similarity between populations' P-matrices is a good indicator of high similarity between P and G". This is certainly not a general rule, as P and G matrices for traits with low heritability can be very different from each other. The classic examples are life history traits.

For drift models, this implementation shares a major deficiency with previous work, in assuming that the expectation is that matrices will remain proportional. This is indeed the large-sample expectation, but any actual population will deviate from that expectation in ways that depend on the unknown parameters of the underlying system – the number of loci, their relative mutability, the underlying M matrix, and of course the effective population size of each part of the genome. For example, Griswold *et al.* 2007 show that even when the underlying M matrix is spherical, realizations based on this will have substantial deviations from sphericity. There is no general treatment of this problem, so all results from such analyses need to be treated with caution. A failure to reject a departure from proportionality is meaningful, but rejections of proportionality do not necessarily indicate that drift is not responsible.

Finally, like much of the software being made freely available, there is no description of what the authors did to validate their implementation of these techniques, and no comparisons with previous analyses to indirectly validate them. One or both of these should really be standard with new software. I know that it is not standard, and do not want to single these authors out on that account. The user, however, should be aware of all unvalidated software, and should perform their own checks. Unfortunately, this is only easy for simple procedures for which a package is not really necessary.

Some very specific issues:

1. Errors: "The proportion of variance not associated with the individuals is called the repeatability."
2. The description of Mantel tests is misleading, as a high correlation does not mean that matrices are the "same," nor does a negative correlation mean that matrices are "opposite."
3. The description of the PCA similarity algorithm is opaque. What is "pondering"?

References

1. Griswold CK, Logsdon B, Gomulkiewicz R: Neutral evolution of multiple quantitative characters: a genealogical approach. *Genetics*. 2007; **176** (1): 455-66 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Hine E, Chenoweth SF, Rundle HD, Blows MW: Characterizing the evolution of genetic variance using genetic covariance tensors. *Philos Trans R Soc Lond B Biol Sci*. 2009; **364** (1523): 1567-78 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Houle D, Meyer K: Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *J Evol Biol*. 2015; **28** (8): 1542-9 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Houle D, Fierst J: Properties of spontaneous mutational variance and covariance for wing size and shape in *Drosophila melanogaster*. *Evolution*. 2013; **67** (4): 1116-30 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Aguirre JD, Hine E, McGuigan K, Blows MW: Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity (Edinb)*. 2014; **112** (1): 21-9 [PubMed Abstract](#) | [Publisher Full Text](#)

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 16 October 2015

doi:[10.5256/f1000research.7623.r10607](https://doi.org/10.5256/f1000research.7623.r10607)

Mark Grabowsky

Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, The Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

This Software Tool is an open source R package for performing a wide variety of evolutionary genetic analyses. It comes at a particularly useful time when large data sets required to use these analytical techniques are more and more available. I predict this package will be used to a greater and greater extent over time.

The team, with lead author Diogo Melo, has made a number of major innovations in the field of evolutionary quantitative genetics and this software tool is well within their purview.

The group describes the potential application of the new software well, cites the sources for the development of the different methods included, and briefly explains each function with an adequate level of discussion. I found the section on Phylogenetic Comparisons most interesting, and would have liked more discussion of these points.

Issues:

1. It would be nice to include data for which the tools could be tried out on and examples of the output included in the text.
2. I also worry about the power of some of these tests, and would appreciate if the authors were to think about adding in caveats where available. There are a few papers out currently that address these issues (e.g. Haber, 2011). For example, I have both heard and performed some analyses using another version of the `DriftTest()` function and it appears about 20 traits are required to ever reject the null hypothesis. As this package is sure to be used by researchers who are less acquainted with these issues they may be unaware of issues with sample sizes, etc.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.