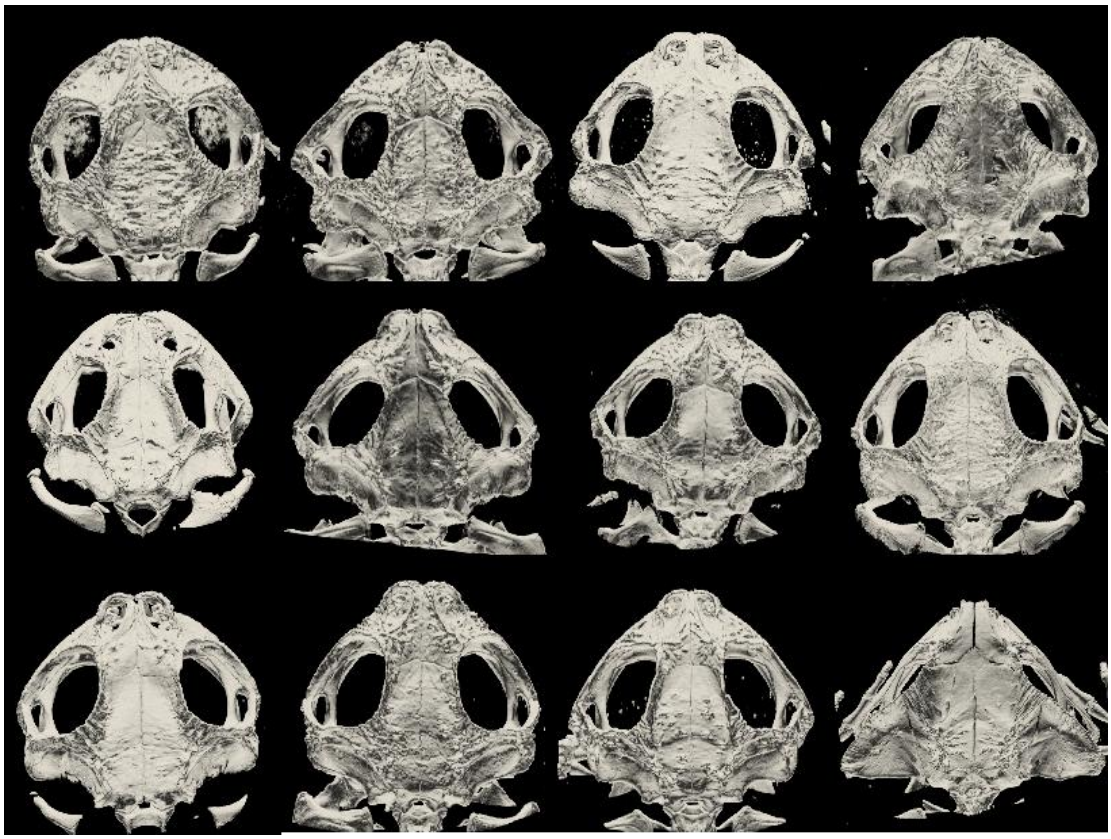
Monique Nouailhetas Simon

Integração morfológica e modularidade em crânios das espécies do grupo *Rhinella granulosa*

Morphological integration and modularity in skulls of the *Rhinella granulosa* species group



São Paulo

2015

Monique Nouailhetas Simon

Integração morfológica e modularidade em crânios das espécies do grupo *Rhinella granulosa*

Morphological integration and modularity in skulls of the *Rhinella granulosa* species group

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

Orientador: Prof. Dr. Gabriel Marroig

São Paulo

2015

Ficha Catalográfica

Simon, Monique Nouailhetas
Integração morfológica e modularidade em crânios das espécies do grupo *Rhinella granulosa*
174 páginas

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

1. Evolução morfológica 2. Genética quantitativa 3. Restrição evolutiva
I. Universidade de São Paulo. Instituto de Biociências.
Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof. Dr. Gabriel Marroig

Orientador

Dedicatória

Dedico esse trabalho ao meu filho Caetano,
uma presença forte e inspiradora.

Sendo um fã de livros como ele é,
a seguir um de seus limeriques favoritos:

“Tem o cardeflon-de-cartola,
que se acha um frajola.
Mas é um bocó,
da até dó!
Nunca foi à escola.”
(Blandina Franco, Passariques do meu quintal)

Epígrafe

"Facts are useless to science unless they are understood. They are to be understood only by theoretical interpretation. The data will never be complete, and their useful, systematic acquisition is dependent upon interpretation of the incomplete data already in hand. The one merit that is claimed for this study is that it suggests new ways of looking at facts and new sorts of fact to look for."

G. G. Simpson ("Tempo and Mode in Evolution")

Agradecimentos

Agradeço ao meu orientador, Gabriel Marroig, por ter me recebido com um projeto sobre sapos no Laboratório de Evolução de Mamíferos (LEM)! Obrigada por ter me incentivado a seguir com meu projeto, mesmo quando estava em dúvida, lá no início, se seria factível. Ainda bem que o foi, e grande parte disso deve-se à você. Obrigada pelas vezes em que discutimos ciência. Obrigada pela atenção e o cuidado com que leu os artigos e a tese, a sua contribuição foi essencial. Finalmente, obrigada pelo rigor científico que você tem. Espero conseguir seguir seu exemplo nesse quesito ao longo de minha carreira (esperando que ela exista!).

Agradeço aos amigos do LEM por diversas vezes discutirem comigo diversos temas, sempre com grande disposição, e por criarem um ambiente pacífico de trabalho: Papete, Ogro, Pato, Paulinha, Alex, Lugar, Aninha, Tafinha, Dani e Wally. Vocês, junto do jardim e do café expresso, formam uma ótima combinação para se trabalhar bem! Consigo nos ver colaborando no futuro em diferentes projetos e crescendo juntos. Agradeço em especial a Diogo Melo (Ogrão) e a Guilherme Garcia (Pato). Obrigada por sempre darem atenção às minhas divagações e questionamentos, conceituais e de programação! Esse trabalho tem grande contribuição de vocês!

Agradeço à minha família pelo apoio a essa empreitada, especialmente à minha querida mãe, Viviane Nouailhetas. Obrigada pelo apoio e amor incondicionais que fizeram e fazem tão bem à minha vida! Não sei se teria feito esse trabalho sem o seu apoio e sua paciência. E obrigada pelas nossas conversas sobre ciência, sobre a vida, sobre a humanidade, sobre filosofia, sobre tudo! Ainda bem que você entende muito bem os sofrimentos e os prazeres da carreira de um pesquisador! Agradeço a meu irmão, Ricardo Nouailhetas Simon, pelo incentivo a seguir como pesquisadora e por ter assistido e lido tanto sobre o mundo natural. Parte da minha escolha em fazer biologia vem da sua admiração pela natureza, algo que compartilhamos. E, é claro, agradeço a meu filho Caetano; é uma alegria imensa ser sua mãe e acompanhar o seu desenvolvimento! Você é lindo, mais que demais, você é lindo sim...

Agradeço à Helena, a tata do Cae, por cuidar tão bem do meu filho enquanto precisei

focar nesse trabalho.

Agradeço aos meus amigos pela companhia, compreensão, divertimento, opiniões, conselhos, broncas, carinho e pensamentos: Mazó, Talita, Cris, Caqui, Juju, Chicão e Érica, Letícia, Michelle, End e Nina. Obrigada por fazerem parte da minha vida e serem pessoas que admiro e respeito.

Agradeço ao Prof. Miguel Trefaut Rodrigues por sua ajuda em contactar vários curadores de coleções científicas e por discussões sobre as espécies e sobre resultados. Agradeço também a Patrícia Narvaes por me mandar seus dados do doutorado e por me ajudar a identificar algumas espécies.

Agradeço aos curadores de coleções científicas brasileiras e internacionais por me emprestarem muitos espécimes de sapo: David Kizirian e Robert Pascocello (AMNH); Célio Haddad e Nadya Pupin (UNESP – Rio Claro); Gustavo H. C. Vieira (CHUFPB); Julián Faivovitch e Santiago J. Nenda (MACN); Márcia F. Renner (MCN-FZB-RS), Glaucia M. F. Pontes (MCT-PUCRS); José P. Pombal Jr. e Pedro Henrique Pinna (MNRJ); Taran Grant e Carolina Mello (MZUSP); Jeremy F. Jacobs e Robert Wilson (USNM); David Cannatella e Travis LaDuc (TNHC).

Agradeço a James Hanken, de Harvard, por dar sua opinião sobre as hipóteses de modularidade. Agradeço a Martín Pereyra e a Julián Faivovitch, do MACN, por me enviarem a filogenia das espécies antes da publicação.

Agradeço ao Departamento de Genética e Biologia Evolutiva e ao Instituto de Biociências da Universidade de São Paulo pela possibilidade de realizar a pós-graduação e pelo acesso ao micro-CT. Agradeço aos professores da pós-graduação que prezam dar uma boa aula e o fazem com gosto. Quero fazer o mesmo!

Agradeço à Fundação de Amparo à Pesquisa de São Paulo, FAPESP, pelo financiamento de minha bolsa de doutorado.

Índice

Resumo

Os conceitos e métodos provindos das teorias de integração morfológica e de genética quantitativa formam o arcabouço teórico para o estudo da evolução de estruturas complexas, compostas de múltiplos caracteres que interagem entre si. Nesse trabalho, utilizamos o crânio como modelo de estrutura complexa e estudamos sua diversificação nas espécies de sapo do grupo *Rhinella granulosa*. As perguntas do trabalho foram: (1) A organização da (co)variação é similar entre as espécies?; (2) A organização da (co)variação é modular nas espécies, conforme expectativas baseadas em desenvolvimento ou função?; (3) Fatores externos, como filogenia e clima, estruturam a similaridade no padrão de covariação entre as espécies?; (4) A diversificação da morfologia média do crânio se deu por deriva ou seleção natural?; (5) A divergência na morfologia média do crânio está associada à variação climática entre as espécies?; e finalmente (6) Restrições evolutivas atuaram na divergência entre as espécies? Os espécimes foram escaneados e validamos o uso de imagens 3D para a mensuração de 21 distâncias lineares. Os crânios das espécies foram representados como matrizes fenotípicas (**P**) de covariância e de correlação entre as distâncias. A similaridade entre as **P** das espécies é alta. As **P** de todas as espécies se conformam a um padrão modular compatível com interações funcionais entre ossos. As diferenças entre as **P** concentram-se no rostro e são associadas a diferenças no clima entre as espécies. Detectamos sinal de seleção natural nos nós mais basais da filogenia e variação local no crânio está associada à variação na sazonalidade da chuva entre as espécies. Restrições evolutivas atuaram na diversificação do crânio das espécies, defletindo as respostas evolutivas para tamanho. Concluímos que tanto seleção estabilizadora e direcional, conectadas à variação climática, quanto restrições evolutivas atuaram na diversificação do crânio das espécies.

Abstract

Concepts and methods within the theories of morphological integration and quantitative genetics characterize the foundation to study the evolution of complex structures, composed of several traits that interact with each other. In this work, we used the skull as a model of complex structure and we studied its diversification in toad species belonging to the *Rhinella granulosa* group. The questions addressed were: (1) Is the (co)variance structure similar across species?; (2) Is the (co)variance structure modular in the species, and compatible with developmental or functional interactions among traits?; (3) Do external factors, such as phylogeny and climate, structure the similarity in covariance pattern across species?; (4) Was the diversification of skull mean morphology driven by drift or natural selection?; (5) Is skull divergence associated to climatic variation across species?; and finally, (6) Is there a role for evolutionary constraints in species skull divergence? We scanned all specimens and we validated the use of 3D images to measure 21 linear distances. The skull was represented as covariance and correlation phenotypic matrices (**P**) among distances. **P** similarity is very high among species. All species' **P** had a modular pattern compatible with functional interactions among bones. Differences in **P** were concentrated in the snout and associated to differences in climate across species. We detected a selection signal in the three most basal nodes of the phylogeny and local variation in the skull is explained by between-species variation in precipitation seasonality. Evolutionary constraints played a major role in species skull diversification, biasing evolutionary responses towards the direction of size. We conclude that stabilizing and directional selection, connected to climatic variation, as well as evolutionary constraints, acted in species skull diversification.

Introdução Geral

A maioria dos caracteres que interessam aos biólogos, como os comportamentais, morfológicos e fisiológicos, são caracteres quantitativos, e portanto, inseridos na teoria da genética quantitativa. O objetivo principal dessa introdução é de apresentar os conceitos básicos da genética quantitativa evolutiva, uma vez que essa teoria fornece os fundamentos para o estudo de caracteres complexos como o crânio, objeto de estudo dessa tese. Escolhi apresentar a genética quantitativa por meio da contextualização histórica das contribuições de grandes nomes da biologia evolutiva. Destaco inicialmente os fundadores da escola biométrica, que tem íntima relação com a genética quantitativa, e a resolução do debate entre os biometricistas e os mendelianos, que culminou na criação da genética de populações. Depois foco em estudos de seleção em caracteres contínuos univariados realizados por Jay L. Lush e Douglas S. Falconer. Posteriormente, abordo a teoria mais recente de seleção natural multivariada elaborada por Russell Lande, que tem conexões com o conceito de paisagem adaptativa, inicialmente elaborado por Sewall Wright e interpretado em termos fenotípicos por G. G. Simpson. Finalmente, esclareço a conexão entre a teoria de Lande com a teoria de integração morfológica e modularidade, elaborada por Everett C. Olson e Robert L. Miller. Aspectos importantes da teoria de modularidade estão explicitados nessa introdução, porém uma abordagem mais detalhada está apresentada no Capítulo 2 da tese. Ao final, apresento os objetivos da tese no contexto do grupo estudado, as espécies de sapos do grupo *Rhinella granulosa*.

A Escola Biométrica

A teoria da genética quantitativa centra-se no estudo da evolução de caracteres que

possuem distribuição fenotípica contínua. O conceito de *semelhança por parentesco* é essencial para a teoria uma vez que a previsão de resposta evolutiva depende de métricas que quantifiquem esse fenômeno (Arnold 1994). Em um trabalho de 1889, Francis Galton exprimiu a semelhança entre a altura dos filhos em relação à altura média dos pais como uma representação gráfica no intuito de mostrar que uma reta ajustada aos dados empíricos desviava de uma linha representado similaridade perfeita (inclinação = 1.0; Figura 1). Galton chamou essa reta de regressão, pois indicava a tendência dos valores empíricos a regredirem a essa reta. Porém, além de criar uma das análises estatísticas mais usadas em ciência (inclusive nessa tese), Galton mostrou que existia um padrão regular na herança de caracteres contínuos. O estatístico Karl Pearson por essa mesma época trabalhou com o zoólogo quantitativo Weldon em questões envolvendo seleção natural, e foi apresentado por este a Galton. Juntos, Galton e Pearson desenvolveram uma parceria no estudo de variação contínua e sua evolução. Pearson, com a invenção da técnica de correlação, mostrou que a mudança na média de um caráter pode ser devida à ação de seleção natural em outro caráter correlacionado com aquele que mudou (Pearson 1903). Essa percepção de mudança correlacionada é de alta relevância em genética quantitativa de caracteres complexos.

As gerações seguintes de biometricistas enfrentariam um histórico debate com os geneticistas mendelianos (Falconer e Mackay 1996). O embate recaia na dúvida de os caracteres discretos estudados pelos mendelianos possuírem propriedades de herança genética e evolução similares às descritas pelos biometricistas para caracteres contínuos (Lynch e Walsh 1998). Os mendelianos utilizavam as leis de segregação independente e o cálculo de razões mendelianas para quantificar as proporções dos fenótipos da prole gerada a partir de cruzamentos

parentais específicos. Os seguidores de Mendel acreditavam que a evolução fenotípica se dava por meio de macro-mutações que produziriam variação nos caracteres discretos. Já os biometricistas argumentavam que a evolução ocorria como resultado de seleção natural atuando nas distribuições contínuas dos caracteres. Esse dilema histórico foi resolvido pelos trabalhos independentes de Ronald Fisher, J. B. S. Haldane e Sewall Wright com a fundação da genética de populações. Os estudos desses pesquisadores mostraram que a seleção natural pode funcionar com a variação contínua observada em populações naturais junto das leis de herança mendeliana (Ridley 1996).

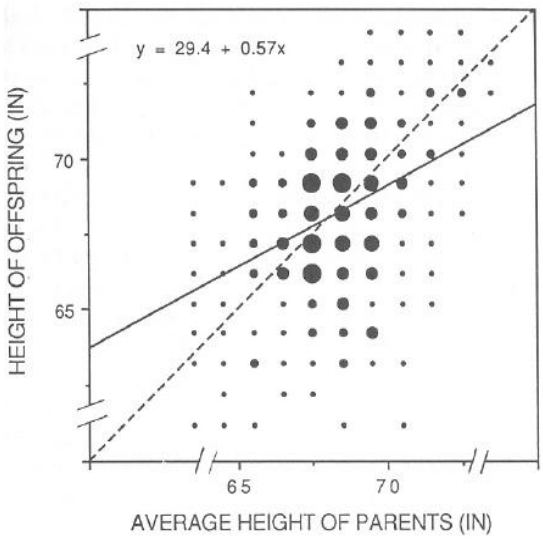


Figura 1. Regressão da altura dos filhos em relação à altura média dos pais de Galton. A reta pontilhada representa herança perfeita dos valores dos filhos diante da altura dos pais. A reta sólida é a reta de regressão ajustada por quadrados mínimos. Extraído de Arnold (1994).

A solução do conflito baseou-se essencialmente na proposta de postular que caracteres contínuos são regulados por múltiplos loci de efeito pequeno, que são herdados segundo as leis de Mendel (Hill e Mackay 2004). A soma dos efeitos destes vários loci, adicionado de efeito ambiental, produz variação fenotípica contínua nos caracteres (Figura 2). Colocando isso em termos da genética quantitativa, o genótipo confere um determinado valor de um caráter ao indivíduo e o ambiente causa um desvio desse valor: P = G + E, sendo P o valor fenotípico, G o valor genotípico e E o

desvio ambiental (Falconer e Mackay 1996). Pensando em termos populacionais, relevante para processos evolutivos, assumimos que o desvio médio ambiental é zero, e que portanto, o valor médio fenotípico corresponde ao valor médio genotípico da população. Essa suposição é fundamental, pois permite que estudemos as propriedades genéticas das populações por meio de seus fenótipos. Entretanto, usar os valores genotípicos não é suficiente para lidar com a herança de caracteres contínuos, pois os pais passam os seus genes e não seus genótipos para sua prole (Falconer e Mackay 1996). A ideia de *efeito médio de um alelo*, elaborado por Fisher (1918; 1930), resolve esse problema: é o desvio fenotípico médio, da média populacional, dos indivíduos que receberam o alelo de um tipo de um dos pais, sendo o outro alelo advindo ao acaso da população. O efeito médio de um alelo é uma propriedade de um gene específico, mas também propriedade da população, uma vez que seu efeito depende das frequências alélicas na população determinando o genótipo completo. O valor genotípico de um caráter na prole resulta da soma dos efeitos médios de todos os alelos parentais (*efeito aditivo*) que afetam esse caráter.

Porém, é difícil determinar quantos genes afetam um caráter contínuo e qual o tamanho do efeito de cada um deles. Na prática, a soma dos efeitos médios dos alelos é medida como o *valor de acasalamento* ("breeding value"): o valor fenotípico de um indivíduo julgado pelo valor médio do caráter em sua prole. Um indivíduo é cruzado com outros indivíduos retirados ao acaso da população, e o desvio fenotípico médio de sua prole (multiplicado por dois), para um dado caráter em relação à média populacional, corresponde ao seu valor de acasalamento. Fisher mostrou que estas duas definições de valor de acasalamento, teórica (soma dos efeitos médios que o indivíduo carrega) e empírica (julgado pelo valor médio da prole produzida) são equivalentes sob acasalamento casualizado.

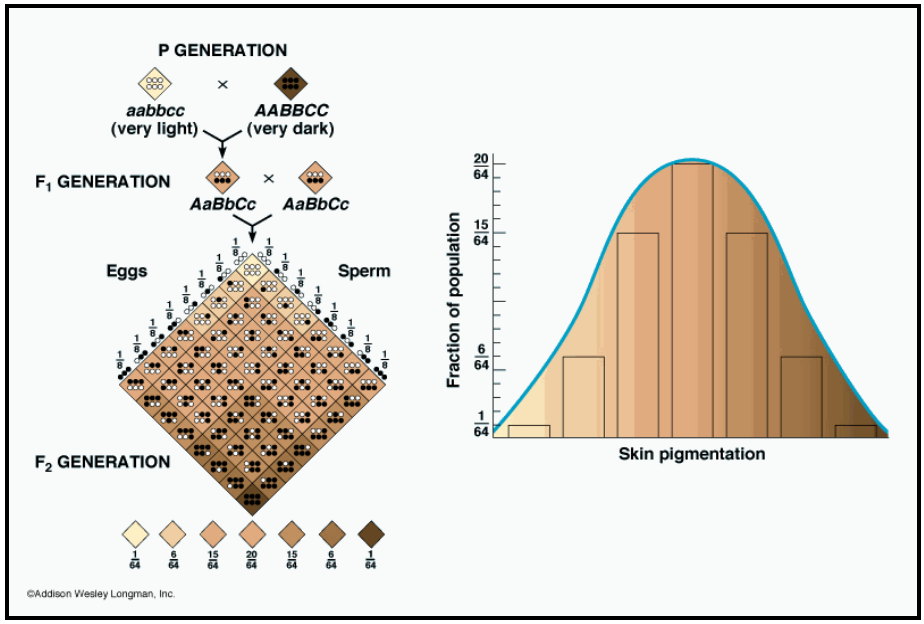


Figura 2. Exemplo de como seis alelos que regulam coloração da pele podem ter herança mendeliana e produzir variação contínua. Diferentes combinações dos alelos em genótipos geram diferentes tonalidades de cor da pele, como representado na distribuição contínua normal à direita. Variação ambental ainda pode adicionar mais variação ao caráter.

Dado que o valor de acasalamento corresponde teoricamente à soma dos efeitos médios dos alelos que regulam a expressão de um determinado caráter, a variância nos valores de acasalamento corresponde à *variância aditiva*, a parte da variância fenotípica que é herdável, e portanto, a principal causa de semelhança entre pais e filhos. Voltando à regressão de Galton (Figura 1), podemos agora interpretá-la como uma regressão entre os valores genéticos aditivos dos filhos e os valores fenotípicos parentais (Arnold 1994). A inclinação dessa reta, portanto, corresponde à *herdabilidade* da altura, como uma proporção entre variação aditiva e variação fenotípica ($h^2 = V_A/V_P$). Mais recentemente, o uso de marcadores moleculares permitiu a determinação de loci que afetam caracteres quantitativos ("QTL = quantative trait loci"), confirmando que múltiplos genes contribuem para a variação nesses caracteres. Por exemplo, um estudo em aproximadamente 2.000 linhagens de

ratos com genes silenciados estimou que por volta de 6.900 alelos afetam peso corpóreo, sendo que a maioria das linhagens afetadas teve menor peso (Reed et al. 2008). Além disso, estudos de QTLs mostraram que existe variação originada de interações entre genes (QTLs epsitáticos; Pavlicev et al. 2008), como Sewall Wright já enfatizava décadas atrás.

Resposta Univariada à Seleção Natural

Outra grande contribuição dos geneticistas populacionais foi diferenciar a seleção natural da resposta evolutiva à seleção natural (Fisher 1930; Haldane 1954). A seleção natural atua na variação fenotípica independente dos mecanismos de herança genética. Em contrapartida, a resposta evolutiva depende de variação herdável. Porém, é possível medir seleção nos fenótipos sem conhecer quais genes determinam esses fenótipos (Arnold e Wade 1994). A teoria evolutiva focada em fenótipos não está fundamentada na frequência de alelos, uma vez que a genética de populações não se estende para caracteres poligênicos (Arnold et al. 2001). Sendo assim, é possível realizar estudos de seleção nos fenótipos e quantificar a resposta evolutiva na prole. Jay L. Lush desenvolveu as primeiras aplicações práticas de genética quantitativa para o melhoramento genético animal, por meio de estudos de seleção artificial e estimativas de herdabilidade (Hill e Mackay 2004). Lush (1937) expressou a mudança em um caráter como a famosa *Equação do Criador* ("Breeder's Equation"):

$$R = h^2 S; \qquad (1)$$

na qual R é a resposta evolutiva (diferença média do caráter na população antes e depois de seleção), h^2 é a herdabilidade do caráter ("narrow sense heritability": variância aditiva dividida pela variância fenotípica) e S é o *diferencial de seleção* (diferença entre a média dos parentais selecionados e a média populacional; Figura

3). A Equação do Criador pode ser usada para estimar a resposta à seleção em sistemas naturais, como nos tentilhões de Darwin estudados pelo casal Grant (Grant e Grant 1995). Entretanto, no caso de estimativa de seleção passada, valores espúrios de *S* podem ser obtidos se a covariação entre o caráter e a aptidão causada pelo ambiente difere da covariação genética entre o caráter e a aptidão (Morrissey et al. 2010).

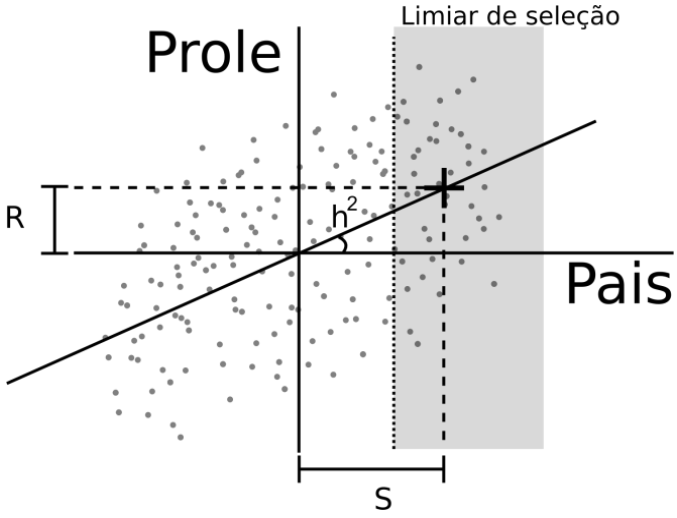


Figura 3. Resposta à seleção ilustrada na regressão pais-prole. Cada ponto corresponde a um par de desvios fenotípicos dos pais e prole em relação à média populacional (ponto 0,0). A área sombreada corresponde à seleção de truncamento e a cruz a média dos pais selecionados. R é a diferença entre a média da prole da média populacional. S é a diferença entre a média dos pais selecionados e a média populacional.

Outro grande nome da genética quantitativa foi Douglas S. Falconer, que junto de outros pesquisadores do Conselho de Pesquisa em Agricultura de Edinburgo, trabalhou em estudos de seleção artificial especialmente em animais de fazenda, coelhos e ratos, mas também em *Drosophila*, aproveitando-se de seu tempo de geração mais rápido (Hill e Mackay 2004). Uma importante proposta de Falconer foi de descrever a resposta à seleção de maneira a levar em conta a força de seleção

aplicada. Para isso, Falconer criou o conceito de *herdabilidade realizada*, calculada por meio de uma regressão da resposta cumulativa no diferencial de seleção cumulativo (*R*/*S* da Equação 1), que pode diferir muito da herdabilidade calculada por meio de correlações de parentes na população (Hill e Mackay 2004). Falconer também trabalhou com efeitos maternos e com doenças de herança complexa, além de publicar o livro referência em genética quantitaiva, "Introduction to Quantitative Genetics", a primeira versão em 1960.

Paisagem Adaptativa Genética e Fenotípica

Em um movimento um tanto oposto ao modelo de efeitos genéticos aditivos de Fisher, Sewall Wright enfatizava a importância de interações entre genes na evolução dos caracteres. Em seus estudos com porcos da Índia, Wright percebeu que efeitos genéticos não aditivos, como epistasia e pleiotropia (um gene afetando vários caracteres) eram fortes e frequentes nos fenótipos (Johnson 2008). Quando as interações entre genes envolvem mais de dois loci, a visualização e a análise desses sistemas torna-se muito difícil. Para auxiliar na interpretação desses sistemas, Wright (1932) desenvolveu a famosa metáfora da *paisagem adaptativa*: uma descrição visual topográfica da relação entre aptidão média da população, usando isolinhas de mesma aptidão, e combinações de frequências de alelos (Figura 4). Os organismos com "combinações mais harmoniosas" de genes ocupariam os picos da paisagem, os pontos de maior aptidão, e a seleção natural tenderia a maximizar o número de organismos no pico adaptativo (Eldredge 2008). Porém, Wright especulava que não somente seleção atuava nas populações, mas também mutações e deriva genética, fazendo com que houvesse um espalhamento dos indivíduos ao redor do pico adaptativo. Wright usou o conceito de paisagem adaptativa para ilustrar suas ideias

sobre a evolução de populações locais de uma espécie (demes) por entre condições ecológicas distintas e a possibilidade de deriva genética permitir a passagem de uma população por um vale na paisagem até outro pico adaptativo ("shifting balance theory"; Svensson e Calsbeek 2012).

A concepção de paisagem adaptativa de Wright foi retomada anos depois pelo paleontólogo G. G. Simpson (1944) em seu livro "Tempo and Mode in Evolution", já na época da chamada Síntese Moderna. Simpson buscava entender a evolução de formas fósseis de mamíferos e lutava para reconciliar a genética mendeliana com a evolução fenotípica ocorrida em tempos evolutivos profundos (Arnold 2014). Simpson usou o conceito de paisagem adaptativa, porém alterando os eixos de combinações alélicas para eixos de valores fenotípicos, uma vez que ele somente tinha acesso aos fenótipos em seus estudos dos fósseis. Simpson elaborou sua teoria de "evolução quântica" na qual uma linhagem muda de *zona adaptativa* rapidamente, mudando também sua morfologia, devido a uma pressão seletiva muito forte. As zonas adaptativas representavam para Simpson espaços morfológicos ocupados ou não por linhagens, mas que correspondiam a morfologias de maior aptidão (Svensson e Calsbeek 2012).

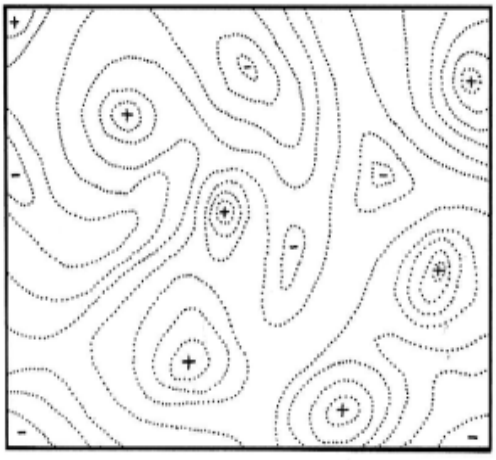


Figura 4. Desenho original de Wright da paisagem adaptativa. Os eixos representam combinações de frequências de alelos (múltiplas dimensões) e o eixo vertical, representado pelas isolinhas, representa a aptidão média de uma população. Os sinais positivos indicam os picos adaptativos locais, enquanto que os negativos indicam os vales, combinações gênicas que reduzem a aptidão média da população. Extraído de Wright (1932).

Simpson considerava a mudança de zonas adaptativas como um evento extraordinário ao longo da evolução das linhagens, sendo que a maioria da evolução se dava dentro das zonas adaptativas, ocorrendo extinções e períodos de estase (Arnold 2014). Simpson explicou a evolução das linhagens de cavalo e sua mudança de dieta, associada à mudança na morfologia dos dentes, como uma mudança entre zonas adaptativas, inclusive postulando que a própria paisagem adaptativa poderia mudar, sendo portanto uma paisagem dinâmica ao longo do tempo (Simpson 1944; Figura 5).

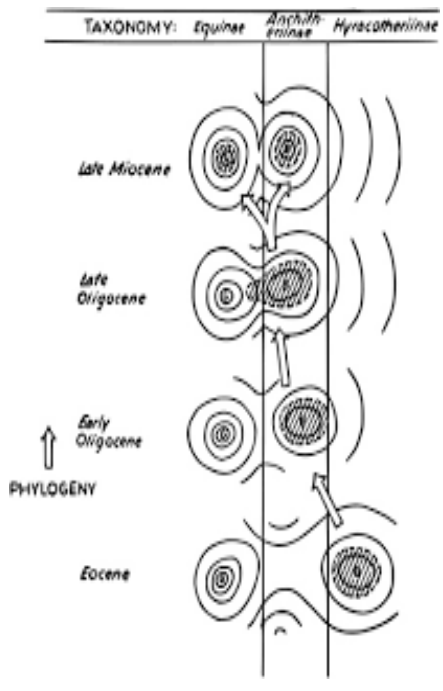


Figura 5. Evolução da dieta na linhagem de cavalos como uma mudança de zonas adaptativas. Simpson concebeu que inicialmente somente uma zona adaptativa estava ocupada (elipse hachurada à direita no desenho mais abaixo), porém já existia uma segunda zona adaptativa de dieta. No Oligoceno, as zonas se aproximaram e uma maior variação fenotípica na lihagem inicial permitiu a transição de parte da população para a nova zona adaptativa, junto de mudança rápida na morfologia dos dentes. Extraído de Simpson (1944).

Resposta à Seleção Multivariada

Nas décadas de 70 e 80, o biólogo evolutivo teórico Russell Lande formalizou o conceito de paisagem adaptativa fenotípica de Simpson na forma de modelos matemáticos (Lande 1976; 1979; 1980), tornando o conceito de valor heurístico em uma teoria solidamente fundamentada na genética quantitativa (Svensson e Calsbeek

2012; Arnold 2014). Lande estava interessado em determinar objetivamente a importância de deriva genética em eventos macro-evolutivos, e para tanto propôs um teste estatístico para a hipótese de deriva, utilizando de tamanho efetivo populacional (Lande 1976; 1980). Lande usou um modelo de ação simultânea de seleção natural e deriva genética para avaliar a hipótese de Wright e Simpson de que a transição entre zonas adaptativas ocorreria por deriva. O seu modelo teórico é baseado na mudança de aptidão média da população gerando uma mudança no fenótipo médio após um episódio de seleção natural. Lande mostrou que a evolução do fenótipo médio por seleção sempre se dá na direção de aumento da aptidão média da população (Lande 1976). A paisagem adaptativa definida por Lande refere-se, portanto, à relação entre valores fenotípicos médios de caracteres e a aptidão média da população.

Em seu artigo de 1979, Lande aponta que faltou o uso de métodos da genética quantitativa em abordagens prévias de evolução e sua conexão com variação intra-populacional. Ao elaborar sobre a evolução de trajetórias alométricas evolutivas, Lande enfatizou que a seleção natural atua em mais de um caráter, ou seja, a seleção natural é multivariada (Lande 1979). Usando da Equação do Criador (equação 1) e o operador matemático de gradiente (∇), que é um vetor de derivadas parciais, Lande deriva a *Equação de Resposta à Seleção Mutivariada* ou *Equação Multivariada do Criador* ("Multivariate Breeder's Equation"):

$$\nabla \ln \overline{W} = \mathbf{P}^{-1}\mathbf{S} = \mathbf{G}^{-1}\Delta\mathbf{z}; \qquad (2)$$

na qual $\nabla \ln \overline{W}$ representa a mudança na aptidão média da população (*gradiente de seleção*), $\mathbf{P}^{-1}$ é a inversa da matriz de variância e covariância fenotípica; **S** é o vetor de diferenciais de seleção para todos os caracteres; $\mathbf{G}^{-1}$ é a inversa da matriz de variância e covariância genética aditiva e $\Delta\mathbf{z}$ indica a mudança no vetor de médias da população após uma geração. A **G** apresenta em sua diaginal valores de variância aditiva dos

caracteres (V_A) e valores de covariância aditiva entre caracteres fora das diagonais, indicando o quanto os caracteres são geneticamente acoplados. O uso do operador de gradiente significa que ocorre uma mudança na aptidão devido a uma pequena mudança no caráter z_i, mantendo todos os demais caracteres fixos. Portanto, o gradiente de seleção pode ser interpretado como a força de seleção direcional atuando somente no caráter i (Lande 1979).

Em um artigo posterior de 1983, Lande e o geneticista quantitativo Stevan J. Arnold demonstraram que o gradiente de seleção direcional é melhor descrito como um vetor de coeficientes de regressão parcial da aptidão relativa em relação aos caracteres (Lande e Arnold 1983). Essa interpretação do gradiente de seleção é derivada do uso da equação de Price (1970):

$$s = \mathrm{Cov}[w, z]; \qquad (3)$$

que expressa que o diferencial de seleção corresponde à covariação entre a aptidão relativa e o caráter z. Price (1970) indicou que essa equação pode ser visualizada como uma regressão linear da aptidão relativa sobre o caráter z, e que a inclinação dessa reta ($\boldsymbol{\beta_{zw}}$) indica o quanto a mudança no caráter afeta a aptidão. Lande e Arnold (1983) generalizaram a equação de Price para múltiplos caracteres e constataram que o vetor $\mathbf{P^{-1}S}$ é um conjunto de coeficientes de regressão parcial da aptidão relativa nos caracteres:

$$\boldsymbol{\Delta z} = \mathbf{GP^{-1}S} = \mathbf{G}\boldsymbol{\beta} \qquad (4)$$

Portanto, o gradiente de seleção também pode ser denominado como um vetor $\boldsymbol{\beta} = \mathbf{P^{-1}S}$. Escrevendo $\mathbf{S} = \boldsymbol{\beta}\mathbf{P}$, podemos notar que o diferencial de seleção tem um termo referente à seleção direta nos caracteres ($\boldsymbol{\beta}$) e um componente referente à seleção indireta nos caracteres determinado por sua covariação fenotípica **P** (Lande e Arnold 1983). O essencial é compreender que ao apresentar o gradiente de seleção em termos

de coeficientes de regressão parciais, Lande e Arnold permitiram que seleção multivariada possa ser estimada em populações naturais por meio de uma técnica estatística.

Para a melhor compreensão da diferença entre o diferencial e o gradiente de seleção, usarei os gráficos de Cheverud (1984), que mostram o efeito de covariação entre caracteres na resposta à seleção. Seleção em um caráter qualquer produz não somente uma resposta direta à seleção, mas também uma resposta indireta na distribuição de caracteres correlacionados com o caráter que efetivamente sofreu seleção (Lande e Arnold 1983). Essa resposta indireta é interpretada como uma *restrição à resposta evolutiva* (Cheverud 1984; Arnold 1992). Isso acontece porque a seleção direta em um caráter X cria um diferencial de seleção S_X, mas também um diferencial de seleção indireto S_Y, causando uma mudança no caráter Y mesmo que ele não tenha sido diretamente selecionado (Figura 6; Cheverud 1984). Porém, ao descontarmos a correlação fenotípica entre os caracteres X e Y, a seleção em X resultará apenas em um diferencial de seleção em X (Figura 6). Considerando diversos caracteres ao mesmo tempo, a covariação entre eles é descontada ao usarmos a inversa da matriz-P para multiplicar o vetor de diferenciais de seleção, resultando no gradiente de seleção ($\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S}$).

O sinal de um elemento do gradiente de seleção indica se o caráter está sob seleção negativa ou positiva, enquanto que sua magnitude indica a força de seleção (Phillips e Arnold 1989). Indo além, considerando a premissa de que tanto os fenótipos quanto a aptidão têm distribuição multivariada normal, Lande e Arnold (1983) mostraram que o gradiente de seleção corresponde à inclinação média da superfície de aptidão. Nesse artigo, os autores também derivam a matriz de seleção não-linear $\boldsymbol{\gamma}$ (seleção estabilizadora na diagonal e seleção correlativa entre caracteres

fora da diagonal), e novamente usando a equação de Price (1970), concluem que essa matriz estima a curvatura média da superfície de aptidão. A inclinação e curvatura médias da superfície de aptidão individual ponderada pela distribuição fenotípica corresponde à inclinação e curvatura médias da paisagem adaptativa (Phillips e Arnold 1989). Portanto, os modelos de Lande e Arnold estimam parâmetros da paisagem adaptativa de Simpson.

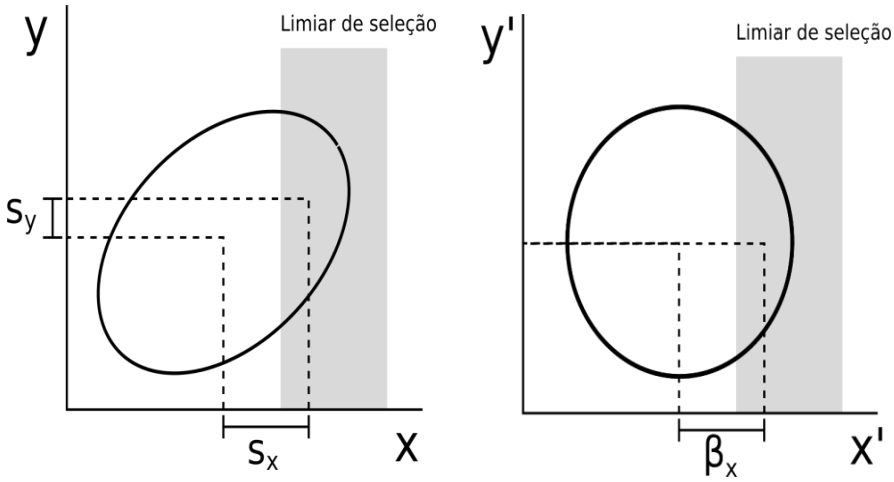


Figura 6. Diferencial e gradiente de seleção. A elipse representa a covariação fenotípica entre os caracteres X e Y. No painel à equerda, X e Y possuem uma covariação positiva, fazendo com que a seleção de truncamento somente em X (Sx) resulte em seleção indireta em Y (S_Y). Ao descontarmos a correlação entre X e Y, usando os resíduos X' e Y' (variação nos caracteres independente um do outro), a seleção de truncamento em X gera seleção apenas em X, correspondente a β_x. Adaptado de Cheverud (1984).

Genética Quantitativa Comparativa

Para se entender a evolução da matriz-G, é preciso comparar matrizes-G entre várias populações ou espécies nas quais conhecemos suas relações filogenéticas (Steppan 1997). Questões centrais referem-se à estabilidade da **G** e à persistência das restrições genéticas ao longo do tempo evolutivo (Arnold 1992; Steppan et al. 2002). Se **G** for

estável, é possível reconstruir a seleção passada que levou à divergência entre populações (Lande 1979). A equação de Lande pode ser rearranjada e estendida para a forma macro-evolutiva:

$$\boldsymbol{\beta} = \mathbf{G}^{-1} (z_i - z_j); \quad (5)$$

sendo que $(z_i - z_j)$ representa a diferença nos vetores de médias entre as espécies *i* e *j*. Na prática, se a equação acima é aplicada para estudos de diferenças entre espécies, o gradiente de seleção representa a seleção cumulativa sofrida por cada linhagem independentemente. Dada a dificuldade de se obter matrizes-G para diversas populações ou espécies aparentadas por conta do número grande de famílias necessário, utiliza-se em geral a matriz-P como substituta da **G**, adotando-se a "Conjectura de Cheverud" na qual uma **P** é em geral melhor estimada que uma **G**; e que **P** bem estimadas são similares às suas correspondentes genéticas (Cheverud 1988). Essa conjectura foi confirmada em diversos trabalhos (e.g. Reusch e Blanckenhorn 1998), assim como similaridade entre **G** e **P**, especialmente em mamíferos (e.g. Cheverud 1995; Marroig e Cheverud 2001; Porto et al. 2009).

A questão de restrições genéticas causada pelas correlações genéticas entre caracteres teve muita atenção nos últimos anos (Blows e Walsh 2009). A resposta à seleção natural tende a ser defletida para direções da matriz-G que concentram a maior proporção de variação. Schluter (1996) propôs chamar essas direções de linhas de menor resistência evolutiva e verificou que a divergência entre populações de diferentes organismos foi na direção do primeiro eixo de maior variação da **G** (g_{max}). A ação de uma restrição genética sob a evolução de um sistema depende da interação entre a **G** e a paisagem adaptativa (Arnold et al. 2001). O alinhamento ou não entre os eixos de máxima variação genética e a direção de seleção vai determinar se a população terá sua resposta dificultada ou facilitada em relação à direção da seleção

(Figura 7).

Recentemente, a ideia de restrição evolutiva vem sendo abordada com métricas como *evolvabilidade* (Hansen e Houle 2008) e *flexibilidade evolutiva* (Marroig et al. 2009). Evolvabilidade indica a quantidade de variação em uma determinada direção, e foi descrita como a magnitude da projeção do vetor resposta na direção de seleção (o quanto da resposta foi na direção de seleção). Porém, a detecção de restrição tem sido avaliada como evolvabilidade na direção da divergência entre populações (e.g. Hansen e Voje 2010; Bolstad et al. 2014), já que poucos estudos calculam gradientes de seleção. Uma vez que a divergência tenha se dado em uma direção de alta evolvabilidade, considera-se que houve a ação de restrição evolutiva. A flexibilidade evolutiva refere-se à correlação entre a resposta evolutiva e a seleção natural e diz o quanto que a divergência entre populações seguiu a direção da seleção. Esses conceitos, junto de reconstrução de seleção, foram abordados no Capítulo 3 da tese.

Integração Morfológica e Modularidade

A teoria evolutiva elaborada por Lande conecta-se com a teoria de integração morfológica e modularidade desenvolvida por Everett Olson e Robert Miller (1958) pelo estudo em comum de caracteres complexos, que apresentam covariação/correlação de seus elementos constituintes. Enquanto a teoria de genética quantitativa fundamenta-se na evolução dos caracteres usando (co)variação genética e fenotípica, a teoria de integração morfológica estuda como essa variação é organizada nos sistemas complexos e por quais fatores ela é moldada. A teoria de genética quantitativa mede o efeito estatístico da variação genética na mecânica do desenvolvimento (Cheverud 1984). Entender como sistemas complexos podem

evoluir de maneira integrada foi uma questão importante em biologia evolutiva desde a época na qual Darwin concebeu a ideia de seleção natural (Cheverud 1984).

Pesquisadores que lidavam com a concepção de um sistema de desenvolvimento epigenético (tradução mecânica do genótipo mais o ambiente no fenótipo), como Conrad Waddington, especulavam que as vias de desenvolvimento que determinavam caracteres complexos eram reguladas por diversos genes que interagiam, como postulava Wright (1930). Waddington (1957) criou o conceito de *paisagem epigenética* para representar vias de desenvolvimento canalizadas, que eram reguladas por uma intricada rede de interações gênicas, sendo que "...cada via depende da interação de um determinado grupo de genes". Uma ideia fundamental levantada por Waddington é que as mutações que ocorrem em genes que regulam complexos funcionais integrados acarretam em uma mudança coordenada do complexo devido à *restrição de desenvolvimento* imposta pela correlação dos caracteres pelo sistema epigenético (Cheverud 1984).

As restrições de desenvolvimento existem porque o sistema epigenético resulta em efeitos fenotípicos pleiotrópicos dos genes individuais, ou seja, os genes afetam mais de um caráter simultaneamente (Wright 1980; Cheverud 1984). Considerar que os caracteres fenotípicos são integrados por efeitos do desenvolvimento e de função é a ideia central da teoria de integração morfológica de Olson e Miller (1958); também elaborada por Raíssa Berg (1960) em plantas. Olson e Miller (1958) propuseram que os caracteres fenotípicos que compartilham uma mesma via de desenvolvimento ou executam uma mesma função são mais integrados entre si do que com os demais caracteres do organismo. Ruppert Riedl, que também trabalhava com desenvolvimento e evolução, achava que seleção para pleiotropia resultaria em um controle conjunto de caracteres fenotípicos que interagiam em uma

mesma via de desenvolvimento ou na execução de uma mesma função (Riedl 1978). A teoria de Riedl, junto da derivação de Cheverud (1982) sobre a evolução de correlações genéticas, e sua conexão com a teoria de Lande (1979; 1980) de ação de seleção estabilizadora sob a matriz-G, estão elaboradas no Capítulo 2 da tese no contexto específico do crânio das espécies de sapo.

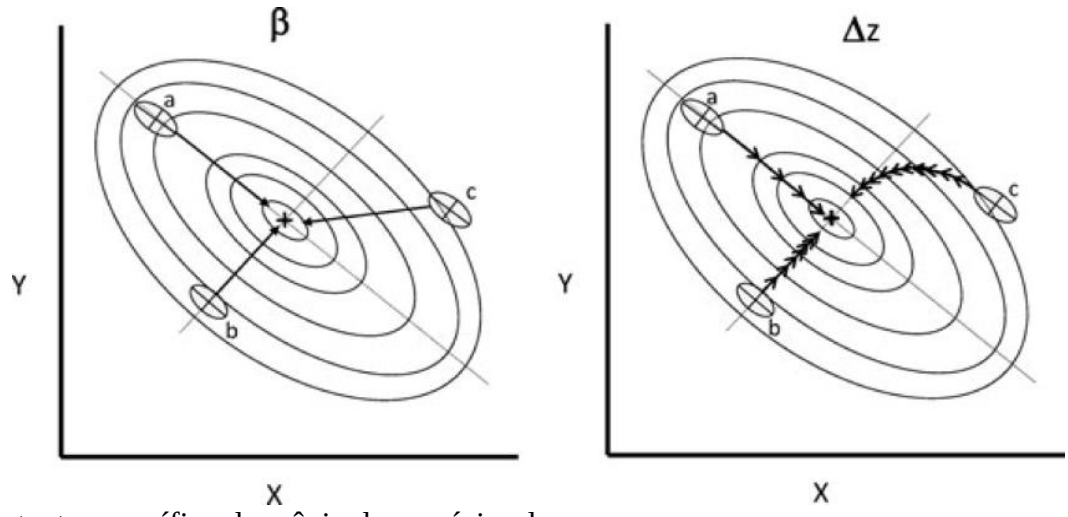


Figura 7. Interação entre populações e a paisagem adaptativa fenotípica. As isolinhas indicam regiões de mesma aptidão na paisagem. A forma da paisagem indica que os caracteres X e Y possuem uma covariação negativa em relação á aptidão. As populações estão representadas como matrizes genéticas em forma de elipse. Os eixos da elipse indicam os dois primeiros eixos de máxima variação genética. O painel à esquerda mostra a direção do gradiente de seleção. O painel à direita mostra a direção e a taxa da resposta evolutiva. Note que para a população c que não está alinhada com a paisagem, a resposta é curva (defletida da direção da seleção). Extraído de Marroig et al. (2009).

Seguindo a proposta de Waddington, a expectativa não é de pleiotropia global, na qual todos os genes influenciam todos os caracteres; mas de um padrão pleiotrópico modular. Um padrão pleiotrópico modular de efeitos genéticos sob o fenótipo corresponde a um *mapa genótipo-fenótipo modular*, como o apresentado por

Günter Wagner e Lee Altenberg em 1996. Segundo Wagner, três critérios precisam ser satisfeitos para um complexo de caracteres ser considerado como uma unidade modular: (1) servir a uma função primária; (2) ser integrado por efeitos pleiotrópicos; e (3) ser relativamente independente de outras unidades (Figura 8). Uma previsão de como a modularidade dos sistemas vivos poderia influenciar sua aptidão é de que uma organização modular pode impedir tanto o aparecimento de potenciais interações desvantajosas entre partes quanto a interferência entre sistemas funcionais distintos, acelerando a taxa de evolução e possibilitando o surgimento de adaptações complexas (Wagner e Altenberg 1996).

Evidências empíricas de uma organização modular da arquitetura genética foram encontradas em estudos de QTLs em elementos ósseos de ratos (Cheverud et al. 1997; Mezey et al. 2000; Cheverud et al. 2004; Kenney-Hunt et al. 2008; Wagner et al. 2008). Esses estudos mostraram que a maioria dos QTLs expressam seus efeitos em apenas alguns caracteres (pleiotropia restrita) e que o padrão de efeitos pleiotrópicos é correlacionado com o padrão de correlações fenotípicas dos caracteres. Isso significa que estudos de modularidade variacional (padrões de covariação no nível populacional; Cheverud 1996; Wagner et al. 2007) podem capturar a arquitetura genética subjacente aos efeitos fenotípicos (Kenney-Hunt et al. 2008).

Modularidade para além dos Mamíferos

Estudos do crânio de mamíferos forneceram evidências contundentes da estabilidade da matriz-P em um espectro bastante amplo de diversidade morfológica (Marroig and Cheverud 2001; Oliveira et al. 2009; Porto et al. 2009). Essa estabilidade ao longo do tempo evolutivo provavelmente ocorre pela manutenção de padrões de seleção estabilizadora comuns nos diversos grupos de mamíferos, relacionados com

desenvolivmento e função compartilhados, refletidos na estrutura modular comum do crânio (especialmente a divisão em face e neurocrânio; Porto et al. 2009). Diante da forte evidência de estabilidade da matriz-P em mamíferos, é interessante investigar se esse mesmo padrão ocorre em outros vertebrados. Recentemente, padrões de integração e hipóteses de modularidade foram medidos/testadas em outros grupos, como cobras (Vincent et al. 2006), salamandras (Ivanovic´ et al. 2005; Ivanovic´ and Kaleizic´ 2010), peixes (Cooper et al. 2011) e lagartos (Kolbe et al. 2011; Sanger et al. 2012; Fabre et al. 2014). As hipóteses são em sua maioria para o crânio e baseadas em aspectos funcionais (biomecânica) mais do que em informação sobre o desenvolvimento (Vincent et al. 2006; Cooper et al. 2011; Fabre et al. 2014). Em geral, esses estudos mostraram integração entre medidas do crânio as quais se supunha que executavam uma mesma função (Vincent et al. 2006; Cooper et al. 2011; Fabre et al. 2014) e algum suporte para modularidade do crânio (Ivanovic´ et al. 2005; Ivanovic´ and Kaleizic´ 2010).

Os dois estudos conduzidos com lagartos do gênero *Anolis* foram os únicos a comparar matrizes-P e hipóteses de modularidade em um maior número de espécies, oito no total (Kolbe et al. 2011; Sanger et al. 2012). Kolbe et al. (2011) reportou uma similaridade moderada a alta entre matrizes-P de caracteres da cabeça, membros e lamelas; e congruência em padrões modulares (apesar da modularidade do crânio não ter sido compartilhada por todas as species). Por outro lado, Sanger et al. (2012) encontraram uma menor similaridade entre matrizes-P e divergência em modularidade entre as espécies. Os autores sugeriram que o módulo rostral apareceu mais de uma vez nos lagartos e que a estabilidade de **P** talvez seja uma exceção em vez da regra (Sanger et al. 2012). Entretanto, uma comparação direta entre os dois estudos em *Anolis* é complicada pelo uso de diferentes métodos para estimar as

matrizes-P (distâncias lineares *versus* distâncias de Procrustes) e para testar modularidade (correlação de matrizes empíricas e teóricas *versus* coeficiente de **RV** de Klingenberg [2009]; que mede a força de covariação entre subconjuntos de landmarks). Ainda, a estabilidade da **P** foi acessada apenas em algumas espécies além de mamíferos. Portanto, consideramos que a estabilidade da **P** em outros vertebrados ainda é uma questão em aberto.

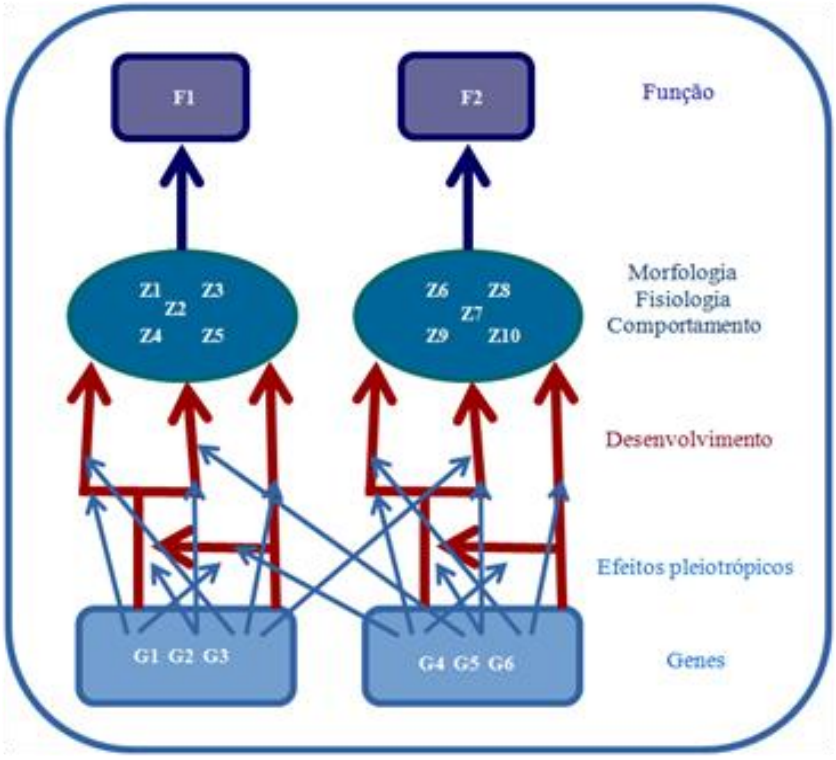


Figura 8. **Mapa genótipo-fenótipo modular.** O esquema mostra dois grupos de genes que apresentam efeitos pleiotrópicos sobre vias de desenvolvimento, que por sua vez originam os complexos de caracteres. Os caracteres dentro de cada complexo interagem para executar uma determinada função (F1 ou F2). Um conjunto de genes apresenta efeitos pleiotróicos praticamente restritos a um único complexo. Modificado de Wagner (1996).

Nessa tese exploramos a estabilidade da matriz-P e sua organização modular em espécies de sapos do grupo *Rhinella granulosa*. Os anfíbios anuros possuem um

desenvolvimento diferenciado em relação aos demais vertebrados, passando por um processo de grande remodelação: a metamorfose. O desenvolvimento distinto dos anfíbios anuros resultou em alterações na origem ontogenética dos ossos do crânio (Piekarski et al. 2014), implicando em expectativas distintas de modularidade em relação aos demais vertebrados (Capítulo 2). Além disso, anuros apresentam vastas modificações cranianas, como perda de elementos ósseos por fusionamento ou falha de desenvolvimento, havendo casos de aparecimento de novos ossos (neopalatinos, prenasais, esfenóide; Trueb 1993). As espécies do grupo *R. granulosa* tem como sinapomorfia a presença de prenasais e apresentam variação no crânio (Pramuk 2006; Narvaes e Rodrigues 2009). É um grupo composto de 13 espécies de sapos, de distribuição geográfica ampla e recentemente a relação filogenética entre as espécies foi esclarecida (Pereyra et al. 2015). Portanto, além do grupo ser interessante para o estudo da modularidade, também possuí os elementos fundamentais para uma investigação da evolução do fenótipo médio do crânio das espécies, associada a potenciais relações com variação climática, dado que são organismos ectotermos (Capítulo 3). Como esse trabalho foi realizado com espécimes de museu preservados em via úmida, optamos por escanear os indivíduos para ter acesso ao crânio. O protocolo de escaneamento junto da validação desse método para a obtenção de medidas do crânio está apresentado no Capítulo 1.

Objetivo Geral e Perguntas

O objetivo geral dessa tese é de investigar a evolução morfológica do crânio das espécies de sapo do grupo *Rhinella granulosa*. Especificamente, buscamos responder as seguintes perguntas:

1. A organização da (co)variação fenotípica dos caracteres do crânio é similar nas espécies de sapo?
2. A organização da (co)variação fenotípica segue um padrão modular conforme as expectativas de desenvolvimento ou de função compartilhados entre caracteres do crânio?
3. Fatores externos relacionados com diferenças filogenéticas, morfológicas médias e climáticas médias entre as espécies estruturam a similaridade na organização da (co)variação fenotípica do crânio?
4. A evolução do fenótipo médio dos caracteres do crânio se deu por deriva genética ou por seleção natural ao longo da filogenia das espécies?
5. Variação morfológica está associada à variação climática entre as espécies de sapo?
6. A divergância morfológica no crânio entre as espécies pode ser explicada pela interação entre seleção natural e restrições genéticas?

Referências

Arnold, S. J. (2014). Phenotypic Evolution: The Ongoing Synthesis: (American Society of Naturalists Address)*. *The American Naturalist*, *183*(6), 729–746.

Arnold, S. J. (1994). Multivariate Inheritance and Evolution: A Review of Concepts. In *Quantitative Genetic Studies of Behavioral Evolution* (ed: Boake, C. R. B., pp. 18–48). The University of Chicago Press.

Arnold, S. J., Pfrender, M. E., & Jones, A. G. (2001). The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica*, *112*(1), 9–32.

Arnold, S. J., & Wade, M. J. (1984). On the Measurement of Natural and Sexual Selection: Theory. *Evolution*, *38*(4), 709.

Berg, R. L. (1960). The ecological significance of correlation pleiades. *Evolution*, 171–180.

Bolstad, G. H., Hansen, T. F., Pelabon, C., Falahati-Anbaran, M., Perez-Barrales, R., & Armbruster, W. S. (2014). Genetic constraints predict evolutionary divergence in Dalechampia blossoms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 1-15.

Cheverud, J. M. (1984). Quantitative Genetics and Developmental Constraints on Evolution by Selection. *Journal of Theoretical Biology*, *110*, 155–171.

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. *The American Naturalist*, *145*(1), 63–89.

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*, *302B*(5), 424–435.

Cheverud, J. M., Routman, E. J., & Irschick, D. J. (1997). Pleiotropic Effects of Individual Gene Loci on Mandibular Morphology. *Evolution*, *51*(6), 2006.

Cooper, W. J., Wernle, J., Mann, K., & Albertson, R. C. (2011). Functional and Genetic Integration in the Skulls of Lake Malawi Cichlids. *Evolutionary Biology*, *38*(3), 316–334.

de Oliveira, F. B., Porto, A., & Marroig, G. (2009). Covariance structure in the skull of Catarrhini: a case of pattern stasis and magnitude evolution. *Journal of Human Evolution*, *56*(4), 417–430.

Eldredge, N. (2008). Some Thoughts on “Adaptive Peaks,” “Dobzhansky’s Dilemma”—and How to Think About Evolution. *Evolution: Education and Outreach*, *1*(3), 243–246.

Fabre, A.-C., Andrade, D. V., Huyghe, K., Cornette, R., & Herrel, A. (2014). Interrelationships Between Bones, Muscles, and Performance: Biting in the Lizard Tupinambis merianae. *Evolutionary Biology*, *41*(4), 518–527.

Falconer, D.S., & Mackay, T. (1996). *Introduction to Quantitative Genetics* (4th ed.). Longman Group.

Fisher, R. A. (1918). IV. The Correlation between Relativeson the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399-433.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection: a Complete Variorium Edition*. Oxford University Press.

Haldane, J. B. S. (1954). The Measurement of Natural Selection. *Caryologia*, 480–487.

Hansen, T. F., & Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, *21*(5), 1201–1219.

Hansen, T. F., & Voje, K. L. (2011). Deviation from the line of least resistance does not exclude genetic constraints: a comment on Berner et al. (2010). *Evolution*, *65*(6), 1821–1822.

Hill, W. G., & Mackay, T. F. (2004). DS Falconer and Introduction to quantitative genetics. *Genetics*, *167*(4), 1529–1536.

Ivanović, A., & Kalezić, M. L. (2010). Testing the hypothesis of morphological integration on a skull of a vertebrate with a biphasic life cycle: a case study of the alpine newt. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *314B*(7), 527–538.

Ivanović, A., Kalezić, M. L., & Aleksić, I. (2005). Morphological integration of cranium and postcranial skeleton during ontogeny of facultative paedomorphic European newts (Triturus vulgaris and T. alpestris). *Amphibia-Reptilia*, *26*(4), 485–495.

Johnson, N. (2008). Sewall Wright and the Development of Shifting Balance Theory. *Nature Education*, *1*, 52.

Kenney-Hunt, J. P., Wang, B., Norgard, E. A., Fawcett, G., Falk, D., Pletscher, L. S. & Cheverud, J. M. (2008). Pleiotropic Patterns of Quantitative Trait Loci for 70 Murine Skeletal Traits. *Genetics*, *178*(4), 2275–2288.

Klingenberg, C. P. 2009. Morphometric integration and modularity in configuration of landmarks: tools for evaluating a priori hypothesis. *Evolution & development*, 11, 405-421.

Kolbe, J. J., Revell, L. J., Szekely, B., Brodie III, E. D., & Losos, J. B. (2011). Convergent evolution of phenotypic integration and its alignment with morphological diversification in Caribbean ecomorphs. *Evolution*, *65*(12), 3608–3624.

Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution*, *30*(2), 314-334.

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*(1), 203–215.

Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution*, *37*(6), 1210-1226.

Lush, J. L. (1937). *Animal Breeding Plans*. Collegiate Press.

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits* (1st ed.).

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., Shirai, L. T., Porto, A., de Oliveira, F. B., & De Conto, V. (2009). The Evolution of Modularity in the Mammalian Skull II: Evolutionary Consequences. *Evolutionary Biology*, *36*(1), 136–148.

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*(1), 305–311.

Morrissey, M. B., Kruuk, L. E. B. & Wilson, A. J. (2010). The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, 23, 1-12.

Olson, E. C., & Miller, R. L. (1958). *Morphological Integration* (1st ed.). The University of Chicago Press.

Pavlicev, M., Kenney-Hunt, J. P., Norgard, E. A., Roseman, C. C., Wolf, J. B., & Cheverud, J. M. (2007). Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. *Evolution*, *62*(1), 199-213.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. XI. On the

influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 1–66.

Pereyra, M. O., Baldo, D., Blotto, B. L., Iglesias, P. P., Thomé, M. T. C., Haddad, C. F. B. & Faivovich, J. (2015). Phylogenetic relationships of toads of the *Rhinella granulosa* group (Anura: Bufonidae): a molecular perspective with comments on hybridization and introgression. *Cladistics*, 31, 1-18.

Phillips, P. C., & Arnold, S. J. (1989). Visualizing Multivariate Selection. *Evolution*, *43*(6), 1209–1222.

Piekarski, N., Gross, J. B., & Hanken, J. (2014). Evolutionary innovation and conservation in the embryonic derivation of the vertebrate skull. *Nature Communications*, *5*, 5661.

Porto, A., de 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.

Price, G. R. (1970). Selection and Covariance. *Nature*, *227*, 520–521.

Reed, D. R., Lawler, M. P. & Tordoff, M. G. (2008). Reduced body weight is a common effect of gene knockout mice. *BMC Genetics*, 9, 1-6.

Reusch, T., & Blanckenhorn, W. U. (1998). Quantitative genetics of the dung fly Sepsis cynipsea: Cheverud's conjecture revisited. *Heredity*, *81*, 111–119.

Ridley, M. (2004). *Evolution* (3rd ed.). Blackwell Publishing.

Riedl, R. (1977). A Systems-Analytical Approach to Macro-Evolutionary Phenomena. *The Quartely Review of Biology*, *52*(4), 351–370.

Schluter, D. (1996). Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution*, *50*(5), 1766-1774.

Simpson, G. G. (1944). *Tempo and Mode in Evolution* (1st ed.). Columbia University Press.

Steppan, S. J. (1997). Phylogenetic Analysis of Phenotypic Covariance Structure. II. Reconstructing Matrix Evolution. *Evolution*, *51*(2), 587-594.

Steppan, S. J., Phillips, P. C., & Houle, D. (2002). Comparative quantitative genetics: evolution of the G matrix. *Trends in Ecology & Evolution*, *17*(7), 320–327.

Svensson, E. I., & Calsbeek, R. (2012). *The Adaptive Landscape in Evolutionary Biology* (1st ed.). Oxford University Press.

Vincent, S. E., Dang, P. D., Herrel, A., & Kley, N. J. (2006). Morphological integration and adaptation in the snake feeding system: a comparative phylogenetic study. *Journal of Evolutionary Biology*, *19*(5), 1545–1554.

Waddington, C. H. (1957). *The Strategy of Genes* (1st ed.). Routledge Press.

Wagner, G. P. (1996). Homologues, natural kinds and the evolution of modularity. *American Zoologist*, *36*(1), 36–43.

Wagner, G. P., & Altenberg, L. (1996). Perspective: Complex Adaptations and the Evolution of Evolvability. *Evolution*, *50*(3), 967-976.

Wagner, G. P., Kenney-Hunt, J. P., Pavlicev, M., Peck, J. R., Waxman, D., & Cheverud, J. M. (2008). Pleiotropic scaling of gene effects and the "cost of complexity." *Nature*, *452*(7186), 470–472.

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, *8*(12), 921–931.

Walsh, B., & Blows, M. W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 41–59.

Wright, S. (1932). The Roles of Mutation, Inbreeding, Crossbreeding ans Selection in Evolution. *Proceedings of The Sixth International Congress Og Genetics*, *1*, 356–366.

Wright, S. (1980). Genic and Organismic Selection. *Evolution*, *34*(5), 825.

Capítulo 1

Landmark Precision and Reliability and Accuracy of Linear Distances estimated by using 3D Computed Micro-tomography

1.1. Abstract

The wider availability of non-destructive and high-resolution methods, such as micro-computed tomography (micro-CT), has prompted its use in anatomical and morphometric studies. Yet, because of the actual scanning procedure and the processing of CT data by software that renders 3D surfaces or volumes, systematic errors might be introduced in placing landmarks as well as in estimating linear distances. Here we assess landmark precision and measurement reliability and accuracy of using micro-CT images of toad skulls and the TINA Manual Landmarking Tool software to place 20 landmarks and extract 24 linear distances. Landmark precision and linear distances calculated from 3D images were compared to the same landmarks and distances obtained with a 3D digitizer in the same skulls. We also compared landmarks and linear distances in 3D images of the same individuals scanned with distinct filters, since we detected variation in bone thickness or density among the individuals used. We show that landmark precision is higher for micro-CT than for the 3D digitizer. Distance reliability was very high within-methods, but decreased in 20% when 3D digitizer and micro-CT data were joined together. Still, we did not find any systematic bias in estimating linear distances with the micro-CT data and the between-methods errors were similar for all distances (around 0.25 mm). We conclude that using micro-CT data for morphometric analysis results in acceptable landmark precision and similar estimates of most linear distances compared to 3D digitizer, although some distances are more prone to discrepancies between-methods. Yet, caution in relation to the scale of the measurements needs to be taken, since the proportional between-method error is higher for smaller distances. Scanning with distinct filters does not introduce a high level of error and is recommended when individuals differ in bone density.

1.2. Introduction

The increasing use of non-destructive and high-resolution data acquisition methods, such as micro-computed tomography (micro-CT), have provided researchers with the opportunity to study the anatomy and morphology of organisms with more detail and at a wider phylogenetic spectrum (e.g. Constantini et al. 2010, Ekdale 2010, Wilkinson et al. 2011, Cuff and Rayfield 2013, Gignac and Kley 2014). Accordingly, 3D image processing software has been developed (e.g. OsiriX: Rosset et al. 2004, Amira: www.amira.com), with some designed to place 3D landmarks for shape or morphometric analysis (e.g. TINA Manual Landmarking Tool: Schunke et al. 2012). However, there is no guarantee that the scanning procedure and the software used to process CT data, by creating 3D surface or 3D volume renderings, do not introduce systematic errors in the data (Kohn and Cheverud 1992; Kim et al. 2012). In addition, the landmark positioning process in the 3D images might also introduce systematic and random errors in the estimation of linear distances (measurements). Thus, the precision of placing landmarks in 3D images with software, as well as the accuracy and reliability of the distances taken by the use of CT data must be validated (Kohn and Cheverud 1992; Kim et al. 2012; Halperin-Sternfeld et al. 2014).

We evaluated the precision of placing the same landmarks in the same individuals of a toad species with two distinct methods: (1) the real skulls and a 3D digitizer to place the landmarks, and (2) 3D images of these skulls, obtained by micro-CT scans at two distinct resolutions, and software to place the landmarks (TINA Manual Landmarking Tool (Schunke et al. 2012; hereafter called TINA-Landmark). The 3D digitizer is an articulated arm that creates a 3D coordinate system in which any point of an object can be identified in relation to a reference point. TINA-Landmark is recently developed open-source software created to enhance the

precision of landmark positioning in 3D images by using volume rendering instead of surface rendering and by showing the cross-section images connected to the 3D volume; Schunke et al. 2012).We chose to compare the landmarks obtained from the 3D images with the ones obtained by the 3D digitizer because, for the latter, the landmarks are taken in the actual skulls, with no processing of data, not the case for the construction of 3D images. Also, in zoological studies, several authors measure the specimens with 3D digitizers (e.g. Zelditch 1988; Cheverud 1995; Young and Hallgrímsson 2005; Porto et al. 2009; just to cite a few), being a widely accepted technique in the morphometry field.

In addition to comparing the landmarks between the two methods, it is also of interest to evaluate the consequence of potential biases in the estimation of linear distances extracted from the landmarks. Although measurement error is intrinsic to the landmarks and independent of the linear distances computed from them, the proportion of error varies with the distance length. That is, if the error in placing landmarks is the same for all landmarks, the proportion of error will be higher for smaller distances than for longer distances (Corner et al. 1992). On the other hand, if the error in placing landmarks varies with landmark type or position in the material, linear distances extracted from these more variable landmarks are expected to have greater error when comparing distinct methods. Therefore, we also calculated reliability and accuracy of linear distances extracted from the landmarks placed with the 3D digitizer and in the 3D images.

Finally, we also compared the same toad specimens scanned with two distinct filters, since we discovered that some individuals differ in bone thickness and/or density. Filters are thin sheets of metal set in front of the material being scanned and can have different thickness. Varying the filter thickness has an effect on the mean X-

ray energy irradiating on the material being scanned. For thinner bones, a lower X-ray energy is necessary to achieve the best 3D volumes and more precise positioning of the landmarks (see Figure 1). Thus, changing filter type might be an additional source of error when placing landmarks and taking linear measurements in 3D images. We consider this last comparison quite relevant in zoological studies because other organisms might present the same variation in bone density and scanning with distinct filters will be indispensable.

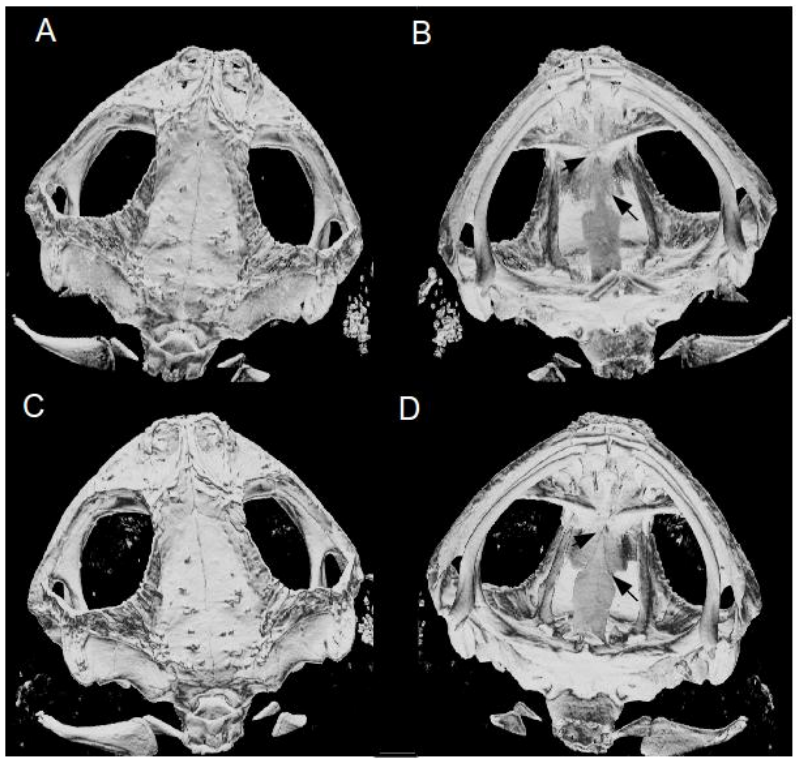


Figure 1. Skull 3D images of a *Rhinella pygmaea* specimen scanned with distinct filters. Several *R. pygmaea* specimens presented transparency in some bones, such as the squamosal, sphenethmoid and the parasphenoid bones, making the precise determination of sutures between these bones difficult (indicated by the arrows). Scanning with a thinner filter corrects for this problem as can be seen in **C** and **D**. **A)** Dorsal view of a skull scanned with an aluminum filter 1.0 mm (AL1.0); **B)** Ventral view of the same skull scanned with AL1.0; **C)** Dorsal view of the same skull scanned with an aluminum filter 0.5 mm (AL0.5); **D)** Ventral view of the same skull scanned with AL0.5.

Although there is no specific theory relating the potential effects on landmarks and on distances when using distinct resolutions, filters and reconstruction algorithms, some expectations based on landmark position and bone thickness can be made. Some of the landmarks that we selected in the toad skulls (Table 1 and Figure 2) were more difficult to visualize in the 3D images than in the real skulls because of their position (landmarks 4, 10, 11, 17 and 20) or because they were placed at thinner bones (landmarks 6, 7, 8 and 13). Thus, we expect more variation in the positioning of these landmarks among methods, and as a consequence, less reliability and accuracy of the linear distances extracted from them (Table 2 and Figure 2).

1.3. Methods

1.3.1. Species and Scanning Procedures

For the micro-CT and 3D digitizer comparisons, as well as for the 3D image compacting factor analysis, we used 20 adult individuals of the toad species *Rhinella granulosa*, collected in January 2012 at a site near the municipality of Angicos (5^o 39´ S, 36^o 36´W), in the state of Rio Grande do Norte, NE Brazil. The toads were sacrificed by peritoneal application of an anesthetic in excess and then were fixed in 70% alcohol. We scanned the toad skulls with an X-ray micro-CT system (SkyScan 1176, Konitch, Belgium) placed at the Instituto de Biociências, Universidade de São Paulo. All individuals were scanned with a 1.0 mm AL filter at two different resolutions: medium (**MED**: 18 μm, 70 kV, 356 μA) and high (**HIGH**: 9 μm, 65 kV, 380 μA). These resolutions correspond to different voxel sizes, the smallest volume unit in the 3D volumes: $3.85\ 10^{-6}$ mm^3 for MED and $1.92\ 10^{-6}$ mm^3 for HIGH, which are much smaller than conventional CTs (voxel sizes ranging from 0.1 to 5 mm^3). Before scanning, toads were wrapped with Parafilm to avoid too much alcohol

evaporation and consequent dehydration, which could lead to blurred images by sample contraction. The scanning time per skull for MED was 20 min. and for HIGH was 1.5 h. After scanning, the skulls were reconstructed using NRecon software (SkyScan, Konitch, Belgium). In this process, 2D projection images are reconstructed to cross-section images by use of a mathematical algorithm (Feldkamp). The first step in the reconstruction process was to choose the lower and upper limits of the threshold for the linear attenuation coefficient (AC). The AC measures how much the intensity of the X-ray beam is reduced as it passes through the material being scanned, and is related to the density of the materials. Based on the histograms of soft tissue and bone density of the skulls, we chose an AC threshold of 0.0 and 0.05 for all specimens scanned. Afterwards, we applied different types of corrections in the reconstruction process to soften some undesirable effects in the images: post-alignment = -5.0 to 2.0; ring artifact reduction = 2.0 to 4.0; beam-hardening correction = 30% and smoothing = 2.0. These corrections are important to avoid blurring and artifacts in the 3D images and for all corrections the values used were small compared to maximum values. Therefore, we do not expect these corrections to interfere in landmark visualization, but instead to improve its identification. The sequences of cross-section images varied from 400 to 600 images for MED (500 MB to 1 GB per skull file size) and from 800 to 1,200 images for HIGH (1 to 5 GB per skull file size) and were all in BMP extension.

For the filter type analysis, we used 20 adult museum specimens of *Rhinella pygmaea* (MZUSP, São Paulo, Brazil). This species was chosen because several specimens presented too much transparency or even holes in their skulls, especially in the squamosal, sphenethmoid and parasphenoid bones, preventing the placement of some landmarks and also suggesting that there were differences in bone density

among individuals . We scanned the 20 toads at medium resolution using two distinct filters: 1.0 mm AL (**AL1.0**: 70 kV, 356 μA) and 0.5 mm AL (**AL0.5**: 50 kV, 500 μA). Filters are thin metal sheets that are set in front of the X-ray source and can have different thickness. Filtration retains a part of the low energy photons of the X-ray, thus increasing mean X-ray energy. The reduction in AL filter thickness from 1.0 mm to 0.5 mm results in a lower mean energy of the X-ray because photons with lower energy traverse the filter and achieve the thinner bones, being retained on them (SkyScan manual). The individuals used in this analysis presented low transparency and no holes since the objective was to evaluate whether the use of distinct filters changed the landmarks positions and the linear distances obtained in the same individuals. In these toads all landmarks could be placed, which would not be the case if we had used specimens with holes or high transparency. CT system, scanning and reconstruction procedures were the same as described above.

1.3.2. Landmarking Procedure and Linear Distances in 3D Images

We placed 20 landmarks at bone sutures (type I landmarks) or bone processes (type II landmarks; 35 landmarks if counting both sides of the skull, left and right) in all views (dorsal, ventral and lateral; Table 1 and Figure 2) using TINA-Landmark software (Schunke et al. 2012). This software provides different views of the data: a 3D volume of the whole skull and three 2D views of the orthogonal cross-sections (axial, sagittal and transversal) throughout the image sequence, enabling landmarks to be precisely placed given that the views are linked (*i.e*. when setting a landmark in the 3D volume, it also appears in all 2D views allowing a refinement of the landmark position). The skull's cross-section sequences were loaded in TINA-Landmark by using the "Sequence Tool" after the BMP files were converted to DICOM files. The

Sequence Tool presents two types of image compacting factors for loading heavy files: “Stride”, which reduces the sequence in the inter-slice direction (*i.e.*, stride = 2.0 and “Stride average” = ON means that an average image will be loaded at every two cross-sections from the sequence, resulting in half of the original size of the sequence file); and “Down-sample”, which reduces the sequence along the *x* and *y* directions (*i.e.*, down-samples the size of the pixels; Bromiley et al. 2012). MED resolution data were loaded using stride = 2.0 whereas HIGH resolution data were loaded using stride = 3.0 and down-sample = 2.0, since they were heavier and failed to load with the same compacting factor than MED data. The down-sampling procedure only affects the visualization of the 3D volume of the skulls, not changing the visualization of the cross-section images (axial, sagittal and transversal views).

Appropriate visualization of 3D volumes and determination of bone threshold value (the average density of soft tissue and bone to guarantee that landmarks are placed exactly where the mouse is pointing in the 3D image) were obtained by following the recommendations contained in the TINA Geometrics Morphometrics Toolkit manual (Bromiley et al. 2012). Finally, after all steps of image adjustments, a landmark list was loaded with the 35 points and each was placed two times (two replicates per individual) in the skull’s images of *R. granulosa* and *R. pygmea* individuals, so we could assess measuring reliability. The corresponding *x*, *y* and *z* landmark coordinates were saved in a TXT file, which was later loaded in the R programming environment (R Core Team 2013) where distances were calculated.

We determined 24 linear distances in the skull of the toads, all distances representing individual bone dimensions (Table 2, except distance 7 which corresponds to orbit size). To obtain distances in mm, the coordinates of MED and HIGH were multiplied by 0.01742 and 0.00871, respectively, which are the size of

the pixels in mm for each resolution. In the case of the *R. pygmea* specimens that were scanned with different filters, landmark coordinates were multiplied by 0.01742 because they were only scanned at medium resolution.

Table 1. Landmark descriptions in the toad skulls. Landmarks are intersections between bone sutures (type I landmarks, 16 in total) or tip of bones (type II landmarks: numbers 1, 14, 15 and 19). Five landmarks are in the medial line and the remaining landmarks are present in both sides of the skull. The landmarks are spread in all three views of the skull: dorsal, lateral and ventral (see Fig. 2). We placed all 20 landmarks with all the methods twice in each individual.

Landmarks	Description	Position	View
1	Anterior tip of nasal bone	midline	dorsal
2	Nasal and frontoparietal suture	midline	dorsal
3	Posterior tip of frontoparietal suture	midline	dorsal
4	Nasal and maxillary suture	right, left	dorsal
5	Nasal and frontoparietal lateral suture	right, left	dorsal
6	Frontoparietal and squamosal suture	right, left	dorsal
7	Frontoparietal, squamosal and occipital suture	right, left	dorsal
8	Squamosal and occipital suture	right, left	dorsal
9	Frontoparietal and occipital suture	right, left	dorsal
10	Prenasal and maxillary suture	right, left	lateral
11	Nasal and maxillary lateral suture	right, left	lateral
12	Squamosal and maxillary suture	right, left	lateral
13	Sphenethmoid and parasphenoid suture	midline	ventral
14	Posterior tip of parasphenoid corpus	midline	ventral
15	Anterior tip of premaxillary bone	right, left	ventral
16	Premaxillary and maxillary suture	right, left	ventral
17	Pterygoid and maxillary suture	right, left	ventral
18	Neopalatine and sphenethmoid suture	right, left	ventral
19	Tip of pterygoid process	right, left	ventral
20	Pterygoid and parasphenoid suture	right, left	ventral

Distances from replicates of the same individual (within each method) were inspected for gross measurement error (difference between replicates above 0.5 mm for most distances, except for distances 5, 8, 10, 11, 13, 17 and 24, which have means smaller than 3.0 mm, and were controlled for error above 0.3 mm), and when detected, the landmarks correspondent were placed again in both replicates and corrected. The reference to consider the magnitude of a gross error was based on the precision of the 3D digitizer, which is 0.01 mm. This procedure was adopted because these gross errors could lead to misleading conclusions about the between-methods analysis, since they are actually referred to gross human error inside each method.

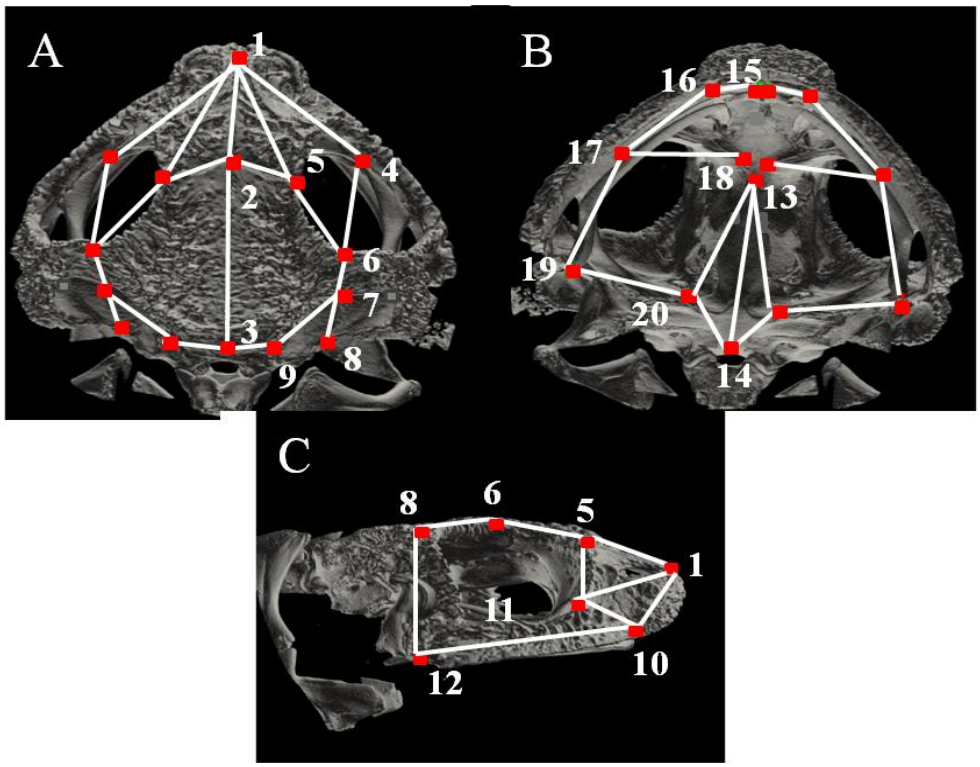


Figure 2. Landmarks and linear distances obtained from the toad skulls. Numbered landmarks in both sides of the skull are shown as red dots in dorsal and ventral views, and only landmarks of the right side of the skull are shown in the lateral view (descriptions in Table 1). Landmarks were placed in bone sutures or bone processes either with TINA-Landmark software in 3D images or with a 3D digitizer in the real skulls. Linear distances are shown as white lines and represent individual bone dimensions, as shown in Table 2.

Table 2. Linear distances determined in the toad skulls. In total, we determined 24 linear distances representing individual dimensions of the bones in the toads' skulls. Distances are spread through the skull in three views (dorsal, ventral and lateral; see Fig.1). We positioned landmarks in both sides of the skull and averaged the distances from both sides.

Distances	Landmarks	Bones
1	1-2	nasal
2	2-3	frontoparietal
3	1-4	nasal
4	1-5	nasal
5	2-5	frontoparietal
6	5-6	frontoparietal
7	4-6	orbit
8	6-8	squamosal
9	7-9	occipital
10	3-9	frontoparietal
11	1-10	prenasal
12	1-11	nasal
13	10-11	nasal
14	5-11	nasal
15	10-12	maxilla
16	8-12	squamosal
17	13-14	parasphenoid
18	13-20	parasphenoid
19	15-16	premaxilla
20	16-17	nasal
21	17-18	neopalatine
22	17-19	pterygoid
23	19-20	pterygoid
24	14-20	parasphenoid

In order to analyze the possible effect on the landmarks and distances of switching the compacting factors (stride and downsample) when loading the cross-section sequences, we placed the landmarks in the *R. granulosa* skulls scanned in medium resolution using the following values in TINA-Landmark: (**S2**) stride = 2.0; (**S3**) stride = 3.0; (**S2/D2**) stride = 2.0 and down-sample = 2.0; and (**S3/D2**) stride = 3.0 and down-sample = 2.0. Again, each individual at each of these four compacting

values were measured twice and gross measurement error between replicates inside each compacting factor was controlled. Thus, total sample size for the compacting factor analysis was 20 individuals for each of the four situations, and 80 in total.

1.3.3. Landmarking Procedure and Linear Distances using 3D Digitizer

The same *R. granulosa* adults that we scanned with MED and HIGH resolutions were cleaned and had their skin removed. In this process, we lost four of the skulls by crushing, 16 skulls remaining. We placed the same landmarks described above in the cleaned skulls by using a 3D digitizer (Microscribe 3DX, IL). Since the toad skulls are quite small (total length around 20.0 mm), we used a binocular loupe to mark the landmarks with a pencil. Then, we placed the digitizer pen in the graphite marks to digitize the landmarks. We placed landmarks in all individuals twice and gross error in the distances from one replicate to the other was also controlled as described before at the time of landmark digitalization. A TXT file was created for the coordinates obtained from DIG and was loaded in R environment in order to calculate the same distances as MED and HIGH. The landmarks and distances obtained by using DIG were considered as the reference for comparisons.

1.3.4. Within-Method Mean Individual Landmark Distances

To assess the differences of placing landmarks in the skull 3D volumes compared to placing them in the real skulls with the 3D digitizer, we used a superimposition method, the Generalized Procrustes Analysis (GPA), to estimate the sample mean landmarks (mean shape) and to calculate individual distances from the mean landmarks within each method. The GPA sumperimposes landmark configurations of several individuals, scaling them to unit centroid size, and uses

least-square estimates for translation and rotation parameters (Bookstein 1990). These transformations of the landmarks are needed because there is no natural coordinate system that is common to all individuals that are digitized and a common shape space is achieved when doing this procedure (Bookstein 1990; Lele 1993). However, this transformation process confounds variation at different homologous landmarks, i.e., the most variable landmarks among individuals (non-isotropic variation) have their variation spread across other landmarks (Richtsmeier et al. 1995; Lele 1993; van der Linde and Houle 2009). Thus, to avoid this problem in our landmark variation analysis, we followed the idea presented in van der Linde and Houle (2009) and excluded from the GPA one landmark at a time to estimate the sample mean shape and the rotation matrices for the 19 remaining landmarks for each individual. By doing that, we precluded that the variation in the landmark of interest got spread through the other landmarks. Afterwards, we multiplied the excluded landmark of each individual by its corresponding rotation matrix (all landmarks of an individual are rotated in the same angle) to have all landmarks in the same new coordinate system, the so called shape space. The last step was to calculate the individual distance of the landmark of interest from the mean landmark (in the shape space, therefore with no scale). This procedure was repeated 20 times for each method, so that all 20 landmarks were excluded at each time and the distances from the individuals and the mean landmark could be calculated. To compare the different methods, we also calculated the mean individual landmark distance from the mean landmarks. Finally, in order to compare the methods of placing landmarks in a more intuitive scale, we multiplied the mean deviations and SD values by its corresponding mean sample centroid size in mm. Only the medial and left side landmarks were used in this analysis and the GPA was done with the "geomorph" (Adams and Otarola-

Castillo 2013) and “shapes” (Dryden 2014) packages in R environment.

1.3.5. Distance Repeatability Within and Between Methods

We calculated repeatability values for all distances in every data subset using the calculations described in Lessells and Boag (1987). Each individual is a group with two replicates and the between-group variance is the sum of squares of the deviations of the group means from the total mean, whereas the within-group variance corresponds to the sum of squares of the deviations of each replicate from its own group mean. The index is calculated as follows:

$$r = \frac{s_A^2}{(s^2 + s_A^2)}$$

s^2_A being the variance among groups and s^2 the residual variance, both calculated from the sum of squares in an ANOVA.

We calculated repeatabilities for MED, HIGH and DIG separately, and also for joined data sets, (MED + HIGH), (MED + DIG), (HIGH + DIG), with the three between-method data sets made up of four replicates per individual. In the same manner, we calculated repeatability for replicates of the different values of stride and down-sample, S2, S3, S2/D2 and S3/D2 separately; and also of joined data sets, (S2 + S3), (S2 + S2/D2), (S2 + S3/D2), (S3 + S2/D2), (S3 + S3/D2) and (S2/D2 + S3/D2). Finally, we did the same calculations for the different filters data sets separately, AL1.0 and AL0.5, and also joining all the replicates (AL1.0 + AL0.5). The distance repeatabilities for separated data sets indicate the within-method reliability of measuring the skulls, whereas the repeatabilities for joined data sets comprise both within and between-method reliability.

1.3.6. Raw and Absolute Differences Within and Between Methods

In addition to the repeatability analysis, we also calculated raw and absolute differences of the distances between replicates, within and between methods. In the first case, we just subtracted the distances between replicates of the same individual and calculated the mean raw and absolute differences across all individuals for each distance. These calculations were made for each data set separately: MED, HIGH, DIG, S2, S3, S2/D2, S3/D3, AL1.0 and AL0.5. The between-method analysis was done by subtracting the distances between the same individuals measured with different methods (*e.g.*, in the case of MED-HIGH, distances from individual 1 of MED were subtracted from the distances of individual 1 of HIGH, and so on for all 20 individuals) and calculating the raw and absolute mean differences across all individuals for each distance. All the differences that involved DIG (MED-DIG and HIGH-DIG) had 16 (4 skulls were lost in the cleaning procedure), while the rest of the differences between methods had 20 individuals. The raw differences show whether any method presents a consistent bias (e.g. under or overestimating the distances) in the data. The absolute differences indicate the magnitude of the error within and between methods, independent if the errors in the distances are in one direction or another. To get an idea of the error in relation to the distances means, we calculated the percentage error by dividing the absolute mean differences by their corresponding means and multiplying by 100.

1.3.7. Statistical Analysis

Correlation tests were all done using Pearson product moment correlation and significance level of 0.05. All ANOVAs in the repeatability analysis, correlation tests and graphics were done in the R environment (R Core Team 2013).

1.4. Results and Discussion

1.4.1. Landmark Precision with Distinct Methods

Mean individual landmark distances from the mean sample landmarks were higher when using the 3D digitizer (DIG) to place the landmarks than when using the 3D images, regardless of the resolutions (medium [MED] or high [HIGH]). This result holds for the shape space (without scale; Table 3), as well as for retaining landmarks' scale information in mm (after multiplying the Procrustes configuration by mean centroid size; Table 4), indicating that placing landmarks in 3D images obtained by the micro-CT is more precise. Even though the precisions of both equipments used in this study are similar (3D digitizer: 0.01 mm; micro-CT resolutions MED: 0.018 mm and HIGH: 0.009 mm), this result may be due to the fact that the 3D images of the skulls are much bigger in the computer screen than the real ones (around tenfold increase), facilitating visualization of several bone structures. The range of deviations of individual landmarks from the mean shape landmarks for all methods (DIG: 0.27 to 0.52 mm; HIGH: 0.14 to 0.33 mm and MED: 0.16 to 0.34 mm) are comparable to the error found by Richtsmeier et al. (1995) when placing the same landmarks in two different 3D images of the same individual (scanned twice with a 1.5 mm slice thickness resolution), which was 0.15 to 0.48 mm. Thus, the precision of both methods in placing landmarks is acceptable for morphometric studies, at least for distances as large as the ones obtained in this study.

The two landmarks with highest mean deviations for DIG (14: posterior tip of parasphenoid process and 19: posterior tip of pterygoid process) were not expected to have higher error, however they both are type II landmarks, suggesting that locating the exact maximum bone curvature was more difficult than locating bone sutures (type I landmarks). The higher error for type II landmarks compared to type I

landmarks is expected when working with the real skulls (Zelditch et al. 2004).

Table 3. Within-methods mean individual landmark distance without scale. The table shows the mean individual landmark distances from the mean landmark. We used Generalized Procrustes Analysis (GPA) to superimpose individual landmarks, but we avoided the spread of variation from any one landmark to the others (see text). The values refer to a shape space and do not have scale. Values in bold are the highest deviations for DIG compared to MED and HIGH.

Mean									
Landmarks	**DIG**	**HIGH**	**MED**	**AL1.0**	**AL0.5**	**S2**	**S2/D2**	**S3**	**S3/D2**
1	0.0135	0.0104	0.0100	0.0116	0.0095	0.0109	0.0106	0.0102	0.0111
2	0.0128	0.0097	0.0111	0.0072	0.0072	0.0110	0.0109	0.0100	0.0115
3	**0.0161**	0.0086	0.0083	0.0105	0.0095	0.0085	0.0089	0.0085	0.0080
4	0.0131	0.0093	0.0115	0.0090	0.0102	0.0125	0.0121	0.0125	0.0125
5	0.0164	0.0112	0.0106	0.0112	0.0182	0.0107	0.0111	0.0115	0.0106
6	**0.0165**	0.0085	0.0093	0.0111	0.0127	0.0098	0.0089	0.0092	0.0099
7	**0.0156**	0.0081	0.0082	0.0074	0.0064	0.0076	0.0078	0.0079	0.0075
8	0.0152	0.0147	0.0122	0.0125	0.0161	0.0130	0.0129	0.0139	0.0136
9	**0.0213**	0.0098	0.0100	0.0117	0.0124	0.0105	0.0096	0.0108	0.0098
10	**0.0172**	0.0083	0.0082	0.0082	0.0089	0.0081	0.0079	0.0084	0.0085
11	0.0137	0.0083	0.0087	0.0078	0.0069	0.0089	0.0085	0.0084	0.0069
12	0.0196	0.0129	0.0138	0.0113	0.0109	0.0137	0.0133	0.0133	0.0124
13	0.0154	0.0126	0.0136	0.0146	0.0164	0.0124	0.0121	0.0126	0.0128
14	**0.0226**	0.0103	0.0094	0.0083	0.0079	0.0091	0.0094	0.0102	0.0102
15	**0.0154**	0.0064	0.0069	0.0072	0.0086	0.0068	0.0065	0.0067	0.0064
16	**0.0114**	0.0063	0.0064	0.0070	0.0077	0.0077	0.0060	0.0068	0.0067
17	0.0120	0.0079	0.0092	0.0094	0.0074	0.0088	0.0087	0.0086	0.0091
18	0.0160	0.0103	0.0106	0.0086	0.0103	0.0104	0.0120	0.0110	0.0104
19	**0.0233**	0.0099	0.0115	0.0122	0.0123	0.0102	0.0100	0.0097	0.0100
20	**0.0217**	0.0090	0.0080	0.0096	0.0097	0.0089	0.0093	0.0089	0.0087

Table 4. Within-methods mean individual landmark distances in mm. The table shows the mean deviation of individual landmarks from the mean landmark. We used Generalized Procrustes Analysis (GPA) to superimpose individual landmarks, but we avoided the spread of variation from any one landmark to the others (see text). To get back to a scale in mm, all the mean deviation values were multiplied by the mean centroid size of the correspondent sample. Values in bold are the highest deviations for DIG compared to MED and HIGH or for AL0.5 compared to AL1.0.

Mean in mm									
Landmarks	**DIG**	**HIGH**	**MED**	**AL1.0**	**AL0.5**	**S2**	**S2/D2**	**S3**	**S3/D2**
1	0.299	0.227	0.219	0.267	0.221	0.234	0.227	0.220	0.239
2	0.284	0.212	0.242	0.165	0.168	0.236	0.234	0.215	0.246
3	**0.357**	0.188	0.182	0.243	0.222	0.182	0.191	0.184	0.172
4	0.290	0.203	0.250	0.208	0.238	0.270	0.261	0.270	0.268
5	0.363	0.245	0.232	0.259	**0.422**	0.230	0.239	0.248	0.227
6	**0.365**	0.187	0.203	0.256	0.296	0.210	0.191	0.199	0.213
7	**0.345**	0.177	0.179	0.170	0.149	0.163	0.167	0.171	0.160
8	0.338	0.323	0.266	0.290	0.374	0.280	0.277	0.300	0.292
9	**0.473**	0.215	0.219	0.271	0.289	0.227	0.207	0.233	0.210
10	**0.380**	0.182	0.179	0.190	0.207	0.174	0.171	0.181	0.182
11	**0.303**	0.182	0.191	0.180	0.160	0.191	0.183	0.181	0.148
12	**0.435**	0.282	0.301	0.262	0.254	0.296	0.285	0.287	0.266
13	0.341	0.275	0.298	0.337	0.383	0.267	0.259	0.271	0.275
14	**0.500**	0.226	0.205	0.192	0.183	0.195	0.202	0.220	0.219
15	**0.342**	0.141	0.152	0.166	0.200	0.147	0.140	0.144	0.137
16	**0.252**	0.137	0.140	0.161	0.179	0.167	0.128	0.147	0.144
17	0.266	0.173	0.201	0.218	0.171	0.189	0.186	0.185	0.196
18	0.355	0.226	0.232	0.199	0.240	0.224	0.257	0.236	0.224
19	**0.516**	0.216	0.252	0.282	0.286	0.220	0.214	0.209	0.214
20	**0.481**	0.196	0.175	0.222	0.226	0.192	0.199	0.191	0.187

Landmarks 3, 5, 6, 7, 9, 10, 11, 12, 15, 16 and 20 presented higher discrepancies in the landmarks for DIG when compared to the micro-CT resolutions (boldface in Table 4). For the landmarks expected to have higher error in the 3D images because of their positions in the skull or because they were located in thinner

bones, only landmarks 4, 8 and 12 corresponded to the expectation. All of them are type I landmarks, agreeing with the findings of other authors working with 3D images of more error in placing landmarks at some particular bone sutures (Richard et al. 2014; Stull et al. 2014).

MED and HIGH presented very similar results, indicating that there is no difference in switching the resolutions to visualize the landmarks. The fact that we had to down-sample the HIGH data by a factor of two to load the cross-section sequences in TINA-Landmark probably did not interfere in landmark precision, since switching the compacting factors with the MED data also produced similar results (comparisons between stride = 2 [S2] and stride = 2 and down-sample = 2 [S2/D2]; and stride = 3 [S3] and stride = 3 and down-sample = 2 [S3/D2]; Table 4). This result is consistent with the fact that the cross-sections are maintained in the original resolution even when down-sampling the data, being possible to refine the landmarks positions in them. Finally, for the filter type analysis, landmarks 6, 8 and 13 presented higher mean individual landmark distances for aluminum 0.5 mm (AL0.5) than for aluminum 1.0 mm (AL1.0), coinciding with the ones expected as more variable for being located at thinner bones. The only exception was landmark 5, with high deviation for AL 0.5. This landmark is located in the nasal and frontoparietal lateral suture, coinciding with a neural crest. It is possible that the presence of the crest caused higher error in the positioning of the landmark.

1.4.2. Linear Distances Reliability and Accuracy

Within-method mean distance repeatabilities (i.e. considering only replicates measured with the same method) were very high, above 0.9 for all methods, as can be seen in Table 5 for micro-CT resolutions and 3D digitizer; and in Table 6 for image

compacting factors and distinct filters. This result shows that measuring procedure is very reliable inside each method. However, between-methods mean distance repeatability is considerably lower (20% lower) when replicates of MED or HIGH were joined with replicates of DIG, but not when replicates of distinct micro-CT resolutions were joined together (Table 5). When replicates of AL0.5 and AL1.0 were joined together, the mean repeatability also reduced (Table 6), yet the drop was less steep (13% lower) than MED+DIG and HIGH+DIG. For the image compacting factors there was no drop in mean distance repeatability when replicates of different values of the compacting factors were considered together (Table 6).

Looking at repeatabilities separately for each distance in the between-methods analysis (Figure. 3B, C), it can be noticed that distances 3, 5, 6, 9, 10, 11, 13, 14, 16, 23 and 24 all have low repeatabilities (below 0.8) and are composed of landmarks detected with higher mean deviations for DIG (landmarks 3, 5, 8, 9, 10, 11, 12 and 19). These distances comprise small as well as large distances, showing that the error is independent of distance length and related to the quality of the landmark. Although we did not make expectations of higher error for landmarks placed with DIG, the landmarks with higher deviations had as a consequence higher discrepancies in the between-methods repeatabilities. However, the distances 17 and 22, composed of the landmarks with the highest error for DIG (14 and 19), did not present low repeatabilities. This is a result of their longer mean lengths (17: 7.14 mm and 22: 5.06 mm) and a smaller proportional error as a consequence.

On the other hand, distances 5, 10, 11, 13 and 24 are small distances (less than 3.0 mm of length) composed of highly discrepant landmarks, and therefore, have a larger proportional error. These differences in proportional error depending on the mean distance length is shown by a significant positive association between mean

distance repeatability and mean distance length for the joined data sets MED+DIG and HIGH+DIG ($r = 0.45$, d.f. = 22, $P = 0.02$ and $r = 0.44$, d.f. = 22, $P = 0.03$, respectively; Figure 3B, C), as well as for S2+S2/D2 and AL0.5+AL1.0 ($r = 0.44$, d.f. = 24, $P = 0.03$ and $r = 0.52$, d.f. = 24, $P = 0.007$, respectively; Figure 3E, I). Although a significant relation between repeatability and distance length exists, practically all distances composed of landmarks with higher differences between-methods presented lower repeatability when compared to within-method repeatability.

Additionally to the distance repeatability analysis, we also compared mean raw and absolute differences in the distances within and between methods. The mean raw differences within and between methods were close to zero for all data sets, being sometimes positive and sometimes negative (Tables 5 and 6), indicating that in average there is no consistent bias when measuring the same individual two times or measuring the same individuals with different methods. Yet, when looking at the raw differences between MED-DIG and HIGH-DIG for the distances separately, we can see that some distances have higher differences, ranging from -0.38 mm to 0.36 mm (Table 7).

These distances correspond to different bones in the toad skulls, indicating that the deviations in the relative landmark positions are not localized in a few bones. We can see that the higher differences (above 0.17 mm in magnitude, the highest difference found for MED-HIGH, values in boldface in Table 7) have approximately the same magnitude in mm for small and long distances. For instance, small distances 5, 8 and 13 (distance means below 3.0 mm) have a mean between-methods difference of 0.23 mm, while long distances have a mean of 0.26 mm.

Table 5. Within and between methods mean distance repeatabilities and mean raw and absolute differences for the 3D digitizer (DIG) versus micro-CT resolutions (MED or HIGH) comparison. Each individual was measured twice by each of the three methods and mean ± s.d. distance repeatabilities were calculated for within (considering only replicates of the same individual measured with the same method) and between-methods (considering the same individual measured with different methods). Within and between methods calculations were also done for raw and absolute differences between linear distances (mean ± s.d.). The last line of the table shows the mean between-method percentage error in relation to distance means.

Mean distance repeatabilities			
Within	**DIG**	**MED**	**HIGH**
	0.94 ± 0.07	0.97 ± 0.02	0.98 ± 0.01
Between	**MED+HIGH**	**DIG+MED**	**DIG+HIGH**
	0.95 ± 0.03	0.76 ± 0.10	0.76 ± 0.12
Mean differences (mm)			
Within	**DIG**	**MED**	**HIGH**
Raw	0.01 ± 0.04	-0.002 ± 0.02	-0.003 ± 0.02
Absolute	0.09 ± 0.03	0.07 ± 0.02	0.06 ± 0.02
Between	**MED-HIGH**	**MED-DIG**	**HIGH-DIG**
Raw	-0.03 ± 0.14	-0.02 ± 0.28	0.01 ± 0.28
Absolute	0.11 ± 0.05	0.27 ± 0.1	0.26 ± 0.1
% of mean	2.5 ± 1.0	6.5 ± 2.3	6.4 ± 3.0

However, as mentioned above, the same magnitude of between-methods error for small and long distances results in a higher proportion of error in relation to the distances means when considering the smaller ones (Table 8).When looking at the between-methods mean percentage error, the error in relation to the distances mean lengths, we can see that the cases around 10% correspond to three small distances: 5 (frontoparietal bone), 13 (nasal bone) and 25 (parasphenoid bone), and just one longer distance: 16 (squamosal bone), which is composed of two landmarks detected with high deviations for DIG (landmarks 8 and 12). The magnitudes of differences that we

found for DIG and the micro-CT resolutions are similar to differences reported by other authors when comparing measurements taken with CT data and digital calipers (Richard et al. 2014; Fernandes et al. 2014), although the last authors did find a systematic bias in the CT data (all distances under-estimated). Yet, when comparing our results with other authors that measured the same specimens with a 3D digitizer (Polhemus 3Space) and CT (Corner et al. 1992; Stull et al. 2014), our error between-methods is much smaller (around five to ten times smaller). This is probably due to the fact that we have used a much higher resolution than these authors in both the micro-CT as well as the digitizer since current equipments are an order of magnitude more accurate than 20 years ago.

The only distances estimated with more error when scanning at different resolutions were distances 17 and 19 (0.17 mm of difference), both distances in the parasphenoid bone, one of thinner bones. It is possible that in this case scanning with HIGH enhanced the visualization of the sutures in the parasphenoid bone compared to MED. The between-methods absolute differences for each distance (Figure 4) also indicate that some distances had higher magnitudes of difference. The worst case was for distance 16 (squamosal bone), the only long distance that presented percentage error around 10% and an error around 0.5 mm. This magnitude of error is very close to what we considered a gross error when measuring the individuals inside each method. When looking at the distinct filter results (AL1.0-AL0.5), we can see that both the raw and absolute differences between individuals scanned with distinct filters are much lower than differences between micro-CT resolutions and 3D digitizer. Similarly, in table 8 we can notice that the highest mean percentage error was around 7%, and all related to small distances.

Table 6. Within and between methods mean distance repeatabilities and mean raw and absolute differences for the image compacting factor comparison and the filter type comparison. The four distinct values of the image compacting factors stride and down-sample were: (S2): stride = 2.0; (S3): stride = 3.0; (S2/D2): stride = 2.0 and down-sample = 2.0 and (S3/D2): stride = 2.0 and down-sample = 3.0. The two different scanning filters were: aluminum 0.5 mm (AL0.5) and aluminum 1.0 mm (AL1.0). Within and between methods mean $\pm$ s.d. distance repeatabilities and mean $\pm$ s.d. raw and absolute difference were calculated, as well as mean percentage error in relation to the distances means of the between methods error.

Mean distance repeatabilities						
Within	**S2**	**S3**	**S2/D2**	**S3/D2**	**AL0.5**	**AL1.0**
	0.97 $\pm$ 0.02	0.98 $\pm$ 0.03	0.97 $\pm$ 0.02	0.97 $\pm$ 0.02	0.95 $\pm$ 0.04	0.93 $\pm$ 0.10
Between	**S2+S3**	**S2+S2/D2**	**S2+S3/D2**	**S3+S2/D2**	**S3+S3/D2**	**AL0.5+AL1.0**
	0.95 $\pm$ 0.03	0.95 $\pm$ 0.03	0.95 $\pm$ 0.04	0.95 $\pm$ 0.03	0.95 $\pm$ 0.04	0.83 $\pm$ 0.16
Mean differences (mm)						
Within	**S2**	**S3**	**S2/D2**	**S3/D2**	**AL0.5**	**AL1.0**
Raw	-0.002 $\pm$ 0.02	-0.003 $\pm$ 0.01	0.008 $\pm$ 0.02	0.006 $\pm$ 0.02	0.0 $\pm$ 0.02	0.003 $\pm$ 0.03
Absolute	0.07 $\pm$ 0.02	0.06 $\pm$ 0.02	0.07 $\pm$ 0.02	0.07 $\pm$ 0.02	0.07 $\pm$ 0.02	0.07 $\pm$ 0.03
Between	**S2-S3**	**S2-S2/D2**	**S2-S3/D2**	**S3-S2/D2**	**S3-S3/D2**	**AL1.0-AL0.5**
Raw	-0.01 $\pm$ 0.13	0.0 $\pm$ 0.12	0.0 $\pm$ 0.13	0.02 $\pm$ 0.13	0.02 $\pm$ 0.13	-0.03 $\pm$ 0.21
Absolute	0.10 $\pm$ 0.04	0.09 $\pm$ 0.04	0.10 $\pm$ 0.04	0.10 $\pm$ 0.06	0.10 $\pm$ 0.05	0.14 $\pm$ 0.05
% of mean	2.3 $\pm$ 0.9	2.3 $\pm$ 1.1	2.4 $\pm$ 1.0	2.4 $\pm$ 1.1	2.3 $\pm$ 1.0	3.4 $\pm$ 2.0

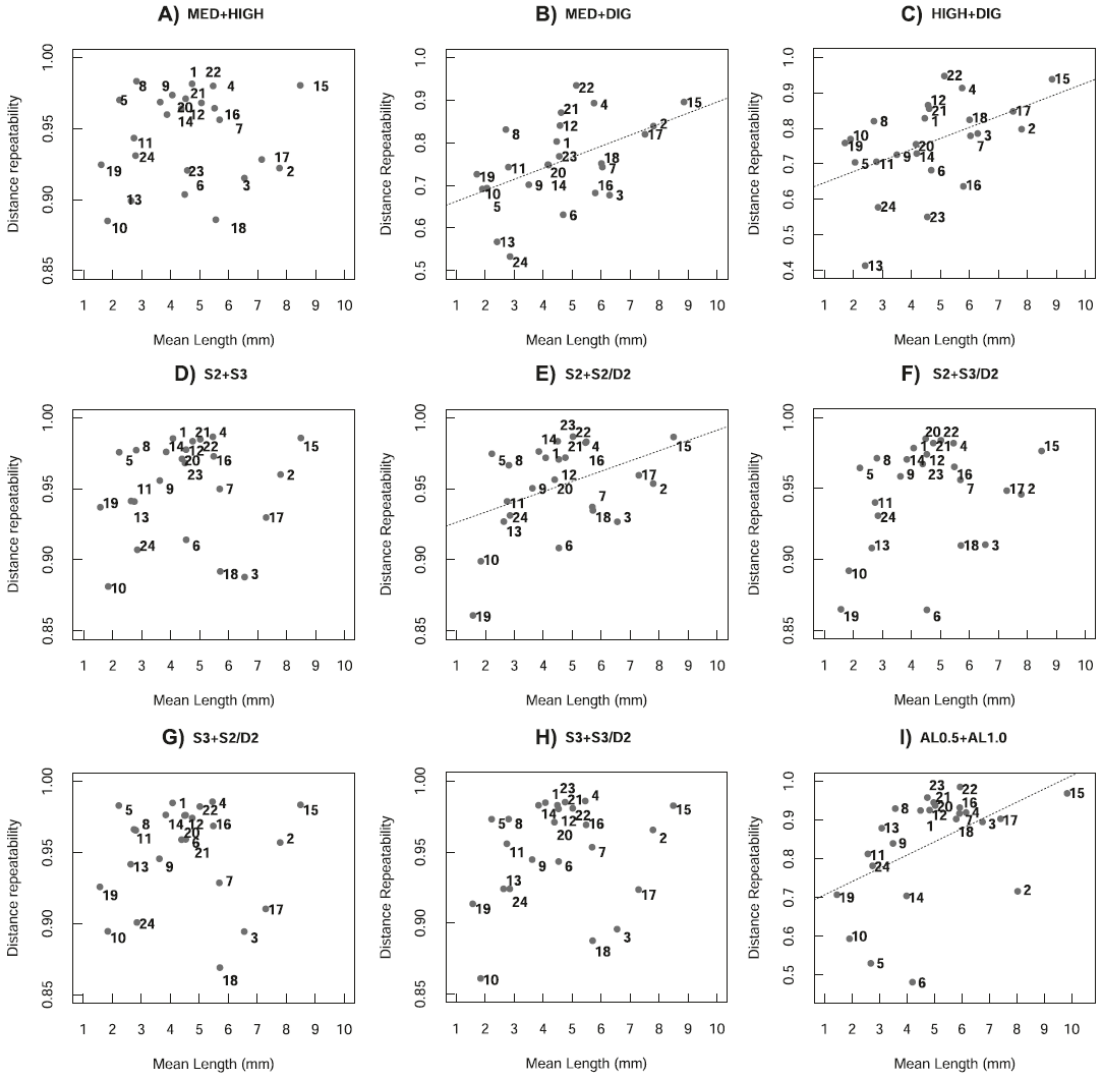


Figure 3. Between-method distance repeatability in relation to distance mean length. Individual toad skulls were each measured twice with different methods: using a 3D digitizer (DIG); using CT scans at medium (MED) or high (HIGH) resolutions; CT scans at MED loaded in TINA-Landmark with distinct values of the compacting factors stride and down-sample (S2, S3, S2/D2 and S3/D2); and CT scans at MED with 0.5 mm aluminum filter (AL0.5) or 1.0 mm aluminum filter (AL1.0). Data from different methods were joined together and distance repeatabilities calculated, representing the reliability in placing landmarks with the different methods. The dashed lines indicate significant correlations between the variables. Numbers 1-24 correspond to the distances as described in Table 2. **A)** MED + HIGH; **B)** DIG + MED; **C)** DIG + HIGH; **D)** S2 + S3; **E)** S2 + S2/D2; **F)** S2 + S3/D2; **G)** S3 + S2/D2; **H)** S3 + S3/D2 and **I)** AL0.5 + AL1.0.

Table 7. Between methods mean raw differences (in mm) in all the distances for joined data sets of micro-CT resolutions and 3D digitizer and of different filter type. The table shows the mean raw differences and their associated standard deviations in the length of the linear distances taken in the skulls between the same individuals scanned at medium (MED) and high (HIGH) resolutions and measured with TINA-Landmark (MED-HIGH); between the same individuals scanned at MED and measured with a 3D digitizer (DIG; MED-DIG); between the same individuals scanned at HIGH and measured with DIG (HIGH-DIG) and between the same individuals scanned with different thickness of an aluminum filter (AL1.0-AL0.5). The mean length of each distance is also shown as well as its correspondence to skull bones. Bold values indicate differences above 0.17 mm in magnitude.

			Between methods raw differences (mm)			
Distances	**Means (mm)**	**Bones**	**MED-HIGH**	**MED-DIG**	**HIGH-DIG**	**AL1.0-AL0.5**
1	**4.06**	**nasal**	-0.03 ± 0.15	**-0.29 ± 0.30**	**-0.26 ± 0.27**	-0.05 ± 0.12
2	7.76	frontoparietal	-0.02 ± 0.23	0.04 ± 0.34	0.06 ± 0.38	0.14 ± 0.63
3	**6.55**	**nasal**	0.08 ± 0.27	**0.33 ± 0.41**	**0.23 ± 0.37**	0.09 ± 0.21
4	5.47	nasal	-0.02 ± 0.14	-0.15 ± 0.28	-0.13 ± 0.25	0.02 ± 0.21
5	**2.24**	**frontoparietal**	0.0 ± 0.06	**0.22 ± 0.13**	**0.22 ± 0.13**	-0.08 ± 0.51
6	**4.49**	**frontoparietal**	-0.03 ± 0.12	**-0.18 ± 0.20**	-0.14 ± 0.21	-0.15 ± 0.53
7	**5.69**	**orbit**	-0.06 ± 0.12	**-0.26 ± 0.27**	**-0.19 ± 0.28**	-0.08 ± 0.14
8	**2.82**	**squamosal**	0.0 ± 0.07	**0.19 ± 0.14**	**0.19 ± 0.15**	-0.06 ± 0.11
9	**3.64**	**occipital**	0.02± 0.11	**0.23 ± 0.23**	**0.21 ± 0.22**	-0.03 ± 0.16
10	1.82	frontoparietal	-0.03 ± 0.11	-0.07 ± 0.21	-0.02 ± 0.19	0.03 ± 0.19
11	2.73	prenasal	0.04 ± 0.08	-0.03 ± 0.20	-0.08 ± 0.21	0.03 ± 0.13
12	4.52	nasal	-0.04 ± 0.11	0.01 ± 0.29	0.05 ± 0.25	0.02 ± 0.13
13	**2.64**	**nasal**	-0.09 ± 0.10	**0.27 ± 0.20**	**0.36 ± 0.19**	-0.06 ± 0.15
14	**3.87**	**nasal**	0.02 ± 0.11	**-0.24 ± 0.23**	**-0.27 ± 0.19**	-0.01 ± 0.35
15	**8.47**	**maxilla**	-0.07 ± 0.16	**-0.26 ± 0.34**	**-0.17 ± 0.27**	-0.07 ± 0.16
16	**5.51**	**squamosal**	-0.02 ± 0.16	**-0.19 ± 0.70**	**-0.18 ± 0.73**	-0.07 ± 0.19
17	**7.14**	**parasphenoid**	**-0.17 ± 0.27**	**-0.27 ± 0.45**	-0.06 ± 0.46	-0.10 ± 0.24
18	**5.56**	**parasphenoid**	**-0.18 ± 0.25**	**-0.38 ± 0.35**	-0.16 ± 0.36	-0.01 ± 0.19
19	1.60	premaxilla	0.02 ± 0.07	-0.08 ± 0.13	-0.10 ± 0.10	-0.08 ± 0.07
20	**4.37**	**nasal**	-0.01 ± 0.11	**0.26 ± 0.22**	**0.27 ± 0.24**	0.07 ± 0.09
21	**4.74**	**neopalatine**	-0.04 ± 0.08	**0.21 ± 0.16**	**0.25 ± 0.15**	0.03 ± 0.10
22	5.06	pterygoid	0.0 ± 0.15	0.01 ± 0.20	0.01 ± 0.17	-0.03 ± 0.08
23	**4.58**	**pterygoid**	-0.06 ± 0.27	0.13 ± 0.39	**0.20 ± 0.63**	-0.03 ± 0.09
24	2.79	parasphenoid	0.01 ± 0.12	-0.05 ± 0.36	-0.04 ± 0.36	-0.05 ± 0.14

Table 8. Mean percentage error in all distances for joined data sets of micro-CT resolutions and 3D digitizer and for distinct filter types. Mean percentage error were calculated from differences in the skull distances between: (1) individuals scanned with medium resolution and the same individuals scanned with high resolution (MED-HIGH); (2) individuals scanned with MED and with a 3D digitizer (MED-DIG); (3) individuals scanned with HIGH and DIG (HIGH-DIG) and (4) individuals scanned with an AL 1.0 mm filter and the same individuals scanned with a AL 0.5 mm filter (AL1.0-AL0.5). Values on bold correspond to mean percentages around 10.0%.

Distances	Means	Between methods mean percentage error			
		MED-HIGH	MED-DIG	HIGH-DIG	AL1.0-AL0.5
	(mm)	(%)	(%)	(%)	(%)
1	4.06	2.86	7.45	6.04	2.35
2	7.76	1.94	3.00	3.66	3.49
3	6.55	3.35	6.75	5.39	2.57
4	5.47	1.93	3.91	3.18	2.19
5	**2.24**	2.30	**10.79**	**10.75**	7.35
6	4.49	1.97	4.39	3.80	5.20
7	5.69	1.85	4.96	4.14	2.11
8	2.82	1.89	7.75	7.93	2.93
9	3.64	2.23	8.15	7.76	3.81
10	1.82	4.57	8.88	7.99	7.28
11	2.73	2.50	5.45	6.23	3.87
12	4.52	1.92	4.14	3.88	1.89
13	**2.64**	4.00	**11.61**	**15.17**	3.93
14	3.87	2.24	5.85	6.42	4.53
15	8.47	1.71	3.67	2.77	1.26
16	**5.51**	2.26	**9.32**	**9.82**	2.67
17	7.14	2.88	5.31	4.34	2.34
18	5.56	3.71	7.11	5.36	2.48
19	1.60	3.65	7.23	6.61	6.54
20	4.37	1.56	7.36	7.72	1.81
21	4.74	1.43	5.11	5.77	1.55
22	5.06	1.54	2.67	2.61	1.14
23	4.58	2.76	6.16	7.80	1.59
24	**2.79**	3.25	**8.65**	**9.19**	4.04

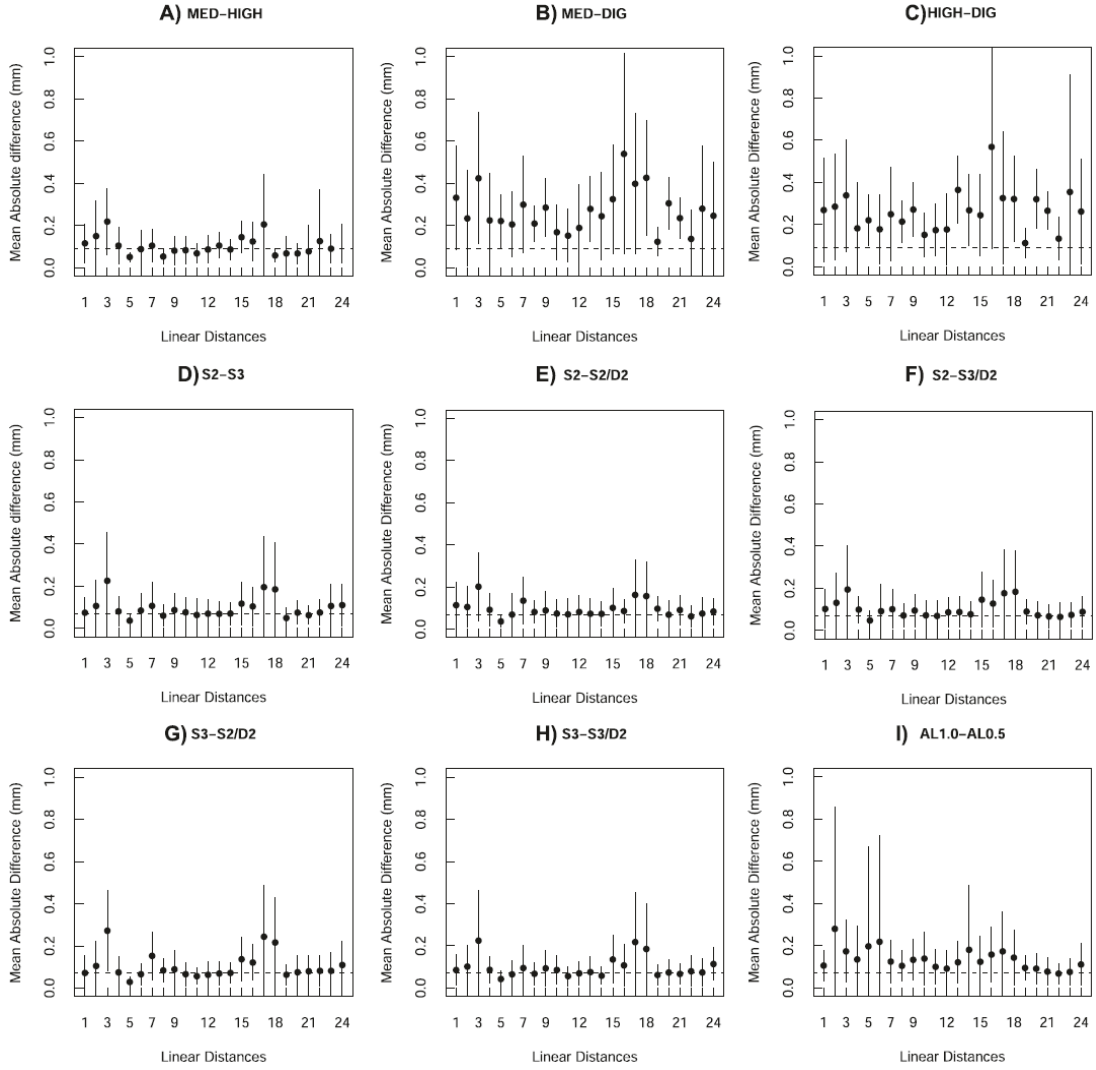


Figure 4. Between-method mean absolute differences in linear distances. Using the same data sets as described in Figure 1, absolute differences in mm between replicates of the same individuals measured with different methods were calculated and averaged across individuals. Circles represent means and vertical lines are the s.d. Dashed lines indicate the mean within-method absolute differences. Numbers 1-24 correspond to the distances as described in Table 2. **A)** MED + HIGH; **B)** DIG + MED; **C)** DIG + HIGH; **D)** S2 + S3; **E)** S2 + S2/D2; **F)** S2 + S3/D2; **G)** S3 + S2/D2; **H)** S3 + S3/D2 and **I)** AL0.5 + AL1.0.

1.5. Conclusions

In order to use 3D images of organisms in morphometric studies, one needs to know if the relative positions among landmarks are kept the same for different scanning procedures or rendering algorithms. By comparing landmark precision and linear distances in toad skulls scanned with two resolutions and measured with the software TINA-Landmark, with the same skulls measured with a 3D digitizer, we conclude that the degree of discrepancy is acceptable in general, although several distances had between-methods discrepancies above 5%. Yet, we must consider that 30% of the distances are below 3.0 mm in length, being quite small distances, thus for several other bigger organisms the error proportion will probably be lower than what we report. Attention needs to be paid in relation to the scale of the distances, as smaller distances might be estimated with proportionally more error. Using distinct micro-CT resolutions, distinct compacting factors for loading 2D cross-section sequences in TINA-Landmark, as well as using distinct filters to scan individuals that differ in bone density do not introduce high errors.

1.6. References

Adams, D.C. & Otarola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4, 393-399.

Bookstein, F.L.(1997). Morphometrics tools for landmark data: *Geometry and Biology*. Cambridge University Press, 425 p.

Bromiley P.A., Ragheb H. & Thacker N.A. (2012). The TINA Morphometrics Geometric Toolkit. *Tina Memo No 2010-007*.

Cheverud, J.M. (1995). Morphological Integration in the Saddle-back tamarin (Saguinus fuscicollis) cranium. *The American Naturalist*, 145,63–89.

Corner, B.D., Lele, S. & Richtsmeier, J.T. (1992). Measuring Precision of Three-Dimensional Landmark Data. *Journal of Quantitative Antrophology*, 3, 347-359.

Costantini D., Alonso M.L., Moazen M. & Bruner E. (2010). The Relationship Between Cephalic Scales and Bones in Lizards: A Preliminary Microtomographic Survey on Three Lacertid Species. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, 293,183–194.

Cuff A.R. & Rayfield E.J. (2013). Feeding Mechanics in Spinosaurid Theropods and Extant Crocodilians. *PloS one*, 8, e65295.

Dryden, I.L. (2015). shapes: Statistical Shape Analysis. URL: *http:// CRAN.R-project.org/packages=shapes*

Ekdale, E.G. (2010). Ontogenetic Variation in the Bony Labyrinth of Monodelphis domestica (Mammalia: Marsupialia) Following Ossification of the Inner Ear Cavities. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, 293, 1896–1912.

Fernandes, T.M.F., Adamczyk, J., Poleti, M.L., Henriques, J.F.C., Friedland, B. & Garib, D.G. (2014). Comparison between 3D volumetric rendering and multiplanar slices on the reliability on linear measurements on CBCT images: an *in vitro* study. *Journal of Applied Oral Science*.

Gignac P.M. & Kley N.J. (2014). Iodine-enhanced micro-CT imaging: Methodological refinements for the study of the soft-tissue anatomy of post-embryonic vertebrates. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 322, 166–176.

Halperin-Sternfeld M., Machtei E. & Horwitz J. (2014). Diagnostic Accuracy of Cone Beam Computed Tomography for Dimensional Linear Measurements in the Mandible. *The International Journal of Oral & Maxillofacial Implants*, 29, 593–599.

Kim G., Jung H.-J., Lee H.-J., Lee J.-S., Koo S. & Chang S.-H. (2012). Accuracy and Reliability of Length Measurements on Three-Dimensional Computed Tomography Using Open-Source OsiriX Software. *Journal of Digital Imaging*, 25, 486–491.

Kohn L.A. & Cheverud J.M. (1992). Issues in evaluating repeatability of an imaging system for use in anthropometry. *Proceedings of the Electronic imaging of the Human bodyworking group.* Crew System Ergonomics Information Analysis Center, Dayton, OH.

Lele, S. (1993). Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference. *Mathematical Geology*, 25, 573-602.

Lessells C.M. & Boag P.T. (1987). Unrepeatable Repeatabilities: A Common Mistake. *The Auk*, 104, 116–121.

Porto A., de 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, 36:118–135.

R Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. URL: http://www.R-project.org/.

Richard, A.H., Parks, C.L. & Monson, K.L. (2014). Accuracy of standard craniometric measurements using multiple data formats. *Forensic Science International*, 242, 177-185.

Richtsmeier, J.T., Paik, C.H., Elfert, P.C., Cole, T.M. & Dahlman, H.R. (2004). Precision, Repeatability and Validation of the Localization of Cranial Landmarks Using Computed Tomography Scans. *Cleft Palate-Craniofacial Journal* 1995, 32, 217-227.

Rosset A., Spadola L. & Ratib O. (2004). OsiriX: An Open-Source Software for Navigating in Multidimensional DICOM Images. *Journal of Digital Imaging*, 17, 205–216.

Schunke A.C., Bromiley P.A., Tautz D. & Thacker N.A. (2012). TINA manual landmarking tool: software for the precise digitization of 3D landmarks. *Frontiers in zoology*, 9, 6.

SkyScan. (2011). SkyScan 1176. *In vivo X-Ray Microtomograph Instruction Manual*.

Stull, K.E., Tise, M.L., Ali, Z. & Fowler, D.R. (2014). Accuracy and Reliability of measurements obtained from computed tomography 3D volume rendered images. *Forensic Science International*, 238, 133-140.

van der Linde, K. & Houle, D. (2009). Inferring the Nature fo Allometry from Geometric Data. *Evolutionary Biology*, 36, 311-322.

Wilkinson M., San Mauro D., Sherratt E. & Gower D.J. (2011). A nine-family classification of caecilians (Amphibia: Gymnophiona). *Zootaxa*, 2874, 41–64.

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

Zelditch, M.L. (1988). Ontogenetic variation in patterns of phenotypic integration in the laboratory rat. *Evolution*, 42, 28-41.

Zeldicth, M.L., Swiderski, D.L., Sheets, H.D. & Fink, W.L. (2004). *Geometric Morphometrics for Biologists: A Primer*. Elsevier Academic, 237 p.

Capítulo 2

Function and Climate shape Covariation Patterns in the Skull of Anuran Amphibians

2.1. Abstract

The evolution of complex characters may be shaped by internal processes, such as development and function, as well as by external processes, related to historical and ecological factors. The unique skull development of anuran amphibians allows for testing the support of developmental *versus* functional processes shaping the associations among traits. We measured phenotypic skull covariance and correlation pattern (P-matrix) similarity and tested modularity hypotheses in the toad species of the *Rhinella granulosa* group. We also investigated whether differences in P-matrices were associated with evolutionary history, mean morphological and mean climatic differences across species. P-matrix similarity is very high among the toad species. We suggest that the high support for functional in contrast to developmental modularity is connected to the biphasic ontogeny of the toads and that function can shape the fitness landscape through internal stabilizing selection. Divergence in species P-matrices is concentrated in the snout and associated to divergence in climate. Differences in snout trait covariance among species may be related to the relevance of snout function in the reproduction of toads. Therefore, there might also be a role for external stabilizing selection associated to different rain patterns across the toad species, in addition to internal stabilizing selection related to skull function.

2.2. Introduction

A central idea in morphological evolution is that changes in one trait may not be independent of changes in other traits within a complex structure (Olson and Miller 1958; Berg 1960). This connectedness among elements of a morphological system is referred as morphological integration. Olson and Miller (1958) proposed that traits sharing a developmental pathway and/or exerting a common function should be more integrated among themselves than with traits from distinct developmental origins or acting in a distinct function. The related concept of modularity refers to the relative independence among the developmental/functional units within a wider system by means of differential pleiotropy, in which a set of genes regulates several traits within a unit, but affects none or only a few traits of other units (Waddington 1957; Wagner 1996; Wagner et al. 2007). Evolutionarily important consequences of integration/modularity are the coordinate change of a complex system (Cheverud 1984) and a higher adaptability of the phenotypes because adaptation of one functional unit could occur with few or no interference in other functional units (Wagner and Altenberg 1996). Empirical evidence for a modular organization of the genetic architecture has been provided by quantitative trait loci (QTL) studies especially in mouse skeletal elements (Cheverud et al. 1997; Mezey et al. 2000; Cheverud et al. 2004; Kenney-Hunt et al. 2008; Wagner et al. 2008).

The end result of the underlying genetic architecture of a structure composed of many traits can be quantified by the additive genetic variance covariance matrix, the G-matrix (Arnold 1994). The G-matrix represents the proportion of variance in multiple traits that is heritable (with the covariances among traits indicating how much traits are genetically coupled) and, thus, is paramount to model the multivariate response of a population to selection and drift (Lande 1979; Lande and Arnold 1983;

Arnold 1992; Arnold et al. 2001). However, in order to make predictions of population response to the evolutionary landscape, the G-matrix must maintain a stable structure through evolutionary time (Arnold et al. 2008). G-matrix stability could be reached if mutation and stabilizing selection patterns were conserved throughout evolution (Lande 1976; Lande 1980), as would be suggested by a common modular organization of the G-matrix across populations and species (Riedl 1978; Cheverud 1982; 1996). We may define two types of stabilizing selection that are important for G-matrix evolution. Internal stabilizing selection is imposed by the epigenetic developmental system, in which traits need to be co-adapted with each other (Cheverud 1984; Maynard Smith et al. 1985). External stabilizing selection is related to the interaction of the traits with their environment (Cheverud 1984; Schwenk and Wagner 2001). Internal stabilizing selection affects the phenotypic variation available for the action of external stabilizing selection and directional selection (Cheverud 1984). Empirical evidence (e.g. Cheverud 1988, 1995, Roff 1995, Arnold and Phillips 1999, Marroig and Cheverud 2001; Porto et al. 2009, Kolbe et al. 2011) as well as simulation studies (Jones et al. 2003, 2004, 2007) support the theoretical expectation of G-matrix stability; even though there are examples of divergence in **G** or phenotypic matrices (**P**) across related species (e.g. Arnold 1981; Lofsvold 1986; Sanger et al. 2012; Haber 2015). Hence, the stability of G and P-matrices may not be a general phenomena and needs to be empirically verified case by case (Turelli 1988; Steppan et al. 2002).

Riedl (1978), followed by Cheverud (1982; 1984), proposed that the genetic integration would evolve to match the shape of the fitness landscape by selection for pleiotropy, in which traits interacting to perform the same function would become controlled by the same set of pleiotropic genes. Yet, the fitness landscape may

resemble developmental or functional integration, or even a combination of both (Cheverud 1984). Distinguishing between developmental and functional integration allows the test of whether development evolved to match function, thus requiring the decoupling between these two phenomena (Young and Badyaev 2006; Zelditch and Swiderski 2011). The mammal mandible is an example of a system in which the relative contribution of developmental and functional interactions has been studied separately (e.g. Monteiro et al. 2005; Young and Badyaev 2006; Zelditch et al. 2008). Yet, in the mammal skull, one of the most studied systems in the modularity literature, this is difficult to accomplish because functional modularity hypotheses are nested within a wider developmental hypothesis (face and neurocranium; see Cheverud 1995). Fortunately, the anuran amphibian skull is a system that provides the opportunity to separate the developmental and functional modularity hypotheses. Recently, Piekarski et al. (2014) published a cranial neural crest (CNC) fate-map study that revealed a remarkable difference in the contribution of the CNC to the bony skull between the anuran *X. laevis* and the rest of the tetrapods. Virtually all skull bones are derived from CNC streams in *X. laevis*, whereas in the other tetrapods (including the axoltl) there is a clear separation between bones derived from the CNC (face) and from the paraxial mesoderm (neurocranium). This difference in skull development in anurans allows for the construction of non-overlapping modularity hypotheses, since the functional units (neurocranium, snout and suspensorium; see Methods) do not coincide with the developmental units (branchial, hyoid and mandibular CNC streams).

In this study we analyzed P-matrix similarity and non-overlapping modularity hypotheses in the skulls of anuran amphibians belonging to the *Rhinella granulosa* species group. This group is composed of thirteen toad species widely distributed in

South America that present well developed cranial crests and prenasal bones as one of the sinapomorphies (Pramuk 2006; Narvaes and Rodrigues 2009; Sanabria et al. 2010). A species comparison of modularity patterns in anuran amphibians has never been done so far. We aimed to answer two main questions: (1) Do toad species present similar P-matrices? and (2) Which processes structures the P-matrices? As internal processes, we tested whether development or function are associated to P-matrix similarity by using the non-overlapping modularity hypotheses. As factors related to external processes, we tested whether evolutionary history, mean morphological or mean climatic differences across species are connected to P-matrix similarity. The *R. granulosa* species group is well suited to answer these questions because a molecular phylogeny for the group has just been published (Pereyra et al. 2015), species vary in skull morphology (Figure 1) and are distributed in different biomes with potentially divergent climates (Narvaes and Rodrigues 2009; Figure 2).

2.3. Methods

2.3.1. Sample, 3D Landmarking and Linear Distances

We used a total of 1,072 specimens belonging to 11 species of the *R. granulosa* group (excluding *R. nattereri* and *R. bernardoi*) plus an out-group species *R. margaritifera*. We could not have access to the two species excluded because there are very few specimens deposited in museum collections. The study design was to scan at least 50 specimens from each locality (when possible) and to use localities that represented the maximum of the species distributions (Table 1). Specimens were loan from several institutions: American Museum of Natural History (AMNH), Coleção Célio F. B. Haddad (UNESP – Rio Claro), Coleção Herpetológica da Universidade Federal da Paraíba (CHUFPB), Museo Argentino de Ciencias Naturales (MACN), Museu de

Ciências Naturais (MCN-FZB-RS), Museu de Ciências e Tecnologia (MCT-PUCRS), Museu Nacional do Rio de Janeiro (MNRJ), Museu de Zoologia da Universidade de São Paulo (MZUSP), National Museum of Natural History (USNM) and Texas Memorial Museum (TNHC). We identified the species following the taxonomic units proposed by Narvaes and Rodrigues (2009). We attributed specimens to their sexes by looking at secondary sexual characteristics. Although size might indicate the age of individuals, we could not determine age classes in our sample because there are no external features in toads that are correlated with age. Young juveniles are easy to distinguish from adults and were not used in this study.

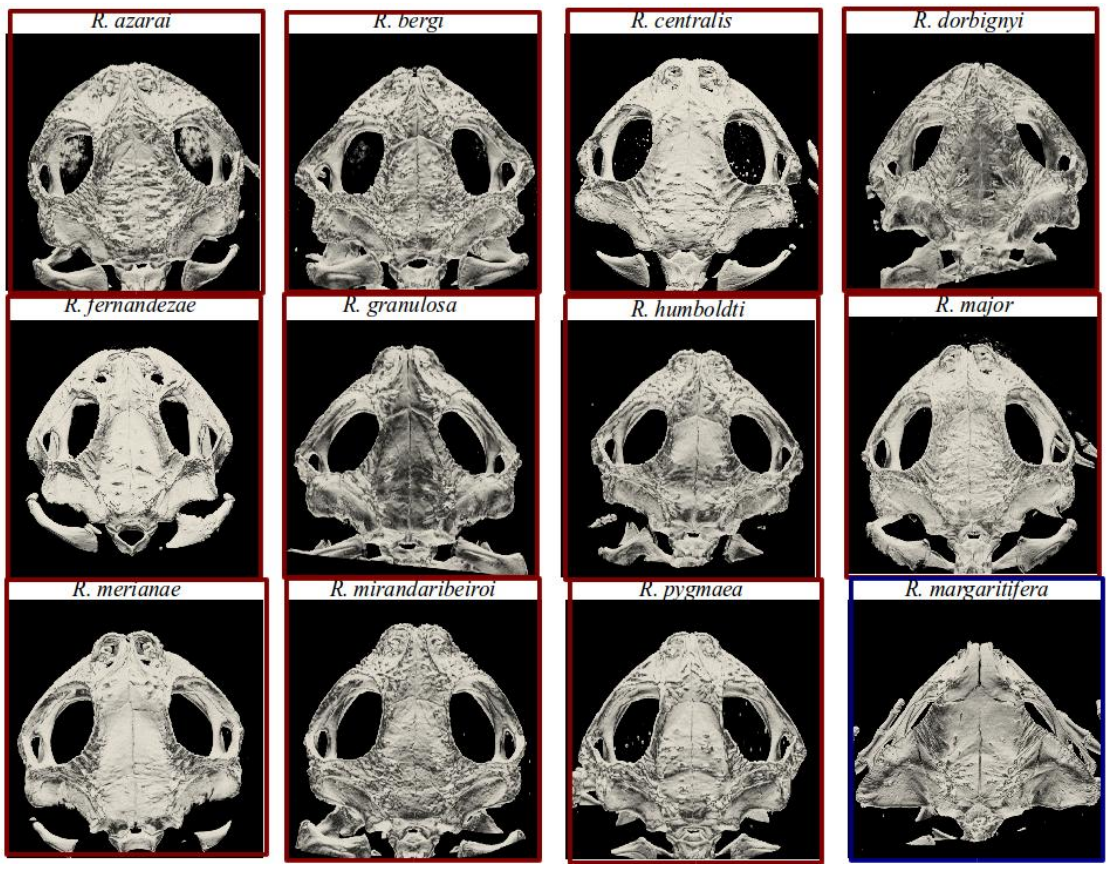


Figure 1. Dorsal views of the species skulls. The figure shows a representative image of the skull of each species from the *R. granulosa* group (red outline), plus the skull of the species used as the external-group *R. margaritifera* (red outline). Images were constructed as 3D volumes in the software TINA-Landmark.

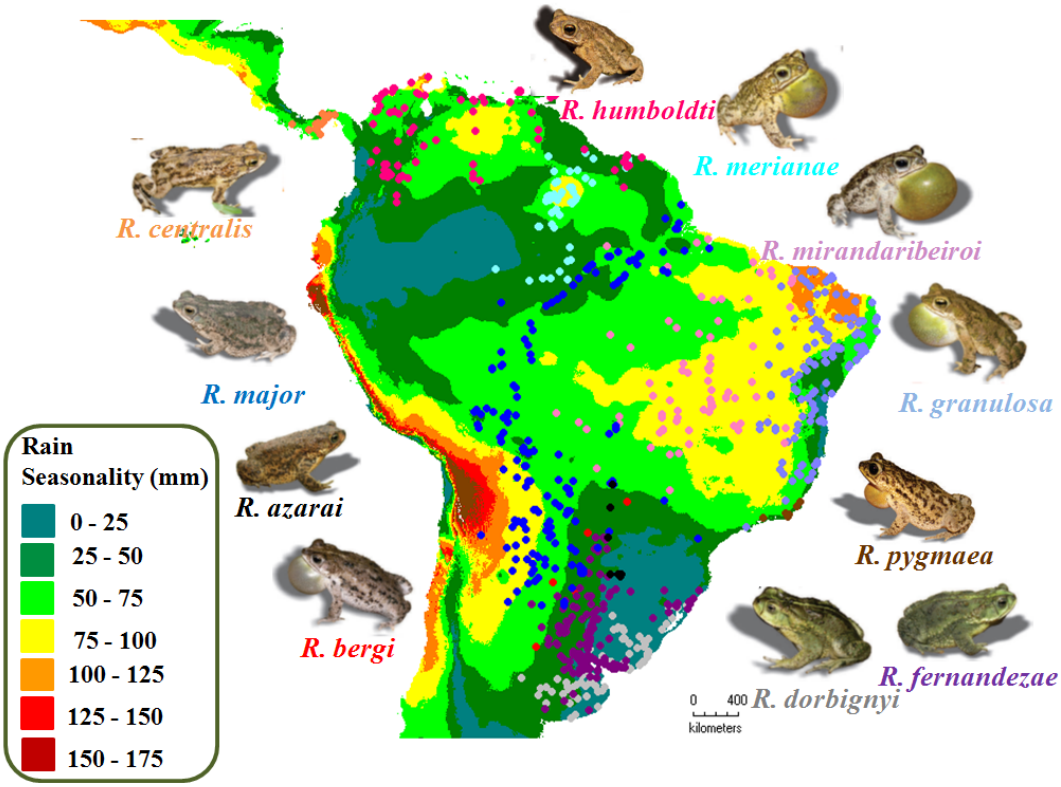


Figure 2. Map of Central and South America showing the distribution of the toad species (dots) and rain seasonality patterns (blue to red colors). Dots with same color indicate the distribution of each species, matching the color of their names below their picture. The climatic data for each species was extracted from its entire distribution using WorldClim database on DIVA-GIS and we chose to show rain seasonality, the coefficient of variation of annual precipitation values. Locality latitude and longitude values were obtained from Narvaes and Rodrigues (2009). Pictures of the toads were extracted from Pereyra et al. 2015.

We scanned all specimens using an X-ray micro-computed tomography system (micro-CT, SkyScan 1176, Konitch, Belgium) at the Instituto de Biociências, Universidade de São Paulo. We scanned specimens using the same resolution (pixel size = 18 μm), but total X-ray energy differed among individuals depending on the thickness of the filter used (AL 1.0 mm = 70 kV, 356 μA; AL 0.5 mm = 50 kV, 500

μA or AL 0.2 mm = 45 kV, 600 μA). We had to use distinct filters while scanning because several specimens varied in bone density. We have tested for an effect of filter type in placing landmarks and estimating linear distances in toad specimens and we concluded that deviations are in an acceptable range (Simon and Marroig 2015; Capítulo 1). Thus, we consider error due to filter type as negligible. We performed the reconstruction process with the NRecon software (SkyScan, Konitch, Belgium) with parameter values as described in Simon and Marroig (2015; Capítulo 1).

Table 1. Localities and sample sizes of all specimens scanned of each species. ARG: Argentina; BRA: Brasil; BOL: Bolívia; COL: Colômbia; GUI: Guianas; PAR: Paraguai.

Species	Localities	Sex	Total
R. azarai	**ARG**: Campo San Juan (4); Parque Provincial Fachinal (2); Ruta Provincial (3)	M (9)	**9**
R. bergi	**ARG**: Resistência (6); **BRA**: Corumbá (21); **PAR**: Concepcion (14), Villa Hayes (4)	F (5); M (40)	**45**
R. centralis	**PAN:** Albrook(9), Bejuco (4), Chitre (3), David (8), desconhecida (11), I. Nueva Gorgona (2)	M (37)	**37**
R. dorbignyi	**ARG**: Chascomus (5); **BRA** – RS: Cassino (19), Jaguarão (23), Montenegro (13), Sta. Vitória do Palmar (37), Tramandaí (10), Viamão (20)	F (36); M (91)	**127**
R. fernandezae	**ARG**: Barraqueras (30), Resistência (6); **BRA** – RS: Gravataí (10), Imbé (50), São Leopoldo (43)	F (65); M (74)	**140**
R. granulosa	**BRA** – BA: Itiúba (25), Jeremoabo (15), Una (18); ES: Fl. Nacional do Rio Preto (20); MG: Marliéria (5), Mocambinhos (23), Pedra Azul (14); PB: Aracaju (24), Cabedelo (9), Faz. Bravo (16), Sítio Junco (28)	F (61); M (136)	**197**
R. humboldti	**COL**: Bolívar (9), La Guajira (11), Magdalena (14), Maracay (3), Meta (6), Tolima (11), Tolú Viejo (16), Villavicencio (11); **VEN**: Puerto Sanariapo (10)	F (26); M (66)	**92**
R. major	**BOL**: S. Jose de Chiquitas (34); **BRA** – AM: Borba (5), Manicoré (29); MS: Est. Cayman (44); PA: S. Luiz (44); RO: P. Velho (32); **PAR**: Concepcion (4), Filadélfia (11), Pq. Nac. Defensores del Chaco (5), Villa Hayes (4)	F (84); M (128)	**212**
R. margaritifera	**BRA** – BA: Una	F (36); M (13)	**49**
R. merianae	**BRA** – AM: Coari (5), Codajás (3), INPA (4), Manaus (6); RR: Ilha de Maracá (3), Tepequém (8), Vila Surumu (51); **GUI** – Aishalton (5), Dubulay (12), Kato (4), Lethem (4)	F (18); M (88)	**106**
R. mirandariberoi	**BRA** – GO: Ilha do Bananal (50); MT: P. Leonardo (23); PA: Cachimbo (42); TO: Goiatins (7), Palmirante (8)	F (54); M (76)	**130**
R. pygmea	**BRA** – RJ: Barra de S. João (20), S. João da Barra (40)	F(16); M (44)	**60**

We placed 22 landmarks in each toad skull (5 midline and 17 on each side of the skull) at bone sutures or bone processes (Figure 3), such that we could assume their homology across all species. We used TINA Manual Landmarking Tool software (Schunke et al. 2012, hereafter called TINA-Landmark) to place landmarks in the 3D skull images. TINA-Landmark creates a 3D volume of the skull that is linked with its 2D orthogonal views allowing a very precise placement of 3D landmarks. From these landmarks, we extracted 21 linear distances (18 of them averaged over both sides of the skull) spread through the whole skull and allocated to specific modularity hypothesis (Table 2). The distances are all individual bone dimensions thought to represent effectively heritable entities (Thomson 1993). We have done the landmarking process twice for each skull so we could detect and correct for gross measurement errors and also calculate distance repeatability (Lessells and Boag 1987), a measure of the proportion of variance among individuals not due to measurement error (Falconer and Mackay 1996).

2.3.2. Species Phenotypic Matrices and Size Variation

We represent the species P-matrices of the skull distances as pooled within-group variance/covariance (V/CV) and Pearson product moment correlation matrices. We used P-matrices because they are much easier to obtain than G-matrices, given that a much smaller sample size is sufficient to estimate a reliable matrix (Cheverud 1988). We first performed outlier and normality analyses (Lilliefors' test with significance level $P < 0.05$) of the distances in all species. For most species we detected none or only one distance with significant departure from normality. To control for sources of variation in the data that are not strictly related to the genotype-phenotype map (Wagner 1996), we used multivariate analysis of variance

(MANOVA) and the Pillai statistic, as well as univariate analyses, to test for effects on the distance's means of the factors sex, geography and the interaction between them. Whenever the interaction or factors were significant, we removed their effect from the distances by extracting the residuals of the appropriate linear model (Table 3).

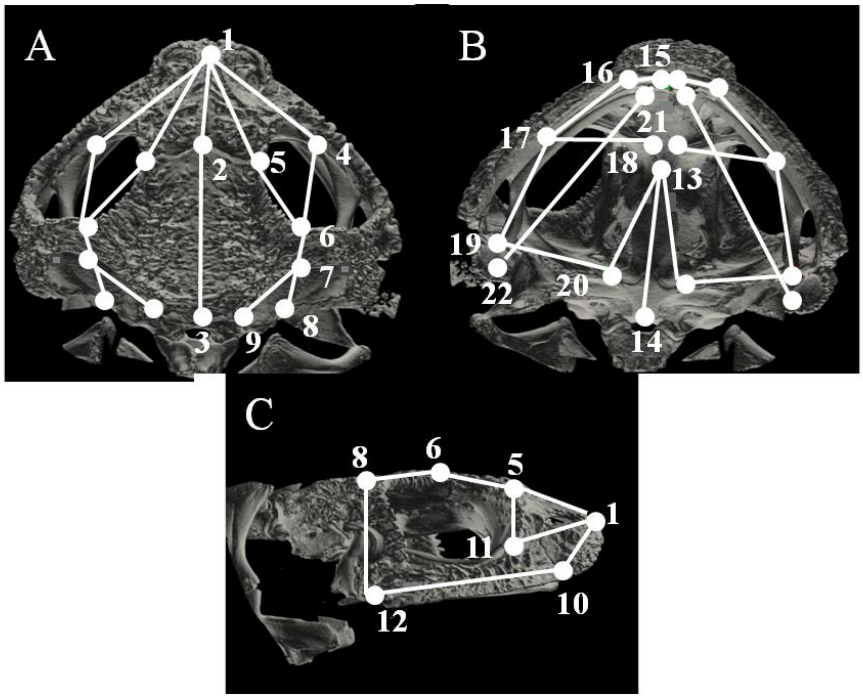


Figure 3. Numbered landmarks and linear distances in the toad skull. We placed 22 landmarks (white dots) spread in the whole skull: dorsal (A), ventral (B) and lateral (C) view. Landmarks are in bone sutures or bone processes to assure homology among species. Lines connecting the landmarks represent the 21 linear distances extracted from each specimen. The distances' descriptions are in Table 2.

Given that size variation obscures modularity patterns (see Marroig et al. 2003, Porto et al. 2013) and toads have indeterminate growth, we also estimated residual V/CV and residual correlation matrices by removing the variation in each species associated with size. We interpret size as the variation in measurement length that is shared by all other measurements, indicating that traits all increase or decrease

together (Porto et al. 2013). We removed from each species V/CV and correlation matrices the size-related eigenvector (**V**, which corresponds to the PC1 and its associated eigenvalue) using the formula:

$$\mathbf{R} = \mathbf{P} - \mathbf{V}^{\mathbf{T}}\,\mathbf{V};$$

in which **P** is the raw P-matrix, and $^{\mathbf{T}}$ indicates the transpose.

Table 2. Linear distances extracted from the landmarks placed in the toad skulls and their associated modularity hypotheses. Developmental hypothesis refers to the cranial neural crest origin of the skull bones represented by the three streams (Piekarski et al. 2014). Hormonal hypothesis is based on bone sensitivity to TH in newts (Ivanovic´and Kaleizic´2010). Functional hypothesis is based on the skull division in the three functional units (Trueb 1993).

			Modularity Hypothesis		
Distances	**Landmarks**	**Bones**	**Developmental**	**Hormonal**	**Functional**
1	1-2	nasal	hyoid	TH-sensitive	snout
2	2-3	frontoparietal	hyoid/mandibular/branchial	TH-unsensitive	neurocranium
3	1-4	nasal	hyoid	TH-sensitive	snout
4	1-5	nasal	hyoid	TH-sensitive	snout
5	5-6	frontoparietal	hyoid/mandibular/branchial	TH-unsensitive	neurocranium
6	4-6	orbit	-	-	neurocranium
7	6-8	squamosal	mandibular	TH-unsensitive	suspensorium
8	7-9	occipital	branchial	TH-unsensitive	neurocranium
9	1-10	prenasal	-	-	snout
10	1-11	nasal	hyoid	TH-sensitive	snout
11	5-11	nasal	hyoid	TH-sensitive	snout
12	10-12	maxilla	mandibular	TH-sensitive	snout
13	8-12	squamosal	mandibular	TH-unsensitive	suspensorium
14	13-14	parasphenoid	hyoid/mandibular	TH-unsensitive	neurocranium
15	13-20	parasphenoid	hyoid/mandibular	TH-unsensitive	neurocranium
16	15-16	premaxilla	hyoid/mandibular	TH-unsensitive	snout
17	16-17	maxilla	mandibular	TH-sensitive	snout
18	17-18	neopalatine	-	TH-unsensitive	snout
19	17-19	pterygoid	mandibular	TH-sensitive	suspensorium
20	19-20	pterygoid	mandibular	TH-sensitive	suspensorium
21	21-22	mandible	-	-	suspensorium

Table 3. Factors controlled for each linear model used to construct species P-matrices and sample size. G is for geography and S is for sex. The effects of the factors on the linear distances were tested with a MANOVA. S x G means that the interaction between factors was significant.

Species	Controlled variation	Sample size
R. bergi	G	45
R. centralis	G	37
R. humboldti	S x G	71
R. merianae	S, G	70
R. granulosa	S, G	189
R. mirandaribeiroi	S x G	125
R. pygmaea	G	58
R. dorbignyi	S, G	100
R. fernandezae	S x G	125
R. major	S, G	205
R. margaritifera	S	38

2.3.3. P-matrix similarity and dissimilarity

We compared raw and residual V/CV and correlation matrices among species using two methods: (1) Random Skewers method (RS; Cheverud 1996; Cheverud and Marroig 2007), which indicates the degree of resemblance in the species' response to random selection; and (2) Krzanowski Projection (Krz; Krzanowski 1979), which essentially test for the correlation of the shared space described by a limited number of principal components. RS is directly derived from Lande's (1979) multivariate selection equation ($\Delta \mathbf{z} = \mathbf{G}\,\boldsymbol{\beta}$; using **P** in the place of **G**), in which we apply 10,000 random selection vectors (21-element vector, extracted from a normal distribution of zero mean and unit variance) to a pair of species matrices being compared and use the average correlation of the species response to selection as the overall similarity index. We then correlate a specific random vector with all other random vectors to construct a random distribution of 21-element vectors. Only correlations below -0.45 or above 0.45 are considered significant. Krz is based on principal component analysis, in

which we used the first 10 PCs ([number of traits/2] - 1) to define a subspace for each species matrix. The first 10 PCs comprise 94% to 98% of the phenotypic within-species variation. The analysis finds the best-matching sets of orthogonal axes for the two subspaces to be compared (*i.e.*, the minimum angles between PCs; Krzanowski, 2000). The similarity index is calculated as the sum of squares of the cosines of the angles between the two sets of PCs (Blows *et al.*, 2004), or in other words, the sum of squares of the vector correlations between the 10 PCs of the two species being compared. This sum may vary from zero (no similarity) to 10 (full similarity: correlation of 1.0 between all PCs), but we present the sum divided by the maximum value (k=10) to have an index ranging from 0 to 1.0 (Marroig & Cheverud, 2010). To take the error in estimating P-matrices into account, we calculated matrix repeatability using bootstraping and auto-correlation (Cheverud et al. 1989). Matrix repeatability (t) indicates the proportion of variance in the matrix due to real population variance (Cheverud 1996). The maximum similarity between two matrices is $(t_1 t_2)^{0.5}$, where t_1 and t_2 are the repeatabilities of matrices 1 and 2, respectively (Cheverud 1996). Thus, we also present adjusted similarity indexes, which are the observable similarity indexes divided by the maximum similarity.

To investigate specific differences in trait covariance structure among species, we used Selection Response Decomposition (SRD) (Marroig et al. 2011). SRD is an extension of the RS method in which the response vector of a certain species matrix subjected to the random selection vectors is unfolded in trait-specific vectors. The trait-specific response vectors from two species are then compared by vector correlations, and the average correlation is the SRD score, specific for each trait. Differences between the global average of the SRD scores (considering all traits) and trait-specific SRD scores indicate the degree of dissimilarity between a pair of species

in the response to selection of a specific trait. We call this difference SRD_{diff}. We present trait-specific SRD_{diff} averaged across all species, for both raw and residual V/CV matrices, to achieve a general pattern of dissimilarity in the P-matrices.

2.3.4. Modularity tests

We constructed three non-overlapping modularity hypotheses (Table 2). For the developmental hypothesis, we used Piekarski et al. (2014) data on the contribution to the bony skull of three distinct CNC streams in *X. laevis*. The correspondence of these CNC streams to specific bones is not straightforward because the premaxillary, frontoparietal and parasphenoid bones are derived from more than one stream. Nevertheless, we could construct a testable developmental hypothesis, although with some overlap between the modules. The hormonal hypothesis was based on thyroxin hormone (TH) control in newts (same hypothesis as Ivánovic and Kalézic 2010), because information on TH sensitivity of anuran skull bones is virtually unavailable. Finally, the functional hypothesis separates the skull in three functional units (Trueb 1993): **1.** Neurocranium: brain and auditory capsule protection and sound reception; **2.** Snout: olfactory capsule protection and support; and **3.** Suspensorium: suspension of the lower mandible and stabilization of the upper mandible with the neurocranium. Next, we created theoretical connectivity matrices (T-matrices) in which associations among distances belonging to the same module were attributed the value of 1.0 (correlated) and associations for distinct modules were attributed the value of zero (uncorrelated; Cheverud et al. 1989). We constructed T-matrices for separate modules, as well as for all modules at once belonging to a single hypothesis (Total hypothesis). The total hypothesis is just the sum of the separate modules in a single matrix with maximum values restricted to 1.0. The T-matrices were then correlated

with the species empirical matrices and the significance tested with the Mantel test (Cheverud et al. 1989). We present modularity results as modularity indexes, composed of the difference between average correlations inside and outside the modules divided by the coefficient of variation of the matrices' eigenvalues (Shirai and Marroig, 2010, Porto et al. 2013). We chose this modularity index because it works for both raw and residual correlation matrices.

2.3.5. Phylogeny, Mean Morphology and Climate

We used the Bayesian molecular phylogeny from Pereyra et al. (2015; Figure 4) constructed with five nuclear and four mitochondrial genes to calculate phylogenetic distances among the toad species as pairwise distances using the branch lengths (each distance is the sum of the branch lengths of two specific tips in the phylogeny). To measure the differences in mean morphology across species, we constructed a morphological distance matrix, using Mahalanobis distance ($\mathbf{D^2}$), a generalized squared distance between groups which adjusts for unequal variances. We used the overall pooled within-group V/CV matrix of the *R. granulosa* species group (**W**; Table 4), weighted by sample size, as the measure of within-group variance (Marroig and Cheverud 2001, Ackermann 2002) and the species mean linear distances to compute each pairwise $\mathbf{D^2}$:

$$\mathbf{D^2_{ij}} = (\text{mean}_i - \text{mean}_j)^{\mathbf{T}}\, \mathbf{W^{-1}}\, (\text{mean}_i - \text{mean}_j);$$

where **i** and **j** are species and $^{\mathbf{T}}$ is the transpose.

For climatic data, we used the WorldClim database (Hijmans et al. 2005) composed of 17 variables (we excluded BIO3and BIO7, both related to thermal amplitude, because they are linear combinations of other variables; but we retained BIO2, the difference between daily maximum and minimum temperature as a

measure of temperature variability; Table 5) calculated from monthly precipitation (mm) and temperature (^{o}C) records obtained from several climatic stations (Hijmans et al. 2005; spatial resolution of approximately 1 Km^2). We extracted the climatic data for each species using DIVA-GIS software. Given that temperature and precipitation data have very different scales (^{o}C and mm) and that variances of precipitation data are much higher than temperature data, potentially biasing results, we opted to transform all climatic data using the z-score transformation (e.g. Duran and Pie 2015).

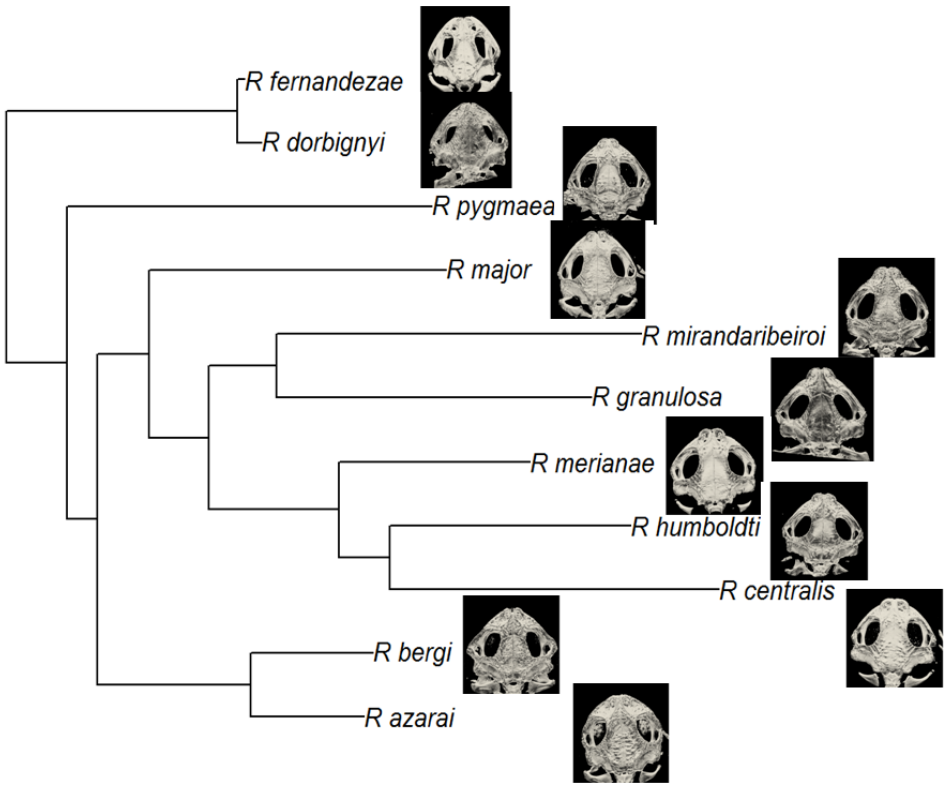


Figure 4. Phylogeny of the *R. granulosa* species group with a representative skull image of each species. The phylogeny is a modified Bayesian tree from Pereyra et al. (2015). We removed all species not used in our study and collapsed individuals of the same species to its most basal node. Branch lengths represent genetic distances among species and ancestors based on nucleotide substitutions.

Similarly to the morphological distance matrix, we constructed an overall pooled within-group V/CV climatic matrix (**W_**clim; Table 6), weighted by locality number for each species, and the species mean climatic variables to compute each climatic pairwise $\mathbf{D}^2$. We correlated the phylogenetic, morphological and climatic distance matrices with V/CV and correlation similarity matrices, for both raw and residual matrices (not including *R. margaritifera*), to investigate whether these factors might structure P-matrix differences. Given that we are correlating a similarity matrix with a dissimilarity one, we expect a negative correlation if the factors are related to the degree of P-matrix similarity. We also correlated the distance matrices among themselves because closely related species could present similar mean morphology and/or similar mean climate, and species with similar mean morphology could also present similar mean climate. The significance of matrix correlations was tested with the Mantel test. All analyses were done using the R programming environment (R Core Team 2014) and the "evolqg" package (Melo et al. 2015).

2.4. Results

2.4.1. P-matrix Similarity/Dissimilarity and Size Variation

Distance repeatabilities range from 0.83 to 0.99, indicating that variance due to measurement error is low. Raw matrix repeatabilities are high for all species, both for V/CV (0.976 $\pm$ 0.016) and correlation matrices (0.981$\pm$ 0.016). Non-adjusted and adjusted P-matrix similarity are remarkably high for raw matrices (V/CV: 0.90 $\pm$ 0.05 and 0.93 $\pm$ 0.04; correlation: 0.92 $\pm$ 0.05 and 0.94 $\pm$ 0.04, respectively), only the external group having slightly smaller similarity (Figures 5A, B; Table 7). Results are similar for Krz comparisons, although similarity is around 15% lower than RS comparisons (Table 8).

Table 4. Overall pooled within-group V/CV (below diagonal) and correlation (above diagonal) matrices.

This matrix corresponds to ancestral matrix (**W**) of the basal-most node in the species phylogeny.

Distances	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	**0.18**	0.58	0.78	0.89	0.48	0.75	0.63	0.61	0.63	0.80	0.73	0.75	0.72	0.67	0.64	0.65	0.69	0.73	0.67	0.68	0.75
2	0.12	**0.25**	0.75	0.75	0.56	0.80	0.74	0.74	0.67	0.75	0.73	0.80	0.78	0.74	0.68	0.68	0.73	0.78	0.70	0.71	0.79
3	0.15	0.17	**0.21**	0.81	0.56	0.76	0.66	0.69	0.75	0.85	0.73	0.84	0.77	0.72	0.69	0.70	0.84	0.86	0.69	0.75	0.84
4	0.20	0.20	0.20	**0.28**	0.38	0.81	0.72	0.67	0.67	0.85	0.86	0.82	0.79	0.75	0.71	0.70	0.75	0.81	0.72	0.74	0.82
5	0.06	0.09	0.08	0.06	**0.10**	0.65	0.24	0.49	0.45	0.51	0.38	0.54	0.50	0.47	0.44	0.44	0.49	0.51	0.45	0.48	0.51
6	0.13	0.16	0.14	0.18	0.08	**0.17**	0.66	0.69	0.70	0.77	0.84	0.84	0.83	0.75	0.71	0.69	0.74	0.81	0.76	0.75	0.83
7	0.09	0.13	0.11	0.13	0.03	0.09	**0.12**	0.62	0.58	0.70	0.70	0.76	0.76	0.68	0.64	0.60	0.67	0.71	0.72	0.67	0.77
8	0.08	0.11	0.10	0.11	0.05	0.09	0.07	**0.09**	0.61	0.67	0.64	0.71	0.68	0.64	0.61	0.58	0.66	0.71	0.65	0.66	0.72
9	0.06	0.08	0.08	0.09	0.03	0.07	0.05	0.04	**0.06**	0.73	0.68	0.62	0.71	0.63	0.59	0.67	0.65	0.72	0.58	0.66	0.69
10	0.12	0.13	0.14	0.16	0.06	0.12	0.09	0.07	0.06	**0.13**	0.80	0.84	0.78	0.73	0.69	0.73	0.79	0.83	0.70	0.74	0.83
11	0.11	0.12	0.11	0.16	0.04	0.12	0.08	0.07	0.06	0.10	**0.12**	0.79	0.78	0.71	0.67	0.66	0.73	0.79	0.70	0.72	0.80
12	0.22	0.27	0.27	0.30	0.12	0.24	0.18	0.15	0.10	0.21	0.19	**0.48**	0.80	0.78	0.75	0.69	0.85	0.85	0.85	0.79	0.94
13	0.16	0.20	0.18	0.21	0.08	0.17	0.14	0.11	0.09	0.14	0.14	0.28	**0.26**	0.76	0.71	0.69	0.75	0.80	0.73	0.81	0.82
14	0.17	0.22	0.20	0.23	0.09	0.18	0.14	0.12	0.09	0.16	0.14	0.32	0.23	**0.35**	0.97	0.63	0.71	0.75	0.69	0.71	0.78
15	0.13	0.16	0.15	0.18	0.07	0.14	0.11	0.09	0.07	0.12	0.11	0.25	0.17	0.28	**0.23**	0.58	0.68	0.72	0.67	0.67	0.74
16	0.04	0.05	0.05	0.05	0.02	0.04	0.03	0.03	0.02	0.04	0.03	0.07	0.05	0.05	0.04	**0.02**	0.59	0.71	0.60	0.65	0.72
17	0.11	0.13	0.14	0.15	0.06	0.11	0.09	0.07	0.06	0.11	0.09	0.22	0.14	0.16	0.12	0.03	**0.14**	0.89	0.59	0.74	0.85
18	0.12	0.15	0.15	0.17	0.06	0.13	0.10	0.08	0.07	0.12	0.10	0.23	0.16	0.17	0.13	0.04	0.13	**0.15**	0.66	0.81	0.85
19	0.12	0.15	0.14	0.16	0.06	0.13	0.11	0.09	0.06	0.11	0.10	0.25	0.16	0.18	0.14	0.04	0.09	0.11	**0.18**	0.71	0.87
20	0.13	0.16	0.16	0.18	0.07	0.14	0.10	0.09	0.07	0.12	0.11	0.25	0.19	0.19	0.15	0.04	0.12	0.14	0.14	**0.20**	0.81
21	0.22	0.28	0.27	0.31	0.11	0.24	0.19	0.16	0.12	0.21	0.19	0.46	0.30	0.33	0.25	0.08	0.22	0.23	0.27	0.26	**0.50**

Table 5. Climatic variables description.

Climatic Variables	Description	Scale
BIO1	Annual Mean Temperature	°C * 10
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	°C * 10
BIO4	Temperature Seasonality (standard deviation)	°C * 10
BIO5	Maximum Temperature of Warmest Month	°C * 10
BIO6	Minimum Temperature of Coldest Month	°C * 10
BIO8	Mean Temperature of Wettest Quarter	°C * 10
BIO9	Mean Temperature of Driest Quarter	°C * 10
BIO10	Mean Temperature of Warmest Quarter	°C * 10
BIO11	Mean Temperature of Coldest Quarter	°C * 10
BIO12	Annual Precipitation	mm
BIO13	Precipitation of Wettest Month	mm
BIO14	Precipitation of Driest Month	mm
BIO15	Precipitation Seasonality (Coefficient of Variation)	mm
BIO16	Precipitation of Wettest Quarter	mm
BIO17	Precipitation of Driest Quarter	mm
BIO18	Precipitation of Warmest Quarter	mm
BIO19	Precipitation of Coldest Quarter	mm

SRD results indicate that the species P-matrices differ very little, practically all distances having SRD_{diff} values averaged across species equal to the global average SRD (Figure. 6A). The only minor differences found are for comparisons against the external group. The first principal components (PC1) of all species raw matrices are size-related and explain a very high variance proportion within species (45% to 80%). Basal-most species present higher average loadings in PC1 for suspensorium traits, whereas the rest of the toad species have higher mean loadings in PC1 for snout traits (Table 9).

Table 6. Climatic overall within-group V/CV matrix.

This matrix was obtained as a weighted average (number of localities) of the species climatic matrices, after climatic variables were z-score transformed.

	BIO1	BIO2	BIO4	BIO5	BIO6	BIO8	BIO9	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19
BIO1	0.25																
BIO2	-0.14	0.75															
BIO4	-0.13	0.24	0.22														
BIO5	0.26	0.30	0.07	0.69													
BIO6	0.23	-0.31	-0.20	0.10	0.28												
BIO8	0.17	0.11	0.03	0.39	0.09	0.42											
BIO9	0.23	-0.28	-0.20	0.10	0.27	0.00	0.36										
BIO10	0.31	-0.03	-0.02	0.51	0.22	0.32	0.21	0.51									
BIO11	0.22	-0.18	-0.17	0.15	0.23	0.10	0.23	0.22	0.21								
BIO12	0.16	-0.31	-0.22	-0.02	0.24	-0.01	0.23	0.08	0.19	0.75							
BIO13	0.15	-0.21	-0.19	0.04	0.21	0.04	0.19	0.11	0.18	0.58	0.54						
BIO14	0.06	-0.34	-0.14	-0.17	0.15	-0.12	0.16	-0.02	0.09	0.44	0.24	0.59					
BIO15	0.03	0.21	0.04	0.16	-0.04	0.11	-0.05	0.08	0.00	-0.19	-0.02	-0.33	0.34				
BIO16	0.15	-0.23	-0.19	0.03	0.21	0.03	0.19	0.10	0.18	0.60	0.54	0.25	-0.04	0.55			
BIO17	0.07	-0.34	-0.15	-0.16	0.16	-0.12	0.17	-0.01	0.10	0.49	0.27	0.59	-0.34	0.29	0.61		
BIO18	-0.01	0.06	0.00	-0.07	-0.06	0.01	-0.05	-0.04	-0.02	0.33	0.20	0.29	-0.16	0.21	0.31	0.90	
BIO19	0.15	-0.35	-0.20	-0.02	0.26	-0.04	0.24	0.10	0.18	0.54	0.44	0.31	-0.13	0.45	0.35	-0.09	0.75

Table 7. Pairwise similarity of species' raw and residual covariance and correlation P-matrices using Random Skewers. The diagonal shows matrix repeatability values for each species. Below diagonal are non-adjusted similarities and above diagonal, adjusted similarities. V/CV indicates variance covariance matrices. All values are significant at $P < 0.05$.

Raw Covariance	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.97**	0.95	0.93	0.94	0.94	0.97	0.98	0.93	0.93	0.89	0.82
2. *R. humboldti*	0.92	**0.98**	0.94	0.96	0.97	0.97	0.96	0.93	0.90	0.90	0.86
3. *R. merianae*	0.91	0.91	**0.97**	0.93	0.98	0.95	0.94	0.95	0.93	0.91	0.90
4. *R. granulosa*	0.92	0.94	0.91	**0.99**	0.97	0.98	0.96	0.92	0.91	0.90	0.83
5. *R. mirandaribeiroi*	0.92	0.95	0.96	0.96	**0.99**	0.98	0.95	0.96	0.93	0.93	0.86
6. *R. major*	0.95	0.95	0.93	0.97	0.97	**0.99**	0.97	0.96	0.95	0.93	0.86
7. *R. bergi*	0.95	0.93	0.91	0.94	0.93	0.95	**0.97**	0.93	0.92	0.90	0.83
8. R. *pygmaea*	0.90	0.90	0.92	0.90	0.93	0.94	0.90	**0.97**	0.96	0.94	0.89
9. R. *dorbignyi*	0.91	0.88	0.91	0.90	0.92	0.94	0.90	0.93	**0.98**	0.97	0.87
10. *R. fernandezae*	0.88	0.88	0.89	0.89	0.92	0.93	0.88	0.92	0.96	**0.99**	0.85
11. *R. margaritifera*	0.78	0.82	0.85	0.80	0.83	0.83	0.79	0.85	0.84	0.82	**0.94**
Raw Correlation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.99**	0.96	0.94	0.98	0.97	0.98	0.98	0.92	0.94	0.95	0.81
2. *R. humboldti*	0.94	**0.98**	0.96	0.96	0.98	0.97	0.97	0.95	0.93	0.95	0.87
3. *R. merianae*	0.92	0.94	**0.98**	0.94	0.97	0.96	0.96	0.96	0.95	0.95	0.89
4. *R. granulosa*	0.97	0.95	0.93	**1.00**	0.98	0.99	0.99	0.92	0.93	0.93	0.82
5. *R. mirandaribeiroi*	0.96	0.96	0.95	0.97	**0.99**	0.99	0.98	0.96	0.94	0.96	0.89
6. *R. major*	0.97	0.96	0.94	0.98	0.98	**0.99**	0.99	0.95	0.95	0.96	0.86
7. *R. bergi*	0.97	0.95	0.94	0.98	0.97	0.98	**0.98**	0.95	0.95	0.97	0.85
8. R. *pygmaea*	0.90	0.92	0.93	0.91	0.94	0.93	0.92	**0.97**	0.95	0.95	0.88
9. R. *dorbignyi*	0.93	0.91	0.93	0.93	0.93	0.94	0.93	0.93	**0.99**	0.97	0.86
10. *R. fernandezae*	0.94	0.94	0.94	0.93	0.95	0.96	0.95	0.93	0.96	**0.99**	0.86
11. *R. margaritifera*	0.78	0.84	0.86	0.79	0.85	0.83	0.82	0.84	0.83	0.83	**0.94**
Residual Covariance	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.92**	0.80	0.80	0.85	0.81	0.88	0.78	0.80	0.80	0.74	0.68
2. *R. humboldti*	0.74	**0.93**	0.87	0.91	0.93	0.92	0.86	0.86	0.85	0.87	0.76
3. *R. merianae*	0.74	0.82	**0.94**	0.94	0.91	0.93	0.90	0.85	0.87	0.84	0.78
4. *R. granulosa*	0.81	0.87	0.90	**0.97**	0.94	0.97	0.89	0.92	0.88	0.88	0.79
5. *R. mirandaribeiroi*	0.76	0.88	0.86	0.91	**0.96**	0.94	0.89	0.88	0.89	0.87	0.80
6. *R. major*	0.83	0.88	0.89	0.94	0.90	**0.97**	0.89	0.90	0.89	0.86	0.78
7. *R. bergi*	0.72	0.79	0.83	0.84	0.83	0.84	**0.91**	0.86	0.87	0.83	0.78
8. R. *pygmaea*	0.73	0.80	0.79	0.87	0.82	0.85	0.78	**0.91**	0.85	0.81	0.74
9. R. *dorbignyi*	0.74	0.81	0.82	0.85	0.85	0.85	0.80	0.79	**0.95**	0.84	0.78
10. *R. fernandezae*	0.70	0.83	0.80	0.85	0.83	0.83	0.77	0.76	0.80	**0.96**	0.78
11. *R. margaritifera*	0.61	0.70	0.71	0.73	0.74	0.73	0.70	0.67	0.71	0.72	**0.89**
Residual Correlation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.91**	0.82	0.75	0.84	0.82	0.85	0.78	0.76	0.77	0.78	0.70
2. *R. humboldti*	0.75	**0.93**	0.87	0.92	0.90	0.90	0.84	0.85	0.82	0.85	0.79
3. *R. merianae*	0.70	0.81	**0.94**	0.92	0.90	0.90	0.88	0.85	0.82	0.81	0.80
4. *R. granulosa*	0.79	0.88	0.88	**0.97**	0.95	0.95	0.90	0.91	0.87	0.89	0.83
5. *R. mirandaribeiroi*	0.76	0.85	0.85	0.91	**0.96**	0.95	0.88	0.89	0.86	0.87	0.83
6. *R. major*	0.80	0.85	0.86	0.92	0.91	**0.97**	0.91	0.88	0.89	0.89	0.80
7. *R. bergi*	0.71	0.77	0.81	0.84	0.82	0.85	**0.90**	0.86	0.84	0.84	0.80
8. R. *pygmaea*	0.70	0.78	0.78	0.85	0.83	0.83	0.78	**0.91**	0.85	0.83	0.76
9. R. *dorbignyi*	0.72	0.77	0.78	0.83	0.83	0.85	0.77	0.79	**0.95**	0.85	0.76
10. *R. fernandezae*	0.73	0.81	0.77	0.86	0.83	0.86	0.78	0.78	0.81	**0.96**	0.79
11. *R. margaritifera*	0.63	0.72	0.74	0.77	0.77	0.75	0.72	0.69	0.70	0.74	**0.89**

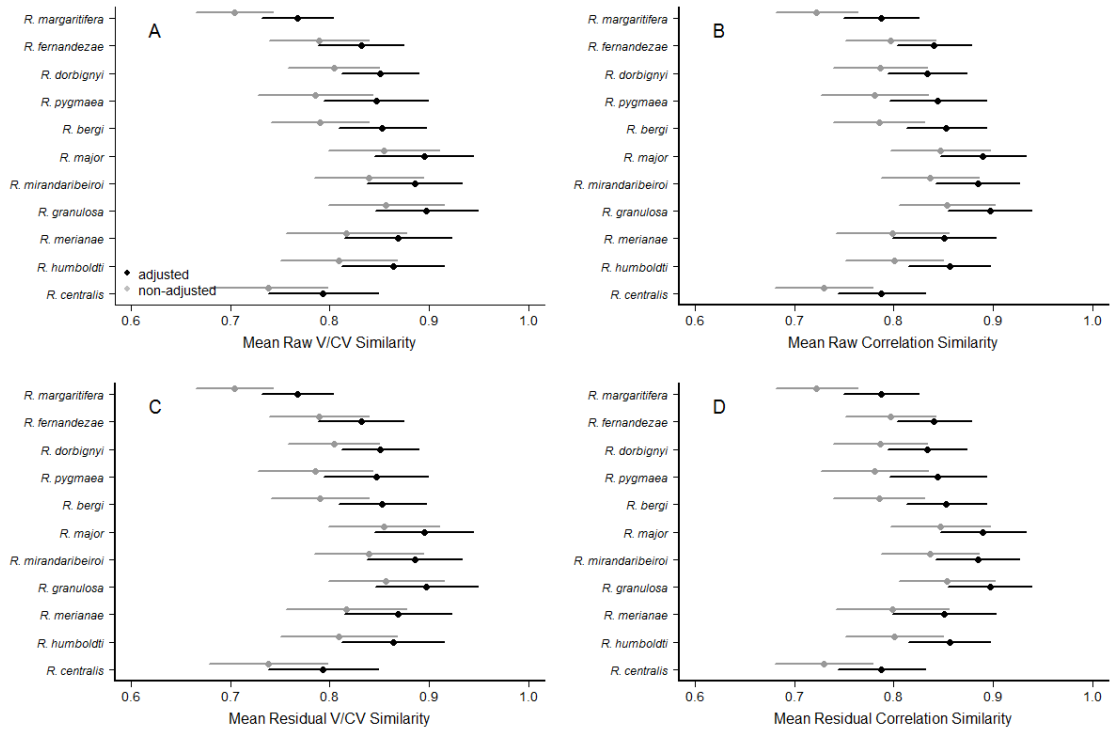


Figure 5. Mean and SD values for species P-matrix similarity compared with Random Skewers (RS). RS method gives an average similarity value between a pair of species P-matrices, indicating the resemblance of the P-matrices in responding to random selection. The similarity values in the x-axis are the mean similarity for each species in relation to all other species, for both raw and residual variance covariance P-matrices (A, C) and for raw and residual correlation matrices (B, D). Residual matrices do not have variance associated to size. Adjusted values take into account matrix repeatability values (see text).

Table 8. Pairwise similarity of species' raw and residual covariance and correlation P-matrices using Krzanowski Projection. The diagonal shows matrix repeatability values for each species. Below diagonal are non-adjusted similarities and above diagonal, adjusted similarities. V/CV indicates variance covariance matrices.All values are significant at $P < 0.05$.

V/CV	1	2	3	4	5	6	7	8	9	10	11
1. *R. centralis*	**0.88**	0.87	0.81	0.85	0.85	0.84	0.87	0.84	0.82	0.87	0.85
2. *R. humboldti*	0.78	**0.92**	0.84	0.92	0.89	0.89	0.88	0.86	0.88	0.90	0.82
3. *R. merianae*	0.73	0.77	**0.92**	0.89	0.86	0.87	0.91	0.80	0.85	0.86	0.77
4. *R. granulosa*	0.77	0.84	0.82	**0.92**	0.91	0.92	0.91	0.87	0.85	0.88	0.81
5. *R. mirandaribeiroi*	0.77	0.82	0.80	0.84	**0.93**	0.86	0.87	0.83	0.87	0.89	0.86
6. *R. major*	0.76	0.83	0.80	0.85	0.80	**0.93**	0.89	0.83	0.85	0.89	0.79
7. *R. bergi*	0.76	0.79	0.81	0.82	0.78	0.81	**0.87**	0.85	0.85	0.86	0.85
8. R. *pygmaea*	0.73	0.77	0.72	0.78	0.75	0.74	0.74	**0.87**	0.88	0.90	0.83
9. R. *dorbignyi*	0.73	0.81	0.77	0.78	0.80	0.79	0.75	0.79	**0.91**	0.91	0.80
10. *R. fernandezae*	0.79	0.83	0.79	0.82	0.83	0.83	0.77	0.81	0.84	**0.93**	0.83
11. *R. margaritifera*	0.75	0.73	0.68	0.72	0.77	0.71	0.74	0.72	0.71	0.74	**0.87**
Correlation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.88**	0.83	0.80	0.82	0.84	0.83	0.82	0.84	0.83	0.82	0.74
2. *R. humboldti*	0.75	**0.93**	0.84	0.87	0.90	0.85	0.86	0.86	0.89	0.86	0.80
3. *R. merianae*	0.71	0.76	**0.89**	0.87	0.89	0.89	0.89	0.85	0.84	0.86	0.75
4. *R. granulosa*	0.75	0.82	0.80	**0.94**	0.97	0.95	0.92	0.89	0.93	0.92	0.78
5. *R. mirandaribeiroi*	0.75	0.83	0.81	0.90	**0.92**	0.92	0.92	0.92	0.91	0.90	0.82
6. *R. major*	0.77	0.80	0.83	0.91	0.87	**0.97**	0.89	0.85	0.89	0.92	0.74
7. *R. bergi*	0.73	0.78	0.79	0.85	0.84	0.83	**0.89**	0.87	0.89	0.89	0.80
8. R. *pygmaea*	0.73	0.77	0.75	0.80	0.82	0.78	0.77	**0.86**	0.89	0.89	0.82
9. R. *dorbignyi*	0.74	0.81	0.75	0.85	0.82	0.83	0.80	0.78	**0.90**	0.93	0.79
10. *R. fernandezae*	0.75	0.80	0.78	0.87	0.84	0.88	0.81	0.80	0.85	**0.94**	0.76
11. *R. margaritifera*	0.65	0.72	0.66	0.71	0.73	0.68	0.70	0.71	0.70	0.69	**0.87**
Residual V/CV	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.87**	0.83	0.80	0.84	0.80	0.82	0.83	0.79	0.80	0.83	0.84
2. *R. humboldti*	0.75	**0.93**	0.88	0.95	0.91	0.91	0.91	0.85	0.88	0.89	0.83
3. *R. merianae*	0.71	0.81	**0.91**	0.91	0.84	0.86	0.92	0.82	0.84	0.84	0.81
4. *R. granulosa*	0.75	0.88	0.84	**0.92**	0.91	0.94	0.93	0.84	0.85	0.90	0.81
5. *R. mirandaribeiroi*	0.72	0.84	0.77	0.84	**0.92**	0.87	0.87	0.82	0.84	0.88	0.86
6. *R. major*	0.74	0.85	0.79	0.87	0.81	**0.94**	0.89	0.83	0.81	0.85	0.82
7. *R. bergi*	0.73	0.82	0.82	0.83	0.79	0.81	**0.87**	0.85	0.83	0.84	0.85
8. R. *pygmaea*	0.69	0.77	0.74	0.76	0.74	0.76	0.75	**0.88**	0.84	0.91	0.82
9. R. *dorbignyi*	0.71	0.81	0.77	0.78	0.77	0.75	0.74	0.75	**0.91**	0.90	0.80
10. *R. fernandezae*	0.74	0.83	0.76	0.83	0.81	0.79	0.75	0.81	0.82	**0.92**	0.83
11. *R. margaritifera*	0.73	0.74	0.72	0.72	0.77	0.74	0.74	0.71	0.71	0.74	**0.86**
Residual Correlation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. *R. centralis*	**0.88**	0.83	0.81	0.83	0.85	0.84	0.85	0.82	0.84	0.83	0.79
2. *R. humboldti*	0.74	**0.93**	0.90	0.91	0.90	0.91	0.91	0.85	0.91	0.89	0.82
3. *R. merianae*	0.72	0.82	**0.90**	0.92	0.90	0.95	0.95	0.88	0.90	0.88	0.82
4. *R. granulosa*	0.76	0.86	0.85	**0.96**	0.95	0.98	0.98	0.88	0.97	0.95	0.82
5. *R. mirandaribeiroi*	0.77	0.84	0.83	0.90	**0.94**	0.90	0.96	0.88	0.95	0.94	0.84
6. *R. major*	0.76	0.84	0.87	0.92	0.84	**0.93**	0.98	0.90	0.94	0.92	0.80
7. *R. bergi*	0.75	0.82	0.85	0.91	0.88	0.89	**0.89**	0.90	0.98	0.94	0.84
8. R. *pygmaea*	0.72	0.76	0.78	0.80	0.79	0.81	0.79	**0.87**	0.89	0.91	0.85
9. R. *dorbignyi*	0.74	0.83	0.81	0.89	0.87	0.85	0.88	0.79	**0.89**	0.98	0.86
10. *R. fernandezae*	0.75	0.82	0.80	0.89	0.87	0.85	0.85	0.82	0.89	**0.92**	0.81
11. *R. margaritifera*	0.68	0.72	0.72	0.74	0.74	0.71	0.73	0.73	0.75	0.71	**0.84**

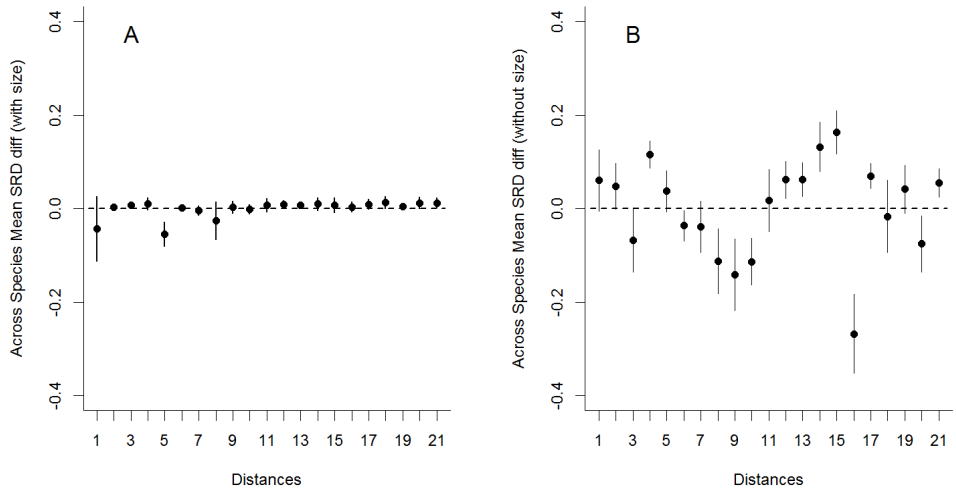


Figure 6. Selection Response Decomposition (SRD) analyses for species raw V/CV matrices (A) and residual V/CV matrices (B). Values in the y-axis indicate the difference between trait-specific SRD scores and the global average SRD score considering all linear distances at once (SRDdiff; see Methods). SRDdiff indicates the dissimilarity in the trait-specific response to random selection. A) SRDdiff for each linear distance for comparisons of raw V/CV P-matrices. B) SRDdiff for comparisons of residual V/CV P-matrices. Note that substantial differences occur only for residual matrices and are concentrated in distances 3 (nasal), 8 (occipital), 9 (prenasal), 10 (nasal) and 16 (premaxilla).

Residual matrix repeatabilities are slightly lower than raw matrix repeatabilities, but still high (V/CV: 0.939 $\pm$ 0.028; correlation: 0.936 $\pm$ 0.028), indicating that matrix estimation was still reliable even after size variation removal. Residual matrix similarity decreased in relation to raw matrix similarity, especially for correlation matrices (V/CV: 0.80 $\pm$ 0.07 and 0.85 $\pm$ 0.06; correlation: 0.79 $\pm$ 0.06 and 0.85 $\pm$ 0.03, for non-adjusted and adjusted values, respectively; Figures 5C, D; Table 7). Residual matrix similarity by Krz does not differ from raw matrices (Table 8). SRD analysis with residual matrices shows that higher SRD_{diff} values are

concentrated in the premaxillary, the occipital, the prenasal and the nasal bones (Figure 6B). The highest the P-matrix similarity, the highest is the pairwise sample size, for both raw and residual matrices and for both methods (Table 10).

Table 9. First principal components of all species correlation matrices and mean coefficients for different functional modularity hypotheses. Table shows the normalized loadings of all linear distances in species PC1. %var PC1 indicates the percentage of total within-species variation explained by each PC1. *R. cent.* = *R. centralis*; *R. humb.*= *R. humboldti*; *R. meria.* = R. *merianae*; *R. gran.* = *R. granulosa*; *R. miran.* = *R. mirandaribeiroi*; *R. maj.* = *R. major*; *R. berg.* = *R. bergi*; *R. pygm.* = *R. pygmaea*; *R. dorb.* = *R. dorbignyi*; *R. fern.* = *R. fernandezae* and *R. marg.* = *R. margaritifera*. In boldface are the highest mean coefficients for each species belonging to a specific modularity hypothesis.

		R. cent.	*R. humb.*	*R. meria.*	*R. gran.*	*R. miran.*	*R. maj.*	*R. berg.*	*R. pygm.*	*R. dorb.*	*R. fern.*	*R. marg.*
	% var PC1	79.5	66.2	60.8	80.6	66.9	74.1	77.4	58.5	71.0	72.6	45.1
Distances	**Bones**	**Loadings**										
1	nasal	0.220	0.219	0.217	0.221	0.215	0.209	0.215	0.217	0.188	0.214	0.089
2	frontoparietal	0.221	0.227	0.226	0.220	0.221	0.222	0.225	0.224	0.216	0.220	0.227
3	nasal	0.232	0.232	0.232	0.225	0.236	0.232	0.233	0.222	0.241	0.228	0.250
4	nasal	0.233	0.238	0.229	0.225	0.232	0.231	0.227	0.238	0.227	0.236	0.217
5	frontoparietal	0.161	0.169	0.099	0.195	0.175	0.165	0.157	0.127	0.048	0.079	0.135
6	orbit	0.234	0.233	0.230	0.229	0.223	0.230	0.229	0.242	0.230	0.235	0.205
7	squamosal	0.203	0.161	0.216	0.211	0.209	0.209	0.197	0.227	0.229	0.204	0.156
8	occipital	0.206	0.193	0.202	0.209	0.203	0.194	0.200	0.180	0.184	0.214	0.076
9	prenasal	0.219	0.229	0.227	0.208	0.199	0.203	0.200	0.169	0.179	0.188	0.236
10	nasal	0.230	0.238	0.248	0.227	0.241	0.229	0.237	0.247	0.231	0.225	0.196
11	nasal	0.234	0.227	0.228	0.220	0.226	0.227	0.224	0.185	0.224	0.227	0.177
12	maxilla	0.235	0.237	0.238	0.233	0.235	0.239	0.232	0.262	0.244	0.242	0.247
13	squamosal	0.229	0.225	0.248	0.225	0.226	0.235	0.223	0.239	0.232	0.230	0.231
14	parasphenoid	0.205	0.216	0.213	0.214	0.216	0.211	0.219	0.209	0.230	0.229	0.252
15	parasphenoid	0.189	0.202	0.198	0.209	0.197	0.203	0.206	0.181	0.222	0.220	0.238
16	premaxilla	0.207	0.206	0.168	0.202	0.183	0.198	0.202	0.171	0.216	0.205	0.197
17	maxilla	0.223	0.222	0.214	0.216	0.214	0.221	0.231	0.231	0.223	0.228	0.244
18	neopalatine	0.228	0.241	0.245	0.231	0.232	0.236	0.235	0.236	0.231	0.232	0.266
19	pterygoid	0.207	0.183	0.203	0.207	0.217	0.222	0.212	0.230	0.223	0.212	0.227
20	pterygoid	0.213	0.213	0.204	0.219	0.219	0.210	0.226	0.231	0.232	0.221	0.274
21	mandible	0.239	0.246	0.246	0.231	0.250	0.240	0.236	0.260	0.248	0.241	0.290
Mean loading	Neurocranium	0.17	0.17	0.16	0.18	0.17	0.17	0.17	0.16	0.16	0.16	0.18
	Snout	**0.226**	**0.229**	**0.225**	**0.221**	**0.221**	**0.223**	**0.224**	0.218	0.220	**0.223**	0.212
	Suspensorium	0.218	0.205	**0.223**	0.219	**0.224**	**0.223**	**0.219**	**0.237**	**0.233**	**0.221**	**0.236**

Table 10. Pearson product-moment correlations between similarity matrices and mean pairwise sample size (harmonic mean). All correlations are significant at $P < 0.05$.

Similarity matrix	
Raw matrices	**Mean sample size**
Covariance (RS)	**0.34**
Covariance (Krz)	**0.58**
Correlation (RS)	**0.30**
Correlation (Krz)	**0.84**
Residual matrices	**Mean sample size**
Covariance (RS)	**0.70**
Covariance (Krz)	**0.50**
Correlation (RS)	**0.75**
Correlation (Krz)	**0.66**

2.4.2. Support for Distinct Modularity Hypotheses

Functional and hormonal hypotheses presented higher support, in terms of number of toad species presenting significant matrix correlations, compared with the developmental hypotheses (Table 11). For raw matrices, the functional and hormonal Total hypotheses, as well as the Snout and TH-sensitive modules were the best supported. The Suspensorium was detected in the more basal species of the group: *R. dorbignyi*, *R. fernandezae* and *R. pygmaea*. More modularity hypotheses were detected in the residual matrices, such as the Neurocranium (Table 11). The greatest change in the developmental hypothesis was the detection of the Branchial module in all species, with even higher modularity indexes than the functional modules.

Table 11. Modularity indexes for raw and residual species matrices. Modularity indexes were calculated as the difference between the average correlation among traits inside a module and the average correlation outside the module divided by the variation coefficient of the P-matrix eigenvalues. Residual matrices do not have size variation. Significance for the correlation of the empirical matrices with the theoretical matrices was given by Mantel tests. Values in italic are for $P < 0.1$ and values in bold are for $P < 0.05$.

Hypothesis	**Raw matrices**				**Residual matrices**			
Developmental	Mandibular	Hyoid	Branchial	Total	Mandibular	Hyoid	Branchial	Total
R. centralis	-0.018	-0.004	-0.033	-0.017	0.001	0.001	0.009	-0.003
R. humboldti	-0.017	0.018	-0.006	-0.001	0.006	0.003	*0.054*	0.005
R. merianae	-0.020	-0.013	-0.003	-0.022	0.014	0.002	**0.190**	0.015
R. granulosa	-0.007	-0.001	-0.008	-0.005	0.004	0.004	**0.033**	0.006
R. mirandaribeiroi	-0.015	-0.003	-0.002	-0.010	0.007	0.005	**0.086**	*0.012*
R. major	-0.008	-0.005	-0.030	-0.009	0.004	0.006	**0.040**	0.004
R. bergi	-0.010	-0.004	-0.011	-0.008	0.002	-0.001	**0.089**	0.003
R. pygmaea	0.003	-0.029	-0.035	-0.012	**0.014**	0.002	**0.060**	*0.017*
R. dorbignyi	-0.003	-0.021	-0.072	-0.020	**0.014**	0.009	**0.132**	0.001
R. fernandezae	-0.013	-0.011	-0.068	-0.016	*0.007*	**0.018**	*0.045*	**0.023**
R. margaritifera	**0.053**	-0.012	-0.021	*0.037*	-0.002	-0.006	**0.119**	-0.004
	Raw matrices				**Residual matrices**			
Functional	Neurocranium	Snout	Suspensorium	Total	Neurocranium	Snout	Suspensorium	Total
R. centralis	-0.023	**0.025**	0.009	**0.013**	**0.014**	**0.011**	*0.013*	**0.015**
R. humboldti	-0.001	**0.040**	-0.004	**0.029**	**0.045**	**0.032**	**0.060**	**0.050**
R. merianae	-0.020	**0.034**	0.024	**0.024**	**0.054**	**0.026**	0.017	**0.040**
R. granulosa	-0.007	**0.012**	0.007	**0.009**	**0.020**	**0.018**	**0.020**	**0.023**
R. mirandaribeiroi	-0.011	*0.017*	0.020	**0.013**	**0.017**	**0.015**	0.017	**0.020**
R. major	-0.018	*0.016*	0.016	**0.010**	**0.029**	**0.013**	0.009	**0.020**
R. bergi	-0.009	*0.017*	0.011	**0.012**	**0.037**	**0.009**	**0.022**	**0.023**
R. pygmaea	-0.016	0.014	**0.053**	**0.017**	**0.053**	**0.021**	0.018	**0.035**
R. dorbignyi	-0.026	0.019	**0.040**	**0.015**	**0.075**	**0.014**	0.008	**0.035**
R. fernandezae	-0.025	0.020	*0.031*	**0.014**	**0.024**	**0.018**	**0.051**	**0.032**
R. margaritifera	-0.009	0.012	0.055	0.018	**0.059**	0.015	0.022	**0.034**
	Raw matrices				**Residual matrices**			
Hormonal	TH-sensitive	TH-unsensitive	Total		TH-sensitive	TH-unsensitive	Total	
R. centralis	**0.029**	-0.024	*0.004*		**0.011**	0.000	0.008	
R. humboldti	*0.026*	-0.020	0.004		*0.012*	0.002	0.010	
R. merianae	*0.033*	-0.018	**0.011**		**0.013**	*0.020*	*0.023*	
R. granulosa	**0.014**	-0.009	**0.004**		**0.009**	*0.007*	**0.012**	
R. mirandaribeiroi	**0.030**	-0.020	**0.007**		**0.016**	0.010	**0.019**	
R. major	**0.026**	-0.019	**0.006**		**0.015**	0.007	**0.016**	
R. bergi	**0.026**	-0.019	**0.006**		**0.017**	0.005	**0.016**	
R. pygmaea	**0.042**	-0.026	**0.012**		*0.013*	**0.019**	**0.023**	
R. dorbignyi	**0.038**	-0.023	**0.011**		**0.021**	**0.023**	**0.032**	
R. fernandezae	*0.028*	-0.023	0.003		*0.009*	0.004	0.009	
R. margaritifera	0.017	-0.016	0.001		0.001	0.006	0.003	

2.4.3. Matrix Similarity and Phylogenetic, Morphological and Climatic Distances

Morphological and phylogenetic distances are shown in Table 12. Climatic distances are shown in Table 13. Closely related species have more similar mean morphology, as well as more similar mean climatic variables, than distantly related species (Figures 7A, C). Also, species differences in mean morphology are associated to differences in mean climate (Figure 7B), even after removing phylogenetic effect from both matrices ($r = 0.49$ $P < 0.05$ d.f. = 18). For both morphological and climatic distances, we discounted the effect of phylogeny by using the residuals of a linear model of morphological and climatic distances on phylogenetic distances, respectively.

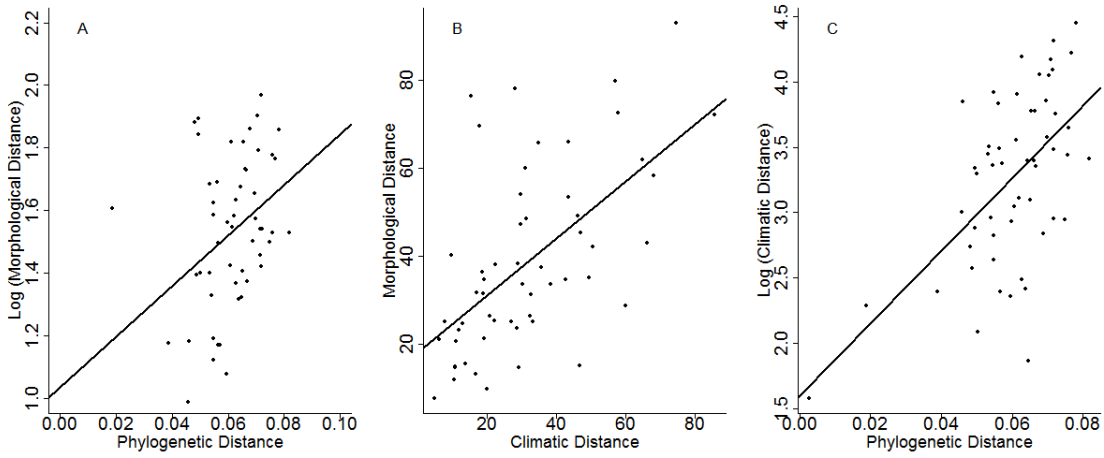


Figure 7. Matrix correlations between morphological, phylogenetic and climatic distances across species. Morphological and climatic distances are pairwise Mahalanobis distances obtained from the 21 linear distances on the skulls and from the 17 z-transformed mean bioclimatic variables of each species, respectively. Phylogenetic distances are the sum of the branch lengths of a pair of tips in the phylogeny. **A)** $r = 0.44$ $P < 0.05$ df =18; **B)** $r = 0.59$ $P < 0.05$ df = 18; **C)** $r = 0.58$ $P < 0.05$ df=18.

Table 12. Morphological (above diagonal) and phylogenetic (below diagonal) distances between the toad species.

	1	2	3	4	5	6	7	8	9	10	11
1. *R. azarai*		40.3	28.8	78.1	76.4	65.9	47.4	69.7	31.3	66.0	48.5
2. *R. bergi*	0.02		34.8	25.2	24.8	38.2	25.4	25.2	14.8	54.1	21.3
3. *R. centralis*	0.07	0.07		72.2	58.3	26.4	9.7	37.5	15.2	33.7	33.7
4. *R. dorbignyi*	0.05	0.05	0.08		7.7	72.6	61.9	49.1	43.1	93.1	42.1
5. *R. fernandezae*	0.05	0.05	0.08	0.00		53.5	45.3	38.4	35.3	79.9	25.1
6. *R. granulosa*	0.06	0.06	0.07	0.07	0.07		21.1	11.9	14.8	13.3	34.7
7. *R. humboldti*	0.06	0.07	0.05	0.07	0.07	0.06		23.3	15.0	31.8	31.6
8. *R. major*	0.05	0.05	0.07	0.06	0.05	0.06	0.06		15.5	20.7	36.4
9. *R. merianae*	0.06	0.06	0.05	0.06	0.06	0.06	0.04	0.05		26.5	23.6
10. *R. mirandaribeiroi*	0.07	0.07	0.08	0.07	0.07	0.05	0.07	0.06	0.06		60.0
11. *R. pygmaea*	0.05	0.05	0.08	0.05	0.05	0.07	0.07	0.06	0.07	0.08	

Table 13. Climatic distances between species (below diagonal) and residuals of climatic distances after discounting for phylogenetic effect (above diagonal).

	1	2	3	4	5	6	7	8	9	10	11
1. *R. azarai*		0.16	0.51	0.37	-0.20	0.25	0.01	-0.10	0.33	0.36	0.37
2. *R. bergi*	9.9		0.15	0.31	-0.38	-0.21	-0.31	-0.91	0.19	-0.04	-0.14
3. *R. centralis*	59.8	42.8		0.69	0.50	-0.11	0.13	0.04	0.97	-0.05	-0.46
4. *R. dorbignyi*	28.3	27.0	85.7		-0.10	0.58	0.61	0.68	0.85	0.72	0.80
5. *R. fernandezae*	15.5	13.1	68.0	4.9		0.33	0.33	0.25	0.60	0.49	0.42
6. *R. granulosa*	34.8	22.5	32.6	57.8	43.6		-1.53	-0.89	-0.77	-0.29	-0.63
7. *R. humboldti*	29.9	22.1	20.2	64.7	47.1	6.4		-0.85	-0.29	-0.67	-0.73
8. *R. major*	17.8	8.0	35.8	46.2	28.9	10.6	12.1		-0.48	-0.95	-0.33
9. *R. merianae*	32.9	29.2	46.9	66.2	49.5	11.0	11.0	13.9		-0.24	-0.09
10. *R. mirandaribeiroi*	43.5	29.8	38.4	74.5	57.1	16.9	17.1	11.2	21.0		-0.26
11. *R. pygmaea*	31.5	19.3	30.4	50.4	33.2	19.1	19.1	18.8	28.7	31.2	

All correlations of P-matrix similarity (raw and residual) with phylogenetic, morphological and climatic distances are shown in Table 14. Similarity of raw matrices has a weak correlation with phylogenetic distances, but only for V/CV matrices. Raw matrix similarity is also related to differences in mean morphology among the species. The relation of raw matrix similarity and morphological distances remained significant even after discounting the effect of phylogeny ('Morphological

residual' in Table 12). The highest correlation of raw matrix similarity is with climatic distances. We tested for a correlation between raw matrix similarity and residuals of climate after removing the effect of phylogeny ('Climatic residual'; Table 14, above diagonal), and the significant association is maintained. Residual matrix similarity is not related to phylogenetic distances or to differences in mean morphology. However, as for raw matrices, the highest correlation of residual matrix similarity is with differences in mean climate, even after removing the effect of phylogeny from climate (Table 14). Relations of similarity matrices, when compared with Krz, and distance matrices are similar to the reported for RS, even though correlations with climate are even stronger for residual matrices (Table 15). When taking into account both the species sample size effect in matrix similarity and the phylogeny effect in climate, the correlation with climatic distances remained significant for both raw (Figures 8A, B) and residual matrices (Figures 8C, D; Table 16).

Table 14. Correlations between similarity matrices obtained by Random Skewers method and distance matrices. Values in bold indicate significant matrix correlation with the Mantel test for $P < 0.05$. Values in italic are for significant correlations at $P < 0.1$.

	Raw Matrices		Residual Matrices	
Distance matrix	Covariance	Correlation	Covariance	Correlation
Phylogeny	*-0.24*	-0.18	-0.07	-0.15
Morphology	**-0.36**	*-0.26*	-0.15	-0.17
Climate	**-0.50**	**-0.41**	**-0.43**	*-0.38*
Morphological residual	**-0.34**	-0.24	-0.15	-0.14
Climatic residual	**-0.45**	*-0.37*	**-0.47**	-0.24

Table 15. Correlations between similarity matrices obtained with Krzanowski Projection method and distance matrices. Values in bold indicate significant matrix correlation with the Mantel test for P < 0.05. Values in italic are for significant correlations at P < 0.1

	Raw Matrices		Residual Matrices	
Distance matrix	Covariance	Correlation	Covariance	Correlation
Phylogeny	*-0.24*	-0.21	-0.24	**-0.34**
Morphology	-0.19	-0.03	*-0.28*	-0.09
Climate	**-0.36**	*-0.33*	**-0.52**	**-0.51**
Morphological residual	-0.11	0.08	-0.23	0.08
Climatic residual	-0.27	-0.26	**-0.47**	**-0.51**

Table 16. Correlations between matrix similarity without the effect of sample size wtih climatic distances (original: Ln (Mean climate) and climatic distances without the effect of phylogeny. Values in bold are significant matrix correlations at P < 0.05 by Mantel test. Values in italic are significant at P < 0.1 by Mantel test.

Matrix similarity without sample size effect		
Raw matrices	**Climate**	**Climate residual**
Covariance (RS)	**-0.48**	**-0.45**
Covariance (Krz)	**-0.34**	*-0.28*
Correlation (RS)	**-0.38**	*-0.36*
Correlation (Krz)	**-0.40**	*-0.36*
Residual matrices	**Climate**	**Climate residual**
Covariance (RS)	**-0.47**	**-0.58**
Covariance (Krz)	**-0.52**	**-0.49**
Correlation (RS)	**-0.61**	**-0.68**
Correlation (Krz)	**-0.39**	-0.24

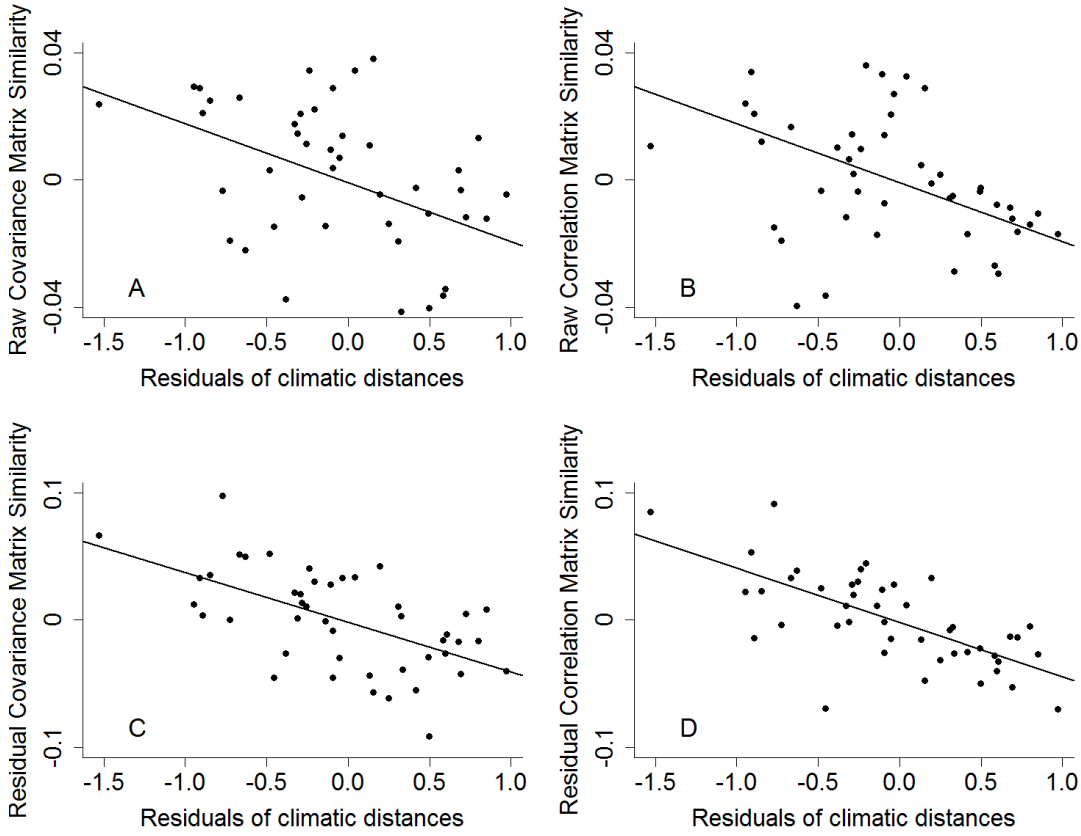


Figure 8. Correlations between raw (A, B) and residual (C, D) similarity matrices (discounting sample size effect) and residual climatic distances (without effect of phylogeny). Similarity matrices of both covariance and correlation matrices were corrected for sample size effect by using the residuals of a linear regression between pairwise matrix similarity and the harmonic mean of number of specimens for each species (see Online Appendix C). Similarly, climatic distances were corrected for phylogenetic effect by using the residuals of climatic distances regressed on phylogenetic distances. A) $r = -0.45$ $P < 0.05$; B) $r = -0.36$ $P < 0.1$; C) $r = -0.58$ $P < 0.05$; D) $r = -0.68$ $P < 0.05$. Matrix correlation significance was assessed with the Mantel test.

2.5. Discussion

The skull of anuran amphibians is a unique system to study modularity due to its distinct development in which all bones are derived from the neural crest (Piekarski et

al. 2014). Piekarski et al. (2014) suggest that this major shift in skull development in relation to other tetrapods evolved in association with the extreme biphasic cranial ontogeny typical of anurans, which present huge cranial remodeling in metamorphosis (Trueb 1985; Rose and Reiss 1993; Kerney et al. 2012). Herein, we show that toad species of the *R. granulosa* group present high P-matrix stability associated mostly to functional modularity. Furthermore, divergence in P-matrices across species is associated to mean climatic differences, which may be connected to the species reproductive ecology.

Covariance and correlation patterns of skull traits are highly conserved in the toad species of the *R. granulosa* group, even when size variation is removed. Along with the high similarity in P-matrices, we also found a common modular organization of the toad species skulls related especially to function, as predicted by the morphological integration theory (Olson and Miller 1958; Berg 1960). All species of the *R. granulosa* group presented a correlation pattern of the skull traits compatible with expected functional interactions (Trueb 1993) and with expected TH-sensitivity of skull bones, even though the hormonal hypothesis is based in newts (Ivanovic´and Kalezic´ 2010).

The differences in the detection of the Snout and Suspensorium modules in raw matrices are related to subtle allometric differences among species. *R. pygmaea*, *R. dorbingyi* and *R. fernandezae* present higher association of Suspensorium traits with PC1, whereas the rest of the species present Snout traits more associated to the PC1. Since size-associated variation is so high in the toad species, these small allometric differences are enough to promote divergence in modularity patterns. This difference disappears after size variation removal, being the Snout module present in all species (including the external group), as well as the Neurocranium and Total

functional modules. No developmental hypothesis was supported in the toads for raw P-matrices (Table 11).

The reduction in similarity when using residual P-matrices indicates that part of the similarity was due to size-related responses of the species when subjected to random selection. This reduction was not seen in Krz comparisons. The response to selection is biased towards the dimensions of maximum variance (Lande 1979; Schluter 1996) and the size dimension (PC1) presents the highest variance among individuals in all toad species. Size variance in toads accounts for at least half the total variation, reaching up to 80%, being a very high relative amount of the total variation in the P-matrices. When compared to mammals, the toad species resemble marsupial species, which present high constraint in skull evolution (Marroig et al. 2009; Shirai and Marroig 2010). As a consequence of this high size variation, the modularity indexes found in the toads is much lower than for most mammals (Porto et al. 2009). Yet, we must be aware that size variance is probably over-estimated in the toads because we could not control for size variation due to age. The higher modularity signal detected in residual compared to raw matrices also occurs in mammal species (Marroig et al. 2004, Porto et al. 2013), confirming that size variation obscures modularity since growth increases correlations among all traits (Zelditch 1988). The higher modularity indexes found for the Branchial developmental module compared to other modules in residual matrices is intriguing, yet this module is only composed of two bones (frontoparietal and occipital bones). Therefore, we argue that the support for developmental modules is still low even after size variation removal. Still, we must take into account that all modularity hypotheses tested are very crude descriptions of complex processes, especially in the case of development. Thus, the use of other methods for detecting modules in the toad skulls

may prove to be interesting.

Considering that the P-matrix is composed of the G-matrix and the environmental matrix (E-matrix; **P** = **G** + **E**, Falconer and Mackay 1996), similarity among P-matrices of different species, distributed in distinct environments, suggests that P and G-matrices are also similar. Otherwise, the E-matrices would have to compensate for the differences in P and G-matrices, which is highly improbable (Marroig and Cheverud 2001). Hence, by assuming that P-matrix stability also indicates G-matrix stability, we may infer that the functional modularity detected in the toads reflects co-inheritance of functional related sets (Riedl 1977; 1978; Wagner 1996). Riedl's (1978) hypothesis is that the genetic and epigenetic systems would 'imitate' the functional relations among traits. Therefore, stability of genetic correlations among skull traits in the toad species suggests conservation of internal stabilizing selection patterns through evolutionary time (Lande 1980; Cheverud 1982; 1984; Arnold et al. 2001). That is, traits that interact in functional performance present fitness interactions, in which the mean fitness of a trait is dependent on the mean phenotypic value of other traits that perform the same function (Cheverud 1982; 1984), causing functional traits to evolve together (evolutionary integration; Cheverud 1996). Other empirical studies have shown an important contribution of functional interactions in shaping integration patterns, with or without a compatibility with developmental expectations (e.g. Zelditch and Carmichael 1989; Monteiro et al. 2005; Young and Badyaev 2006; Zelditch et al. 2008; Monteiro and Nogueira 2010; Klingenberg et al. 2010). Nevertheless, the anuran skull presents a unique example of a mismatch between developmental and functional modularity associated to a dramatically different development.

The relative importance of functional and hormonal modularity in the toads in

contrast to developmental modularity might be related to the two-phase ontogeny of anurans. Embryonic and larval development in most anurans happen in an aquatic phase, in which CNC migration of skull cells from different streams occurs and the skull performs tadpole functions. Later on, metamorphosis begins, and in the case of anurans, it is an extreme remodeling process in which several cartilages are re-absorbed or ossified by induction of TH regulation (Trueb 1985; Rose and Reiss 1993). Metamorphosis mediates the transition of the individuals from the aquatic phase to a terrestrial phase, in which the skull performs adult functions. Consequently, we must consider which processes can we capture by measuring adult toads (in analogy to the Palimpsest model in mammals proposed by Hallgrímsson et al. 2009). Thus, we suggest that the CNC developmental modularity signal gets erased by metamorphosis, and by measuring adult skulls, we capture the signal of hormonal modularity connected to metamorphosis and functional modularity related to the functions of adult toads. This scenario does not completely exclude a role for development after metamorphosis, given that skull bones continue to change after metamorphosis (Trueb and Hanken 1992) and epigenetic interactions, such as muscle-bone interactions (Swiderski and Zelditch 2011), must be important to determine the skull variation and its interaction with the fitness landscape. Still, the toad skull provides a good example of the importance of function in structuring the P-matrix and possibly the G-matrix as well.

Although toad species P-matrices are very similar, they are not identical. Differences in matrix similarity are only related to phylogeny in V/CV raw matrices, indicating that these differences are due to distinct proportions of size variation among species. After size variation is removed, matrix similarity is not related to either phylogenetic or mean morphological distances. Yet, mean morphology is

related to phylogeny in the toad species. This association shows that the phenotypic means were evolving along with the species diversification of the *R. granulosa* toad species group and yet covariance patterns were kept stable. The same pattern was found in New World monkeys cranial diversification (Marroig and Cheverud 2001) and in *Anolis* lizards morphological diversification (Revel et al. 2007; Kolbe et al. 2011), supporting Lande's (1980) selection-mutation balance model. We also found significant associations between climatic distances with phylogeny, as well as with morphological distances. These relations are interesting because Narvaes and Rodrigues (2009) have speculated that the diversification of the *R. granulosa* group could be associated to climatic changes, especially because species distributions are predominantly allopatric and partly associated to specific open habitats.

Climate also seems to be the main factor structuring P-matrix similarity in the toad species. This result suggests a role for external stabilizing selection related to climatic variables underlying P-matrix structure across the toad species. As SRD results show, differences in residual matrices are concentrated in the snout module. Therefore, species in more divergent climates have more divergent covariance structure for the snout bones, which is probably associated to snout function. The snout region of amphibians protects and supports the olfactory capsule (Duellman and Trueb 1994), which includes the vomeronasal organ. Vomeronasal function is related to sampling water cues (Doving et al. 1993), such as detection of conspecific chemical cues in tadpoles (Halpern and Martínez-Marcos 2003). However, Jungblut et al. (2011) have shown that a new sensorial epithelium develops at metamorphic climax in the olfactory system of *Rhinella arenarum*, which is probably related to the detection of water-born odors in juveniles and adult toads. The authors suggest that this new epithelium may be used by adult toads to detect water molecules at the time

of reproduction (Jungblut et al. 2011). The species of the *R. granulosa* group are known for being ground-dwellers, remaining in holes especially at the dry season (Gallardo 1965; Narvaes and Rodrigues 2009; Pereyra et al. 2015). They are also known to have explosive reproduction, spawning in ephemeral pools after strong episodes of rainfall in which they emerge from the holes. Accordingly, we may hypothesize that difference in snout covariance pattern among the toad species is due to the relative importance of snout function in detecting water cues for reproduction, depending on the rain patterns of the different habitats the species inhabit. Hence, species in more arid habitats, such as *R. granulosa* in the Caatinga and *R. mirandaribeiroi* in the Cerrado, could be subjected to a distinct pattern of external stabilizing selection in the snout traits than species distributed in more humid habitats, such as R. *dorbignyi*, *R. fernandezae* and *R. bergi* in the Chaco. The shape of the snout fitness landscape might depend on the climate the species are exposed to, in which reproductive success is more influenced by snout function in more arid environments. Yet, we cannot rule out the hypothesis that directional selection acting on trait means of the snout region could have produced a change in covariance patterns across species in distinct habitats. Directional selection can potentially produce changes in the G-matrix (Melo and Marroig 2015) and the association that may exist between snout function and differential reproduction favors an action of directional selection for the toad species. Thus, examining the trait means, especially of the snout region, may provide support for directional selection causing the differences in toad P-matrices. Also, behavioral tests comparing toad species from more arid and less arid habitats might confirm whether the species actually differ in detecting water pools.

2.6. Conclusions

In conclusion, the toad skulls are an example of the impact that functional interactions among traits may have in the fitness landscape. The high support for functional modularity in contrast to the low support for developmental modularity is probably related to the biphasic ontogeny of the toads and metamorphosis. The high P-matrix similarity across toad species is probably associated to conservation of internal stabilizing selection, related to function and hormonal control in metamorphosis. Differences in P-matrix similarity are structured by differences in climate, suggesting the action of external stabilizing selection related to snout function. The skull of toads is a good example of the interplay between internal and external processes acting on the characters covariance matrix.

2.7. References

Ackermann, R.R. 2002. Patterns of covariation in the hominoid craniofacial skeleton: implications for paleoanthropological models. *Journal of Human Evolution* 43, 167–187.

Arnold, S.J. 1981. Behavioral Variation in Natural Populations. I. Phenotypic, Genetic and Environmental Correlations Between Chemoreceptive Responses to Prey in the Garter Snake, Thamnophis elegans. *Evolution*, 35, 489.

Arnold, S.J., Bürger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. 2008. Understading the evolution and stability of the G-matrix. *Evolution*, 62, 2451–2461.

Arnold, S.J., Pfrender, M.E. & Jones, A.G. 2001. The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica*, 112, 9–32.

Arnold, S. J., 1994. Multivariate Inheritance and Evolution: A Review of Concepts. In Boake, Cristine R. B.

Arnold, S. J. & Phillips, P. C. 1999. Hierarchical Comparison of Genetic Variance-Covariance Matrices. II. Coastal-Inland Divergence in the Garter Snake, Tamnophis elegans. *Evolution*, 53, 1516–1527.

Berg, R. L. 1960. The Ecological Significance of Correlation Pleiades. *Evolution*, 14, 171–180.

Cheverud, J. M. 1995. Morphological Integration in the Saddle-Back Tamarin (Saguinus fuscicollis) Cranium. *American Naturalist*, 145, 63–89.

Cheverud, J. M. 1988. A Comparison of Genetic and Phenotypic Correlations. *Evolution*, 42, 958–958.

Cheverud, J. M. 1984. Quantitative Genetics and Developmental Constraints on Evolution by Selection. *Journal of Theoretical Biology* , 110, 155–171.

Cheverud, J.M. 1982. Phenotypic, Genetic, and Environmental Morphological Integration in the Cranium. *Evolution*, 36, 499.

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* 302B, 424–435.

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., Routman, E.J. & Irschick, D.J. 1997. Pleiotropic Effects of Individual Gene Loci on Mandibular Morphology. *Evolution*, 51, 2006.

Cheverud, J.M., Wagner, G.P. & Dow, M.M. 1989. Methods for the Comparative Analysis of Variation Patterns. *Systematic Zoology* , 38: 201.

Døving, K.B., Trotier, D., Rosin, J.-F. & Holley, A. 1993. Functional architecture of the vomeronasal organ of the frog (genus Rana). *Acta Zoologica* , 74, 173–180.

Duellman, W. E. & Trueb, L. 1994. *Biology of Amphibians*, 2nd ed. The Johns Hopkins University Press. Blatimore, Maryland.

Duran, A. & Pie, M. R., 2015. Tempo and mode of climatic niche evolution in Primates. *Evolution*, 69.

Falconer, D. S. & Mackay, T. F. C. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman Group Ltd. Essex, England.

Gallardo, J. M. 1965. The species Bufo granulosus Spix (Salientia: Bufonidae) and its geographic variation. *Bulletin of the Museum of Comparative Zoology*, 134, 107–138.

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

Hallgrímsson, B., Jamniczky, H., Young, N.M., Rolian, C., Parsons, T.E., Boughner, J.C. & Marcucio, R.S. 2009. Deciphering the Palimpsest: Studying the Relationship Between

Morphological Integration and Phenotypic Covariation. *Evolutionary Biology*, 36, 355–376.

Halpern, M. 2003. Structure and function of the vomeronasal system: an update. *Progress in Neurobiology*, 70, 245–318.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

Ivanović, A. & Kalezić, M.L. 2010. Testing the hypothesis of morphological integration on a skull of a vertebrate with a biphasic life cycle: a case study of the alpine newt. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 314B, 527–538.

Jones, A.G., Arnold, S.J. & Bürger, R. 2007. The mutation matrix and the evolution of evolvability. *Evolution*, 61, 727–745.

Jones, A.G., Arnold, S.J. & Burger, R. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution*, 58, 1639–1654.

Jones, A.G., Arnold, S.J., Bürger, R. & Houle, D. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution*, 57, 1747–1760.

Jungblut, L.D., Pozzi, A.G. & Paz, D.A. 2011. Larval development and metamorphosis of the olfactory and vomeronasal organs in the toad Rhinella (Bufo) arenarum (Hensel, 1867): Nasal sensory epithelia ontogeny in Rhinella arenarum. *Acta Zoologica*, 92, 305–315.

Kenney-Hunt, J.P., Wang, B., Norgard, E.A., Fawcett, G., Falk, D., Pletscher, L.S., Jarvis, J.P., Roseman, C., Wolf, J. & Cheverud, J.M. 2008. Pleiotropic Patterns of Quantitative Trait Loci for 70 Murine Skeletal Traits. *Genetics*, 178, 2275–2288.

Kerney, R.R., Brittain, A.L., Hall, B.K. & Buchholz, D.R. 2012. Cartilage on the move: Cartilage lineage tracing during tadpole metamorphosis. *Development, Growth & Differentiation*, 54, 739–752.

Klingenberg, C.P., Debat, V. & Roff, D.A. 2010. Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. *Evolution*, 74, 481-490.

Kolbe, J.J., Revell, L.J., Szekely, 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.

Krzanowski, W.J. 1979. Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74, 703–707.

Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. *Genetics*, 94, 203–215.

Lande, R., 1979. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. *Evolution*, 33, 402.

Lande, R. & Arnold, S.J. 1983. The Measurement of Selection on Correlated Characters. *Evolution* 37, 1210.

Lessells, C. M. & Boag, P. T. 1987. Unrepeatable Repeatabilities: A Common Mistake. *The Auk*, 104, 116–121.

Lofsvold, D. 1986. Quantitative Genetics of Morphological Differentiation in Peromyscus. I. Tests of the Homogeneity of Genetic Covariance Structure Among Species and Subspecies. *Evolution*, 40, 559–573.

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., 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., Vivo, M. & Cheverud, J.M. 2003. Cranial evolution in sakis (Pithecia, Platyrrhini) II: evolutionary processes and morphological integration: Cranial evolution in sakis. *Journal of Evolutionary Biology*, 17, 144–155.

Melo, D., Garcia, G., Hubbe, A., Assis, A. P. & Marroig, G. 2015. EvolQG - An R package for evolutionary quantitative genetics. *F1000 Research*, 4, 925.

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.

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.

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.

Narvaes, P. & Rodrigues, M.T. 2009. Taxonomic revision of Rhinella granulosa species group (Amphibia, Anura, Bufonidae), with a description of a new species. *Arquivos de Zoologia*, 40, 1–73.

Pereyra, M.O., Baldo, D., Blotto, B.L., Iglesias, P.P., Thomé, M.T.C., Haddad, C.F.B., Barrio-Amorós, C., Ibáñez, R. & Faivovich, J. 2015. Phylogenetic relationships of toads of the *Rhinella granulosa* group (Anura: Bufonidae): a molecular perspective with comments on hybridization and introgression. *Cladistics*, 31, 1-18..

Piekarski, N., Gross, J.B. & Hanken, J. 2014. Evolutionary innovation and conservation in the embryonic derivation of the vertebrate skull. *Nature Communications*, 5, 5661.

Porto, A., de 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, 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, 3305–3322.

Pramuk, J.B. 2006. Phylogeny of south American Bufo (Anura: Bufonidae) inferred from combined evidence. *Zoological Journal of the Linnean Society*, 146, 407–452.

Revell, L.J., Harmon, L.J., Langerhans, R.B. & Kolbe, J.J. 2007. A phylogenetic approach to determining the importance of constraint on phenotypic evolution in the neotropical lizard Anolis cristatellus. *Evolutionary Ecology Research*, 9, 261–282.

Riedl, R. 1977. A systems-analytical approach to macro-evolutionary phenomena. *Quarterly Review of Biology*, 52, 351–370.

Riedl, R. 1978. *Order in Living Organisms: Systems Analysis of Evolution*. John Wiley and Sons Ltd. New York, USA.

Roff, D. A. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. *Heredity*, 74, 481–490.

Rose, C. S. & Reiss, J. O., 1993. Metamorphosis and the Vertebrate Skull: Ontogenetic Pattern and Developmental Mechanisms. *in* Hanken, J. and Hall, B. K. eds. *The Skull: Development*. Chicago, USA.

Sanabria, E., Quiroga L., Arias, F. & Cortez, R. 2010. A new species of Rhinella (Anura: Bufonidae) from Ischigualasto Provincial Park, San Juan, Argentina. *Zootaxa*, 2396, 50-60.

Sanger, T.J., Mahler, D.L., Abzhanov, A. & Losos, J.B. 2012. Roles for modularity and constraint in the evolution of cranial diversity among *Anolis*. *Evolution*, 66, 1525–1542.

Schluter, D. 1996. Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution*, 50, 1766.

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.

Simon, M.N. & Marroig, G. 2015. Landmark precision and reliability and accuracy of linear distances estimated by using 3D computed micro-tomography and the open-source TINA Manual Landmarking Tool software. *Frontiers in Zoology* , 12.

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.

Thomson, K. S. 1993. Segmentation, the Adult Skull and the problem of Homology. *in* Hanken, J. and Hall, B. K. eds. *The Skull: Functional and Evolutionary Mechanics*. Chicago, USA.

Trueb, L. & Hanken, J. 1992. Skeletal development in Xenopus laevis (Anura: Pipidae). *Journal of morphology*, 214, 1–41.

Trueb, L. 1993. Patterns of Cranial Diversity among the Lissamphibia. *in*: Hanken, J. and Hall, B. K. eds. *The Skull: Functional and Evolutionary Mechanics*. Chicago, USA.

Trueb, L. 1985. A summary of osteocranial development in anurans with notes on the sequence of cranial in Rhinophrynus dorsalis (Anura: Pipoidea: Rhinophrynidae). *South African Journal of Science*, 81, 181–185.

Turelli, M., 1988. Phenotypic Evolution, Constant Covariances, and the Maintenance of Additive Variance. *Evolution*, 42, 1342.

Waddington, C. H. 1957. *The Strategy of the Genes: a Discussion of Some Aspects of Theoretical Biology*. *in* Allen, G. ed.

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

Wagner, G.P. & Altenberg, L. 1996. Perspective: Complex Adaptations and the Evolution of Evolvability. *Evolution*, 50, 967.

Wagner, G.P., Kenney-Hunt, J.P., Pavlicev, M., Peck, J.R., Waxman, D. & Cheverud, J.M. 2008. Pleiotropic scaling of gene effects and the "cost of complexity." *Nature*, 452, 470–472.

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

Young, R.L. & Badyaev, A.V. 2006. Evolutionary persistence of phenotypic integration: influence of developmental and functional relationships on complex trait evolution. *Evolution*, 60, 1291–1299.

Zelditch, M. L., 1988. Ontogenetic variation in patterns of phenotypic integration in the laboratory rat. *Evolution*, 42, 28–41.

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* Hallgrímsson, Benedikt and Hall, Brian K., eds. *Epigenetics: Linking Genotype and Phenotype in Development and Evolution*. Berkeley, California.

Zelditch, M.L., Wood, A.R., Bonett, R.M. & Swiderski, D.L. 2008. Modularity of the rodent mandible: integrating bones, muscles, and teeth. *Evolution & development*, 10, 756–768.

Capítulo 3

High Evolutionary Constraints Limited Adaptive Responses to Past Climatic Changes in Toad Skulls

3.1. Abstract

Correlations among traits interacting to build a complex structure may act as constraints, deflecting the evolutionary response from the direction of natural selection. Populations with high genetic constraints may have low ability to respond to environmental alterations and become more vulnerable to local extinction. Here we investigated the interplay of direction and strength of selection with evolutionary constraints and climatic variation driving skull divergence of related toad species. By combining approaches from both quantitative genetics and comparative methods, we show that past selection was probably linked to changes in precipitation seasonality across species, favouring local morphological changes in the skull. However, most evolutionary responses were in the direction of size, the dimension of highest within-species phenotypic variance. Our results indicate that species with high variance in size have low evolutionary potential to respond to climatic changes, unless the selective pressures are orthogonal to or aligned with directions of high genetic variance.

3.2. Introduction

All organisms are complex systems composed of many traits that interact with each other in some way. The study of genetic constraints is fundamental to the understanding of the evolution of such complex systems (Arnold 1992; Futuyma 2010). The relations among traits within a complex structure are represented by the additive genetic variance covariance matrix (the G-matrix; Lande 1979), and its interaction with the adaptive landscape is the main phenomenon treated by quantitative genetics theory (Arnold et al. 2001; Steppan et al. 2002). Genetic correlations among traits can constrain the evolutionary trajectories along the adaptive landscape, biasing the multivariate response to selection towards directions of high within-population variance (Lande 1979; Arnold et al. 2001; Walsh and Blows 2009). The first axis of maximum variance of the G-matrix (g_{max}; combination of traits with highest genetic variance) is predicted to act as a line of least evolutionary resistance, facilitating population divergence in its direction (Schluter 1996). Empirical studies have shown this effect in a variety of natural systems (Marroig and Cheverud 2005; Renaud et al. 2006; Chenoweth et al. 2010), indicating that within-species variational properties might indeed interfere with divergence across species.

Quantitative genetics theory has been used to model the response of populations to environmental fluctuations (Lande & Lynch 1993; Hellman & Pineda-Krch 2007; Chevin 2012). These models posit that environmental alterations may produce shifts in the adaptive landscape, as indeed has been seen in nature (Gibbs & Grant 1987; Linhart & Grant 1996). In the context of complex characters, combinations of traits with low genetic variance limit adaptive evolution. Thus, genetic constraints may facilitate, hamper or even prevent populations from adapting

to these new selective pressures (Hellman & Pineda-Krch 2007; Chevin 2012). Which scenario will occur depends on the degree of alignment between the direction of selection and directions of high genetic variance (Arnold et al. 2001; Hellman & Pineda-Krch 2007; Chevin 2012). Climate change is expected to act as a new and probably strong selective pressure, given that it influences several aspects of species biology, on ecological as well as on evolutionary timescales (Bradshaw & Holzapfel 2006; Parmesan 2006; Hoffmann & Sgrò 2011). In fact, in the last decade, the threat imposed by climate change on biodiversity has received increased attention in the scientific community (Groves et al. 2012; Gillson et al. 2013; Urban 2015). Nevertheless, the interaction of climate and constraints has mainly been investigated in ecological timescales (Etterson & Shaw 2001; Etterson 2004; Teplitsky et al. 2001), overlooking the potential interplay between climate and constraints on phenotypic evolution at a wider phylogenetic scale. Knowledge of how evolutionary forces arising from climate change interacted with genetic properties of organisms in past divergence may hint about the evolutionary potential of species in face of future climatic change (Teplitsky et al. 2014).

Ectothermic vertebrates are especially threatened by climatic alterations. Iconic examples are amphibian species and their high susceptibility to environmental changes, as evidenced by their worldwide population declines (Pounds et al. 2006). We therefore chose to study a group of closely related toad species (*Rhinella granulosa* species group), thus incorporating the dimension of phylogeny into the analysis, and we used skull morphology as a model of complex character evolution associated with climatic variation. Specifically, we aimed to investigate the relative contributions of natural selection and constraints in shaping skull diversity in the studied toad species and the potential interplay of those factors with climate. This is

the first empirical study to evaluate the effect of evolutionary constraints on a high-dimensional system (21 linear distances in the skull) interacting with climate changes. We estimated past directional selection along a species phylogeny and successfully linked its direction with climatic variation across species. However, the major role that evolutionary constraints played in toad skull diversification indicates that these species will not respond to selection due to climate change unless selection is orthogonal to or aligned with directions of high genetic variance.

3.3. Methods

3.3.1. Sample and Linear Distances

We analysed 1,034 specimens of 11 toad species belonging to the *Rhinella granulosa* group. We scanned all the specimens with an X-ray micro-tomography system (SkyScan 1176, Konitch, Belgium) with methods and parameters described in Simon & Marroig (2015; Capítulo 1). We placed 22 three-dimensional landmarks in the toad skulls using TINA Manual Landmarking Tool software (Schunke et al. 2012) in order to extract 21 linear distances allocated into three distinct functional sets (Figure 1).

3.3.2. Testing for Random Drift: Regression and PC Correlation Tests

We used a molecular Bayesian phylogeny of the toad species (Pereyra et al. 2015) to investigate the relevance of random drift in driving skull diversification with two distinct tests. The principle behind the random drift test is that drift may be modeled as a diffusion process with no directionality, in which the variation among descendant populations will be proportional to the variation within the ancestral population (Lande 1979; Lofsvold 1988; Felsenstein 1988; Arnold et al. 2001; Hohenlohe &

Arnold 2008).

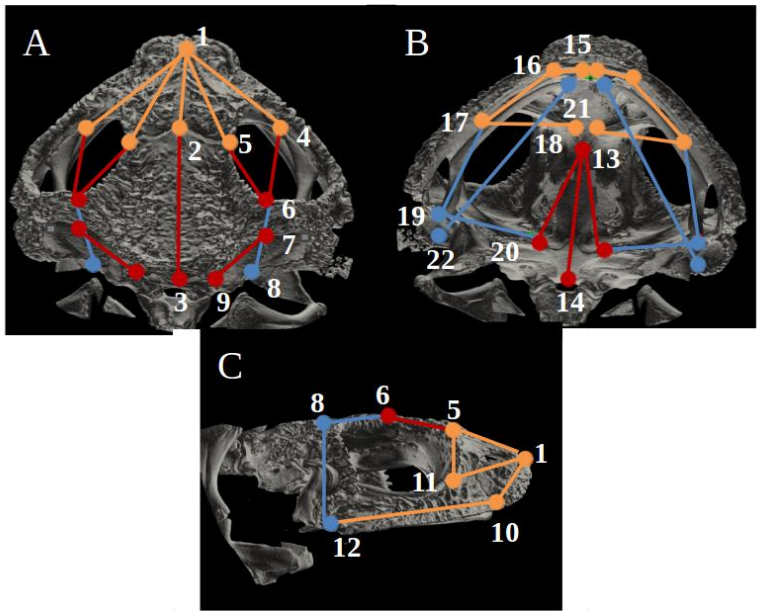


Figure 1. Numbered landmarks and distances in the toad skull. Different colours indicate hypothetical functional sets to which distances belong: orange – snout; red – neurocranium; blue – suspensorium.

The evolutionary landscape of populations under drift is a flat landscape, with no adaptive peaks (Figure 2). The regression test rejects drift as the null hypothesis when variation in the ancestral population (within-species: **W**) is not proportional to the variation between-species (**B**). Here we modeled morphological variation of ancestral populations as the pooled within-group variance and covariance phenotypic matrix (**W**). Ackermann and Cheverud (2002) developed a simplified regression test to compare the within-group variation (**W**-matrix) and the between-group variation (**B**-matrix) by using the principal components of **W**, which are maximum variance directions (eigenvectors: linear combinations of the original traits) in the morphospace and uncorrelated to each other. Therefore, **W** is represented by its

eigenvalues, i.e. the variance explained by each of its PCs. The **B** matrix corresponds to the variance of the population (species) scores in **W**'s PCs, calculated by multiplying species 21-trait means by **W**'s eigenvectors (normalized PC loadings). Hence, **B** indicates the variation among species in each eigenvector of **W**. To test whether **W** and **B** are proportional, we transformed the relation between both matrices into a linear regression in logarithmic scale:

$$lnB_i = \ln\left(\frac{t}{N_e}\right) + \beta \ln W_i; \quad (1)$$

where B_i is the between-population variance, W_i is the within-population variance for the *i*th eigenvector, *t* is the time in generations and N_e is the effective population size (Ackermann & Cheverud 2002). If population divergence was due to random drift, the regression slope (β) will be equal to one, meaning that **W** and **B** are proportional. If β is different from one, the divergence pattern does not correspond to the expectation by random drift. We performed five tests in total, following the phylogeny for clades with four or more taxa only, given that statistical power is determined by the number of taxa in the regression. For each test, we used the corresponding node **W** matrix, weighted by sample size of the descendant clade (terminal and/or node matrices). We called the 95% confidence interval (CI) of the empirical regression slope the '95% CI theoretical', and if it did not contain 1.0, we rejected drift. The phylogeny does not include *R. nattereri*, which was recovered as polyphyletic with *R. merianae*; and we had to exclude *R. bernardoi* because we could not access any specimens from this species. However, *R. bernardoi* is a sister taxon to *R. fernandezae* + *R. dorbignyi* (Pereyra et al. 2015), which form a three taxa group that would not be tested for random drift due to low sample size.

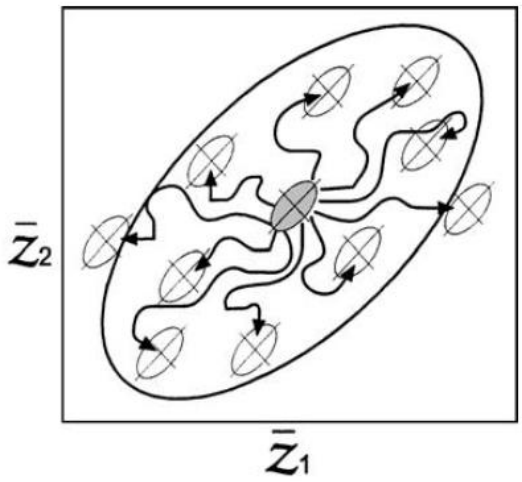


Figure 2. Flat bivariate adaptive landscape. The scheme shows an ancestral G-matrix (gray ellipse in the center; lines inside the ellipse represent two directions of highest genetic variance) and several descendant populations, also represented by G-matrices, on a flat adaptive landscape. z_1 and z_2 are the averages of any two traits and crosses inside ellipses indicate their specific averages. Solid curve arrows indicate the evolutionary trajectories of the descendant populations, which are random walks. The larger ellipse is the 95% confidence ellipse of descendant population average traits (between-groups G-matrix). We expect the between-groups matrix to be proportional to the ancestral G-matrix under random drift. Extracted from Arnold et al. (2001).

We also constructed a 95% CI for the empirical regression slope under the assumption of drift. We used a simulation approach developed by Prôa et al. (2013) that takes into account the empirical ancestral matrix (**W**) and the empirical between-species matrix (**B**), but also a range of values of t/Ne. The B-matrix was calculated as the difference between the total cross-product sum of squares (considering all specimens of all species, but not using species as a factor in the model) and the within-group cross-product sum of squares (using species as a factor). In the B-matrix calculation, we used the residuals of linear models for each species, controlling for sex and locality differences among individuals, and we added to each individual

within a species its average distances so we could maintain between-species variation. We used a range of t/Ne values (100, 1,000 and 10,000) based on the estimated divergence time of the toad species (12 to 35 Ma; Rowe & Beebee 2004) and estimated generation time and Ne in *Bufo calamita* (4.5 years and 100, respectively; Van Boxclaer et al. 2010). Descendant populations were simulated from the ancestral matrix using a multivariate normal distribution with zero mean and t/Ne*$\mathbf{B}$ variance (drift expectation). Then, the simulated populations' mean vectors were calculated, projected in $\mathbf{W}$ eigenvectors to obtain the simulated $\mathbf{B}$. The simulations were done 1,000 times, and we used the average 95% CI for the slopes of the regressions.

In addition to the random drift test, we also computed correlations between species scores in $\mathbf{W}$ eigenvectors. These PCs are by definition orthogonal to each other at the within-group level, and therefore independent. Under diversification through random drift, we expect the between-group scores in the PCs to remain independent. However, under the alternative hypothesis of natural selection producing population divergence, species PC scores could be correlated if selective pressures on these dimensions were also correlated ('selective covariance'; Felsenstein 1988). Selective covariance implies that different sets of traits confer higher fitness in relation to different aspects of the same environment (or to different functional performances), hence co-selection of these sets of traits will enhance average population fitness as a whole. We correlated n-1 number of PCs, with n corresponding to the number of taxa in each node tested. We performed both Bartlett's χ^2 test for significant correlations of the PC correlation matrix and univariate tests of PC pairwise correlations. We rejected drift when at least one PC correlation was significantly different than zero. Drift tests rely on the assumption that the G or P-matrices are stable through evolutionary time (Felsenstein 1988). We

constructed variance covariance P-matrices for all species in the *R. granulosa* species group (except for *R. azarai*), controlling for sex and locality mean differences. The species P-matrices are highly similar (mean similarity = 0.90 $\pm$ 0.05; Capítulo 2). Thus, these toad species are suitable for the drift tests.

3.3.3. Evolutionary Regression of Morphology on Climate

To explore whether variation in species skull morphology (species scores on the first four morphological PCs of **W**) can be explained by variation in climatic variables, we performed evolutionary regressions (Hansen et al. 2008). This analysis assumes an Ornstein-Uhlenbeck process, which fits a scenario where we expect populations to be subjected to one or many selective regimes (Hansen 1997; Butler and King, 2004; Hansen et al. 2008). The expected trait value of a species is a weighted average of the effects of the adaptive optimum and the ancestral state (i.e. phylogenetic inertia). In order to estimate the regression intercept and coefficients, the analysis uses two parameters: adaptation half-life ($t_{1/2}$), the time it takes for an adaptation to a new regime to overcome the influence of ancestral states (i.e. the strength of phylogenetic inertia); and v_y, the expected variance between species that evolved for a period of time under the same selective regime (Hansen et al. 2008). Given that we have only 11 species, we calculated the regression parameters for weak phylogenetic inertia ($t_{1/2}$ reaching only 10% of total phylogeny length) and strong phylogenetic inertia ($t_{1/2}$ reaching the full phylogeny length). We calculated evolutionary regressions for the first six morphological PCs on climatic variables. We constructed a within-group V/CV climatic matrix by using z-score transformed bioclimatic variables (*Worldclim* database; Hijmans et al. 2005) obtained for the entire distribution range of the *R. granulosa* species group (species climatic means in Table 1). We extracted the PCs of

the climatic matrix and by inspecting the first and second climatic PCs (Table 2), we selected as predictors only climatic variables with the highest loadings. The first climatic PC is a contrast between temperature seasonality (mean diurnal range) and almost all precipitation variables (especially annual precipitation). The second climatic PC contrasts mean and maximum temperatures and precipitation seasonality with the amount of rain in the warmest and driest periods

3.3.4. Evolutionary Response and Selection Gradient Reconstruction

The evolutionary response is a vector of mean differences in skull traits between ancestral-descendant populations ($\Delta\mathbf{z}$). We estimated ancestral trait means using the maximum likelihood approach developed by Schluter et al. (1997). We also estimated ancestral states with linear parsimony (Swofford & Maddison 1987) and the results were similar. Selection gradient reconstruction was done by rearranging the multivariate breeder's equation (Lande 1979; Lande & Arnold 1983):

$$\beta = \mathbf{W}^{-1} \Delta\mathbf{z}, \qquad (2)$$

where $\mathbf{W}^{-1}$ is the inverse of the pooled within-group V/CV matrix for each node in the phylogeny. The inversion of $\mathbf{W}$ is problematic because it is estimated with error, accumulating noise in the smallest eigenvalues (Marroig et al. 2012). Given that the smallest eigenvalues dominate the inverted matrix, causing biases in the estimation of β, we controlled for noise in the ancestral matrices using the approach delineated in Marroig et al. (2012). We retained the first seven eigenvalues and extended the seventh one to the next 14 eigenvalues of the ancestral matrices. To test the hypothesis that the evolutionary response is aligned with the direction of selection, we correlated the normalized $\Delta\mathbf{z}$ with the normalized β vectors for each branch in the phylogeny (Marroig & Cheverud 2005).

Table 1. Species means for the climatic variables. Species means were calculated after extracting the climatic variables from all localities where the species were ever collected (Narvaes and Rodrigues 2009). R. aza. = *R. azarai*; R. cent. = *R. centralis*; R. humb. = *R. humboldti*; R. meri. = *R. merianae*; R. mira. = *R. mirandaribeiroi*; R. maj. = *R. major*; R. berg. = *R. bergi*; R. pygm. = *R. pygmaea*; R. dorb. = *R. dorbignyi*; R. fern. = *R. fernandezae*.

Variables	Description	*R. aza.*	*R. cent.*	*R. humb.*	*R. meri.*	*R. gran.*	*R. mira.*	*R. maj.*	*R. berg.*	*R. pygm.*	*R. dorb.*	*R. fern.*
BIO1	Annual Mean Temperature (°C)	22.1	26.9	26.1	26.5	24.0	25.4	24.1	21.4	23.1	16.4	18.0
BIO2	Mean Diurnal Range (°C)	11.8	8.3	9.6	9.4	10.0	11.8	11.7	12.8	8.3	11.1	11.2
BIO4	Temperature Seasonality (°C)	33.4	5.8	5.9	5.1	13.2	9.0	23.4	38.3	18.6	43.8	43.6
BIO5	Maximum Temperature of Warmest Month (°C)	32.7	32.8	32.3	32.4	31.0	33.3	33.3	33.4	30.0	29.6	30.9
BIO6	Minimum Temperature of Coldest Month (°C)	11.4	21.9	20.3	21.2	16.8	16.8	14.2	9.6	15.8	5.9	7.1
BIO8	Mean Temperature of Wettest Quarter (°C)	23.5	26.3	25.8	26.0	24.4	25.6	26.1	25.3	24.6	17.8	21.7
BIO9	Mean Temperature of Driest Quarter (°C)	18.2	27.0	26.1	26.8	23.2	24.7	21.4	16.5	20.7	14.7	13.8
BIO10	Mean Temperature of Warmest Quarter (°C)	26.1	27.7	26.8	27.1	25.4	26.4	26.8	26.1	25.5	22.1	23.6
BIO11	Mean Temperature of Coldest Quarter (°C)	17.7	26.2	25.3	25.8	22.1	24.1	20.9	16.4	20.7	11.0	12.5
BIO12	Annual Precipitation (mm)	1560.8	1896.8	1706.8	1959.4	960.9	1616.9	1418.0	1099.7	1079.0	1070.3	1118.6
BIO13	Precipitation of Wettest Month (mm)	190.3	322.8	265.4	326.7	180.2	295.9	225.5	149.4	156.3	112.3	128.9
BIO14	Precipitation of Driest Month (mm)	76.5	11.4	37.6	60.3	18.3	8.9	29.7	26.6	35.0	65.8	59.2
BIO15	Precipitation Seasonality (mm)	25.5	66.8	56.5	58.6	72.0	78.7	61.5	46.9	43.4	16.7	23.5
BIO16	Precipitation of Wettest Quarter (mm)	490.0	816.2	700.7	889.5	477.0	810.5	621.1	406.7	421.3	310.4	347.2
BIO17	Precipitation of Driest Quarter (mm)	271.6	60.0	150.3	209.4	67.6	37.7	109.9	106.4	115.1	216.6	197.6
BIO18	Precipitation of Warmest Quarter (mm)	444.1	290.0	285.5	278.4	239.2	276.1	380.7	387.6	354.7	279.9	313.1
BIO19	Precipitation of Coldest Quarter (mm)	275.4	669.3	482.7	735.7	191.4	242.0	281.3	116.7	115.1	237.6	202.8

Table 2. Coefficients for the first and second climatic PCs. Coefficients in bold are the highest and were used as predictor variables in evolutionary regressions.

Variables	Description	Scale	PC1	PC2
BIO1	Annual Mean Temperature	°C * 10	-0.13	**0.28**
BIO2	Mean Diurnal Range	°C * 10	**0.29**	0.10
BIO4	Temperature Seasonality	°C * 10	0.17	-0.03
BIO5	Maximum Temperature of Warmest Month	°C * 10	0.03	**0.54**
BIO6	Minimum Temperature of Coldest Month	°C * 10	-0.20	0.18
BIO8	Mean Temperature of Wettest Quarter	°C * 10	0.01	**0.35**
BIO9	Mean Temperature of Driest Quarter	°C * 10	-0.20	0.16
BIO10	Mean Temperature of Warmest Quarter	°C * 10	-0.08	**0.46**
BIO11	Mean Temperature of Coldest Quarter	°C * 10	-0.16	**0.20**
BIO12	Annual Precipitation	mm	**-0.42**	-0.01
BIO13	Precipitation of Wettest Month	mm	**-0.32**	0.09
BIO14	Precipitation of Driest Month	mm	**-0.31**	**-0.21**
BIO15	Precipitation Seasonality	mm	0.14	**0.21**
BIO16	Precipitation of Wettest Quarter	mm	**-0.33**	0.08
BIO17	Precipitation of Driest Quarter	mm	**-0.33**	**-0.21**
BIO18	Precipitation of Warmest Quarter	mm	-0.15	**-0.20**
BIO19	Precipitation of Coldest Quarter	mm	**-0.35**	0.06
		% variation	27.0	37.0

3.3.5. Amount of Divergence and Evolutionary Constraints

We estimated the amount of divergence as the norm of $\Delta\mathbf{z}$ (the vector length) for all branches in the phylogeny (20 in total). To evaluate constraints, we investigated whether the normalized $\Delta\mathbf{z}$ and β were correlated with the first four PCs of the ancestral matrices[9]. We used a multivariate linear regression model to investigate factors influencing the variation in the amount of divergence and in constraint related to p_{max} ($\Delta\mathbf{z}$ x PC1). In the full model, in addition to testing for an effect of branch lengths and strength of selection (magnitude of the β vectors), we included two classes of predictors: (1) factors related to the direction of selection and (2) factors related to the amount of variation in the direction of selection. In the first class, we used the correlation of the direction of selection with PC1 and with PC4 (β x PC1 and

β x PC4, respectively), plus the interactions of β x PC1 and β x PC4 with strength of selection. In the second class, we used evolvability in the direction of selection ($e(\beta)$) plus its interactions with strength of selection and with β x PC1 and β x PC4. Evolvability was calculated as in Hansen and Houle (2008), i.e. the length of the projection of each $\Delta\mathbf{z}$ onto each corresponding β (i.e. the amount of variance in the direction of selection) using the ancestral matrices of each node in the phylogeny standardized by centroid size (to control for scale). We also calculated evolvability in the direction of divergence to compare with evolvability in the direction of selection. Thus initially we had 10 predictors; however, most factors were not significant and were removed from the final model. The factors that were not significant in the full model were strength of selection (F = 0.04 P = 0.96), branch length (F = 0.19 P = 0.83), evolvability (F = 0.92 P = 0.43) and the interactions between evolvability and strength of selection (F = 0.95 P = 0.42), evolvability and $\boldsymbol{\beta}$ x p_{max} (F = 2.6 P = 0.11) and evolvability and $\boldsymbol{\beta}$ x p_{max} (F = 3.52 P = 0.06). Given that the selection reconstruction process implies that selection is responsible for divergence, the aim of the model is to investigate the relative importance of the direction of selection in reference to the axes of highest variance in the P-matrix to account for divergence across species.

3.3.6. Threshold of the Alignment of Selection with p_{max} to Deflect Evolutionary Responses

We simulated 10,000 random selection vectors (β) from a multivariate distribution of zero mean and unit variance and applied them onto the species variance covariance matrices to produce 10,000 evolutionary responses ($\Delta\mathbf{z}$)**.** We then correlated the

normalized $\Delta \mathbf{z}$s and the random βs with the species matrices PC1 to construct $\Delta \mathbf{z}$ x PC1 *versus* β x PC1 graphs for each species. On these graphs we detected the threshold of the β x PC1 correlation which determined a $\Delta \mathbf{z}$ x PC1 correlation above 0.5 (i.e. the minimum alignment of selection with PC1 to cause a evolutionary response deflected to PC1). All analyses were performed using the R programming environment (R Core Team 2015).

3.4. Results

3.4.1. Climate-Related Selection Signature in Toad Skulls

We rejected drift in the three most basal nodes of the phylogeny with both tests (Figure 3; Table 3). Interestingly, the between-species variance in the morphological PC4 (i.e. species means projected onto within-species PC4, yielding species scores) is almost four times higher than expected by drift alone, suggesting selection on this dimension (Figure 4; Table 4). PC4 contrasts distances from the squamosal, pterygoid, maxillary and mandible bones with distances from the frontoparietal and nasal bones (Figure 5A). On the other hand, the within- and between-species variances were proportional for PC1 in all nodes (Figure 4). This axis is size related (Table 4) and explains the highest phenotypic variance within-species (~70%; also known as p_{max}).

The PC correlation tests suggest that some morphological PCs were co-selected along species diversification in the skull. The highest PC correlation was between PC4 and PC6 for nodes 13 and 14 in the phylogeny (Figure 6; Table 3). Species scores on these PCs should remain linearly uncorrelated if drift was the only process acting on skull divergence across species. Accordingly, PC6 also had higher

between- than within-species variation (Figure 4).

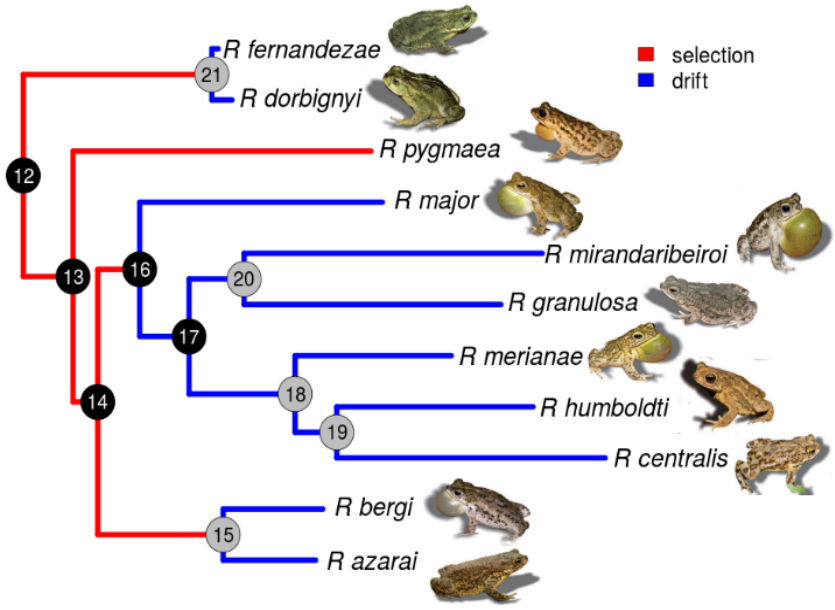


Figure 3. *R. granulosa* species group phylogeny showing the ancestors (numbered circles) and the random drift test results. Black circles indicate clades in which regression and PC correlation tests were done. Toad images were extracted from Pereyra et al. (2015). We rejected drift in the three basal-most nodes with both tests.

Variation in precipitation seasonality explains 85% of the between-species variation in the morphological PC4 (Figure 5B; Table 5). Between-species variation in both PC1 and PC6 were also explained by variation in precipitation seasonality (Figure 7; Table 5). Other climatic variables such as precipitation of driest quarter also explained significant variation in PC4 and PC6, but not as much as precipitation seasonality. Variation in annual precipitation explained some variation in morphological PC3, yet, given that we did not find a sign of selection on this PC, we argue that is not evidence enough for climate-related selection.

Table 3. Results of drift tests. Regression coefficient values and 95% confidence intervals (theoretical and simulated) for the five nodes in the phylogeny tested (see Figure 3); N = the number of taxa used in each test. The Bartlett test indicates whether there are significant correlations between morphological PCs. Df = degrees of freedom; r = the highest PC correlation; P = the significance value; average |r| = the average PC correlation for each node.

Regression tests							
		95% CI theoretical		**95% CI simulations**			
Nodes	***B***	**Lower**	**Upper**	**Lower**	**Upper**	**PCs**	**N**
12	**1.23**	0.96	1.49	**0.83**	**1.17**	4	11
13	**1.17**	0.91	1.42	**0.77**	**1.13**	4, 6	9
14	**1.16**	0.93	1.39	**0.74**	**1.13**	6	8
16	1.09	0.80	1.39	0.90	1.33	...	6
17	1.18	0.89	1.48	0.85	1.34	...	5

PC correlation tests							
						Average	
Nodes	**Bartlett**	**df**	**P**	**r**	**P**	**|r|**	**PCs**
12	**97.4**	**45**	**<0.001**	**-0.78**	**0.004**	**0.34**	**4-9; 3-10**
13	38.4	28	0.09	**-0.75**	**0.02**	0.31	**4-6; 3-5; 1-6**
14	26.9	21	0.17	**-0.83**	**0.01**	0.37	**4-6; 3-6; 1-6**
16	6.53	10	0.77	0.72	0.11	0.32	...
17	1.37	6	0.97	0.6	0.29	0.3	...

3.4.2. Interaction Between Selection and p_{max} Explains Species Divergence in the Skull

The reconstructed selection gradients (**β**) are on average more aligned with the morphological PC4 (mean vector correlation = 0.31 $\pm$ 0.23) than with any other morphological PC (PC1: 0.06 $\pm$ 0.05; PC2: 0.10 $\pm$ 0.09; PC3: 0.20 $\pm$ 0.16; Table 5). We considered only the first four morphological PCs because they account for 90% of species divergence.

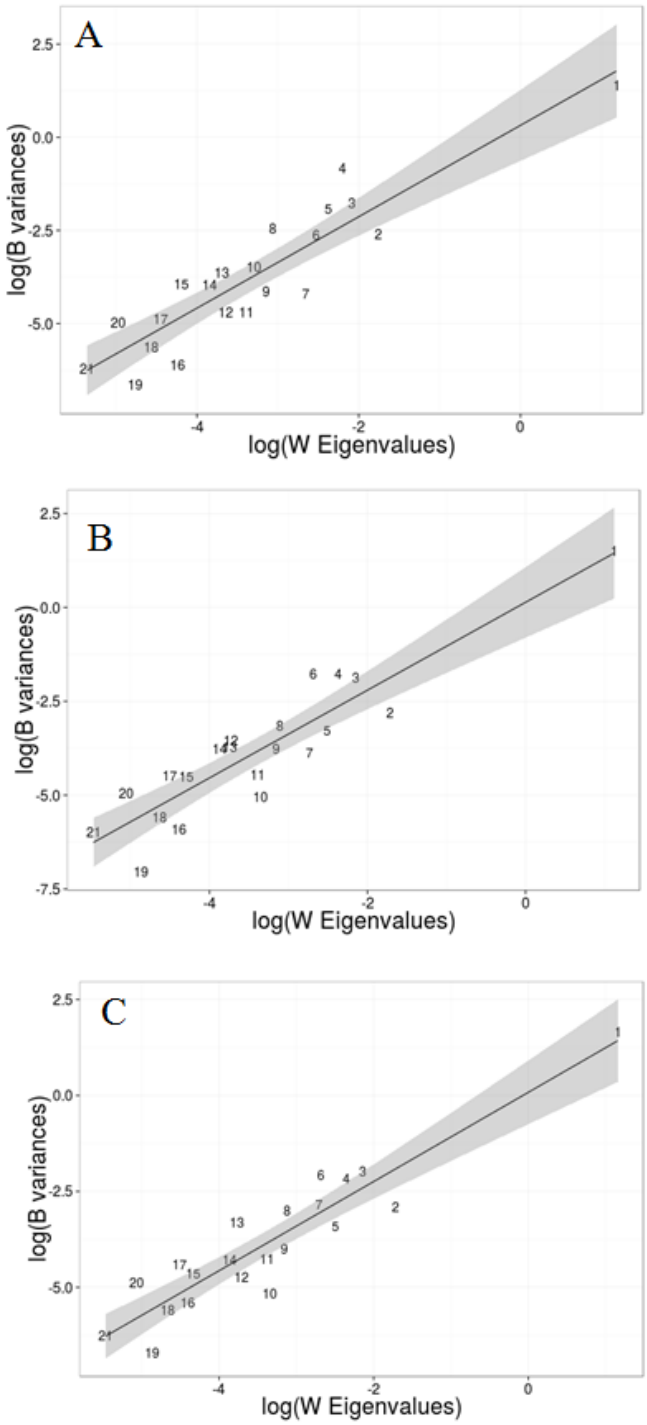


Figure 4. Regression plots comparing between- (B variances) and within-species (W Eigenvalues) variances for the three most basal nodes in the toad phylogeny. Numbers indicate the PCs and the grey shade indicates the 95% confidence interval of the regression line .**A)** Node 12**. B)** Node 13**. C)** Node 14**.**

Table 4. Normalized loadings for the first six morphological PCs. Normalized loadings in bold are above 0.2 (except for PC1). % variation indicates the relative variance explained by each PC within- (using the within-group pooled ancestral matrix) and between-species (using species scores on the PCs of the within-species matrix only for node 12).

		Functional	Within-group Principal Components					
Distances	**Bones**	**hypothesis**	**1**	**2**	**3**	**4**	**5**	**6**
1	nasal	snout	-0.19	-0.13	**-0.50**	-0.10	0.18	**-0.31**
2	frontoparietal	neurocranium	-0.24	0.01	**0.24**	**0.38**	**-0.42**	0.07
3	nasal	snout	-0.23	-0.13	-0.14	**0.22**	**0.29**	0.19
4	nasal	snout	-0.26	-0.10	**-0.52**	-0.17	-0.12	-0.09
5	frontoparietal	neurocranium	-0.10	-0.03	0.18	**0.53**	**0.27**	**-0.48**
6	orbit	neurocranium	-0.20	-0.05	0.00	0.18	-0.10	**-0.36**
7	squamosal	suspensorium	-0.15	-0.02	0.05	**-0.22**	**-0.37**	**0.23**
8	occipital	neurocranium	-0.13	-0.03	0.06	0.16	-0.11	-0.02
9	prenasal	snout	-0.10	-0.02	-0.10	0.16	-0.08	0.03
10	nasal	snout	-0.18	-0.09	-0.18	0.05	0.08	0.04
11	nasal	snout	-0.16	-0.06	-0.18	-0.05	-0.18	-0.03
12	maxilla	snout	-0.36	-0.17	**0.30**	**-0.21**	**0.29**	0.03
13	squamosal	suspensorium	-0.25	-0.01	-0.06	0.14	**-0.43**	-0.03
14	parasphenoid	neurocranium	-0.29	**0.69**	-0.02	-0.03	0.07	0.00
15	parasphenoid	neurocranium	-0.22	**0.61**	-0.02	-0.07	0.13	0.00
16	premaxilla	snout	-0.06	-0.03	-0.03	0.04	-0.03	-0.01
17	maxilla	snout	-0.18	-0.09	-0.01	0.14	**0.24**	**0.41**
18	neopalatine	snout	-0.19	-0.07	-0.07	0.18	0.09	**0.30**
19	pterygoid	suspensorium	-0.20	-0.08	**0.30**	**-0.38**	-0.10	**-0.40**
20	pterygoid	suspensorium	-0.21	-0.06	0.04	0.09	-0.12	0.12
21	mandible	suspensorium	-0.37	-0.21	**0.30**	**-0.27**	0.16	0.06
		Eigenvalues	3.26	0.17	0.13	0.11	0.09	0.08
	% variation	**Within**	77.4	4.1	3.0	2.6	1.9	1.7
		Between	77.7	**1.4**	3.3	**8.4**	2.8	1.4

Several evolutionary responses (**Δz**) are correlated with the direction of **β** (mean vector correlation = 0.58 $\pm$ 0.21), yet most of them are also correlated with p_{max} (**Δz** x PC1 mean vector correlation = 0.75 $\pm$ 0.27; Table 6). Interestingly, only when **β** is orthogonal to p_{max} (vector correlation < 0.02) is the response highly aligned with **β** (**Δz** x **β** vector correlation $\geq$ 0.88; Table 6). Simulation results show that the

threshold of the alignment between selection and PC1 to deflect evolutionary responses also to PC1 is a vector correlation around 0.05 and 0.1 (Figure 8).

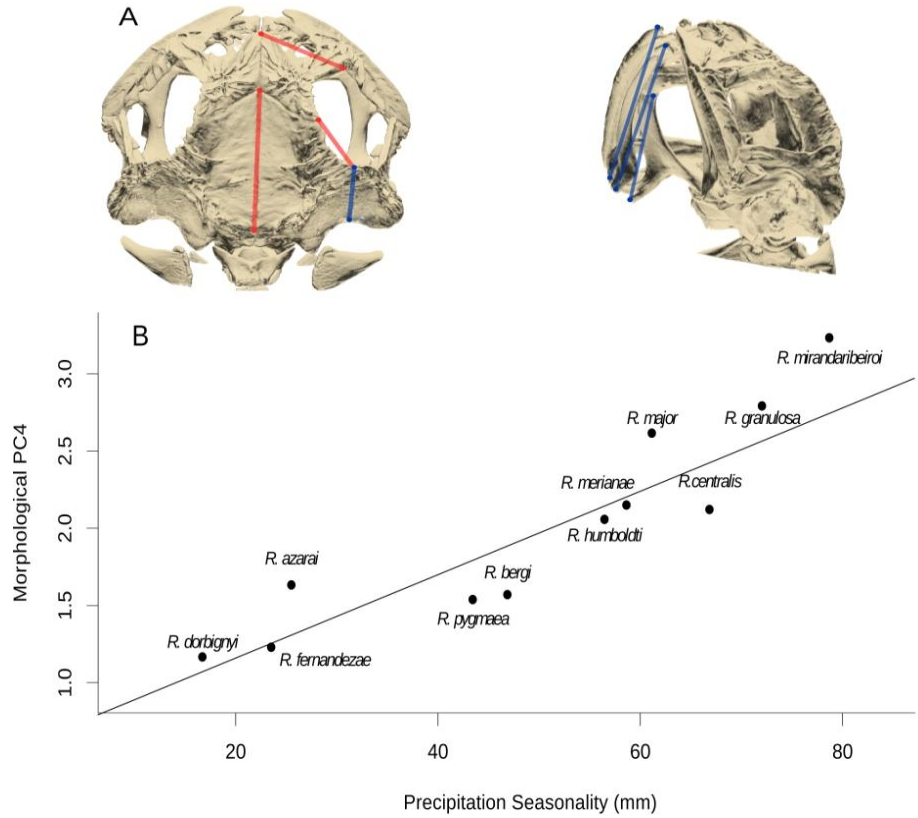


Figure 5. Description of morphological PC4 (A) and its relation with precipitation seasonality (B). A) The linear distances indicated on the skull image have the highest contribution to PC4. Different colours reflect the contrast in the PC4 dimension: snout and frontoparietal bones (red) against squamosal, maxillary, pterygoid and mandible bones (blue). **B)** Almost all variation in PC4 is explained by variation in precipitation seasonality (evolutionary regression: r^2 = 85%).

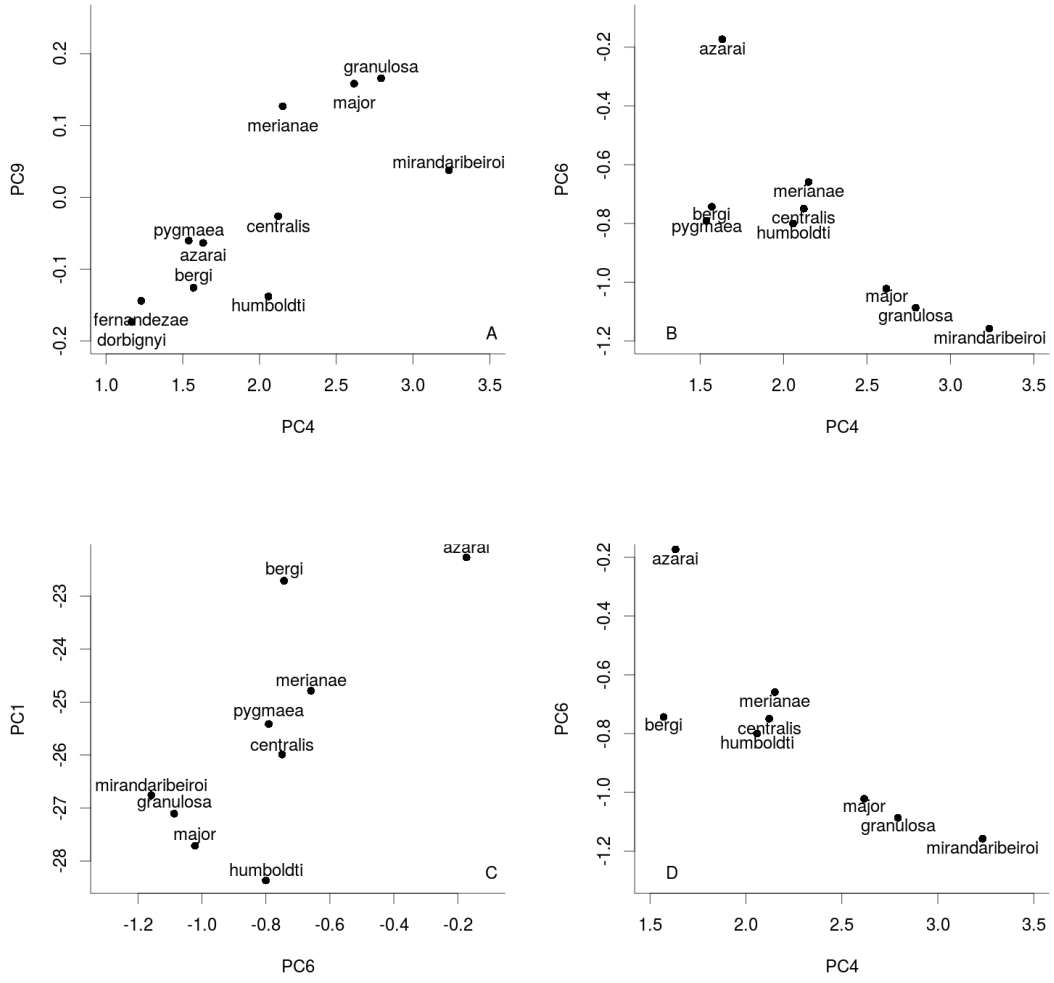


Figure 6. Correlations between species scores on the principal components of the within-group ancestral covariance matrix. A) Node 12: r = 0.77, N = 11, P < 0.05; **B)** Node 13: r = -0.76, N = 9, P < 0.05; **C)** Node 13: r = 0.74, N = 9, P < 0.05; **D)** Node 14: r = -0.83, N = 8, P < 0.05.

High proportions of the amount of divergence (r^2 = 0.95) and of constraints ($\mathbf{\Delta z}$ x p_{max}; r^2 = 0.90) were explained by the reduced model (Table 6). The interaction between $\boldsymbol{\beta}$ x p_{max} and strength of selection has the strongest effect on divergence and $\mathbf{\Delta z}$ x p_{max} values (Figures 9A, B). The more selection is aligned with p_{max} (higher values of $\boldsymbol{\beta}$ x p_{max}), the higher is the influence of strength of selection on the amount

of divergence (Figure 9A). High values of **β** x p_{max} also increase the constraint in the responses (**Δz** x p_{max}), yet this effect is more pronounced when strength of selection also increases (Figure 9B). The effect of the interaction between **β** x PC4 and strength of selection on the amount of divergence and on **Δz** x p_{max} is much weaker than the **β** x p_{max} interaction with selection (Figure 9C, D; respectively).

Evolvabilities in the direction of selection were all lower than evolvabilities in random directions. In contrast, evolvabilities in the direction of divergence (**Δz**) were several times above evolvabilities in random directions, except for branches 7 and 16 of the phylogeny, which also have the highest **Δz** x **β** vector correlations (see Table 6). Although significant in the model, the interaction between **β** x PC4 and evolvability in the direction of selection had the lowest effect on the dependent variables (Figure 9E, F; respectively; Table 6). Indeed, removing this interaction from the final model did not greatly change its explanatory power (amount of divergence: $r^2 = 0.94$ and **Δz** x p_{max}: $r^2 = 0.88$).

Table 5. Evolutionary regressions of between-species variation in morphological PCs on climatic variables. Half-life indicates the strength of phylogenetic inertia (PI). r^2% refers to the amount of variation in morphological PCs explained by each climatic variable. The intercept and coefficients are presented with their standard deviation in parenthesis. Values in bold correspond to the preferred models based on AICc. Note that only for PC1 is strong phylogenetic inertia the preferred model.

Evolutionary Regressions		
Morphological PC1	**Weak PI**	**Strong PI**
half-life	0.006	0.01
AICc	98.9	**87.6**
r^2 %	45.8	**44.7**
Intercept	-24.2 (0.27)	-23.8 (0.36)
Rain seasonality	-0.037 (0.005)	**-0.042 (0.007)**
Morphological PC4	**Weak PI**	**Strong PI**
half-life	0.001	0.01
AICc	**32.7**	33.5
r^2 %	**85.0**	81.0
Intercept	**0.65 (0.15)**	0.64 (0.19)
Rain seasonality	**0.027 (0.003)**	0.027 (0.004)
Morphological PC6	**Weak PI**	**Strong PI**
half-life	0.001	0.01
AICc	**22.7**	26.3
r^2 %	**60.50**	47.70
Intercept	**-0.57 (0.07)**	-0.47 (0.14)
Rain seasonality	**-0.005 (0.001)**	-0.0070 (0.0025)

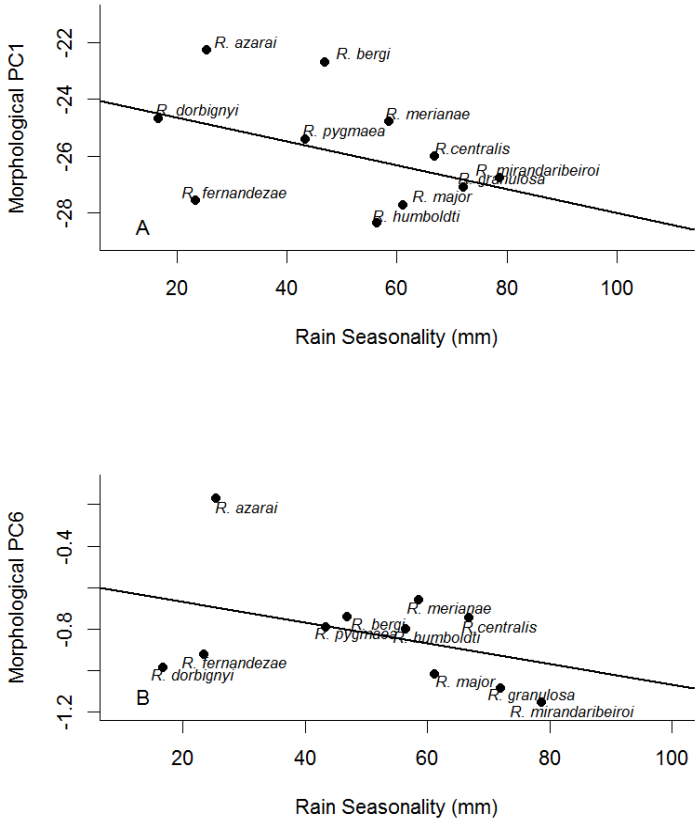


Figure 7. Evolutionary regressions for morphological PC1 (A) and PC6 (B) on rain seasonality. PC1 is a size-related dimension, whereas PC6 is a contrast between groups of bones (see Table S4). While PC6 was probably co-selected with PC4 by selection related to rain seasonality, the relation of PC1 with rain seasonality is probably spurious due to the action of genetic constraints. **A)** r^2 = 44.7%; **B)** r^2 = 60.5%.

Table 6. Vector correlations between evolutionary response (Δz), selection gradients (β) and within-species morphological PCs.

Phylogeny branches can be visualized in Fig. 3. Values in bold indicate significant correlations for 21-trait vector length.

	Vector Correlations									
Phylogeny branches	**ΔZ X β**	**ΔZ X PC1**	**PC1 X β**	**ΔZ X PC2**	**PC2 X β**	**ΔZ X PC3**	**PC3 X β**	**ΔZ X PC4**	**PC4 X β**	**amount**
1. Node 19 - *R. centralis*	**0.75**	**0.56**	0.03	**0.62**	0.39	0.11	0.11	0.16	0.21	0.58
2. Node 19 - *R. humboldti*	0.33	**0.99**	0.17	0.03	0.07	0.01	0.02	0.05	0.24	**3.82**
3. Node 18 - Node 19	0.40	**0.95**	0.12	0.16	0.24	0.12	0.29	0.00	0.00	**0.13**
4. Node 18 - *R. merianae*	0.36	**0.97**	0.14	0.09	0.17	0.04	0.11	0.05	0.14	**1.79**
5. Node 20 - *R. granulosa*	**0.57**	**0.83**	0.03	0.03	0.02	0.30	0.30	0.06	0.08	**0.92**
6. Node 20 - *R. mirandaribeiroi*	**0.88**	0.45	0.01	0.06	0.03	0.09	0.06	**0.60**	**0.54**	1.02
7. Node 17 - Node 18	**0.95**	0.07	0.00	0.19	0.09	**0.67**	0.45	0.23	0.20	0.05
8. Node 17 - Node 20	**0.68**	**0.74**	0.03	0.03	0.02	0.18	0.16	0.30	0.34	0.12
9. Node 16 - Node 17	**0.65**	**0.77**	0.03	0.10	0.06	0.04	0.04	0.21	0.25	0.08
10. Node 16 - *R. major*	0.37	**0.95**	0.07	0.06	0.08	0.09	0.18	0.13	0.32	**3.74**
11. Node 15 - *R. azarai*	**0.57**	**0.80**	0.02	0.21	0.14	0.20	0.16	0.06	0.05	**1.58**
12. Node 15 - *R. bergi*	**0.67**	**0.73**	0.02	0.06	0.03	0.02	0.01	0.03	0.02	0.62
13. Node 14 - Node 15	0.33	**0.98**	0.13	0.02	0.05	0.16	**0.57**	0.03	0.14	**4.70**
14. Node 14 - Node 16	0.45	**0.92**	0.05	0.00	0.00	0.04	0.06	0.19	0.35	**0.30**
15. Node 13 - Node 14	**0.75**	**0.66**	0.02	0.04	0.02	0.32	0.28	0.26	0.28	0.03
16. Node 13 - *R. pygmaea*	**0.95**	0.14	0.00	0.43	0.22	0.44	0.36	**0.64**	**0.64**	0.79
17. Node 12 - Node 13	**0.64**	**0.78**	0.04	0.13	0.12	0.26	0.31	0.35	0.48	0.12
18. Node 12- Node 21	**0.64**	**0.78**	0.04	0.13	0.12	0.26	0.31	0.35	0.48	1.78
19. Node 21 - *R. fernandezae*	0.34	**0.97**	0.12	0.02	0.06	0.02	0.07	0.20	**0.75**	**0.53**
20. Node 21 - *R. dorbignyi*	0.35	**0.96**	0.10	0.05	0.11	0.02	0.05	0.22	**0.73**	**4.84**

Table 7. Evolvabilities in random directions and in the directions of selection and divergence.

The matrices used for the calculation on each branch were pooled within-group matrices weighted by sample size, with each element divided by centroid size.

Note that evolvabilities in the direction of selection are lower than those in random directions, whereas evolvabilities in the direction of divergence are in general higher than those in random directions, except for the two cases in which the response was highly aligned with selection.

	Random directions	Selection direction	Mean ratio	Random directions	Divergence direction	Mean ratio
Phylogeny branches	e_{mean}	e_{mean}	Sel/Random	e_{mean}	e_{mean}	Div/Random
1. Node 19 - *R. centralis*	0.000134	0.000047	0.35	0.000129	0.000705	5.5
2. Node 19 - *R. humboldti*	0.000134	0.000114	0.85	0.000114	0.001978	17.4
3. Node 18 - Node 19	0.000124	0.000065	0.53	0.000095	0.001644	17.3
4. Node 18 - *R. merianae*	0.000120	0.000081	0.67	0.000099	0.001709	17.3
5. Node 20 - *R. granulosa*	0.000145	0.000043	0.30	0.000127	0.001597	12.6
6. Node 20 - *R. mirandaribeiroi*	0.000141	0.000049	0.34	0.000137	0.000518	3.8
7. Node 17 - Node 18	0.000125	0.000040	0.32	0.000129	0.000064	0.5
8. Node 17 - Node 20	0.000120	0.000044	0.37	0.000113	0.001121	9.9
9. Node 16 - Node 17	0.000128	0.000046	0.36	0.000119	0.001253	10.6
10. Node 16 - *R. major*	0.000124	0.000052	0.42	0.000111	0.001861	16.7
11. Node 15 - *R. azarai*	0.000139	0.000043	0.31	0.000126	0.001455	11.6
12. Node 15 - *R. bergi*	0.000136	0.000036	0.27	0.000123	0.001204	9.7
13. Node 14 - Node 15	0.000122	0.000076	0.62	0.000110	0.001922	17.5
14. Node 14 - Node 16	0.000132	0.000054	0.41	0.000108	0.001719	16.0
15. Node 13 - Node 14	0.000120	0.000048	0.40	0.000101	0.000817	8.1
16. Node 13 - *R. pygmaea*	0.000112	0.000054	0.48	0.000129	0.000111	0.9
17. Node 12 - Node 13	0.000138	0.000073	0.53	0.000123	0.001375	11.2
18. Node 12- Node 21	0.000147	0.000073	0.49	0.000122	0.001375	11.2
19. Node 21 - *R. fernandezae*	0.000194	0.000127	0.66	0.000158	0.003076	19.5
20. Node 21 - *R. dorbignyi*	0.000192	0.000115	0.60	0.000160	0.003025	18.9

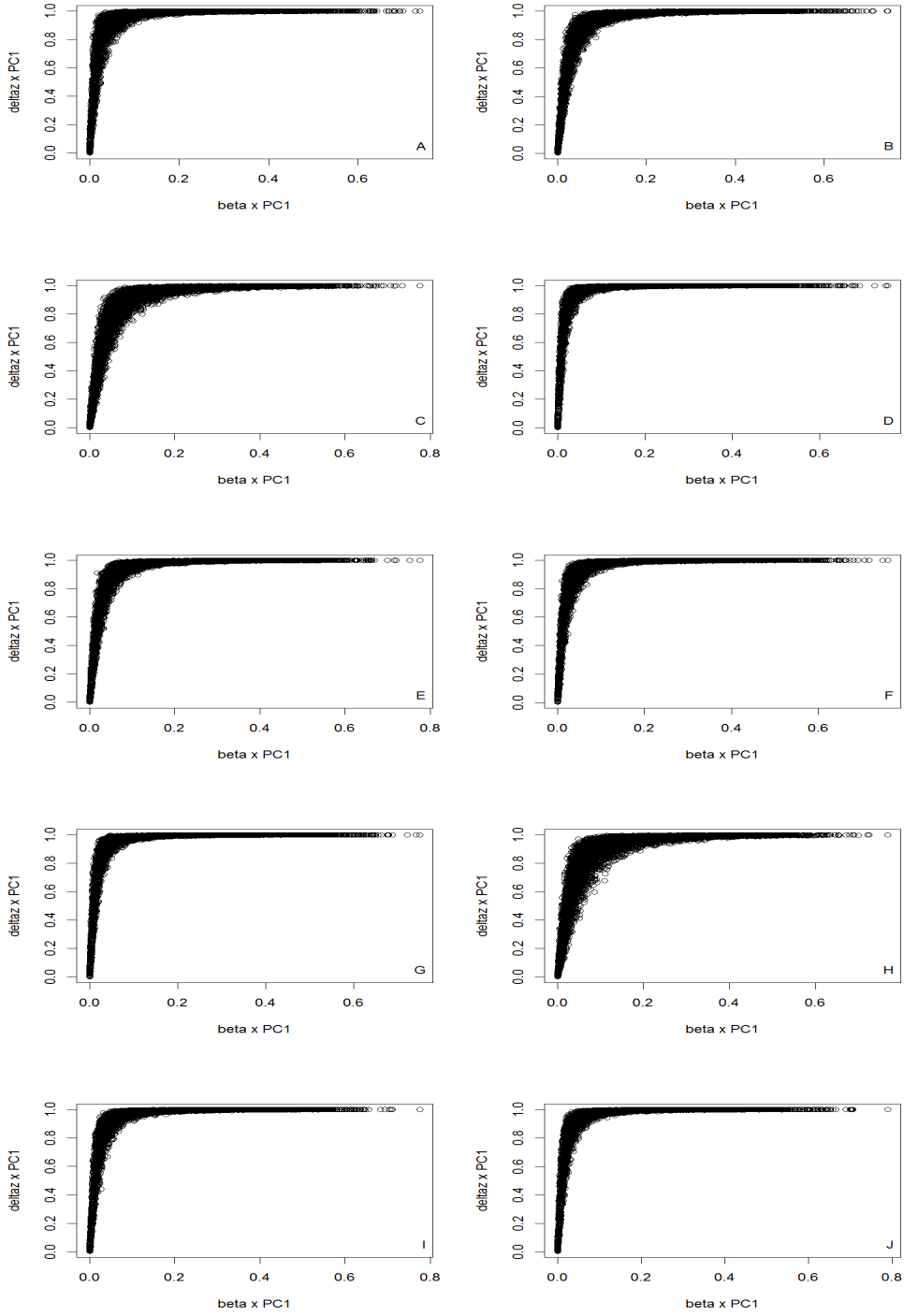


Figure 8. Threshold for the alignment of selection (beta) with PC1 to deflect evolutionary responses (deltaz) to the PC1 direction. Graphs show the correlations between 10,000 random selection vectors with species matrices' PC1 and their relation with the correlations between the evolutionary responses and species' matrices PC1. **A)** *R. centralis*; **B)** *R. humboldti*; **C)** *R. merianae*; **D)** *R. granulosa*; **E)** *R. major*; **F)** *R. bergi*; **G)** *R. dorbignyi*; **H)** *R. fernandezae* and **I)** *R. pygmaea*.

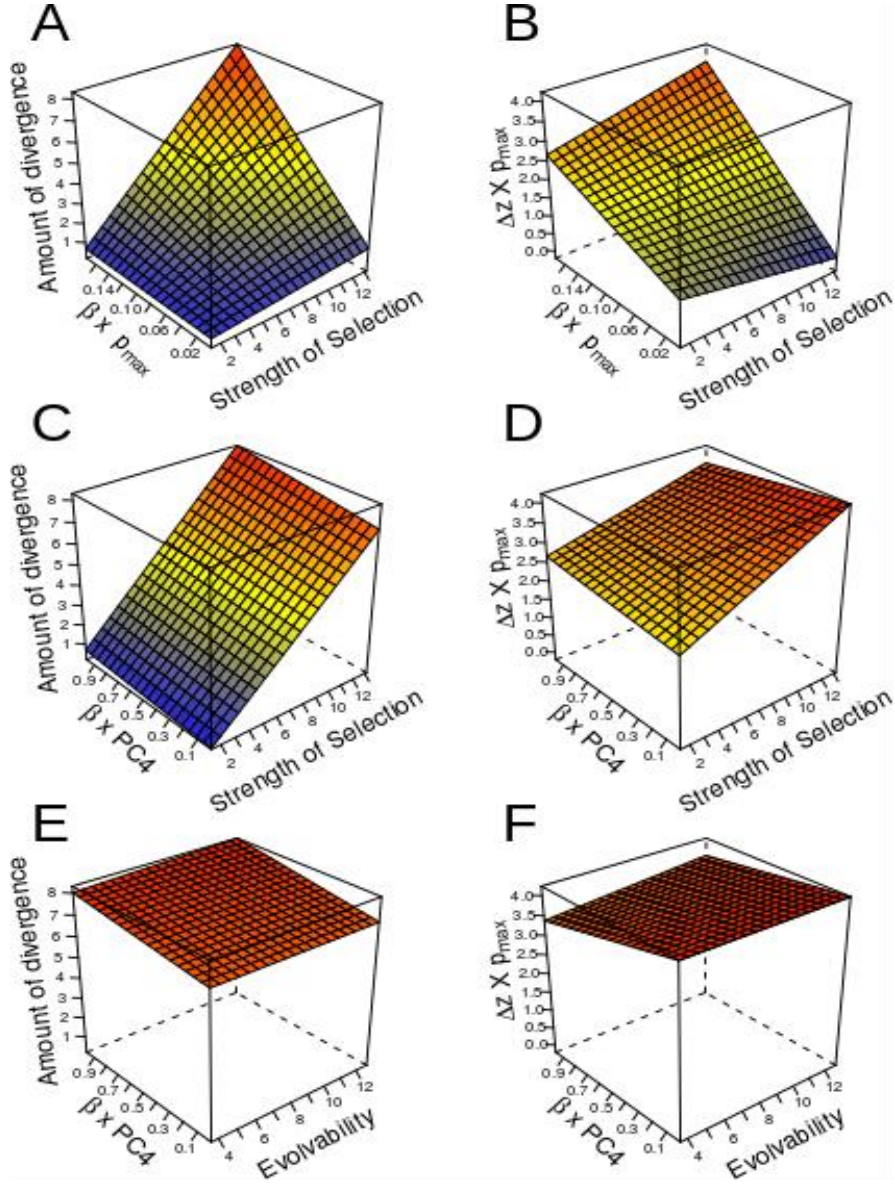


Figure 9. Interaction plots for the multivariate linear regression model. A) Effect of the interaction between the alignment of selection with the first axis of highest phenotypic variance (p_{max}) and strength of selection on amount of divergence. **B)** Effect of the same interaction as in **A** on constraint in the evolutionary response. **C)** Effect of the interaction between the alignment of selection with PC4 and strength of selection on the amount of divergence. **D)** Effect of the same interaction as in **C** on constraint in the evolutionary response. **E)** Effect of the interaction between the alignment of selection with PC4 and evolvability in the direction of selection on the amount of divergence. **F)** Effect of the same interaction as in **E** on constraint in the evolutionary response.

3.5. Discussion

Climate is a fundamental dimension of species' ecological niches (Grinnell 1917; Hutchinson 1957). Accordingly, the effects of climatic variation have been studied in several natural systems, with potentially important implications for biological conservation (Bradshaw & Holzapfel 2006; Parmesan 2006; Hoffmann & Sgrò 2011). Still, we are far from understanding the interplay between micro-evolutionary processes, such as selection and genetic constraints, and climatic effects in shaping species diversification on a phylogenetic scale (Teplitsky et al. 2014). To our knowledge, this is the first study integrating quantitative genetics and phylogenetic methods to investigate divergence in morphology and its potential relation with climate. An understanding of the interaction between genetic, ecological and evolutionary processes in the past is critical to the evaluation of the evolutionary potential of species to respond to future climatic changes.

Our results strongly suggest that natural selection has acted on toad skull diversity. Changes in precipitation patterns across species seem to be important past selective pressures, especially precipitation seasonality. Species subjected to more precipitation seasonality have shorter suspensorium bones and longer frontoparietal and nasal bones; whereas the opposite pattern is true for species exposed to less precipitation seasonality (variation in PC4). Other parts of the skull, such as the neopalatines and the orbit area (variation in PC6) were probably co-selected with traits associated to PC4. A mechanistic connection between the adult toad skull and precipitation patterns can be made, especially for snout bones. The snout region protects and supports the olfactory epithelium, which has a role in the detection of water cues in adult toads (Jungblunt et al. 2011). The species from the *R. granulosa* group are explosive breeders dependent on ephemeral pools to reproduce (Narvaes &

Rodrigues 2009), and the ability to detect sites for reproduction is probably fundamental to population persistence.

Even though we also found an association between precipitation seasonality and size variation between species, this is most likely a spurious relation. The PC1-precipitation seasonality regression was probably driven by the action of selection on traits associated with PC4, but given that several responses involved size, this relation appeared as significant. Considering that the between-species variation in PC1 is compatible with the expected by drift, we argue that size was not under the action of direct selection. However, we must be aware that size variation may be over-estimated in our data because we could not control for age. Since toads have indeterminate growth, age heterogeneity can inflate the size variance within-species. Even so, caution needs to be taken when judging phenotype-environment associations as adaptive, since it may be an indirect response to the environment due to genetic constraints.

In addition to finding a signature of selection in skull diversification, we also show that phenotypic/genetic constraints played a major role in toad species divergence. Several studies have shown the effect of constraints in natural systems and in population/species divergence (Schluter 1996; Caruso 2004; Marroig & Cheverud 2005; Renaud et al. 2006; Hunt 2007; Chenoweth et al. 2010; Bolstad et al. 2014). However, few empirical studies on evolutionary constraints have estimated selection gradients (Etterson & Shaw 2001; Marroig & Cheverud 2005; Chenoweth et al. 2010), even though information on the direction of selection might change the interpretation of constraints (i.e. if selection is aligned with g_{max}, the response is actually facilitated). In our system, selection was not aligned with p_{max}, while the evolutionary responses were severely deflected in its direction. Interestingly, even a

very low alignment of selection with p_{max} (vector correlations between 0.05 and 0.1) was enough to bias the evolutionary responses against following the direction of selection. Although this pattern might be an artefact due to the high variance in PC1 in all species, possibly enhancing the error in the reconstruction of selection gradients (Marroig et al. 2012), low $\boldsymbol{\beta}$ x p_{max} values have also been observed in studies that estimated selection independently of matrices or population divergence (Chenoweth et al. 2010). Moreover, the fact that selection gradients were substantially aligned with PC4 and that we controlled for noise in their estimation (Marroig et al. 2012) strengthens our confidence in those selection estimates. Thus, evolutionary constraints are extremely high in these toad species.

Constraints are expected to influence divergence when there are high amounts of variation concentrated in few dimensions of the G-matrix (Schluter 1996; Kirkpatrick 2009) associated with low evolvability in the direction of selection (Hansen & Houle 2008; Hansen et al. 2011). Though evolvability in the directions of the reconstructed selection gradients are indeed smaller than values obtained in random direction, they only acted as constraints to evolutionary responses when selection was more aligned with PC4, having actually a small effect on species divergence. The low contribution to divergence of evolvability in the direction of selection is made even clearer by the fact that its removal from the regression model only mildly altered explanatory power. Hence, by estimating the direction of past selection, we were able to effectively calculate its evolvability and test its constraining effect, as opposed to calculating evolvability in the direction of divergence (e.g. Bolstad et al. 2011; Hansen & Voje 2011). By so doing, we were able to conclude that the alignment of selection with p_{max} and its interaction with strength of selection was more important than the amount of variation in the direction

of selection, and that the former is enough to explain species divergence.

3.6. Conclusions

Jointly, the results presented here show that past selection in these toad skulls was related to changes in precipitation seasonality, which affected traits related to the snout, frontoparietal and suspensorium bones. Even though species responded to some extent to this selective pressure (as variation in PC4 and PC6 indicates), high genetic correlations among skull traits acted as constraints, deflecting evolution towards the size dimension. Divergence in size is a common phenomenon of several amphibian species (e.g. Kozak et al. 2009; Wollenberg et al. 2011; Caruso et al. 2011), suggesting that variation within-species might have also acted as a constraint in other systems. Even though body size has been shown to be an important fitness component of amphibians (e.g. Semlistch et al. 1988; Altwegg & Reyer 2003), the study of within-species variation is essential to test whether evolution of size was indeed adaptive or the consequence of genetic constraints (Marroig & Cheverud 2010). Considering that future climate change will alter precipitation patterns, probably enhancing precipitation seasonality and aridity, we expect that several amphibian species will not respond in the direction of change, but instead respond mainly in the size dimension due to constraints. It is critical that conservation efforts take into account not only the direct effects of climate change on populations, but also the interaction between strength and direction of selection and genetic constraints, as well as the potential impact of this interaction on population resilience (Lande & Lynch 1993; Hellmann & Pineda-Krch 2006; Chevin 2012). In this work, we add another dimension, morphological changes in complex traits, to models that predict the impact of climate change on biodiversity based on physiological tolerances

(Sinervo et al. 2010).

3.7. References

Ackermann, R. R. & Cheverud, J. M. 2002. Discerning Evolutionary Porcesses in Patterns of Tamarin (Genus Saguinus) Cranofacial Variation. *American Journal of Physical Anthropology*, 117, 260-271.

Arnold, S. J. 1992. Constraints on Phenotypic Evolution. *The American Naturalist*, 140**,** S85–S107.

Altwegg, R. and Reyer, H.-U. 2003. Patterns of Natural Selection on Size at Metamorphosis in Water Frogs. *Evolution*, 57, 872-882.

Arnold, S. J., Pfrender, M. E. & Jones, A. G. 2001. The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica*, 112**,** 9–32.

Bolstad, G. H. *et al.* 2014. Genetic constraints predict evolutionary divergence in Dalechampia blossoms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369**,** 1-15.

Butler, M. A. & King, A. A. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist*, 164, 683-695.

Caruso, C.M. 2004. The Quantitative Genetics of Floral Trait Variation in Lobelia: Potential Constraints on Adaptive Evolution. *Evolution*, 58, 732-740.

Caruso, N. M., Sears, M. W., Adams, D. C. & Lips, K. R. 2014. Widespread rapid reductions in body size of adult salamanders in response to climate change. *Global Change Biology*, 20, 1751-1759.

Chenoweth, S. F., Rundle, H. D. & Blows, M. W. 2010. The Contribution of Selection and Genetic Constraints to Phenotypic Divergence. *The American Naturalist*, 175**,** 186–196.

Chevin, L.-M. 2012. Genetic Constraints on Adaptation to a Changing Environment: Genetic Constraints in a Changing Environment. *Evolution*, 67**,** 708–721.

Etterson, J. R. Evolutionary Potential of Chamaecrista Fasciculata in Relation to Climate Change. II. Genetic Architecture of Three Populations Reciprocally Planted Along an Environmental Gradient in the Great Plains. *Evolution*, 58**,** 1459–1471.

Etterson, J. R. & Shaw, R. 2001. Constraint to Adaptive Evolution in Response to Global Warming. *Science*, 294**,** 151–154.

Felsenstein, J. 1988. Phylogenies and Quantitative Characters. *Annual Review of Ecology, Evolution and Systematics*, 19, 445–471.

Futuyma, D. J. 2010. Evolutionary Constraint and Ecological Consequences. *Evolution*, 64**,** 1865–1884.

Gibbs, H. L. & Grant P. R. 1987. Oscillating selection on Darwin's finches. *Nature*, 321, 511-512.

Gillson, L., Dawson, T. P., Jack, S. & McGeoch, M. A. 2013. Accommodating climate change contingencies in conservation strategy. *Trends in Ecology & Evolution*, 28**,** 135–142.

Grinnell, J. 1917. The Niche-Relationships of the California Thrasher. *The Auk*, 34,427-433.

Groves, C. R. *et al.* 2012. Incorporating climate change into systematic conservation planning. *Biodiversity and Conservation*, 21**,** 1651–1671.

Hansen, T. F. 1997. Satbilizing Selection and the Comparative Analysis of Adaptation. *Evolution*, 51, 1341-1351.

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

Hansen, T. F., Pélabon, C. & Houle, D. 2011. Heritability is not Evolvability. *Evolutionary Biology*, 38, 258–277.

Hansen, T. F., Pienaar, J. & Orzack, S. H. 2008. A Comparative Method for Studying Adaptation to a Randomly Evolving Environment. *Evolution*, 62, 1965-1977.

Hansen, T. F. & Voje, K. L. 2011. Deviation from the Line of Least Resistance does not exclude Genetic Constraints: A Comment on Berner et al. (2010). *Evolution*, 65, 1821–1822.

Hellmann, J. & Pineda-Krch, M. 2006. Constraints and reinforcement on adaptation under climate change: Selection of genetically correlated traits. *Biological Conservation*, 137**,** 599–609.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

Hoffmann, A. A. & Sgrò, C. M. 2011. Climate change and evolutionary adaptation. *Nature*, 470**,** 479–485.

Hohenlohe, P. A. and Arnold, S. J. 2008. MIPoD: A Hhypothesis-Testing Framework for Microevolutionary Inference from Patterns of Divergence. *The American Naturalist*, 171,

366-385.

Hunt, G. 2007. Evolutionary Divergence in Directions of High Phenotypic Variance in the Ostracode Genus Poseidonamicus. *Evolution*, 61, 1560-1576.

Hutchinson, G. E. 1957. The multivariate niche. *Cold Spring Harbor Symposium on Quantitative Biology*, 22, 415-421.

Jungblunt, L.D., Pozzi, A. G. & Paz, D. A. 2011. Larval development and metamorphosis of the olfactory and vomeronasal organs in the toad Rhinella (Bufo) arenarum (Hensel, 1867). *Acta Zoologica*, 92, 305-315.

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, 136**,** 271–284.

Kozak, K. H., Mendyk, R. W. & Wiens, J. J. 2009. Can Parallel Diversification occur in Sympatry? Repeated Patters of Body-Size Evolution in Coexisting Clades of North American Salamanders. *Evolution*, 63, 1769–1784.

Lande, R. 1979. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. *Evolution*, 33**,** 402–416.

Lande, R. & Arnold, S. J. 1983. The Measurement of Selection on Correlated Characters. *Evolution*, 37, 1210–1226.

Lande, R. & Lynch, M. 1993. in: *Biotic Interactions and Global Change*. (eds. Kareiva, P., Kingsolver, J. & Huey, R.) p. 234–250 (Sinauer Associates Inc.).

Linhart, Y. B. & Grant, M. C. 1996. Evolutionary Significance of Local Genetic Differentiation in Plants. *Annual Review of Ecology and Systematics*, 27**,** 237–277.

Lofsvold, D. 1988. Quantitative Genetics of Morphological Differentiation in Peromyscus. II. Analysis of Selection and Drift. *Evolution*, 42, 54-67.

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

Narvaes, P. & Rodrigues, M. T. 2009. Taxonomic revision of Rhinella granulosa species group (Amphibia, Anura, Bufonidae), with a description of a new species. *Arquivos de*

Zoologia, 40, 1–73.

Parmesan, C. 2006. Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution and Systematics*, 37**,** 637–669.

Pereyra, M. O. *et al.* 2015. Phylogenetic relationships of toads of the *Rhinella granulosa* group (Anura: Bufonidae): a molecular perspective with comments on hybridization and introgression. *Cladistics*, 31, 1–18

Pounds *et al.* 2006. Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature,* 439, 161-167.

Prôa, M., O'Higgins, P. & Monteiro, L. R. 2013. Type I Error Rates for Testing Genetic Drift with Phenotypic Covariance Matrices: A Simulation Study. *Evolution,* 67, 185–195.

Renaud, S., Auffray, J.-C. & Michaux, J. 2006. Conserved Phenotypic Variation Patterns, Evolution along Lines of Least Resistance, and Departure due to Selection in Fossil Rodents. *Evolution,* 60**,** 1701–1717.

Rowe, G. & Beebee T. J. 2004. Reconciling genetic and demographic estimators of effective population size in the anuran amphibian *Bufo calamita*. *Conservation Genetics*, 5, 287-298.

Schluter, D. 1996. Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution*, 50**,** 1774–1766.

Schluter, D., Price, T., Mooers, A. O. & Ludwig, D. 1997. Likelihood of Ancestor States in Adaptive Radiation. *Evolution,* 51, 1699–1711.

Schunke, A. C., Bromiley, P. A., Tautz, D. & Thacker, N. A. 2012. TINA manual landmarking tool: software for the precise digitization of 3D landmarks. *Frontiers in Zoology,* 9, 6.

Semlitsch, R. D., Scott,D. E. & Pechmann, J. H. K. 1988. Time and size at metamorphosis related to adult fitness in Ambystoma talpoideum. *Ecology,* 69,184-192.

Simon, M. & Marroig, G. 2015. Landmark precision and reliability and accuracy of linear distances estimated by using 3D computed micro-tomagraphy and the open-source TINA Manual Landmarking Tool software. *Frontiers in Zoology,* 12, 12.

Sinervo, B. et al. 2010. Erosion of Lizard Diversity by Climate Change and Altered Thermal Niches. *Science,* 328, 894-899.

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.

Swofford, D. L. & Maddison, W. P. 1987. Reconstructing Ancestral Character States Under Wagner Parsimony. *Mathematical Biosciences,* 87, 199-229.

Teplitsky, C. et al. 2011. Quantitative Genetics of Migration Syndromes: a study of two barn swallow populations. *Journal of Evolutionary Biology*, 24, 2025-2039.

Teplitsky, C., Robinson, M.R. & Merilä, J. 2014. in: *Quantitative Genetics in the Wild* (eds. Charmantier, A., Garant, D. & Kruuk, E.B.) 190-208 (Oxford University Press).

Urban, M. C. 2015. Accelerating extinction risk from climate change. *Science,* 348**,** 571–573.

Van Bocxlaer, I. et al. 2010. Gradual Adaptation Toward a Range Expansion Phenotype Initiated the Global Radiation of Toads. *Science,* 327, 679-682.

Walsh, B. & Blows, M. W. Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. 2009. *Annual Review of Ecology, Evolution, and Systematics*, 40**,** 41–59.

Wollenberg, K. C., Vieites, D. R., Glaw, F. & Vences, M. 2011. Speciation in little: the role of range and body size in the diversification of Malagasy mantellid frogs. *BMC evolutionary biology,* 11, 217

Considerações Finais

Os principais resultados dessa tese indicam que diferentes processos evolutivos atuaram nas relações entre caracteres (seleção estabilizadora interna e externa) e na diversificação da morfologia média (seleção direcional e restrições evolutivas) do crânio das espécies de sapo do grupo *Rhinella granulosa*. Contudo, tanto as diferenças inter-específicas nas correlações entre caracteres quanto nas médias dos caracteres foram determinadas por mudanças climáticas, especialmente nos padrões de chuva. Sugerimos, portanto, que a paisagem adaptativa na qual a população ancestral do grupo *R. granulosa* evoluiu tenha mudado conforme os padrões de chuva mudaram, tanto em sua curvatura média, refletindo alterações em seleção estabilizadora externa, quanto em sua inclinação média, refletindo alterações na força de seleção direcional. Especificamente, é provável que uma mudança maior na paisagem tenha ocorrido na separação das linhagens que hoje habitam o Chaco úmido e as demais linhagens que habitam ambientes mais áridos. De acordo com a ideia de que ambientes mais úmidos fornecem mais oportunidades de reprodução e que o crânio tenha uma relevância funcional na detecção de poças para a reprodução, tanto a seleção estabilizadora quanto a direcional sob a morfologia do crânio seriam mais fracas nesse contexto do que em ambientes mais áridos, nos quais a detecção de poças pode ser mais difícil. Em ambientes mais áridos é possível que a competição por poças também seja mais intensa.

Narvaes & Rodrigues (2009) especularam que a diversificação do grupo se deu a partir de um estoque ancestral de ampla distribuição em regiões de áreas abertas na América do Sul. Com as alterações climáticas ocorridas no Quaternário (por volta de 2 Ma), as áreas abertas tornaram-se mais fragmentadas, separadas por formações florestais que emergiram nos ciclos úmidos, e populações derivadas da ancestral podem ter se isolado uma das outras e se diferenciado (Narvaes 2003). Evidências a favor dessa hipótese de especiação são a associação entre diferenças no clima e distâncias filogenéticas, indicando que espécies mais

próximas filogeneticamente estão em climas mais similares; e a associação entre diferenças no clima e diferenças na morfolgia, indicando que espécies de morfologia similar estão sujeitas a regimes climáticos similares. Além disso, a relação entre seleção direcional, que provavelmente atuou apenas em alguns ossos do crânio das espécies, e a sazonalidade da chuva, também é uma evidência de diversificação atrelada à mudanças climáticas. Esse tipo de especiação é chamada especiação ecológica, pois é um processo direcionado pelas diferenças no ambiente entre populações. Porém, as evidências mais fortes a favor dessa hipótese envolvem o surgimento de isolamento reprodutivo entre as espécies ou ao menos a constatação de que um híbrido possuí aptidão reduzida nos ambientes específicos das linhagens parentais (Schluter 1996). Portanto, estudos futuros que lidem com a aptidão de híbridos entre as espécies do grupo *R. granulosa* em diferentes ambientes, ou mesmo a aptidão de diferentes espécies em diferentes ambientes, podem ser muito promissores em esclarecer se especiação ecológica foi o processo que originou as espécies.

Entretanto, várias das espécies do grupo *R. granulosa* hibridizam, e possuem, portanto, um enfraquecimento dos mecanismos de isolamento reprodutivo. Apesar da ocorrência de híbridos não invalidar as espécies como tais, provavelmente existe fluxo gênico entre populações de diferentes espécies. A produção de híbridos a partir de espécies que possuem médias fenotípicas distintas resulta na manutenção de variação genética na direção da divergência entre as espécies. Isso porque a hibridização cria desequilíbrio de ligação entre alelos responsáveis pela divergência entre as espécies (Schluter 1996). Dado que grande parte das respostas evolutivas estimadas na filogenia das espécies foi na direção de p_{max}, hibridização entre espécies pode ser um dos fatores mantendo variação na direção de p_{max}, ou seja, na direção de tamanho. Porém, Narvaes & Rodrigues (2009) reportam somente 7% de simpatria entre as espécies ao longo das localidades estudadas. Portanto, se existe manutenção de variação em tamanho entre espécies por hibridização, a contribuição provavelmente é baixa. Além disso, é bastante plausível que a alta variação em tamanho dentro de cada espécie seja causada pela grande quantidade de alelos que regulam o processo

de crescimento. Quanto mais fatores, genéticos e ambientais, interferirem em taxas de crescimento, maior a variação em tamanho.

Existe a possibilidade de que a variação em tamanho dentro de cada espécie esteja super-estimada. Isso porque não pudemos controlar para o fator idade dos espécimes. Dado que os anfíbios possuem crescimento indeterminado, esperamos que indivíduos mais velhos sejam maiores que indivíduos mais jovens. Se a variação no PC1 dentro da população ancestral for na verdade menor que os 70% estimados do total, a variação entre espécies no PC1 pode ser maior que dentro das espécies (em vez de ser proporcional, considerando a variação dentro e entre espécies nos demais PCs). Esse padrão de variação no PC1 sugeriria três situações possíveis de processos evolutivos: (1) seleção direcional em tamanho causando divergência entre espécies; (2) seleção estabilizadora nas demais dimensões morfológicas, reduzindo a variação entre espécies nesses outros PCs ou (3) ação conjunta de seleção direcional no PC1 e estabilizadora nos demais PCs. Em relação ao primeiro cenário, uma vez que os gradientes de seleção estimados realmente não sejam na direção de tamanho, é improvável que tenha havido seleção direcional em tamanho. Continuaríamos, portanto, a interpretar a relação PC1-sazonalidade da chuva como espúria. Entretanto, essa informação sobre os gradientes de seleção não invalida a hipótese de seleção estabilizadora nos demais PCs, caso a variação no PC1 entre espécies fosse maior que entre espécies. Uma estimativa mais precisa de variação em tamanho dentro das espécies pode ser determinada no futuro com o uso da técnica de esqueleto-cronologia (Castanet & Smirina 1990).

Finalmente, uma conexão clara entre variação na dimensão morfológica (PC4) que esteve sob seleção direcional e as diferenças encontradas no padrão de covariação entre as espécies (concentradas no rostro) não pode ser feita. O PC4 não reflete um contraste entre distâncias pertencentes ao módulo rostro e distâncias dos demais módulos, e nem mesmo possuí coeficientes maiores para as distâncias detectadas como dissimilares em covariação. Isso sugere que não foi a ação de seleção direcional que levou à divergência na estrutura de covariação do rostro entre as espécies. Porém, o PC3 define um contraste mais claro entre distâncias do nasal (rostro) e distâncias nos demais módulos. Portanto, o terceiro eixo de

maior variação dentro da população ancestral reflete um contraste entre rostro e os demais módulos.

Apesar das ressalvas levantadas acima, que refletem as limitações do trabalho, podemos concluir que a variação em padrões de chuva provavelmente foi muito relevante na diversificação das espécies do grupo *R. granulosa*, causando divergência tanto na estrutura de covariação quanto nas médias dos caracteres do crânio. A magnitude das restrições evolutivas detectadas para as espécies, relacionadas com a quantidade de variação em tamanho, podem estar super-estimadas, porém é muito provável que tamanho seja de fato a primeira linha de menor resistência evolutiva, como é o caso em diversos grupos de mamíferos. Além disso, a forte estabilidade de matrizes-P que encontramos está de acordo com o encontrado nos mamíferos, fortalecendo a noção de que a matriz-G pode ser estável por longos períodos de tempo.

Referências

Castanet, J. & Smirina, E. 1984. Introduction to the skeletochronological method in amphibians and reptiles. *Annales de sSciences Naturelles, Zoologie et Biologie Animale*, 11, 191-196.

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.

Narvaes, P. & Rodrigues, M. T. 2009. Taxonomic revision of Rhinella granulosa species group (Amphibia, Anura, Bufonidae) with a description of a new species. *Arquivos em Zoologia*, 40, 1-73.

Narvaes, P. 2003. Revisão taxonômica das espécies de *Bufo* do complexo *Rhinella granulosa* (Amphibia, Anura, Bufonidae). *Tese defendida no Instituto de Biociências, USP*.

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, 3305–3322.

Schluter, D. 1996. Ecological causes of adaptive radiation. *The American Naturalist*, 148, S40-S64.

Anexo 1

Artigo publicado durante o doutorado: Simon, M. & Marroig, G. 2015. Landmark precision and reliability and accuracy of linear distances estimated by using 3D computed micro-tomagraphy and the open-source TINA Manual Landmarking Tool software. *Frontiers in Zoology*, 12, 12.

Simon and Marroig *Frontiers in Zoology* (2015) 12:12
DOI 10.1186/s12983-015-0101-5

FRONTIERS IN ZOOLOGY

METHODOLOGY Open Access

Landmark precision and reliability and accuracy of linear distances estimated by using 3D computed micro-tomography and the open-source TINA Manual Landmarking Tool software

Monique Nouailhetas Simon*† and Gabriel Marroig†

Abstract

Introduction: The wider availability of non-destructive and high-resolution methods, such as micro-computed tomography (micro-CT), has prompted its use in anatomical and morphometric studies. Yet, because of the actual scanning procedure and the processing of CT data by software that renders 3D surfaces or volumes, systematic errors might be introduced in placing landmarks as well as in estimating linear distances. Here we assess landmark precision and measurement reliability and accuracy of using micro-CT images of toad skulls and the TINA Manual Landmarking Tool software to place 20 landmarks and extract 24 linear distances. Landmark precision and linear distances calculated from 3D images were compared to the same landmarks and distances obtained with a 3D digitizer in the same skulls. We also compared landmarks and linear distances in 3D images of the same individuals scanned with distinct filters, since we detected variation in bone thickness or density among the individuals used.

Results: We show that landmark precision is higher for micro-CT than for the 3D digitizer. Distance reliability was very high within-methods, but decreased in 20 % when 3D digitizer and micro-CT data were joined together. Still, we did not find any systematic bias in estimating linear distances with the micro-CT data and the between-methods errors were similar for all distances (around 0.25 mm). Absolute errors correspond to about 6.5 % of the distance's means for micro-CT resolutions and 3D digitizer comparisons, and to 3 % for the filter type analysis.

Conclusions: We conclude that using micro-CT data for morphometric analysis results in acceptable landmark precision and similar estimates of most linear distances compared to 3D digitizer, although some distances are more prone to discrepancies between-methods. Yet, caution in relation to the scale of the measurements needs to be taken, since the proportional between-method error is higher for smaller distances. Scanning with distinct filters does not introduce a high level of error and is recommended when individuals differ in bone density.

Keywords: Bone density, Distance repeatability, Geometric morphometrics, Traditional morphometry, Toad skulls

Introduction

The increasing use of non-destructive and high-resolution data acquisition methods, such as micro-computed tomography (micro-CT), have provided researchers with the opportunity to study the anatomy and morphology of organisms with more detail and at a wider phylogenetic spectrum (e.g. [1–5]). Accordingly, 3D image processing software has been developed (e.g. OsiriX: [6], Amira: www.amira.com), with some designed to place 3D landmarks for shape or morphometric analysis (e.g. TINA Manual Landmarking Tool: [7]). However, there is no guarantee that the scanning procedure and the software used to process CT data, by creating 3D surface or 3D volume renderings, do not introduce systematic errors in the data [8, 9]. In addition, the landmark positioning process in the 3D images might also introduce systematic and random errors in the estimation of linear distances (measurements). Thus, the precision of placing landmarks in 3D

* Correspondence: monique.simon@usp.br
†Equal contributors
Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, Rua do Matão, 277, 05508-090 São Paulo-SP, Brazil

images with software, as well as the accuracy and reliability of the distances taken by the use of CT data must be validated [8–10].

In this study, we evaluated the precision of placing the same landmarks in the same individuals of a toad species with two distinct methods: (1) the real skulls and a 3D digitizer to place the landmarks, and (2) 3D images of these skulls, obtained by micro-CT scans at two distinct resolutions, and software to place the landmarks (TINA Manual Landmarking Tool ([7]; hereafter called TINA-Landmark). The 3D digitizer is an articulated arm that creates a 3D coordinate system in which any point of an object can be identified in relation to a reference point. TINA-Landmark is recently developed open-source software created to enhance the precision of landmark positioning in 3D images by using volume rendering instead of surface rendering and by showing the cross-section images connected to the 3D volume [7]).We chose to compare the landmarks obtained from the 3D images with the ones obtained by the 3D digitizer because, for the latter, the landmarks are taken in the actual skulls, with no processing of data, not the case for the construction of 3D images. Also, in zoological studies, several authors measure the specimens with 3D digitizers (e.g. [11–14], just to cite a few), being a widely accepted technique in the morphometry field.

In addition to comparing the landmarks between the two methods, it is also of interest to evaluate the consequence of potential biases in the estimation of linear distances extracted from the landmarks. Although measurement error is intrinsic to the landmarks and independent of the linear distances computed from them, the proportion of error varies with the distance length. That is, if the error in placing landmarks is the same for all landmarks, the proportion of error will be higher for smaller distances than for longer distances [15]. On the other hand, if the error in placing landmarks varies with landmark type or position in the material, linear distances extracted from these more variable landmarks are expected to have greater error when comparing distinct methods, although an association between distance length and error proportion will still exist. Therefore, we also calculated reliability and accuracy of linear distances extracted from the landmarks placed with the 3D digitizer and in the 3D images.

Finally, we also compared the same toad specimens scanned with two distinct filters, since we discovered that some individuals differ in bone thickness and/or density. Filters are thin sheets of metal set in front of the material being scanned and can have different thickness. Varying the filter thickness has an effect on the mean X-ray energy irradiating on the material being scanned. For thinner bones, a lower X-ray energy is necessary to achieve the best 3D volumes and more precise positioning of the landmarks (see Fig. 1). Thus, changing filter type might be an additional source of error when placing landmarks and taking linear measurements in 3D images. We consider this last comparison quite relevant in zoological studies because other organisms might present the same variation in bone density and scanning with distinct filters will be indispensable.

Although there is no specific theory relating the potential effects on landmarks and on distances when using distinct resolutions, filters and reconstruction algorithms, some expectations based on landmark position and bone thickness can be made. Some of the landmarks that we selected in the toad skulls (Table 1 and Fig. 2) were more difficult to visualize in the 3D images than in the real skulls because of their position (landmarks 4, 10, 11, 17 and 20) or because they were placed at thinner bones (landmarks 6, 7, 8 and 13). Thus, we expect more variation in the positioning of these landmarks among methods, and as a consequence, less reliability and accuracy of the linear distances extracted from them (Table 2 and Fig. 2). We report that placing landmarks in 3D images obtained by micro-CT scanning is more precise than placing the same landmarks with the 3D digitizer. Yet, average differences in linear distances among methods are acceptable and represent a low error proportion in relation to the distances lengths. Scanning with distinct resolutions and distinct filters do not introduce high errors.

Results and discussion

Landmark precision with distinct methods

We compared landmark precision between-methods by calculating the mean distance of each landmark in the individuals from the same landmarks of a mean shape estimated with a slightly modified General Procrustes Analysis, for all methods (see Methods section). Mean individual landmark distances from the mean sample landmarks were higher when using the 3D digitizer (DIG) to place the landmarks than when using the 3D images, regardless of the resolutions (medium [MED] or high [HIGH]). This result holds for the shape space (without scale, Table A1 in Additional file 1), as well as for retaining landmark scale information in mm (after multiplying the Procrustes configuration by mean centroid size, Table 3), indicating that placing landmarks in 3D images obtained by the micro-CT is more precise. Even though the precisions of both equipments used in this study are similar (3D digitizer: 0.01 mm; micro-CT resolutions MED: 0.018 mm and HIGH: 0.009 mm), this result may be due to the fact that the 3D images of the skulls are much bigger in the screen than the real ones (around tenfold increase), facilitating visualization of several bone structures. The range of deviations of individual landmarks from the mean shape landmarks for all methods (DIG: 0.27 to 0.52 mm; HIGH: 0.14 to

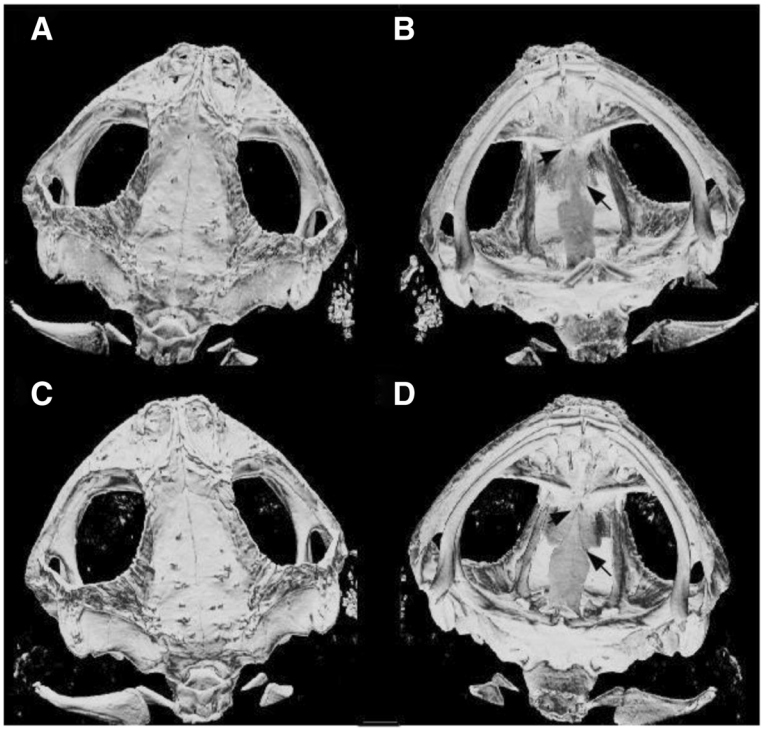


Fig. 1 Skull 3D images of a *Rhinella pygmaea* specimen scanned with distinct filters. Several *R. pygmaea* specimens presented transparency in some bones, such as the squamosal, sphenethmoid and the parasphenoid bones, making the precise determination of sutures between these bones difficult (indicated by the arrows). Scanning with a thinner filter corrects for this problem as can be seen in **c** and **d**. **a** Dorsal view of a skull scanned with an aluminum filter 1.0 mm (AL1.0); **b)** Ventral view of the same skull scanned with AL1.0; **c)** Dorsal view of the same skull scanned with an aluminum filter 0.5 mm (AL0.5); **d)** Ventral view of the same skull scanned with AL0.5

0.33 mm and MED: 0.16 to 0.34 mm) are comparable to the error found by Richtsmeier et al. (1995) [16] when placing the same landmarks in two different 3D images of the same individual (scanned twice with a 1.5 mm slice thickness resolution), which was 0.15 to 0.48 mm. Thus, the precision of both methods in placing landmarks is acceptable for morphometric studies, at least for distances as large as the ones obtained in this study.

The two landmarks with highest mean deviations for DIG (14: posterior tip of parasphenoid process and 19: posterior tip of pterygoid process) were not expected to have higher error, however they both are type II landmarks, suggesting that locating the exact maximum bone curvature was more difficult than locating bone sutures (type I landmarks). The higher error for type II landmarks compared to type I landmarks is expected when working with the real skulls [17]. Landmarks 3, 5, 6, 7, 9, 10, 11, 12, 15, 16 and 20 presented higher discrepancies in the landmarks for DIG when compared to the micro-CT resolutions (boldface in Table 3). For the landmarks expected to have higher error in the 3D images because of their positions in the skull or because they were located in thinner bones, only landmarks 4, 8 and 12 corresponded to the expectation. All of them are type I landmarks, agreeing with the findings of other authors working with 3D images of more error in placing landmarks at some particular bone sutures [18, 19].

MED and HIGH presented very similar results, indicating that there is no difference in switching the resolutions to visualize the landmarks. The fact that we had to down-sample the HIGH data by a factor of two to load the cross-section sequences in TINA-Landmark probably did not interfere in landmark precision, since switching the compacting factors with the MED data also produced similar results (comparisons between stride = 2 [S2] and stride = 2 and down-sample = 2 [S2/D2]; and stride = 3 [S3] and stride = 3 and down-sample = 2 [S3/D2]; Table 3). This result is consistent with the fact that the cross-sections are maintained in the original resolution even when down-sampling the data (which affects only the resolution of the 3D volume), being possible to refine the landmarks positions in them. Finally, for the filter type analysis, landmarks 6, 8 and 13 presented higher mean individual landmark distances for aluminum 0.5 mm (AL0.5) than for aluminum 1.0 mm (AL1.0), coinciding with the ones expected as more variable for being located at thinner bones. The only exception was landmark 5, with high deviation

Table 1 Landmark descriptions in the toad skulls. Landmarks are intersections between bone sutures (type I landmarks, 16 in total) or tip of bones (type II landmarks: numbers 1, 14, 15 and 19). Five landmarks are in the medial line and the remaining landmarks are present in both sides of the skull. The landmarks are spread in all three views of the skull: dorsal, lateral and ventral (see Fig. 1). We placed all 20 landmarks with all the methods twice in each individual

Landmarks	Description	Position	View
1	Anterior tip of nasal bone	midline	dorsal
2	Nasal and frontoparietal suture	midline	dorsal
3	Posterior tip of frontoparietal suture	midline	dorsal
4	Nasal and maxillary suture	right, left	dorsal
5	Nasal and frontoparietal lateral suture	right, left	dorsal
6	Frontoparietal and squamosal suture	right, left	dorsal
7	Frontoparietal, squamosal and occipital suture	right, left	dorsal
8	Squamosal and occipital suture	right, left	dorsal
9	Frontoparietal and occipital suture	right, left	dorsal
10	Prenasal and maxillary lateral suture	right, left	lateral
11	Nasal and maxillary lateral suture	right, left	lateral
12	Squamosal and maxillary suture	right, left	lateral
13	Sphenethmoid and parasphenoid suture	midline	ventral
14	Posterior tip of parasphenoid corpus	midline	ventral
15	Anterior tip of premaxillary bone	right, left	ventral
16	Premaxillary and maxillary suture	right, left	ventral
17	Pterygoid and maxillary suture	right, left	ventral
18	Neopalatine and sphenethmoid suture	right, left	ventral
19	Tip of pterygoid process	right, left	ventral
20	Pterygoid and parasophenoid suture	right, left	ventral

for AL 0.5. This landmark is located in the nasal and frontoparietal lateral suture, coinciding with a neural crest. It is possible that the presence of the crest caused higher error in the positioning of the landmark.

Linear distances reliability and accuracy

The reliability of the linear distances was obtained by using distance repeatability, which indicates the reliability of multiple measurements in the same individuals. It describes the proportion of variance due to differences among individuals (between-group variance) in relation to the residual variance plus the between-group variance [20]. Within-method mean distance repeatabilities (i.e. considering only replicates measured with the same method) were very high, above 0.9 for all methods, as can be seen in Table 4 for micro-CT resolutions and 3D digitizer; and in Table 5 for image compacting factors and distinct filters. This result shows that measuring procedure is very reliable inside each method. However, between-methods mean distance repeatability is considerably lower (20 % lower) when replicates of MED or HIGH were joined with replicates of DIG, but not when replicates of distinct micro-CT resolutions were joined together (Table 4). When replicates of AL0.5 and AL1.0 were joined together, the mean repeatability also reduced (Table 5), yet the drop was less steep (13 % lower) than MED + DIG and HIGH + DIG. For the image compacting factors there was no drop in mean distance repeatability when replicates of different values of the compacting factors were considered together (Table 5). Looking at repeatabilities separately for each distance in the between-methods analysis (Fig. 3b,c), it can be noticed that distances 3, 5, 6, 9, 10, 11, 13, 14, 16, 23 and 24 all have low repeatabilities (below 0.8) and are composed of landmarks detected with higher mean deviations for DIG (landmarks 3, 5, 8, 9, 10, 11, 12 and 19). These distances comprise small as well as large distances, showing that the error is independent of distance length and related to the quality of the landmark. Although we did not make expectations of higher error for landmarks placed with DIG, the landmarks with higher deviations had as a consequence higher discrepancies in the between-methods repeatabilities. However, the distances 17 and 22, composed of the landmarks with the highest error for DIG (14 and 19), did not present low repeatabilities. This is a consequence of their longer mean lengths (17: 7.14 mm and 22: 5.06 mm) and a smaller proportional error as a consequence. On the other hand, distances 5, 10, 11, 13 and 24 are small distances (less than 3.0 mm of length) composed of highly discrepant landmarks, and therefore, have a larger proportional error. These differences in proportional error depending on the mean distance length is shown by a significant positive association between mean distance repeatability and mean distance length for the joined data sets MED + DIG and HIGH + DIG ($r = 0.45$, d.f. = 22, $P = 0.02$ and $r = 0.44$, d.f. = 22, $P = 0.03$, respectively; Fig. 3b,c), as well as for S2 + S2/D2 and AL0.5 + AL1.0 ($r = 0.44$, d.f. = 24, $P = 0.03$ and $r = 0.52$, d.f. = 24, $P = 0.007$, respectively; Fig. 3e,i). Although a significant relation between repeatability and distance length exists, practically all distances composed of landmarks with higher differences between-methods presented lower repeatability when compared to within-method repeatability.

Additionally to the distance repeatability analysis, we also compared mean raw and absolute differences in the distances within and between methods. The mean raw differences within and between methods were close to zero for all data sets, being sometimes positive and sometimes negative (Tables 4 and 5), indicating that in average there is no consistent bias when measuring the same individual two times or measuring the same individuals with different methods. Yet, when looking at the raw differences between MED-DIG and HIGH-DIG for

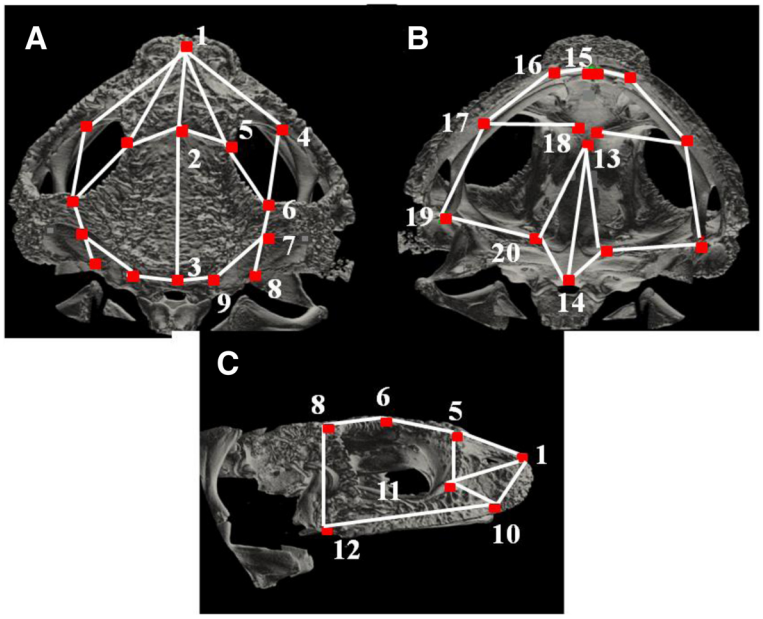


Fig. 2 Landmarks and linear distances used in the toad skulls. Numbered landmarks in both sides of the skull are shown as red dots in dorsal (**a**) and ventral (**b**) views, and only landmarks of the right side of the skull are shown in the lateral (**c**) view (descriptions in Table 1). Landmarks were placed in bone sutures or bone processes either with TINA-Landmark software in 3D images or with a 3D digitizer in the real skulls. Linear distances are shown as white lines and represent individual bone dimensions, as shown in Table 2

the distances separately, we can see that some distances have higher differences, ranging from −0.38 mm to 0.36 mm (Table A2 in the Additional file 1). These distances correspond to different bones in the toad skulls, indicating that the deviations in the relative landmark positions are not localized in a few bones. We can see that the higher differences (above 0.17 mm in magnitude, the highest difference found for MED-HIGH, values in boldface in Table A2) have approximately the same magnitude in mm for small and long distances. For instance, small distances 5, 8 and 13 (distance means below 3.0 mm) have a mean between-methods difference of 0.23 mm, while long distances have a mean of 0.26 mm. However, as mentioned above, the same magnitude of between-methods error for small and long distances results in a higher proportion of error in relation to the distances means when considering the smaller ones (Table A3 in Additional file 1). When looking at the between-methods mean percentage error, the error in relation to the distances mean lengths, we can see that the cases around 10 % correspond to three small distances: 5 (frontoparietal bone), 13 (nasal bone) and 25 (parasphenoid bone), and just one longer distance: 16 (squamosal bone), which is composed of two landmarks detected with high deviations for DIG (landmarks 8 and 12). The magnitudes of differences that we found for DIG and the micro-CT resolutions are similar to differences reported by other authors when comparing measurements taken with CT data and digital calipers [18, 21], although the last authors did find a systematic bias in the CT data (all distances under-estimated). Yet, when comparing our results with other authors that measured the same specimens with a 3D digitizer (Polhemus 3Space) and CT [15, 19], our error between-methods is much smaller (around five to ten times smaller). This is probably due to the fact that we have used a much higher resolution than these authors in both the micro-CT as well as the digitizer since current equipments are an order of magnitude more accurate than 20 years ago.

The only distances estimated with more error when scanning at different resolutions were distances17 and 19 (0.17 mm of difference), both distances in the parasphenoid bone, one of the thinner bones in the toad skull. It is possible that in this case scanning with HIGH enhanced the visualization of the sutures in the parasphenoid bone compared to MED. Scanning with HIGH instead of with MED increases the scanning time per skull over 4 times (20 min versus 1.5 h), so the advantage in enhancing the accuracy of some distances by using higher resolution depends on the question being asked and the time available. The between-methods absolute differences for each distance (Fig. 4) also indicate that some distances had higher magnitudes of difference. The worst case was for distance 16 (squamosal bone), the only long distance that presented percentage error around 10 % and an error around 0.5 mm. This magnitude of error is very close to what we considered a gross

Table 2 Linear distances determined in the toad skulls. In total, we determined 24 linear distances representing individual dimensions of the bones (or the orbit) in the toads' skulls. Distances are spread through the skull in three views (dorsal, ventral and lateral; see Fig. 1). We positioned landmarks in both sides of the skull and averaged the distances from both sides

Distances	Landmarks	Bones
1	1-2	nasal
2	2-3	frontoparietal
3	1-4	nasal
4	1-5	nasal
5	2-5	frontoparietal
6	5-6	frontoparietal
7	4-6	orbit
8	6-8	squamosal
9	7-9	occipital
10	3-9	frontoparietal
11	1-10	prenasal
12	1-11	nasal
13	10-11	nasal
14	5-11	nasal
15	10-12	maxilla
16	8-12	squamosal
17	13-14	parasphenoid
18	13-20	parasphenoid
19	15-16	premaxilla
20	16-17	nasal
21	17-18	neopalatine
22	17-19	pterygoid
23	19-20	pterygoid
24	14-20	parasphenoid

error when measuring the individuals inside each method. When looking at the distinct filter results (AL1.0-AL0.5), we can see that both the raw and absolute differences between individuals scanned with distinct filters are much lower than differences between micro-CT resolutions and 3D digitizer. Similarly, in table A3 we can notice that the highest mean percentage error for filter type comparisons was around 7 %, and all related to small distances.

Conclusions

In order to use 3D images of organisms in morphometric studies, one needs to know if the relative positions among landmarks are kept the same for different scanning procedures or rendering algorithms. By comparing landmark precision and linear distances in toad skulls scanned with two resolutions and measured with the software TINA-Landmark, with the same skulls measured with a 3D digitizer, we conclude that the degree of discrepancy is acceptable in general, although several distances had between-methods discrepancies above 5 %. Yet, we must consider that 30 % of the distances are below 3.0 mm in length, being quite small distances, thus for several other bigger organisms the error proportion will probably be lower than what we report. Attention needs to be paid in relation to the scale of the distances, as smaller distances might be estimated with proportionally more error. Using distinct micro-CT resolutions, distinct compacting factors for loading 2D cross-section sequences in TINA-Landmark, as well as using distinct filters to scan individuals that differ in bone density do not introduce high errors. We recommend the switching of filters when individuals to be compared differ in bone density, at least with the resolutions that we have tested.

Methods

Species and scanning procedures

For the micro-CT and 3D digitizer comparisons, as well as for the 3D image compacting factor analysis, we used 20 adult individuals of the toad species *Rhinella granulosa*, collected in January 2012 at a site near the municipality of Angicos (5° 39′ S, 36° 36′W), in the state of Rio Grande do Norte, NE Brazil. The toads were sacrificed by peritoneal application of an anesthetic in excess and then were fixed in 70 % alcohol. We scanned the toad skulls with an X-ray micro-CT system (SkyScan 1176, Konitch, Belgium) placed at the Instituto de Biociências, Universidade de São Paulo. All individuals were scanned with a 1.0 mm AL filter at two different resolutions: medium (**MED**: 18 μm, 70 kV, 356 μA) and high (**HIGH**: 9 μm, 65 kV, 380 μA). These resolutions correspond to different voxel sizes, the smallest volume unit in the 3D volumes: 3.85 10^{-6} mm^3 for MED and 1.92 10^{-6} mm^3 for HIGH, which are much smaller than conventional CTs (voxel sizes ranging from 0.1 to 5 mm^3). Before scanning, toads were wrapped with Parafilm to avoid too much alcohol evaporation and consequent dehydration, which could lead to blurred images by sample contraction. The scanning time per skull for MED was 20 min. and for HIGH was 1.5 h. After scanning, the skulls were reconstructed using NRecon software (SkyScan, Konitch, Belgium). In this process, 2D projection images are reconstructed to cross-section images by use of a mathematical algorithm (Feldkamp). The first step in the reconstruction process was to choose the lower and upper limits of the threshold for the linear attenuation coefficient (AC). The AC measures how much the intensity of the X-ray beam is reduced as it passes through the material being scanned, and is related to the density of the materials. Based on the histograms of soft tissue and bone density of the skulls, we chose an AC threshold of 0.0 and 0.05 for all specimens scanned. Afterwards, we applied different types of

Table 3 Within-methods mean individual landmark distances in mm

Mean in mm									
Landmarks	DIG	HIGH	MED	AL1.0	AL0.5	S2	S2/D2	S3	S3/D2
1	0.299	0.227	0.219	0.267	0.221	0.234	0.227	0.220	0.239
2	0.284	0.212	0.242	0.165	0.168	0.236	0.234	0.215	0.246
3	**0.357**	0.188	0.182	0.243	0.222	0.182	0.191	0.184	0.172
4	0.290	0.203	0.205	0.208	0.238	0.270	0.261	0.270	0.268
5	0.363	0.245	0.232	0.259	**0.422**	0.230	0.239	0.248	0.227
6	**0.365**	0.187	0.203	0.256	0.296	0.210	0.191	0.199	0.213
7	**0.345**	0.177	0.179	0.170	0.149	0.163	0.167	0.171	0.160
8	0.338	0.323	0.266	0.290	0.374	0.289	0.227	0.300	0.292
9	**0.473**	0.215	0.219	0.271	0.289	0.227	0.207	0.233	0.210
10	**0.380**	0.182	0.179	0.190	0.209	0.174	0.171	0.181	0.182
11	**0.303**	0.182	0.191	0.180	0.160	0.191	0.183	0.181	0.148
12	**0.435**	0.282	0.301	0.262	0.254	0.296	0.285	0.287	0.266
13	0.341	0.275	0.298	0.337	0.383	0.267	0.259	0.271	0.275
14	**0.500**	0.226	0.205	0.192	0.183	0.195	0.202	0.220	0.219
15	**0.342**	0.141	0.152	0.166	0.200	0.147	0.140	0.144	0.137
16	**0.252**	0.137	0.140	0.161	0.179	0.167	0.128	0.147	0.144
17	0.266	0.173	0.201	0.218	0.171	0.189	0.186	0.185	0.196
18	0.355	0.226	0.232	0.199	0.240	0.224	0.257	0.236	0.224
19	**0.516**	0.216	0.252	0.282	0.286	0.220	0.214	0.209	0.214
20	**0.481**	0.196	0.175	0.222	0.226	0.192	0.199	0.191	0.187

The table shows the mean deviation of individual landmarks from the mean landmark. We used Generalized Procrustes Analysis (GPA) to superimpose individual landmarks, but we avoided the spread of variation from any one landmark to the others (see text). To get back to a scale in mm, all the mean deviation values were multiplied by the mean centroid size of the correspondent sample. Values in bold are the highest deviations for DIG compared to MED and HIGH or for AL0.5 compared to AL1.0

corrections in the reconstruction process to soften some undesirable effects in the images: post-alignment = −5.0 to 2.0; ring artifact reduction = 2.0 to 4.0; beam-hardening correction = 30 % and smoothing = 2.0. These corrections are important to avoid blurring and artifacts in the 3D images and for all corrections the values used were small compared to maximum values. Therefore, we do not expect these corrections to interfere in landmark visualization, but instead to improve its identification. The sequences of cross-section images varied from 400 to 600 images for MED (500 MB to 1 GB per skull file size) and from 800 to 1,200 images for HIGH (1 to 5 GB per skull file size) and were all in BMP extension.

For the filter type analysis, we used 20 adult museum specimens of *Rhinella pygmaea* (MZUSP, São Paulo, Brazil). This species was chosen because several specimens presented too much transparency or even holes in their skulls, especially in the squamosal, sphenethmoid and parasphenoid bones, preventing the placement of some landmarks and also suggesting that there were differences in bone density among individuals (Fig. 1). We scanned the 20 toads at medium resolution using two distinct filters: 1.0 mm AL (**AL1.0**: 70 kV, 356 μA) and 0.5 mm AL (**AL0.5**: 50 kV, 500 μA). Filters are thin metal sheets that are set in front of the X-ray source and can have different thickness. Filtration retains a part of the low energy photons of the X-ray, thus increasing mean X-ray energy. The reduction in AL filter thickness from 1.0 mm to 0.5 mm results in a lower mean energy of the X-ray because photons with lower energy traverse the filter and achieve the thinner bones, being retained on them [22]. The individuals used in this analysis presented low transparency and no holes since the objective was to evaluate whether the use of distinct filters changed the landmarks positions and the linear distances obtained in the same individuals. In these toads all landmarks could be placed, which would not be the case if we had used specimens with holes or high transparency. CT system, scanning and reconstruction procedures were the same as described above.

Landmarking procedure and linear distances in 3D images

We placed 20 landmarks at bone sutures (type I landmarks) or bone processes (type II landmarks; 35

Table 4 Within and between methods mean distance repeatabilities and mean raw and absolute differences for the 3D digitizer (DIG) versus micro-CT resolutions (MED or HIGH) comparison. Each individual was measured twice by each of the three methods and mean ± s.d. distance repeatabilities were calculated for within (considering only replicates of the same individual measured with the same method) and between-methods (considering the same individual measured with different methods). Within and between methods calculations were also done for raw and absolute differences between linear distances (mean ± s.d.). The last line of the table shows the mean between-method percentage error in relation to distance means

Mean distance repeatabilities			
Within	**DIG**	**MED**	**HIGH**
	0.94 ± 0.07	0.97 ± 0.02	0.98 ± 0.01
Between	**MED + HIGH**	**DIG + MED**	**DIG + HIGH**
	0.95 ± 0.03	0.76 ± 0.10	0.76 ± 0.12
Mean differences (mm)			
Within	**DIG**	**MED**	**HIGH**
Raw	0.01 ± 0.04	−0.002 ± 0.02	−0.003 ± 0.02
Absolute	0.09 ± 0.03	0.07 ± 0.02	0.06 ± 0.02
Between	**MED-HIGH**	**DIG-MED**	**DIG-HIGH**
Raw	−0.03 ± 0.14	−0.02 ± 0.28	0.01 ± 0.28
Absolute	0.11 ± 0.05	0.27 ± 0.1	0.26 ± 0.1
% of mean	2.5 ± 1.0	6.5 ± 2.3	6.4 ± 3.0

landmarks if counting both sides of the skull, left and right) in all views (dorsal, ventral and lateral; Table 1 and Fig. 1) using TINA-Landmark software [7]. This software provides different views of the data: a 3D volume of the whole skull and three 2D views of the orthogonal cross-sections (axial, sagittal and transversal) throughout the image sequence, enabling landmarks to be precisely placed given that the views are linked (*i.e.* when setting a landmark in the 3D volume, it also appears in all 2D views allowing a refinement of the landmark position). The skull's cross-section sequences were loaded in TINA-Landmark by using the "Sequence Tool" after the BMP files were converted to DICOM files. The Sequence Tool presents two types of image compacting factors for loading heavy files: "Stride", which reduces the sequence in the inter-slice direction (*i.e.*, stride = 2.0 and "Stride average" = ON means that an average image will be loaded at every two cross-sections from the sequence, resulting in half of the original size of the sequence file); and "Down-sample", which reduces the sequence along the x and y directions (*i.e.*, down-samples the size of the pixels; [23]). MED resolution data were loaded using stride = 2.0 whereas HIGH resolution data were loaded using stride = 3.0 and down-sample = 2.0, since they were heavier and failed to load with the same compacting factor than MED data. The down-sampling procedure only affects the visualization of the 3D volume of the skulls, not changing the visualization of the cross-section images (axial, sagittal and transversal views).

Appropriate visualization of 3D volumes and determination of bone threshold value (the average density of soft tissue and bone to guarantee that landmarks are placed exactly where the mouse is pointing in the 3D image) were obtained by following the recommendations contained in the TINA Geometrics Morphometrics Toolkit manual [23]. Finally, after all steps of image adjustments, a landmark list was loaded with the 35 points and

Table 5 Within and between methods mean distance repeatabilities and mean raw and absolute differences for the image compacting factor comparison and the filter type comparison

Mean distance repeatabilities						
Within	**S2**	**S3**	**S2/D2**	**S3/D2**	**AL0.5**	**AL1.0**
	0.97 ± 0.02	0.98 ± 0.03	0.97 ± 0.02	0.97 ± 0.02	0.95 ± 0.04	0.93 ± 0.10
Between	**S2 + S3**	**S2 + S2/D2**	**S2 + S3/D2**	**S3 + S2/D2**	**S3 + S3/D2**	**AL0.5 ± AL1.0**
	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.04	0.95 ± 0.03	0.95 ± 0.04	0.83 ± 0.16
Mean differences (mm)						
Within	**S2**	**S3**	**S2/D2**	**S3/D2**	**AL0.5**	**AL1.0**
Raw	−0.002 ± 0.02	−0.003 ± 0.01	0.008 ± 0.02	0.006 ± 0.02	0.0 ± 0.02	0.003 ± 0.03
Absolute	0.7 ± 0.02	0.06 ± 0.02	0.07 ± 0.02	0.07 ± 0.02	0.07 ± 0.02	0.07 ± 0.03
Between	**S2-S3**	**S2-S2/D2**	**S2-S3/D2**	**S3-S2/D2**	**S3-S3/D2**	**AL0.5-AL1.0**
Raw	−0.01 ± 0.13	0.0 ± 0.12	0.0 ± 0.13	0.02 ± 0.13	0.02 ± 0.13	−0.03 ± 0.21
Absolute	0.10 ± 0.04	0.09 ± 0.04	0.10 ± 0.04	0.10 ± 0.06	0.10 ± 0.05	0.14 ± 0.05
% of mean	2.3 ± 0.9	2.3 ± 1.1	2.4 ± 1.0	2.4 ± 1.1	2.3 ± 1.0	3.4 ± 2.0

The four distinct values of the image compacting factors stride and down- sample were: (S2): stride = 2.0; (S3): stride = 3.0; (S2/D2): stride = 2.0 and down-sample = 2.0 and (S3/D2): stride = 2.0 and down-sample = 3.0. The two different scanning filters were: aluminum 0.5 mm (AL0.5) and aluminum 1.0 mm (AL1.0). Within and between methods mean ± s.d. distance repeatabilities and mean ± s.d. raw and absolute difference were calculated, as well as mean percentage error in relation to the distances means of the between methods error

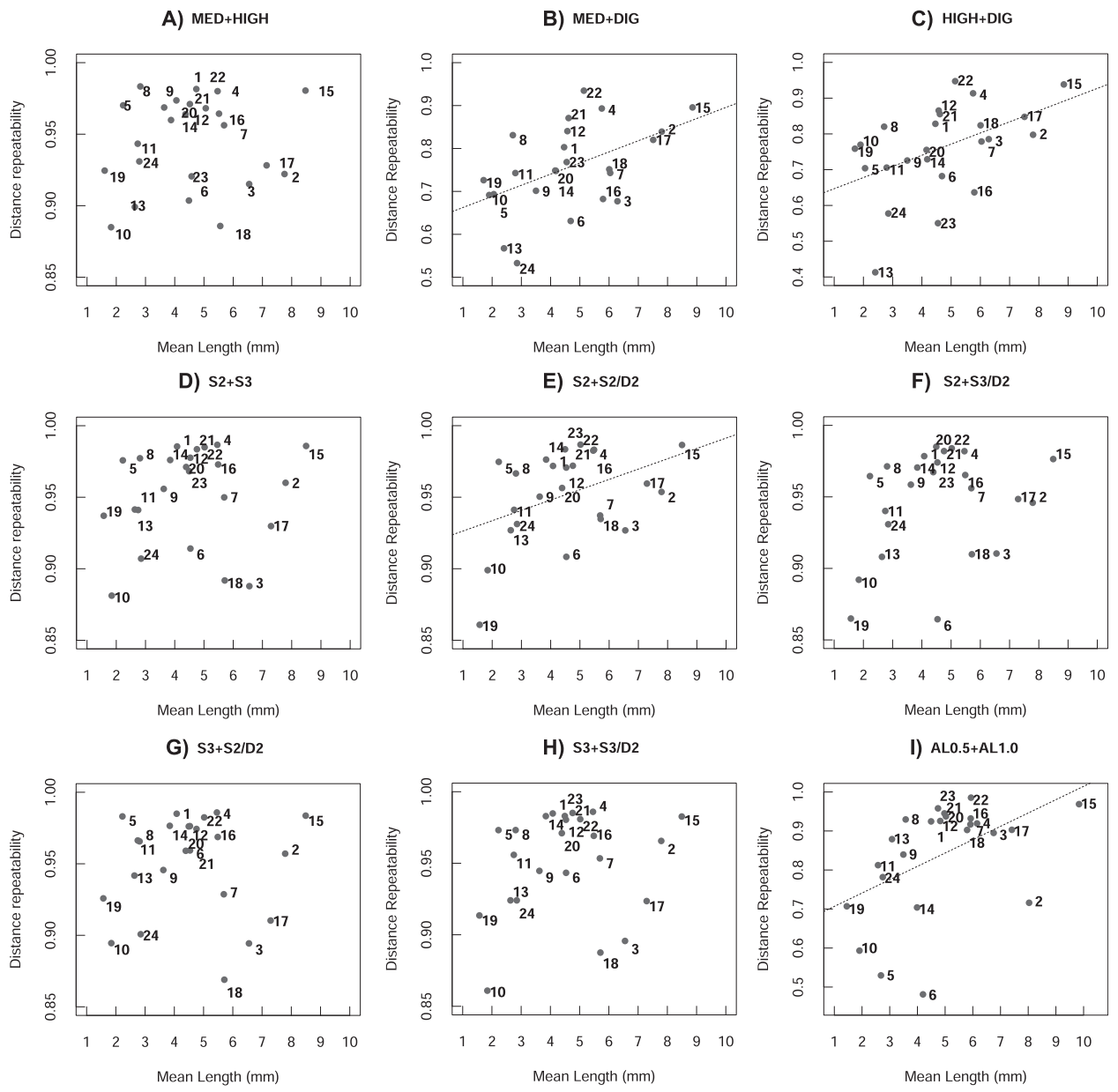


Fig. 3 Between-method distance repeatability in relation to distance mean length. Individual toad skulls were each measured twice with different methods: using a 3D digitizer (DIG); using CT scans at medium (MED) or high (HIGH) resolutions; CT scans at MED loaded in TINA-Landmark with distinct values of the compacting factors stride and down-sample (S2, S3, S2/D2 and S3/D2); and CT scans at MED with 0.5 mm aluminum filter (AL0.5) or 1.0 mm aluminum filter (AL1.0). Data from different methods were joined together and distance repeatabilities calculated, representing the reliability in placing landmarks with the different methods. The dashed lines indicate significant correlations between the variables. Numbers 1–24 correspond to the distances as described in Table 2. **a)** MED + HIGH; **b)** DIG + MED; **c)** DIG + HIGH; **d)** S2 + S3; **e)** S2 + S2/D2; **f)** S2 + S3/D2; **g)** S3 + S2/D2; **h)** S3 + S3/D2 and **i)** AL0.5 + AL1.0

each was placed two times (two replicates per individual) in the skull's images of *R. granulosa* and *R. pygmea* individuals, so we could assess measuring reliability. The corresponding *x*, *y* and *z* landmark coordinates were saved in a TXT file, which was later loaded in the R programming environment [24] where distances were calculated.

We determined 24 linear distances in the skull of the toads, all distances representing individual bone dimensions (Table 2, except distance 7 which corresponds to orbit size). To obtain distances in mm, the coordinates of MED and HIGH were multiplied by 0.01742 and 0.00871, respectively, which are the size of the pixels in mm for each resolution. In the case of the *R. pygmea*

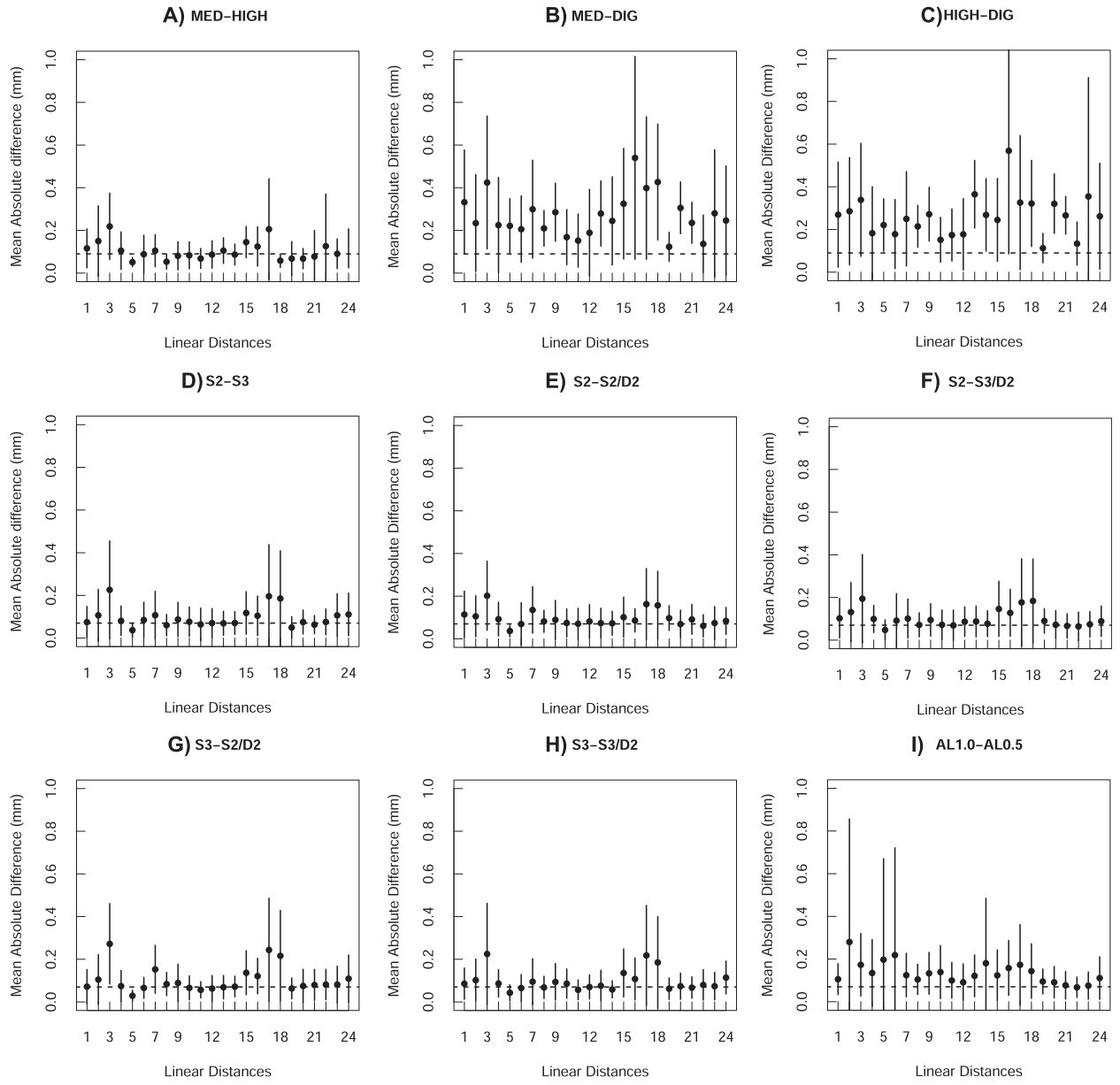


Fig. 4 Between-method mean absolute differences in linear distances. Using the same data sets as described in Fig. 3, absolute differences in mm between replicates of the same individuals measured with different methods were calculated and averaged across individuals. Circles represent means and vertical lines are the s.d. Dashed lines indicate the mean within-method absolute differences. Numbers 1–24 correspond to the distances as described in Table 2. **a)** MED + HIGH; **b)** DIG + MED; **c)** DIG + HIGH; **d)** S2 + S3; **e)** S2 + S2/D2; **f)** S2 + S3/D2; **g)** S3 + S2/D2; **h)** S3 + S3/D2 and **i)** AL0.5 + AL1.0

specimens that were scanned with different filters, landmark coordinates were multiplied by 0.01742 because they were only scanned at medium resolution. Distances from replicates of the same individual (within each method) were inspected for gross measurement error (difference between replicates above 0.5 mm for most distances, except for distances 5, 8, 10, 11, 13, 17 and 24, which have means smaller than 3.0 mm, and were controlled for error above 0.3 mm), and when detected, the landmarks correspondent were placed again in both replicates and corrected. The reference to consider the magnitude of a gross error was based on the precision of the 3D digitizer, which is 0.01 mm. This procedure was adopted because these gross errors could lead to misleading conclusions about the between-methods analysis, since they are actually referred to gross human error inside each method. In order to analyze the possible effect on the landmarks and distances of switching the

compacting factors (stride and downsample) when loading the cross-section sequences, we placed the landmarks in the *R. granulosa* skulls scanned in medium resolution using the following values in TINA-Landmark: (**S2**) stride = 2.0; (**S3**) stride = 3.0; (**S2/D2**) stride = 2.0 and down-sample = 2.0; and (**S3/D2**) stride = 3.0 and down-sample = 2.0. Again, each individual at each of these four compacting values were measured twice and gross measurement error between replicates inside each compacting factor was controlled. Thus, total sample size for the compacting factor analysis was 20 individuals for each of the four situations, and 80 in total.

Landmarking procedure and linear distances using 3D digitizer

The same *R. granulosa* adults that we scanned with MED and HIGH resolutions were cleaned and had their skin removed. In this process, we lost four of the skulls by crushing, 16 skulls remaining. We placed the same landmarks described above in the cleaned skulls by using a 3D digitizer (Microscribe 3DX, IL). Since the toad skulls are quite small (total length around 20.0 mm), we used a binocular loupe to mark the landmarks with a pencil. Then, we placed the digitizer pen in the graphite marks to digitize the landmarks. We placed landmarks in all individuals twice and gross error in the distances from one replicate to the other was also controlled as described before at the time of landmark digitalization. A TXT file was created for the coordinates obtained from DIG and was loaded in R environment in order to calculate the same distances as MED and HIGH. The landmarks and distances obtained by using DIG were considered as the reference for comparisons.

Within-method mean individual landmark distances

To assess the differences of placing landmarks in the skulls 3D volumes compared to placing them in the real skulls with the 3D digitizer, we used a superimposition method, the Generalized Procrustes Analysis (GPA), to estimate the sample mean landmarks (mean shape) and to calculate individual distances from the mean landmarks within each method. The GPA sumperimposes landmark configurations of several individuals, scaling them to unit centroid size, and uses least-square estimates for translation and rotation parameters [25]. These transformations of the landmarks are needed because there is no natural coordinate system that is common to all individuals that are digitized and a common shape space is achieved when doing this procedure [26, 27]. However, this transformation process confounds variation at different homologous landmarks, i.e., the most variable landmarks among individuals (non-isotropic variation) have their variation spread across other landmarks [16, 26, 27]. Thus, to avoid this problem in our landmark variation analysis, we followed the idea presented in van der Linde and Houle (2009) [27] and excluded from the GPA one landmark at a time to estimate the sample mean shape and the rotation matrices for the 19 remaining landmarks for each individual. By doing that, we precluded that the variation in the landmark of interest got spread through the other landmarks. Afterwards, we multiplied the excluded landmark of each individual by its corresponding rotation matrix (all landmarks of an individual are rotated in the same angle) to have all landmarks in the same new coordinate system, the so called shape space. The last step was to calculate the individual distance of the landmark of interest from the mean landmark (in the shape space, therefore with no scale). This procedure was repeated 20 times for each method, so that all 20 landmarks were excluded at each time and the distances from the individuals and the mean landmark could be calculated. To compare the different methods, we also calculated the mean individual landmark distance from the mean landmarks. Finally, in order to compare the methods of placing landmarks in a more intuitive scale, we multiplied the mean deviations and SD values by its corresponding mean sample centroid size in mm. Only the medial and left side landmarks were used in this analysis and the GPA was done with the "geomorph" [28] and "shapes" [29] packages in R environment.

Distance repeatability within and between methods

We calculated repeatability values for all distances in every data subset using the calculations described in [20]. Each individual is a group with two replicates and the between-group variance is the sum of squares of the deviations of the group means from the total mean, whereas the within-group variance corresponds to the sum of squares of the deviations of each replicate from its own group mean. The index is calculated as follows:

$$r = \frac{s_A^2}{\left(s^2 + s_A^2\right)}$$

s_A^2 being the variance among groups and s^2 the residual variance, both calculated from the sum of squares in an ANOVA.

We calculated repeatabilities for MED, HIGH and DIG separately, and also for joined data sets, (MED + HIGH), (MED + DIG), (HIGH + DIG), with the three between-method data sets made up of four replicates per individual. In the same manner, we calculated repeatability for replicates of the different values of stride and down-sample, S2, S3, S2/D2 and S3/D2 separately; and also of joined data sets, (S2 + S3), (S2 + S2/D2), (S2 + S3/D2), (S3 + S2/D2), (S3 + S3/D2) and (S2/D2 + S3/D2). Finally, we did the same calculations for the different

filters data sets separately, AL1.0 and AL0.5, and also joining all the replicates (AL1.0 + AL0.5). The distance repeatabilities for separated data sets indicate the within-method reliability of measuring the skulls, whereas the repeatabilities for joined data sets comprise both within and between-method reliability.

Raw and absolute differences within and between methods

In addition to the repeatability analysis, we also calculated raw and absolute differences of the distances between replicates, within and between methods. In the first case, we just subtracted the distances between replicates of the same individual and calculated the mean raw and absolute differences across all individuals for each distance. These calculations were made for each data set separately: MED, HIGH, DIG, S2, S3, S2/D2, S3/D3, AL1.0 and AL0.5. The between-method analysis was done by subtracting the distances between the same individuals measured with different methods (*e.g.*, in the case of MED-HIGH, distances from individual 1 of MED were subtracted from the distances of individual 1 of HIGH, and so on for all 20 individuals) and calculating the raw and absolute mean differences across all individuals for each distance. All the differences that involved DIG (MED-DIG and HIGH-DIG) had 16 (4 skulls were lost in the cleaning procedure), while the rest of the differences between methods had 20 individuals. The raw differences show whether any method presents a consistent bias (e.g. under or overestimating the distances) in the data. The absolute differences indicate the magnitude of the error within and between methods, independent if the errors in the distances are in one direction or another. To get an idea of the error in relation to the distances means, we calculated the percentage error by dividing the absolute mean differences by their corresponding means and multiplying by 100.

Statistical analysis

Correlation tests were all done using Pearson product moment correlation and significance level of 0.05. All ANOVAs in the repeatability analysis, correlation tests and graphics were done in the R environment [23].

Additional file

Additional file 1: Table A1. Within-methods mean individual landmark distance without scale. **Table A2.** Between methods mean raw differences (in mm) in all the distances for joined data sets of micro-CT resolutions and 3D digitizer and of different filter type. **Table A3.** Mean percentage error in all distances for joined data sets of micro-CT resolutions and 3D digitizer and for distinct filter types.

Abbreviations

AL0.5: Aluminum filter 0.5 mm; AL1.0: Aluminum filter 1.0 mm; DIG: 3 D digitizer; HIGH: High resolution; MED: Medium resolution; D2: Down-sample = 2.0; S2: Stride = 2.0; S3: Stride = 3.0; s^2: Within-group variance; s^2_A: Among-group variance.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MNS conceived the general idea of the study, collected and analyzed the data and drafted the manuscript. GM conceived the design of the study, analyzed and discussed the data and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Dr. Taran Grant from MZUSP for the loan of the *R. pygmaea* specimens used in this study. MNS also thanks Francisco de Assis and family for providing house in the field, and Pedro Leite Ribeiro for helping to collect and preserve *R. granulosa* specimens used in the study. We thank the technician Marco Antônio Puodizius for helping cleaning the toad skulls and the Departamento de Zoologia at Instituto de Biociências, USP, for allowing the use of their facilities. We also thank Guilherme Garcia for assistance with the landmark analysis and the two anonymous reviewers for their very careful revision of our manuscript and for their important criticisms. This study was supported by Fundação de Amparo à Pesquisa de São Paulo (FAPESP), Ph.D. scholarship 2011/07584-2, multi-user project 2009/54203-4 and thematic project2011/14295-7.

Received: 10 September 2014 Accepted: 10 April 2015
Published online: 03 June 2015

References

1. Costantini D, Alonso ML, Moazen M, Bruner E. The Relationship Between Cephalic Scales and Bones in Lizards: A Preliminary Microtomographic Survey on Three Lacertid Species. Anat Rec Adv Integr Anat Evol Biol. 2010;293:183–94.
2. Ekdale EG. Ontogenetic Variation in the Bony Labyrinth of Monodelphis domestica (Mammalia: Marsupialia) Following Ossification of the Inner Ear Cavities. Anat Rec Adv Integr Anat Evol Biol. 2010;293:1896–912.
3. Wilkinson M, San Mauro D, Sherratt E, Gower DJ. A nine-family classification of caecilians (Amphibia: Gymnophiona). Zootaxa. 2011;2874:41–64.
4. Cuff AR, Rayfield EJ. Feeding Mechanics in Spinosaurid Theropods and Extant Crocodilians. PLoS One. 2013;8:e65295.
5. Gignac PM, Kley NJ. Iodine-enhanced micro-CT imaging: Methodological refinements for the study of the soft-tissue anatomy of post-embryonic vertebrates. J Exp Zool B Mol Dev Evol. 2014;322:166–76.
6. Rosset A, Spadola L, Ratib O. OsiriX: An Open-Source Software for Navigating in Multidimensional DICOM Images. J Digit Imaging. 2004;17:205–16.
7. Schunke AC, Bromiley PA, Tautz D, Thacker NA. TINA manual landmarking tool: software for the precise digitization of 3D landmarks. Front Zool. 2012;9:6.
8. Kohn LA, Cheverud JM. Issues in evaluating repeatability of an imaging system for use in anthropometry, Proceedings of the Electronic imaging of the Human bodyworking group. OH: Crew System Ergonomics Information Analysis Center, Dayton; 1992.
9. Kim G, Jung H-J, Lee H-J, Lee J-S, Koo S, Chang S-H. Accuracy and Reliability of Length Measurements on Three-Dimensional Computed Tomography Using Open-Source OsiriX Software. J Digit Imaging. 2012;25:486–91.
10. Halperin-Sternfeld M, Machtei E, Horwitz J. Diagnostic Accuracy of Cone Beam Computed Tomography for Dimensional Linear Measurements in the Mandible. Int J Oral Maxillofac Implants. 2014;29:593–9.
11. Zelditch ML. Ontogenetic variation in patterns of phenotypic integration in the laboratory rat. Evolution. 1988;42:28–41.
12. Cheverud JM. Morphological Integration in the Saddle-back tamarin (Saguinus fuscicollis) cranium. Am Nat. 1995;145:63–89.
13. Young NM, Hallgrímsson B. Serial homology and the evolution of mammalian limb covariation structure. Evolution. 2005;59:2691–704.
14. Porto A, de Oliveira FB, Shirai LT, De Conto V, Marroig G. The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes. Evol Biol. 2009;36:118–35.
15. Corner BD, Lele S, Richtsmeier JT. Measuring Precision of Three-Dimensional Landmark Data. J Quant Antrophology. 1992;3:347–59.

16. Richtsmeier JT, Paik CH, Elfert PC, Cole TM, Dahlman HR. Precision, Repeatability and Validation of the Localization of Cranial Landmarks Using Computed Tomography Scans. Cleft Palate Craniofac J. 1995;32:217–27.
17. Zeldicth ML, Swiderski DL, Sheets HD, Fink WL. Geometric Morphometrics for Biologists: A Primer. New York and London: Elsevier Academic Press; 2004. p. 237.
18. Richard AH, Parks CL, Monson KL. Accuracy of standard craniometric measurements using multiple data formats. Forensic Sci Int. 2014;242:177–85.
19. Stull KE, Tise ML, Ali Z, Fowler DR. Accuracy and Reliability of measurements obtained from computed tomography 3D volume rendered images. Forensic Sci Int. 2014;238:133–40.
20. Lessells CM, Boag PT. Unrepeatable Repeatabilities: A Common Mistake. Auk. 1987;104:116–21.
21. Fernandes, TMF, Adamczyk, J, Poleti, ML, Henriques, JFC, Friedland, B, Garib, DG: Comparison between 3D volumetric rendering and multiplanar slices on the reliability on linear measurements on CBCT images: an in vitro study. Journal of Applied Oral Science 2014.
22. SkyScan. SkyScan 1176. In vivo X-Ray Microtomograph Instruction Manual. 2011.
23. Bromiley PA, Ragheb H, Thacker NA. The TINA Morphometrics Geometric Toolkit, Tina Memo No 2010–007. 2012.
24. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2013. URL: http://www.R-project.org/.
25. Bookstein, FL: Morphometrics tools for landmark data: Geometry and Biology. Cambridge University Press. 1997. p. 425.
26. Lele S. Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference. Math Geol. 1993;25:573–602.
27. van der Linde K, Houle D. Inferring the Nature fo Allometry from Geometric Data. Evol Biol. 2009;36:311–22.
28. Adams DC, Otarola-Castillo E. geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol Evol. 2013;4:393–9.
29. Dryden, IL: shapes: Statistical Shape Analysis. URL: http://cran.r-project.org/web/packages/shapes/index.html.