Tereza Cristina Giannini

Distribuição geográfica de abelhas e plantas
associadas através de modelagem
computacional

*Geographical distribution of associated bees
and plants through computational modeling*

Orientadora: Isabel Alves-dos-Santos

Co-Orientador: Antonio Mauro Saraiva

São Paulo

2011

Tereza Cristina Giannini

Distribuição geográfica de abelhas e plantas
associadas através de modelagem
computacional

*Geographical distribution of associated bees
and plants through computational modeling*

Tese apresentada ao Instituto de
Biociências da Universidade de São
Paulo, para a obtenção de Título de
Doutor em Ecologia, na Área de
Ecossistemas Terrestres.

Orientadora: Isabel Alves-dos-Santos

Co-Orientador: Antonio Mauro Saraiva

São Paulo

2011

Ficha Catalográfica

Giannini, Tereza Cristina
Distribuição geográfica de abelhas e plantas
associadas através de modelagem computacional
*Geographical distribution of associated bees and
plants through computational modeling*

Número de páginas: 140

Tese (Doutorado) - Instituto de Biociências
da Universidade de São Paulo. Departamento de
Ecologia.

1. Palavra-Chave Principal : distribuição
geográfica; 2. Palavra-Chave Secundária:
modelagem preditiva; 3. Palavra-Chave
Secundária: abelhas

I. Universidade de São Paulo. Instituto de
Biociências. Departamento de Ecologia.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Orientador(a)

*Dedico este trabalho aos meus pais, José e Deolinda
- os primeiros que me ensinaram que a beleza está ao redor -*

e também à Rita (in memoriam), Lourdes, Luciana, Marcelo, Gabriel e Caio

os amores da minha vida.

Naquele Império, a Arte da Cartografia atingiu tal perfeição que o mapa duma só província ocupava toda uma cidade, e o mapa do Império, toda uma província. Com o tempo, esses mapas desmedidos não mais satisfaziam e os Colégios de Cartógrafos levantaram um mapa do Império que tinha o tamanho do Império e coincidia ponto por ponto com ele. Apegadas ao estudo da Cartografia, as gerações seguintes entenderam que esse extenso mapa era inútil e não sem impiedade o entregaram às inclemências do sol e dos invernos. Nos desertos do oeste subsistem despedaçadas ruínas do mapa, habitadas por animais e por mendigos. Em todo o país não resta outra relíquia das disciplinas geográficas.

Citado por Jorge Luis Borges

*Se eu posso atingir o coração de Dublin
Posso atingir o coração de todas as cidades do mundo.
No particular está contido o universal.*

James Joyce

Agradecimentos

Aos meus orientadores, Isabel Alves dos Santos e Antonio Mauro Saraiva, pelo apoio, oportunidades e orientação ao longo do projeto.

Ao Professor Jacobus Biesmeijer da Universidade de Leeds (Leeds, Inglaterra) e ao Dr. Daniel S. Chapman do *Centre for Ecology and Hydrology* (Edimburgo, Escócia) por terem me recebido em seus laboratórios e pelas valiosas discussões e aprendizados.

À Professora Vera Lucia Imperatriz Fonseca por seu incentivo e entusiasmo constantes.

A todos os Professores do Departamento de Ecologia do Instituto de Biociências, especialmente Astrid M. P. Kleinert, Roberto Shimizu, Gisela Shimizu, Silvana Buzato, Pêrsio S. Santos Filho, Paulo Inácio K. L. Prado e Paulo R. Guimarães Jr. pelo apoio e discussões. Agradeço também ao Professor José Rubens Pirani do Departamento de Botânica do IBUSP por ter me apresentado o gênero *Krameria*.

Aos Professores do Laboratório de Automação Agrícola da Escola Politécnica, especialmente Pedro L. P. Correa, Carlos E. Cugnasca, André R. Hirakawa e ao Professor João José Neto do Laboratório de Linguagem e Técnicas Adaptativas, também da EPUSP, pela convivência e entusiasmo acerca dos temas da Biologia.

Aos Professores da Universidade Federal do Paraná, Gabriel Melo e Danuncia Urban pela ajuda com os espécimes de *Peponapis fervens* da coleção entomológica Padre Jesus Santiago Moure.

À Marinez Ferreira Siqueira do Jardim Botânico do Rio de Janeiro e Renato de Giovanni do Centro de Referência em Informação Ambiental, pelo apoio inicial com as ferramentas de modelagem.

À Maria Aldete J. F. Ferreira (Embrapa), Luiz Soledad (Federação dos Trabalhadores na Agricultura do Estado de Goiás), Antonio Aguiar (Museu de Zoologia), Rafael Lira-Saade (*Universidad Nacional Autónoma de México*), Michael Nee (*New York Botanical Garden*), Andreas Taeger (*Deutsches Entomologisches Institut*), Robert L. Zuparko (*California Academy of Sciences*), John Ascher (*American Museum of*

National History) e Brian Harris (*Smithsonian Institution National Museum of Natural History*) pela cessão de parte dos dados para montagem do banco de dados utilizado nesse projeto. Agradeço igualmente a Rafael Lira-Saade e Ricardo Ayala (ambos da *Universidad Nacional Autónoma de México*) e James Cane (*United States Department of Agriculture*) pela checagem e correção dos pontos de ocorrência de *Peponapis* e *Cucurbita*. Agradeço também aos amigos Adriana Takahasi, Maria Cláudia M. P. Medeiros e Jaime Passarini Jr. pela participação na montagem do banco de dados sobre *Krameria* e à Mariana Taniguchi pela troca de dados de *Centris* e *Krameria*.

Aos colegas do Laboratório de Abelhas (Eduardo, Sheina, Samuel, Maria, Aline, Tarcila, Renata, Guará, Charles, Carina, Patrícia, Bruno e Paola) e aos do Laboratório de Automação Agrícola (Etienne, Fabiana, Renata, Wilian, Eurípedes, Eduardo, Renato, Fabrício e Elisângela) pela interação e amizade. Um agradecimento especial à Marilda, Sérgio e Denise pelo ‘tempo de casa,’ e ao André, pelas intermináveis (e muito apreciadas!) discussões sobre modelagem.

Aos amigos de Edimburgo, Leeds e Leiden, especialmente, Janice, Alan, Luiza, John, Stephanie, Amy, Joseph e Katie. Agradeço também à Judith Slaa e Jacobus Biesmeijer pela gentileza de terem me recebido em sua casa em Leiden, e às crianças Lars e Channa pela doçura e alegria.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela bolsa de estudos concedida.

Índice

Introdução	1
Capítulo 1. Modelagem preditiva de distribuição de espécies: usos, aplicações e desafios atuais	10
Capítulo 2. Ecological niche modeling and geographical distribution of pollinator and plants: a case study of <i>Peponapis fervens</i> (Smith, 1879) (Eucerini: Apidae) and <i>Cucurbita</i> species (Cucurbitaceae)	42
Capítulo 3. Ecological niche similarities of <i>Peponapis</i> Robertson (Hymenoptera: Apidae) and non-domesticated <i>Cucurbita</i> L. (Violales: Cucurbitaceae) of North and Central America	67
Capítulo 4. Ecological niche modeling and principal component analysis of <i>Krameria</i> Loefl. (Krameriaceae)	91
Capítulo 5. Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants	107
Conclusão	129
Resumo	131
Abstract	132

Distribuição geográfica de abelhas e plantas associadas através de modelagem computacional

Introdução

Interações entre abelhas e plantas associadas

Desde Darwin (1859) sugere-se que a diversidade seja resultante das interações entre organismos. Para Thompson (2005) a riqueza de espécies é consequência de muitos processos de coevolução, nos quais as espécies estreitamente relacionadas apresentam evolução mútua. Especificamente em relação às plantas e seus visitantes florais, Waser (1983) sugeriu que a diversidade encontrada entre as formas, cores e tamanho das flores parece ser correlata à morfologia e comportamento dos visitantes florais com os quais elas interagem.

Especialmente as interações entre plantas e polinizadores especialistas têm sido consideradas como relações de evolução mútua, muitas vezes ocorrendo entre pares de espécies altamente especializadas (Thompson, 1994, Fenster *et al.*, 2004).

Para as abelhas, Wcislo e Cane (1996) sugeriram que a especialização varia entre as espécies que restringem seu forrageamento em uma (monolética) ou poucas (oligoléticas) espécies de plantas relacionadas. Segundo os mesmos autores, muitas das espécies solitárias são especialistas para coleta de pólen ou óleo devido às necessidades alimentares de seus imaturos, e que estas são mais abundantes em climas temperados do que em tropicais, ideia essa também defendida por Michener (1979; 2007) e Roubik (1992).

Mitchell *et al.* (2009) resumiram a complexidade dos sistemas generalista e especialista da seguinte forma: se uma espécie interage com visitantes de diferentes *taxa* que oferecem serviços similares de polinização, e se os custos da interação são comparáveis, os benefícios para as plantas deveriam ser similares e há pouco incentivo para que elas se especializem em atrair um grupo particular de polinizadores. Por outro lado, se alguns visitantes florais são mais efetivos na transferência de pólen, a seleção poderia favorecer certos caracteres que promoveriam esses polinizadores efetivos, selecionando espécies particulares de polinizadores, o que poderia explicar porque certas plantas têm

caracteres que parecem restringir seus visitantes. Assim, a relação custo-benefício das interações entre plantas e polinizadores teria um papel importante em determinar se a interação é mais generalista ou especialista.

Distribuição geográfica de espécies associadas

A interação obrigatória entre espécies ao longo de suas áreas de distribuição foi analisada por Godsoe *et al.* (2009) em duas variedades de *Yucca brevifolia* e seus polinizadores especialistas (*Tegeticula synthetica* e *T. antithetica*). O pareamento encontrado entre os fenótipos das flores e dos polinizadores ao longo da distribuição sugeriu que as diferenças entre as duas variedades de *Y. brevifolia* poderiam ter surgido como resultado de interações com os polinizadores ou como resultado de diferenças no ambiente. Os autores testaram essas hipóteses alternativas e concluíram que fatores não climáticos, tais como a interação entre as espécies, teriam sido os responsáveis pelas divergências encontradas. Craig *et al.* (2007) também encontraram variações no formato e tamanho das galhas de uma mosca (*Eurosta solidaginis*) em sua planta hospedeira (*Solidago altissima*) devido à complexas interações com diferentes inimigos naturais encontrados ao longo de sua distribuição geográfica.

Por outro lado, Toju e Sota (2006) ao analisarem a interação entre a camélia japonesa (*Camellia japonica*) e o besouro (*Curculio camelliae*) predador de suas sementes encontraram variações no comprimento da tromba do besouro e da espessura do fruto ao longo de sua distribuição, provavelmente, devido às condições climáticas presentes em diferentes latitudes de suas áreas de ocorrência.

A combinação entre os fatores bióticos e abióticos, ao longo da distribuição de espécies associadas, foi também analisada por Espíndola *et al.* (2010), que investigaram as diferentes espécies de polinizadores associados à *Arum maculatum* ao longo de sua distribuição na Europa. As diferenças encontradas nos polinizadores foram explicadas pelos fatores climáticos nas diferentes paisagens onde as espécies ocorriam. Esses fatores dirigiam a especificidade da interação uma vez que afetavam a fenologia das espécies interagentes. Já Pellissier *et al.* (2010) investigaram o efeito da altitude na distribuição de tipos florais e na diversidade funcional com relação ao padrão de atividade dos polinizadores, analisando parcelas de vegetação. Em altas altitudes as flores permanecem sempre abertas, disponíveis a polinizadores que nessas condições

mais adversas, diminuem sua atividade. Já os botões bilabiados são abundantes em altitudes mais baixas, que são caracterizadas por grande diversidade de abelhas.

Assim, as diferentes pressões seletivas modificam localmente os padrões de interação entre espécies associadas, estruturando as comunidades em diferentes áreas geográficas (Laine, 2009).

Modelagem preditiva de distribuição geográfica

Algumas ferramentas podem ser utilizadas para auxiliar na compreensão da complexidade da distribuição geográfica de espécies que interagem. A modelagem preditiva de distribuição de espécies (também conhecida como modelagem de nicho ecológico) consiste em um método computacional que apresenta múltiplos objetivos, tais como prever áreas de ocorrência potencial e estimar impactos na distribuição de espécies causados por mudanças globais (Franklin, 2009).

Basicamente, a modelagem de distribuição combina dados de ocorrência com variáveis ambientais, construindo uma representação das necessidades ecológicas da espécie (Anderson *et al.*, 2003) e determinando áreas adequadas à sua ocorrência (Phillips, 2008). Vários algoritmos têm sido aplicados para gerar tais modelos, cujos resultados podem ser projetados em um mapa que mostra áreas similares àquelas onde a espécie ocorre e representa suas áreas potenciais de ocorrência (Stockwell e Peters, 1999).

De acordo com Soberón (2010), a modelagem de distribuição baseia-se em alguns aspectos fundamentais que determinam a ocorrência das espécies: 1) os abióticos, que basicamente definem as características ambientais tais como clima, solo e altitude; 2) os bióticos, que correspondem às interações entre a espécie e seus parceiros mutualistas, seus predadores, competidores, entre outros, que ocupam o mesmo ambiente; 3) os movimentos espaciais dos indivíduos e populações, que determinam sua dinâmica de ocupação e dispersão. Os dois primeiros aspectos foram detalhados nos trabalhos pioneiros de Hutchinson (1957, 1978) sobre nicho ecológico, e o último, principalmente na revisão de Pulliam (2000). Hortal *et al.* (2010) sugeriram um aspecto adicional associado à biogeografia, que também atuaria na distribuição das espécies devido aos aspectos intrínsecos de suas áreas geográficas de ocorrência, mas também devido ao tempo e seu efeito na história evolutiva das espécies.

Os quatro aspectos citados acima moldariam a distribuição em diferentes escalas (Figura 1). Em escalas globais e continentais, os fatores abióticos e biogeográficos seriam os mais importantes. Por sua vez, em escalas menores, os fatores bióticos e de dinâmica de ocupação seriam os mais atuantes. Em escalas intermediárias, os quatro fatores atuariam conjuntamente, plasmando a distribuição de forma dinâmica ao longo de um gradiente ambiental de áreas mais ou menos adequadas à espécie.

Até recentemente, a modelagem usava exclusivamente dados abióticos para construir os cenários de distribuição. Porém, com o desenvolvimento de novas técnicas, tornou-se possível a inclusão de outros conjuntos de dados, tais como as interações entre espécies e a dispersão (Thuiller *et al.*, 2009). Assim, o avanço da técnica, apoiado num embasamento e aprofundamento conceitual, tem possibilitado diversos estudos que relacionam vários requisitos ecológicos das espécies com suas áreas de ocupação.

Figura 1. Importância relativa dos quatro tipos de fatores que afetam a distribuição das espécies através da escala espacial. As áreas coloridas indicam a importância relativa de cada fator (modificado de Hortal *et al.*, 2010).

Objetivo geral, justificativa e organização do presente trabalho

O objetivo geral do presente trabalho consistiu em analisar a distribuição geográfica e as características de nicho de abelhas que apresentam interações estreitas, com ênfase nas plantas por elas visitadas.

O estudo é justificado pelo reconhecimento crescente sobre a importância dos polinizadores na manutenção de serviços ecossistêmicos (Daily, 1997) e na agricultura (Klein *et al.*, 2007). O declínio global dos polinizadores tem sido alvo de crescente divulgação e investigação (Buchmann e Nablan, 1996; Kearns *et al.*, 1998; Biesmeijer *et al.*, 2006). A complexidade das múltiplas causas que atuam nesse declínio (Potts *et al.*, 2010) implica em diferentes estratégias de estudo, inclusive no desenvolvimento e aplicação de tecnologias mais robustas para análise.

O primeiro capítulo desse trabalho consiste em uma revisão sobre a modelagem preditiva de distribuição geográfica, onde são apresentados os fundamentos ecológicos desse tipo de modelagem, bem como uma descrição sucinta dos passos envolvidos nesse processo.

A seguir, foi analisada a similaridade dos componentes abióticos do nicho e das áreas de distribuição de um grupo de plantas e abelhas especialistas (*Cucurbita* x *Peponapis*) (Hurd *et al.*, 1971). O segundo capítulo enfoca as espécies da América do Sul e o terceiro, as espécies da América do Norte.

O quarto capítulo analisa a aplicação da modelagem em espécies de uma família botânica (Krameriaceae) que apresenta uma interação especializada com seus polinizadores (*Centris*, Apidae) (Simpson, 1989). Essa aplicação teve como principal objetivo fazer uma análise sobre a influência dos fatores climáticos na distribuição dessas espécies e determinar algumas das características de nicho ecológico.

O quinto capítulo testa os efeitos da inclusão de informações sobre as ocorrências das espécies interagentes (fatores bióticos) no processo de modelagem. Foram analisados dois tipos de interação: o que ocorre entre abelhas que apresentam relação do tipo hospedeiro-parasita (*Bombus* spp) (Benton, 2006) e entre plantas hospedeiras e suas abelhas polinizadoras (*Krameria* x *Centris*) (Simpson, 1989). Foram determinados os impactos da inclusão desse tipo de informação na acurácia do modelo e também nas projeções de distribuição futura, considerando-se um dos cenários de mudanças climáticas elaborado pelo *Intergovernmental Panel on Climate Change* (IPCC).

Referências

- ANDERSON, R. P.; LEW, D.; PETERSON, A. T. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* v. 162, p. 211-232. 2003.
- BENTON, T. *Bumblebees: the natural history and identification of the species found in Britain*. United Kingdom, Collins. 2006.
- BIESMEIJER, J. C.; ROBERTS, S. P. M.; REEMER, M.; OHLEMULLER, R.; EDWARDS, M.; PEETERS, T.; SCHAFFERS, A. P.; POTTS, S. G.; KLEUKERS, R.; THOMAS, C. D.; SETTELE, J.; KUNI, W. E. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* v. 313, p. 351-354. 2006.
- BUCHMANN, S. L.; NABHAN, G. P. *The Forgotten Pollinators*. Washington, Island. 1996.
- CRAIG, T. P.; ITAMI, J. K.; HORNER, J. D. Geographic variation in the evolution and coevolution of a tritrophic interaction. *Evolution* v. 61, p. 1137–1152. 2007.
- DAILY, G. C. *Nature's Services: Societal Dependence on Natural Ecosystems*. Washington, Island. 1997.
- DARWIN, C. *On the Origin of Species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, John Murray. 1859.
- ESPÍNDOLA, A.; PELLISSIER, L.; ALVAREZ, N. Variation in the proportion of flower visitors of *Arum maculatum* along its distributional range in relation with community-based climatic niche analyses. *Oikos* v. 120, p. 728–734. 2011.
- FENSTER, C. B.; ARMBRUSTER, W. S.; WILSON, P.; DUDASH, M. R.; THOMSON, J. D. Pollination syndromes and floral specialization. *Annual Review of Ecology and Systematics* v. 35, p. 375–403. 2004.
- FRANKLIN, J. *Mapping species distribution. Spatial inference and prediction*. New York, Cambridge. 2009.

- GODSOE, W.; STRAND, E.; SMITH, C. I.; YODER, J. B.; ESQUE, T. C.; PELLMYR, O. Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytologist* v. 183, p. 589-99. 2009
- HORTAL, J.; ROURA PASCUAL, N.; SANDERS, N.; RAHBK, C. Understanding (insect) species distributions across spatial scales. *Ecography* v. 33, p. 51-53, 2010.
- HURD, P. D. JR.; LINSLEY, E. G.; WHITAKER, T. W. Squash and gourd bees (*Peponapis*, *Xenoglossa*) and the origin of the cultivated *Cucurbita*. *Evolution* v. 25, p. 218-234. 1971.
- HUTCHINSON, G. Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology* v. 22, p. 415-427. 1957.
- HUTCHINSON, G. E. *An introduction to population ecology*. New Haven, Yale University. 1978.
- KEARNS, C. A.; INOUE, D. W.; WASER, N. M. Endangered mutualisms: the conservation of plant-pollinator interactions. *Annual Review of Ecology and Systematics* v. 29, p. 83–112. 1998.
- KLEIN, A. M.; VASSIERE, B. E.; CANE, J. H.; STEFFAN-DEWENTER, I.; CUNNINGHAM, S. A.; KREMEN, C.; TSCHRNTKE, T. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London B* v. 274, p. 303-313. 2007.
- LAINE, A. L. Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *Journal of Experimental Botany* v. 60, p. 2957–2970. 2009.
- MICHENER, C. D. Biogeography of the bees. *Annals of Missouri Botanical Garden* v. 66, p.277-347. 1979.
- MICHENER, C. D. *The bees of the world*. 2^a edição. Baltimore, The Johns Hopkins University. 2007.
- MITCHELL, R. J.; IRWIN, R. E.; FLANAGAN, R. J.; KARRON, J. D. Ecology and evolution of plant-pollinator interactions. *Annals of Botany* v. 103, p. 1355-1363. 2009.

- PELLISSIER , L.; POTTIER , J.; VITTOZ , P.; DUBUIS, A.; GUISAN, A. Spatial pattern of floral morphology: possible insight into the effects of pollinators on plant distributions. *Oikos* v. 119, p. 1805–1813. 2010.
- PHILLIPS, S. Transferability, sample selection bias and background data in presence only modelling: a response to Peterson et al. (2007). *Ecography* v. 31, p. 272-278, 2008.
- POTTS, S. G.; BIESMEIJER, J. C.; KREMEN, C.; NEUMANN. P.; SCHWEIGER, O.; KUNIN, W. E. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution* v. 25, p. 345-353. 2010.
- PULLIAM, H. On the relationship between niche and distribution. *Ecology Letters*, v. 3, p. 349-361. 2000.
- ROUBIK, D. W. Loose niches in tropical communities: why are there so few bees and so many trees? In: HUNTER, M. D.; OHGUSHI, T.; PRICE, P. W. *Effects of resource distribution on animal-plant interactions*. New York, Academic, pp. 327-54. 1992.
- SIMPSON, B. B. Krameriaceae. *Flora Neotropica* v. 49, p. 1-109. 1989.
- SOBERÓN, J. Niche and area of distribution modeling: a population ecology perspective. *Ecography* v. 33, p. 159-167. 2010.
- STOCKWELL D. R. B.; PETERS, D. The GARP Modeling System: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* v. 13, p. 143-158. 1999.
- THOMPSON, J. N. *The coevolutionary process*. Chicago, University of Chicago. 1994.
- THOMPSON, J. N. *The geographic mosaic of coevolution*. Chicago, University of Chicago. 2005.
- THUILLER, W.; LAFOURCADE, B.; ENGLER, R.; ARAUJO, M. B. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography* v. 32, p. 369-373. 2009.
- TOJU, H.; SOTA, T. Imbalance of predator and prey armament: geographic clines in phenotypic interface and natural selection. *American Naturalist* v. 167, p. 105–117. 2006.

- WASER, N. M. Competition for pollination and floral character differences among sympatric plant species: a review of evidence. In: JONES, C. E.; LITTLE, R. J. *Handbook of pollination biology*. New York, Scientific and Academic Editions, pp. 277-293. 1983.
- WCISLO, W. T.; CANE, J. H. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. *Annual Review of Entomology* v. 41, p. 257-286. 1996.

CAPÍTULO 1

Modelagem preditiva de distribuição de espécies: usos, aplicações e desafios atuais

Tereza C. Giannini; Marinez F. Siqueira; André L. Acosta; Francisco C. C. Barreto; Antonio M. Saraiva; Isabel Alves-dos-Santos

***Rodriguesia* (submetido)**

Exemplo de modelagem de *Centris trigonoides* (Apidae)

Elaboração: T. C. Giannini

Modelagem preditiva de distribuição de espécies: usos, aplicações e desafios atuais

Tereza C. Giannini¹; Marinez F. Siqueira²; André L. Acosta¹; Francisco C. C. Barreto³;
Antonio M. Saraiva⁴; Isabel Alves-dos-Santos¹

1. Instituto de Biociências. Universidade de São Paulo. Rua do Matão trav. 14. 05508-900. São Paulo, SP.
2. Instituto de Pesquisa Jardim Botânico do Rio de Janeiro. Unidade de Botânica Sistemática. Rua Pacheco Leão, 915. 22460-030. Rio de Janeiro, RJ.
3. Centro de Ciências Humanas e Naturais. Universidade Federal do Espírito Santo. Av. Fernando Ferrari. 29075-910. Vitória, ES.
4. Escola Politécnica. Universidade de São Paulo. Av. Prof. Luciano Gualberto, 380 - 05508-970. São Paulo, SP.

Resumo

A modelagem preditiva tem sido aplicada para analisar a distribuição geográfica de espécies, a partir de extrapolações das características ambientais dos locais conhecidos de ocorrência da espécie. O interesse por esse tipo de modelagem deve-se à necessidade de repostas rápidas e fundamentadas para as ameaças que as espécies têm enfrentado, devido à perda de habitat, invasão de espécies exóticas, mudanças climáticas, entre outros. Este artigo oferece uma visão geral dos avanços recentes no campo da modelagem e visa incentivar a discussão e aplicação do método, que pode auxiliar tanto na aquisição de conhecimento básico sobre a biologia das espécies, quanto na análise e formulação de políticas para sua conservação.

Palavras chaves: nicho ecológico, áreas potenciais de ocorrência, algoritmos de modelagem.

Introdução

A perda e fragmentação de habitat e as mudanças climáticas são apenas dois exemplos de alterações ambientais causadas por fatores antropogênicos, com consequências

diretas sobre a distribuição das espécies. Essas ameaças crescentes demandam novas tecnologias e ferramentas de análise, para que se possa adquirir ou aprofundar o conhecimento existente sobre as espécies e auxiliar em sua proteção e conservação.

Em linhas gerais, a modelagem de distribuição de espécies consiste em um processamento computacional que combina dados de ocorrência de uma ou mais espécies com variáveis ambientais dos locais de ocorrência, construindo assim uma representação das condições requeridas pelas espécies (Anderson *et al.*, 2003). Alguns algoritmos têm sido aplicados para criar modelos que representam essas condições e que podem ser projetados sobre um mapa que exhibe as áreas potenciais de ocorrência dessas espécies.

A modelagem de distribuição tem sido amplamente utilizada com múltiplos objetivos, tais como: utilização de modelos de distribuição potencial em análises biogeográficas (Siqueira e Durigan, 2007); conservação de espécies raras ou ameaçadas (Araújo e Williams, 2000; Engler *et al.*, 2004); reintrodução de espécies (Hirzel *et al.*, 2002); perda de biodiversidade (Polasky e Solow, 2001); impactos de mudanças climáticas (Peterson *et al.*, 2002; Oberhauser e Peterson, 2003; Siqueira e Peterson, 2003; Thomas *et al.*, 2004; Pearson *et al.*, 2006; Araújo *et al.*, 2006; Araújo *et al.*, 2008; Wiens *et al.*, 2009); avaliação do potencial invasivo de espécie exóticas (Peterson *et al.* 2003; Peterson e Robins, 2003; Peterson *et al.*, 2006a); estudo das possíveis rotas de disseminação de doenças infecciosas (Costa *et al.*, 2002; Peterson *et al.*, 2006b; Levine *et al.*, 2007); auxílio na determinação de áreas prioritárias para conservação (Ortega-Huerta e Peterson, 2004; Chen, 2009), entre outros.

Este artigo apresenta os fundamentos ecológicos que estão na base da conceituação e aplicação da modelagem preditiva da distribuição de espécies. Na sequência apresenta também um breve histórico de seu desenvolvimento com algumas das principais contribuições ao avanço dessa técnica. Com o objetivo de promover a aplicação deste tipo de estudo, as etapas do processo de modelagem também são discutidas de forma sucinta, e finalmente, são apresentados os principais desafios ainda existentes neste campo.

Fundamentos ecológicos da modelagem

A modelagem preditiva tem sido chamada por alguns autores de modelagem de distribuição geográfica (Austin, 2002) e por outros, de modelagem de nicho ecológico (Austin, 1992; Soberón e Peterson, 2005). A ausência de consenso foi relacionada, entre outros fatores discutidos adiante, com a dificuldade de se definir, de forma consistente, o conceito de nicho (Elith e Leathwick, 2009). Esse conceito é um dos elementos centrais em ecologia, e costuma ser aplicado em diversas áreas, desde estudos de comportamento, morfologia e fisiologia de indivíduos até abordagens que avaliam o papel desempenhado por uma determinada espécie em um ecossistema (Chase e Leibold, 2003).

A primazia da aplicação do conceito de nicho é atribuída frequentemente a Joseph Grinnell que discutiu, em uma série de artigos (Grinnell 1904, 1917, 1924), o nicho de várias espécies de aves e mamíferos. De acordo com suas descrições, o nicho consistiria não exatamente em um atributo da espécie, mas sim do espaço (Schoener, 1989), e buscava definir as condições ou situações ambientais nas quais a espécie seria capaz de sobreviver e se reproduzir.

Posteriormente, Elton (1927) apresentou o nicho de uma espécie como seu papel dentro da cadeia alimentar, analisando tanto o impacto sofrido pela espécie, quanto o que ela causa em seu ambiente. Seu enfoque é primariamente voltado ao comportamento e às interações tróficas. De acordo com o autor, o nicho seria definido pelo tamanho corporal e pelos hábitos alimentares da espécie, enfatizando que o nicho corresponderia a “pequenas subdivisões” de categorias já estabelecidas, tais como: carnivorina, herbivorina ou insetivorina (Elton, 1927, p. 64).

A conceituação de nicho ecológico deu um grande salto a partir dos trabalhos de George Evelyn Hutchinson, que define o termo nicho como a soma de todos os fatores ambientais que atuam sobre o organismo: “uma região de um hiper-espaco n -dimensional¹” (Hutchinson, 1944), sendo que o termo “hiper-espaco” foi substituído posteriormente por “hiper-volume” (Hutchinson, 1957). O autor define um número qualquer (n) de fatores limitantes (por exemplo, luminosidade, temperatura e recursos

¹ No original: “The term niche is here defined as the sum of all the environmental factors acting on the organism; the niche thus defined is a region of an n -dimensional hyper-space.” (HUTCHINSON, 1944),

alimentares) à sobrevivência do organismo. A amplitude de tolerância a estes fatores, ou seja, os valores mínimos e máximos dos limites de tolerância para cada fator, constitui um hiper-volume com n dimensões. Portanto, o espaço abrangido por este hiper-volume representaria a faixa de variação dos fatores ambientais na qual a espécie é capaz de sobreviver e se reproduzir.

Além de reestruturar o conceito de nicho, Hutchinson (1957) também introduziu a diferenciação entre nicho fundamental e nicho efetivo (ou realizado). A intersecção das múltiplas variáveis ambientais, que definem as condições e recursos necessários à espécie, constituiria seu nicho fundamental. Porém, embora as condições ambientais adequadas à sobrevivência da espécie possam existir em muitas regiões, outras características biológicas atuam no sentido de restringir sua área de ocorrência. Por exemplo, a utilização dos recursos alimentares pode ser alterada devido a interações com outras espécies (tais como competição, predação, parasitismo) reduzindo o nicho fundamental a uma fração que pode ser efetivamente explorada, e a essa fração foi atribuído o nome de nicho efetivo (Hutchinson, 1957).

O mesmo autor sugeriu ainda que se faça uma clara distinção entre as variáveis de nicho enfocadas como condição e como recurso. Condição é definida por variáveis não-interativas e por requisitos ambientais em larga escala, que são relevantes para compreender as propriedades geográficas e ecológicas da espécie. Recurso aborda as interações bióticas e a dinâmica entre consumidores e o que é consumido (Hutchinson, 1978).

Recentemente, buscando uma abordagem quantitativa e mais rigorosa em termos estatísticos, Chase e Leibold (2003) propuseram uma definição de nicho que resgata a conceituação dos três principais autores citados acima (Grinnel, Elton e Hutchinson). O nicho corresponderia às “condições ambientais que permitem à espécie suprir suas necessidades mínimas, de tal forma que a taxa de natalidade de uma população local seja igual ou maior à sua taxa de mortalidade, junto com o conjunto de efeitos *per capita* da espécie nessas condições ambientais².” Os autores utilizaram a isóclina de taxa de crescimento populacional zero (baseando-se nos trabalhos de Tilman, 1982) para avaliar as respostas de uma espécie frente a um ou mais fatores, e examinar como

² No original: “the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions.” (Chase e Leibold, 2003, p. 15).

suas taxas de natalidade e mortalidade variam ao longo do gradiente do fator utilizado. Por exemplo, a taxa de crescimento populacional pode aumentar diante de maior disponibilidade de recursos, ou diminuir frente à presença de predadores (consistente com os conceitos de Grinnel e Hutchinson). Além das necessidades da espécie por um fator em particular, a espécie também pode causar um impacto sobre este fator (o que seria mais consistente com o conceito de Elton). Para quantificar esse efeito, os autores utilizam um “vetor de impacto”. Por exemplo, se o fator for um recurso trófico, sua concentração ou densidade poderá reduzir conforme a taxa de forrageamento da espécie que o utiliza, e assim, o vetor de impacto apontará para sua diminuição (Chase e Leibold, 2003).

Por outro lado, a distribuição geográfica de uma dada espécie foi considerada como uma expressão complexa de sua ecologia e história evolutiva, influenciada por vários fatores que operam em intensidades e escalas diferentes tais como, a história natural da espécie e sua capacidade de dispersão (Brown e Lomolino, 1996). Por exemplo, Pulliam (2000) discutiu a complexidade da dinâmica de dispersão diferenciando habitats do tipo fonte e dreno. Os habitats fonte são aqueles nos quais a reprodução local excede a mortalidade, e nos drenos ocorre o oposto. Por definição, os drenos não apresentam condições necessárias para a espécie sobreviver e se reproduzir, embora os indivíduos possam ocorrer nessas áreas devido aos processos de imigração a partir de suas fontes. Assim, é possível encontrar indivíduos em locais onde as condições não são inteiramente adequadas à sobrevivência e à reprodução da espécie. Por outro lado, as espécies podem apresentar limites de dispersão, e consequentemente, não chegar aos locais adequados. Portanto, a dinâmica fonte-dreno prevê que uma espécie possa ser encontrada fora dos limites de seu nicho fundamental, enquanto que, ao contrário, o limite na capacidade de dispersão prevê que certos locais, potencialmente adequados, possam não ser ocupados por determinadas espécies.

Considerando essa dinâmica, Soberón (2010) sugeriu três fatores principais que determinam a distribuição geográfica, sendo os dois primeiros uma função direta do nicho (Figura 1): 1) condições abióticas (por exemplo, o clima); 2) fatores bióticos (por exemplo, interações entre espécies); 3) capacidade de dispersão, através de sua locomoção ou através de agentes externos, o que determina o grau de acessibilidade da espécie a outras áreas (ver também Soberón e Peterson, 2005; Soberón, 2007).

Assim, Soberón (2010) sugere que a distribuição geográfica não é relacionada apenas com o nicho ecológico, mas também inclui outros fatores, os quais nem sempre podem ser incluídos na modelagem. Portanto, segundo o autor, “os modelos de nicho têm sido validados universalmente através de comparações entre ocorrências no espaço geográfico, embora eles sejam, estritamente falando, modelos de nicho ecológico e não de distribuição geográfica³” (Soberón e Peterson, 2005, p. 8). Já Phillips (2008) discute que a modelagem, em alguns casos, tenta estimar o quão satisfatório um local é para a espécie considerando-se suas condições ambientais. Ou seja, em termos estatísticos, as variáveis que estão sendo modeladas indicam se um local é satisfatório para a espécie e não exatamente se ele está sendo ocupado.

Figura 1. Fatores que determinam a distribuição geográfica de uma espécie. **G**: região geográfica considerada; **A**: área onde as condições abióticas são favoráveis para a espécie. Esta é uma região potencial e, provavelmente, não ocupada pela espécie em todas as suas seções; **B**: região onde as condições bióticas são favoráveis; **M**: área que a espécie foi capaz de alcançar em um dado período; **G_I**: região que pode ser potencialmente invadida porque os dois tipos de condições (A e B) são favoráveis, mas a espécie ainda não foi capaz de alcançá-la; **G_O**: área real de distribuição da espécie. Os círculos fechados representam a população fonte e os abertos, a população dreno. (Soberón, 2010, p. 161)

Jiménez-Valverde *et al.* (2008b) afirmam que o termo modelagem de nicho deveria ser evitado, pois além de ser complexo, implica em considerar os efeitos tanto dos fatores bióticos quanto dos abióticos que atuam sobre o organismo. Segundo os autores, os modelos são capazes de projetar simulações da distribuição das espécies no espaço

³ No original: “Niche models have universally been validated via comparisons with occurrences in geographic space, even though they are, strictly speaking, models of ecological niches and not geographic distributions” (Soberón e Peterson, 2005, p. 8).

geográfico, mas não oferecem uma descrição do nicho das espécies, pois, embora haja disponibilidade de dados relacionados com condições ambientais, os dados sobre interações ainda são escassos ou de difícil interpretação. Como resultado, a maioria dos exemplos na literatura tem utilizado apenas as variáveis abióticas em suas análises e, assim, a modelagem deveria ser considerada como modelagem de distribuição e não de nicho (Araújo e Guisan, 2006; Jiménez-Valverde *et al.*, 2008b; Phillips, 2008; Elith e Leathwick, 2009).

Soberón e Nakamura (2009, p. 19644) sintetizaram a questão da seguinte forma: “Essencialmente, essa é uma técnica usada para estimar áreas de distribuição reais ou potenciais, ou conjuntos de habitats favoráveis para uma dada espécie, com base em sua presença e (às vezes) ausência observadas. Estes métodos relacionam ‘nicho’ com ‘área de distribuição’. As aspas são usadas para indicar que definições rigorosas desses conceitos ainda não foram apresentadas⁴”.

Principais contribuições ao desenvolvimento da modelagem preditiva

Segundo Guisan e Thuiller (2005) as técnicas computacionais desenvolvidas para analisar o impacto das variáveis ambientais na distribuição de uma espécie surgiram, principalmente nas décadas de 1970 com os trabalhos de Austin (1971) e Nix *et al.* (1977). Nas décadas de 1980-1990 as simulações de Ferrier (1984) e os livros de Verner *et al.* (1986) e Margules e Austin (1991) foram importantes contribuições para promover essa nova abordagem, bem como as revisões de Franklin (1995) e Austin (1998).

Atualmente, alguns grupos têm se destacado na discussão e aplicação da modelagem, entre eles, o grupo da Universidade de Kansas (EUA) (por exemplo, Peterson, 2006; Peterson *et al.*, 2008; Soberón, 2007 e 2010); o da Universidade de Madri (Espanha) (Lobo, 2008; Lobo *et al.* 2008; Jimenez-Valverde *et al.* 2008a e 2008b); o NCEAS (*National Center for Ecology Analysis and Synthesis*) com o projeto *Predicting Species Distributions Working Group* (Graham *et al.*, 2004; Guisan *et al.*, 2007; Elith *et al.*, 2006; Wisz *et al.*, 2008); e o da Universidade de Melbourne (Austrália) (Elith e

⁴ No original: “Essentially, it is a technique used to estimate actual or potential areas of distribution, or sets of favorable habitats for a given species, on the basis of its observed presences and (sometimes) absences. These methods relate ‘niches’ to ‘areas of distribution.’ The quotes are used to indicate that rigorous definitions of those concepts have not as yet been presented.” (Soberón e Nakamura, 2009, p. 19644).

Leathwick, 2007 e 2009; Elith *et al.*, 2010; Elith *et al.*, 2011). No Brasil, o Laboratório de Ecologia Teórica e Ecologia Aplicada e Conservação da Universidade Federal de Goiás tem produzido vários artigos sobre a problemática associada ao tema e apresentado soluções metodológicas (De Marco *et al.*, 2008; Diniz-Filho *et al.*, 2009b, c, d.; Terribile *et al.*, 2010). Também é preciso salientar o papel da Escola Politécnica da Universidade de São Paulo, do CRIA (Centro de Referência em Informação Ambiental) e do INPE (Instituto Nacional de Pesquisas Espaciais) especialmente no que se refere ao desenvolvimento do programa openModeller (Santana *et al.*, 2008; Muñoz *et al.*, 2009), que disponibiliza vários algoritmos para modelagem preditiva.

O desenvolvimento da modelagem aconteceu concomitante com iniciativas internacionais que visavam, principalmente, a padronização, compartilhamento e disponibilização de dados primários de coleções biológicas, museus e herbários (Graham *et al.*, 2004). Atualmente, as principais iniciativas são a do BIS/TDWG (*Biodiversity Information Standards / Taxonomic Database Working Group* - desde 1985), voltado principalmente para o desenvolvimento de padrões para compartilhar dados; ITIS (*The Integrated Taxonomic Information System* - desde meados da década de 90), com a meta de criar um banco de dados sobre informações taxonômicas; GBIF (*Global Biodiversity Information Facility* - desde 1996), cujo objetivo primário é o de tornar disponíveis os dados sobre biodiversidade e; IABIN (*Inter-American Biodiversity Information Network* – desde 1996) com a finalidade de promover colaboração técnica sobre o armazenamento, uso e compartilhamento de informações sobre biodiversidade. No Brasil foi criada a rede speciesLink, uma iniciativa do CRIA, com o objetivo geral de disponibilizar dados de coleções biológicas brasileiras. Outras iniciativas importantes são o programa Biota FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo) e o Sis-Biota Brasil, ambos com o objetivo de inventariar e caracterizar a biodiversidade, respectivamente, desse estado e do país.

Além disso, é importante também destacar os avanços em métodos que permitiram maior robustez e detalhamento nos modelos geográficos (Elith e Leathwick, 2009). Houve um avanço muito grande na geração e disponibilização de variáveis preditoras (Turner *et al.*, 2003), tais como: 1) Variáveis climáticas: Worldclim (Hijmans *et al.*, 2005) e Köppen-Geiger *climate classification* (Rubel e Kottek, 2010); 2) Ecorregiões: WWF (*World Wildlife Fund* - Olson *et al.*, 2001); 3) Edáficas: USGS (*United States Geological Survey*) e 4) Modelos de elevação: USGS/EROS (*United States Geological*

Survey Center for Earth Resources Observation and Science). Destaca-se no Brasil, a iniciativa do IBGE (Instituto Brasileiro de Geografia e Estatística) em prover dados associados a diversas características ambientais.

Em duas revisões recentes sobre modelagem (Elith e Leathwick, 2009; Soberón e Nakamura, 2009), os autores sugeriram que entre os desafios ainda existentes para a modelagem está a inclusão de interações bióticas em seu processo. Assim, merecem destaque as iniciativas dedicadas à disponibilização de dados de interações entre organismos, especialmente plantas e polinizadores. Exemplos importantes são o desenvolvimento de extensões de padrões de troca de dados originalmente propostos pelo BIS/TDWG (Cartolano Junior *et al.*, 2007; Saraiva *et al.*, 2009b), a iniciativa da IABIN-PTN (*Inter-American Biodiversity Information Network - Pollinators Thematic Network*) (Saraiva *et al.*, 2009a) e a Rede de Polinizadores (Canhos *et al.*, 2010). Os dados e análises subsequentes, tornadas acessíveis a partir dessas iniciativas, poderão, num futuro próximo, auxiliar a pesquisa sobre interações, bem como, expandir a aplicação da modelagem em estudos de nicho e distribuição geográfica de espécies.

Ainda existem poucos exemplos na literatura de aplicações da modelagem para analisar organismos que apresentam algum tipo de interação. São encontrados exemplos deste tipo de aplicação em interações de competição (Leathwick e Austin, 2001; Anderson *et al.*, 2002), predação (Sánchez-Cordero e Martinez-Meyer, 2000) e entre polinizadores e plantas hospedeiras (Jiménez-Valverde *et al.*, 2008a; Godsoe *et al.*, 2009; Giannini *et al.*, 2010; Giannini *et al.*, 2011). Uma análise onde foram incluídos os dados de interação na modelagem apresentou como resultado a melhoria na acurácia dos modelos obtidos (Heikkinen *et al.*, 2007). Já Araujo e Luoto (2007) demonstraram a importância de se incluir dados de interação quando são analisados os possíveis efeitos das mudanças climáticas sobre a distribuição de espécies.

Considerando a importância das interações entre organismos e a influência das características geográficas sobre essas interações, conclui-se que este é um campo fértil e em aberto no desenvolvimento da modelagem.

Etapas do processo de modelagem

A modelagem de distribuição de espécies é um processo complexo que compreende várias etapas, as quais exigem o conhecimento de diversos conceitos e técnicas ligados

não só a biologia, mas também a biodiversidade, geografia e clima, e aos processamentos requeridos antes, durante e depois da obtenção do modelo. As seguintes etapas são sugeridas para executar a modelagem (baseado em Santana *et al.*, 2008).

1) *Tratamento dos pontos de ocorrência da(s) espécie(s)*

A primeira etapa consiste na obtenção, avaliação e preparo dos pontos de presença e ausência (quando disponíveis) das espécies a serem modeladas, o que pode incluir o georreferenciamento dos pontos e a exclusão de dados imprecisos ou duvidosos. Os dados de ausência são muito raros e difíceis de serem obtidos, e por conta disto, alguns algoritmos (Bioclim, Environmental Distance, Maxent) têm sido utilizados na modelagem por demandarem apenas dados de presença (Elith e Leathwick, 2009). Os dados podem ser adquiridos *in locus*, através de levantamentos em áreas alvo, fornecendo dados inéditos à análise e/ou através de fontes de dados já publicados em formato digital como o GBIF (<http://www.gbif.org>), speciesLink (<http://splink.cria.org.br>), IABIN (<http://www.iabin.net>) e ainda, por meio de literatura acadêmica especializada.

Existem também algumas propostas para a geração de dados de pseudo-ausência para uso com algoritmos que obrigatoriamente precisam desse tipo de dado. Duas dessas propostas podem ser encontradas através da aplicação de ENFA (*Ecological-Niche Factor Analysis*) (Hirzel *et al.*, 2002) ou das estratégias presentes no pacote BIOMOD (Thuiller, 2003). Tem sido sugerido que a definição dos pontos de pseudo-ausência deve seguir alguma estratégia e não ser feita aleatoriamente (Engler *et al.*, 2004, McPherson *et al.*, 2004). Em relação à localização dos pontos, VanDerWal *et al.* (2009) sugeriu que a performance do modelo é menor quando os pontos de pseudo-ausência são definidos em uma região muito restrita ou muito ampla em relação às presenças registradas, sugerindo que o melhor seria considerar regiões intermediárias para essa escolha. Porém, mais recentemente, Lobo e Tognelli (2011) demonstraram que a localização dos pontos de pseudo-ausência afeta menos o modelo que o número de pontos. De acordo com a análise feita por estes autores, os resultados sugeriram que o aumento de número de pontos de pseudo-ausência aumenta a acurácia dos modelos, sendo sugerida uma relação de 100 pontos de pseudo-ausência para cada ponto de presença. Além disso, um maior número de pontos acarretaria também em uma diminuição na taxa de sobreprevisão, uma tendência que o modelo pode apresentar de determinar certas áreas para as quais não existem ocorrências reportadas, como sendo

potencialmente adequadas para a espécie (ver detalhes acerca de sobreprevisão mais abaixo).

Os dados de coleções biológicas e museus podem apresentar limitações devido à imprecisão da informação ou viés na escolha de áreas de coletas que, às vezes, privilegiam locais onde o acesso é mais fácil e menos custoso, ou ainda, onde já se tem certeza da ocorrência da espécie em estudo (Hortal *et al.*, 2008; Lobo, 2008). Apesar disso, o uso desse tipo de dado é frequentemente justificado devido à falta de levantamentos sistemáticos, além da demanda crescente por mapas de distribuição de espécies e estimativas sobre mudanças potenciais nessa distribuição devido às alterações climáticas (Thuiller *et al.*, 2008; Elith e Leathwick, 2009).

Visando estimar um número mínimo de pontos de ocorrência para modelagem, Wisz *et al.* (2008) utilizaram três classes em suas análises, com 10, 30 e 100 pontos cada. Os autores concluíram que os algoritmos Maxent e GARP (versão disponibilizada pelo aplicativo openModeller; Muñoz *et al.*, 2009) são os melhores para analisar 10 pontos, mas modelos com maior acurácia foram obtidos com pelo menos 30 pontos, considerado pelos autores como um número mínimo. No entanto, outros autores sugeriram números menores (Stockwell e Peterson, 2002) e outros, discutiram a utilização de métodos específicos para análises com número muito baixo de ocorrências, como no caso de espécies raras (Siqueira *et al.*, 2009) ou crípticas (Pearson *et al.*, 2007).

2) Tratamento e escolha das camadas ambientais

Essa etapa visa identificar, adquirir e converter as camadas ambientais, que geralmente, são utilizadas como “*raster*”. Esse formato é comum aos programas de SIG (Sistema de Informação Geográfica) e consiste em imagens digitalizadas e exibidas como uma matriz de pontos. Em alguns casos, a conversão precisa ser feita através de um programa de SIG como o ArcGIS (ESRI Inc.) ou o DIVA GIS (LizardTech, Inc. and University of California), sendo este último, um programa de código aberto disponível na internet (outros programas gratuitos são o GRASS GIS e o Quantum GIS, ambos da Open Source Geospatial Foundation).

É importante também, definir adequadamente a resolução das camadas ambientais. Elementos importantes a serem considerados na escolha são: a precisão dos pontos de ocorrência, as características do terreno e da espécie (Elith e Leathwick, 2009). Por

outro lado, Guisan *et al.* (2007) demonstraram que a resolução das camadas utilizadas tem pouco efeito sobre os modelos, enfatizando a dificuldade que existe em estimar adequadamente esse fator. Mais recentemente, Soberón (2010) demonstrou que o uso de alta resolução justifica-se especialmente se informações bióticas (interações entre espécies, por exemplo) forem consideradas na modelagem – do contrário, caso sejam usados apenas dados abióticos, o uso de resolução mais baixa é o suficiente para produzir bons modelos. De maneira geral, é indicado evitar-se o uso de camadas de variáveis que sejam altamente correlacionadas, sendo recomendada a aplicação de técnicas específicas de seleção de variáveis. Alguns autores utilizam Análises de Componentes Principais (De Marco *et al.*, 2008) e Índices de Correlação ou então técnicas de mineração de dados. Quanto ao número dessas variáveis, é reconhecido entre os estatísticos que o tamanho da amostra em testes de regressão logística deve ser no mínimo, 10 vezes maior que o número de variáveis preditoras utilizado (Peduzzi *et al.*, 1996). Para os algoritmos de aprendizagem de máquina, o número ideal de variáveis não é tão bem definido; ainda faltam estudos sobre esse tema, mas é fato que se pode aplicar uma proporção bem menor.

3) Algoritmos de modelagem

Algoritmos são sequências finitas de comandos, executados através de códigos escritos em alguma linguagem computacional. Um dos algoritmos frequentemente utilizados para modelagem preditiva é o Maxent (Phillips *et al.*, 2006). Este algoritmo estima a distribuição da espécie buscando a distribuição de probabilidade da entropia máxima (isto é, próximo à uniformidade) que considera que o valor esperado de cada um dos conjuntos de características (no caso, variáveis ambientais) se aproxima da média. Vários algoritmos de modelagem podem ser encontrados em alguns sistemas computacionais, tais como o *openModeller* (Muñoz *et al.*, 2009), que conta com uma ampla variedade de algoritmos integrados, incluindo *Genetic Algorithm for Rule set Production* (GARP), o *Ecological Niche Factor Analysis* (ENFA), o *Support Vector Machine* (SVM), o Bioclim, e o *Environmental Distance* (Carpenter *et al.*, 1993). Outro sistema disponível é o BIOMOD (Thuiller, 2003; Thuiller *et al.*, 2009), acrônimo de *Biodiversity Modelling*, desenvolvido para a plataforma R (*The R Foundation for Statistical Computing*). Nesse pacote podem ser encontrados nove algoritmos, entre eles, o *Generalised Linear Models* (GLM), *Generalised Additive Models* (GAM),

Classification Tree Analysis (CTA), *Artificial Neural Networks* (ANN) e o *Random Forest* (RF).

Elith e Graham (2009) e Elith e Leathwick (2009) apresentaram revisões onde foram comparados alguns algoritmos, e discutidas as características, vantagens e desvantagens de cada um deles.

4) Avaliação dos modelos

4.1. Geração de um conjunto independente de dados

Para avaliar a qualidade do modelo é preciso gerar um conjunto independente de dados. Existem pelo menos duas formas de se fazer isso: 1) Coletando novos dados em trabalho de campo ou levantamento de literatura; 2) Dividindo seu conjunto de dados em conjuntos de treino e teste antes de realizar a modelagem.

Independente da estratégia escolhida é fundamental ter dois (ou mais) conjuntos de dados diferentes, uma vez que, um conjunto será utilizado para gerar o modelo (dados de treino) e outro irá testá-lo (dados de teste).

A forma mais comum de se dividir o conjunto de dados é através da aplicação de números aleatórios à tabela de registros de ocorrências, geralmente 50% dos dados para gerar o modelo e 50% para testar (Fielding e Bell, 1997; Hirzel e Guisan, 2002) ou 70% para treino e 30% para teste. Outra forma é dividir os dados espacialmente (geograficamente) e utilizar uma parte como treino e outra como teste (Peterson *et al.*, 2008).

Atualmente tem-se usado técnicas de *bootstrap* (Efron, 1979) para gerar diferentes conjuntos de dados e avaliar assim a precisão dos modelos de distribuição. Esta técnica envolve a partição dos dados aleatoriamente, com reposição, em vários conjuntos de treino e teste. Esta técnica de validação tem sido frequentemente utilizada, juntamente com *Cross-validation*, *Subsamples* e *Jackknife*. A *Cross-validation* consiste basicamente em dividir aleatoriamente o banco de dados em k partições, dessas, $k-1$ são utilizadas para treinar o algoritmo e a partição restante é utilizada para o teste. O processo é repetido k vezes. Independente do método, o erro obtido em cada uma das partições é calculado obtendo-se um valor médio da qualidade do modelo gerado. Basicamente, o que definirá a técnica a ser aplicada é o número de dados disponíveis. *Cross-validation* e *Subsamples* precisam de uma quantidade maior de dados, enquanto *bootstrap* e *Jackknife* podem ser aplicadas para poucos dados.

4.2. Erros de omissão (subprevisão) e sobreprevisão (ou comissão)

Para avaliar o resultado da modelagem, dois tipos de erros são mensurados, baseados na matriz de confusão (Figura 2). São eles:

4.2.a) Erro de omissão: esse tipo de erro é geralmente considerado como um erro verdadeiro, representado pela letra C na matriz de confusão (Figura 2). Contudo, sob algumas circunstâncias, um registro de presença pode não ser muito confiável quanto a sua identificação taxonômica e/ou georrefenciamento. Estes pontos podem representar um *outlier* para o algoritmo, ou seja, pontos com informação ambiental muito fora do padrão gerado pelos demais pontos do conjunto de treino. Nessas circunstâncias, um erro de omissão não seria um erro, e sim uma forma do algoritmo conferir menos importância para pontos considerados ‘ruins’. Uma consequência direta disso, caso se tenha dúvida sobre a qualidade dos registros de ocorrência da espécie, é que não é recomendado usar taxa de omissão de 0% ao se rodar um modelo. É preferível deixar uma margem de segurança (por exemplo, até 10%) para que o algoritmo possa trabalhar melhor essa questão. Porém, na literatura atual há uma predominância de trabalhos utilizando 5% de taxa de omissão (Elith *et al.*, 2006; Phillips *et al.*, 2006).

4.2.b) Erro de sobreprevisão (comissão): este pode ou não ser um erro verdadeiro. Consiste na previsão de ocorrência em uma área na qual não se tem registro da presença da espécie, e é representado pela letra B na matriz de confusão (Figura 2).

4.3. Validação dos modelos

A avaliação da qualidade dos modelos gerados é ainda um tema em evolução, pois a forma como as avaliações são feitas estatisticamente ainda é incompleta e não totalmente resolvida na literatura (Fielding e Bell, 1997; Peterson *et al.*, 2008). A validação de um modelo é também conhecida como teste externo e consiste da confrontação de seus resultados contra dados reais da distribuição da espécie. Isso distingue esse processo do chamado teste interno, também conhecido como verificação, no jargão da modelagem matemática. Discussões sobre a teoria para validação de modelos pode ser encontrada em Pearce e Ferrier (2000) e Lobo *et al.* (2008).

Essa é uma das etapas mais importantes do processo de modelagem. Sem a validação, a interpretação de um modelo perde seu sentido, visto que tudo o que está representado pode estar incorreto ou com graus inaceitáveis de imprecisão.

Os métodos mais comuns para avaliar a qualidade do modelo são baseados na matriz de confusão (Figura 2), mas ao comparar resultados de modelagens realizadas por diferentes algoritmos, é necessário utilizar também métodos que sejam independentes de limites de corte. Portanto, temos dois tipos de avaliação, a dependente e a independente de um limite de corte específico.

4.3.a) Avaliação dependente do limite de corte

As medidas tradicionais de capacidade discriminatória de um modelo dependem da adoção de um limite de decisão arbitrário (*threshold*) o que acaba inserindo mais um fator de subjetividade na interpretação do modelo. Uma boa revisão sobre limites de decisão pode ser encontrada em Liu *et al.* (2005).

A escolha desse limite de decisão é usualmente baseada no conhecimento da probabilidade da ocorrência da espécie de interesse (por exemplo, o conhecimento sobre a amplitude da distribuição esperada da espécie, ampla ou restrita) assim como, no julgamento das consequências em tomar decisões erradas no processo e, além disso, está também diretamente relacionada com o propósito da modelagem. Por exemplo, se uma espécie é ameaçada e o objetivo do modelo é identificar áreas potenciais para maximizar o sucesso de sua reintrodução na natureza, é preciso identificar áreas adequadas que minimizem a chance de erro, ou seja, modelos menos inclusivos (com baixa sobreprevisão). Isso requer um limite de decisão alto, que selecione apenas as áreas com altos valores de adequabilidade ambiental para a ocorrência da espécie. Já quando o interesse é avaliar o potencial invasivo de espécies exóticas, ou aumentar o conhecimento sobre o nicho de uma espécie pouco coletada, pode-se adotar um limite de corte mais baixo (maiores áreas de sobreprevisão) para aumentar a área de interesse em relação à ocorrência das espécies. O limiar atualmente mais utilizado na literatura é o MPT (*Minimum Presence Training* ou ainda *Lowest Presence Training*), mas cada valor de corte tem vantagens e desvantagens de uso. A escolha deve sempre estar alinhada com o que se pretende fazer com o modelo gerado, ou seja, depende do propósito da modelagem.

4.3.b) Matriz de confusão

Todo modelo apresenta seus erros e acertos, que são avaliados em conjunto para determinar a qualidade do mesmo. A Figura 2 apresenta a matriz de confusão, ou

seja, o esquema que reúne as possíveis formas de acerto e erro em relação às previsões do modelo e à distribuição “real” da espécie na natureza.

Para se calcular os valores da matriz de confusão é necessário aplicar um limite de corte para avaliar o modelo. Depois de estabelecido o limite de corte, testes estatísticos podem ser aplicados para avaliar a qualidade do modelo gerado. Estes testes avaliam a probabilidade do acerto dos pontos de teste (taxa de omissão) ser diferente do acaso em relação à área prevista do modelo gerado. Um modelo pode “acertar” todos os pontos de teste, mas em uma área tão grande que o resultado do teste pode não ser significativo. A recíproca é verdadeira, ou seja, o modelo pode não acertar todos os pontos, mas a área do modelo é tão pequena (em relação à área de estudo) que a chance de acertar aqueles pontos é significativamente diferente do acaso. Um teste binomial uni-caudal pode ser usado para se determinar se um modelo prediz os pontos de teste significativamente melhor do que o acaso (Anderson *et al.*, 2002). Esta medida não pode ser usada diretamente para comparar modelos gerados por diferentes algoritmos porque diferentes algoritmos (principalmente algoritmos com saídas diferentes) podem gerar diferentes áreas previstas como presente, ficando difícil uma comparação entre eles.

	presença real	ausência real
presença prevista	a	b
ausência prevista	c	d

a = verdadeiro positivo
b = falso positivo
c = falso negativo
d = verdadeiro negativo
a e d são acertos
b e c são erros
taxa de falso positivo = erro de sobreprevisão = $b/(b+d)$
taxa de falso negativo = erro de comissão = $c/(a+c)$
sensibilidade = $a/(a+c)$
especificidade = $d/(b+d)$
acurácia = $a+d/(a+b+c+d)$

Figura 2. Matriz de confusão, onde a e d são previsões corretas; b é erro de sobreprevisão (falsos positivos) e c é erro de omissão (falsos negativos).

Visando solucionar esse problema surgiu a ideia de usar as curvas características de operação (ROC), que são amplamente utilizadas na área de controle de qualidade, em processos industriais, na área de saúde, na padronização de doses-resposta, etc. O cálculo da área sob a curva (AUC) fornece uma medida única do desempenho do modelo, independente da escolha prévia de um limiar de corte específico, pois a curva é construída a partir de vários limiares (cada limite de corte é responsável por

um ponto da curva) e o cálculo da área sob a curva elimina então, a escolha de um limite específico. Este valor (AUC) mede a capacidade discriminatória do modelo, permitindo interpretar seu resultado como a probabilidade de que, ao se sortear dois pontos - um do conjunto de presença e outro do conjunto de ausência - o modelo consiga prever os dois corretamente. Uma das melhores fontes de explicação sobre as curvas características de operação pode ser encontrada no relatório de Fawcett (2003).

A curva ROC é obtida plotando-se a sensibilidade no eixo y e o valor 1-especificidade no eixo x para todos os possíveis limites de corte. A sensibilidade também é conhecida como a taxa de verdadeiros positivos, e representa ausência de erro de omissão. Já a especificidade, ou a variante 1-especificidade, também é conhecida como a taxa de falso positivo, e representa o erro de sobreprevisão. A área abaixo da curva (AUC) é normalmente determinada conectando os pontos com linhas diretas e o valor da área é calculado pelo método de trapezóide. Esta área sob a curva (AUC) tem a seguinte interpretação: quanto mais próximo do valor um, ou seja, quanto mais distante o resultado do modelo for da previsão aleatória, melhor o desempenho do modelo (Rushton *et al.* 2004, Phillips *et al.* 2006).

Esta análise caracteriza-se por avaliar o desempenho do modelo através de todos os possíveis limites de corte, gerando um único valor, que representa a área sob a curva (AUC), que pode então ser usado para comparações entre diferentes algoritmos.

Com apenas dados de presença o uso do ROC parece inaplicável, pois sem dados de ausência, não há como calcular a especificidade. Este problema é contornado considerando um problema distinto de classificação. Ao invés de tentar avaliar o modelo por sua capacidade de discernir presenças e ausências, sua capacidade de distinguir presença é estimada contra o acaso. Isto é feito sorteando-se, para cada pixel x na área de estudo, um ponto x-aleatório e para cada pixel x dentro da área de distribuição geográfica da espécie é definido um ponto x-presente. O modelo de distribuição potencial de espécies fará suas previsões para cada pixel correspondente a esses exemplos, mas sem ter informação sobre se este é uma presença ou um ponto escolhido ao acaso da área de estudo (*background points*).

Logo, é possível obter-se previsões tanto para amostras positivas (x-presentes) quanto negativas (x-aleatório, pixels selecionados uniformemente ao acaso do

background da área de modelagem). Juntas, essas informações são suficientes para definir uma curva ROC.

O uso do ROC com dados de presença e ausência difere do cálculo feito apenas com dados de presença em relação ao valor máximo do AUC, que é nesse caso, menor do que 1 (Wiley *et al.*, 2003). Se a distribuição da espécie cobre uma taxa a de pixels, então o máximo que a AUC pode ter é exatamente $1 - a/2$. Como não sabemos o valor de a , não podemos dizer qual o valor ótimo que a AUC pode chegar sob essas circunstâncias (Wiley *et al.*, 2003).

Uma área igual a 1 representaria o “modelo perfeito”; uma área de 0,5 indica que o modelo seleciona ao acaso. De um ponto de vista prático, um teste de validação pode adotar os valores de AUC a seguir como indicadores da qualidade do modelo (Metz, 1986): excelente (0,90 - 1,0); bom (0,80 - 0,90); médio (0,70 - 0,80); ruim (0,60 - 0,70); muito ruim (0,50 - 0,60).

4.3.c) Problemas com o uso da AUC

Atualmente há uma grande discussão na literatura quanto à aplicação da AUC nessas circunstâncias (Lobo *et al.*, 2008). Neste trabalho, os autores apontaram cinco principais problemas associados ao uso da AUC. Destes, os mais importantes são o fato da análise ROC dar o mesmo peso aos erros de omissão e sobreprevisão (Peterson *et al.*, 2008), e o fato do tamanho da área de estudo em relação à distribuição conhecida da espécie influenciar grandemente na taxa de ausência corretamente predita pelo modelo. Em relação a esta última situação, é possível diminuir a área de estudo para o mais próximo possível da distribuição conhecida (ou esperada) da espécie, visando minimizar este problema. Apesar disso, ainda são utilizadas medidas de AUC, principalmente na comparação entre algoritmos, como também estimativas em testes como o da proporção binomial baseado na matriz de confusão associada a algum limiar de corte.

Outros testes estatísticos têm sido propostos como meios alternativos de avaliar o desempenho de modelos, tais como Kappa (*Cohen's Kappa Statistic*) e o TSS (*True Skill Statistics*) (Thuiller *et al.*, 2009) porém, até o presente momento, o AUC mostra-se o teste mais amplamente utilizado pelos pesquisadores.

A validação de modelos ainda é uma área de pesquisa ativa e novas técnicas podem surgir tão rápido quanto outras podem se tornar obsoletas. Portanto, é necessário

manter-se continuamente informado sobre o assunto, acompanhando o ritmo das publicações mais recentes sobre este tópico.

4.3.d) Validação de Campo

Aplicar o modelo de volta no campo é o teste mais robusto que se pode fazer com os resultados de modelagem, visando avaliar a capacidade do modelo em acertar a distribuição da espécie em áreas de alta (como presença) e baixa (como ausência) adequabilidade ambiental. No entanto, alguns cuidados devem ser tomados quanto à validação de campo. O ideal é planejar o levantamento em diferentes níveis de adequabilidade ambiental para se obter informações mais completas sobre o tamanho da área de ocorrência da espécie.

Outra forma de validação é a avaliação do especialista na espécie modelada. É o que se chama de avaliação biológica e é feita através de interpretação visual (Phillips *et al.*, 2006) pelo especialista na espécie modelada. Esta avaliação é muito importante, pois é bastante comum atingir um resultado estatisticamente muito bom, mas que biologicamente não o é.

É importante ressaltar que os modelos são ferramentas intermediárias em um estudo biogeográfico ou ecológico. O uso de informações adequadas e a interpretação correta dos resultados irão determinar a confiabilidade de um estudo baseado em modelos para estimar as características do relacionamento entre espécies e seu meio. Por isso o apoio de um especialista no grupo focado, tanto para eventualmente corrigir a base de dados de ocorrência quanto para auxiliar na interpretação dos dados, é fundamental.

5) Procedimentos pós-modelagem

Com o objetivo de aumentar a confiabilidade dos modelos gerados, ou visando diminuir a incerteza de se utilizar modelos gerados por diferentes algoritmos, técnicas de projeção combinada (*ensemble forecasting*) podem ser aplicadas aos resultados da modelagem oriunda de diferentes algoritmos e/ou de diferentes cenários (diferentes modelos e cenários climáticos, por exemplo) (ver revisão em Araújo e New, 2007). Basicamente, esta técnica consiste em gerar um modelo de consenso onde áreas de alta adequabilidade serão as áreas que todos (ou a maioria) dos modelos indicaram como sendo propícias para a ocorrência da espécie. Diferentes estratégias podem ser aplicadas, como por exemplo, análises de componentes principais (PCA) enfocando

diferentes algoritmos ou camadas ambientais (Thuiller, 2004; Araújo *et al.*, 2006), médias ponderadas dos resultados obtidos através dos valores de acurácia (AUC ou TSS ou Kappa) (Thuiller *et al.*, 2009) ou estratégias múltiplas com níveis crescentes de complexidade, desde combinação dos mapas resultantes até análises multidimensionais (Diniz-Filho *et al.*, 2009a, 2010a, 2010b).

Além disso, na literatura existem algumas sugestões de como melhorar o resultado da modelagem retirando-se do modelo as áreas na qual a espécie não ocorre por fatores históricos (barreiras geográficas) (Anderson *et al.*, 2003). Pode-se ainda retirar áreas não habitadas devido a interações bióticas (por exemplo, competição) (Anderson *et al.*, 2002). O modelo ainda pode ser ajustado utilizando-se dados de uso da terra para excluir áreas desflorestadas ou alteradas, áreas urbanas, ou ainda áreas de cultivo (Anderson e Martinez-Meyer, 2004).

Desafios

Vários desafios se apresentam ainda à técnica de modelagem de distribuição de espécies. Entre eles são destacados os seguintes (Elith e Leathwick, 2009; Santana e Saraiva, 2010):

- Além da AUC, ainda não existe um método efetivo para avaliar a acurácia dos modelos preditivos. Apesar de existirem vários algoritmos disponíveis, nem sempre é possível comparar o resultado de cada um deles e estimar as diferenças de alguma forma. Assim, outras ferramentas de pós-análise deveriam ser desenvolvidas de tal forma a proporcionar métodos mais acurados para qualificar os modelos obtidos.
- O uso apenas de dados de presença continuará, uma vez que dados de ausência são difíceis de obter. Assim, métodos mais robustos que avaliem os vieses na amostragem e os resultados obtidos precisam ser aplicados.
- São também necessários o desenvolvimento e evolução de novas ferramentas de pré-análise, especialmente que auxiliem no processo de escolha das camadas ambientais a serem utilizadas no processo. Nem sempre o conhecimento *a priori* das variáveis que influenciam a distribuição de um grupo está disponível. Além disso, o desenvolvimento e disponibilização de dados abióticos e bióticos em

formato digital e em diferentes escalas são também ações fundamentais para a melhoria da qualidade dos modelos obtidos.

- A modelagem de distribuição é diretamente dependente da qualidade dos dados de ocorrência das espécies. Apesar do envolvimento de várias iniciativas internacionais ainda existe pouca disponibilidade de dados, pois muito ainda não está digitalizado ou disponibilizado em portais da internet. Além disso, muitos dados disponibilizados ainda apresentam erros seja de identificação ou de digitalização e necessitariam ser corrigidos por especialistas nos grupos enfocados. Neste contexto, seria muito importante o desenvolvimento e a implementação de filtros de qualidade de dados, tanto taxonômicos como de georeferenciamento, para podermos ter acesso a dados com maior garantia de qualidade.
- A modelagem de distribuição preditiva precisa ser utilizada como um meio e não um fim em si. O propósito deveria ser mais focado na aquisição de mais informação sobre a ocorrência da espécie para que outras técnicas (mais robustas) possam ser também aplicadas. Especificamente em aplicações de cunho conservacionista, esta técnica passa a ser um importante mecanismo de auxílio especialmente na ausência de uma quantidade maior de dados ou na urgência de tomada de decisão.
- Finalmente, o uso de modelagem preditiva, com este enfoque, deveria ser mais divulgado para estimular os pesquisadores, especialmente taxonomistas e ecólogos, que conhecem de forma mais profunda a relação espécie e ambiente, a ampliarem seus conhecimentos. Na medida em que mais pesquisadores se debruçarem sobre o tema, com o uso correto da técnica, seu desenvolvimento, agora de forma mais focada, poderá ser aprimorado.

Agradecimentos

Os autores agradecem ao CNPq (proc. 385212/2009-7; 575069/2008-2), à FAPESP (proc. 04/15801-0) e à CAPES (proc. 3030-10-5) pelos apoios concedidos a esta proposta.

Referências bibliográficas

- Anderson, R. P.; Peterson, A. T.; Gómez-Laverde, M. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* 98: 3–16.
- Anderson, R. P.; Lew, D.; Peterson, A. T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162: 211–232.
- Anderson, R. P.; Martinez-Meyer, E. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation* 116: 167–179.
- Araújo, M. B.; Williams, P. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96: 331–345.
- Araújo, M. B.; Guisan, A. 2006. Five (or so) challenges for species distribution modeling. *Journal of Biogeography* 33: 1677–1688.
- Araújo, M. B.; Thuiller, W.; Pearson, R. G. 2006. Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* 33: 1712–1728.
- Araújo, M. B.; Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16: 743–753.
- Araújo, M. B.; New, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42–47.
- Araújo, M. B.; Nogues-Bravo, D.; Reginster, I.; Rounsevell, M.; Whittaker, R. J. 2008. Exposure of European biodiversity to changes in human-induced pressures. *Environmental Science & Policy* 11: 38–45.
- Austin, M. P. 1971. Role of regression analysis in plant ecology. *The Proceedings of the Ecological Society of Australia* 6: 63–75.
- Austin, M. P. 1992. Modelling the environmental niche of plants: implications for plant community response to elevated CO₂ levels. *Australian Journal of Botany* 40: 615–30.
- Austin, M. P. 1998. An ecological perspective on biodiversity investigations: example from Australian eucalypt forests. *Annals of the Missouri Botanical Garden* 85: 2–17.
- Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101–18.
- Brown, J. H.; Lomolino, M. V. 2006. *Biogeografia*. 2nd edição. Ribeirão Preto: Funpec.

- Canhos, D. A. L.; Canhos, V. P.; Marino, A. 2010. Polinizadores: desafios e oportunidades no desenvolvimento de sistemas de informação on-line. *Oecologia Australis* 14: 307-316.
- Carpenter, G.; Gillison, A. N.; Winter, J. 1993. DOMAIN: A flexible modeling procedure for mapping potential distributions of animals and plants. *Biodiversity and Conservation* 2: 667-680.
- Cartolano JR., E. A.; Saraiva, A. M.; Correa, P. L. P.; Giannini, T. C.; Giovanni, R. 2007. Uma proposta de esquema de dados de relacionamento entre espécimes. In: XXXIII Conferencia Latinoamericana de Informática (CLEI 2007) e Taller Latinoamericano de Informática para la Biodiversidad (INBI 2007), San Jose.
- Chase J. M.; Leibold, M. A. 2003. *Ecological niches: linking classical and contemporary approaches*. Chicago: The University of Chicago Press.
- Chen, Y. H. Conservation biogeography of the snake family Colubridae of China. *North-Western Journal of Zoology* 5: 251-262.
- Costa, J.; Peterson, A. T.; Beard, C. B. 2002. Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis neiva*, 1911, the most important Chagas' disease vector in northeastern Brazil (Hemiptera, Reduviidae, Triatominae). *American Journal of Tropical Medicine And Hygiene* 67: 516-520.
- De Marco, P.; Diniz-Filho, J. A. F.; Bini, L. M. 2008. Spatial analysis improves species distribution modelling during range expansion. *Biology Letters* 4: 577-580.
- Diniz-Filho, J. A. F.; Bini, L. M.; Oliveira, G.; Barreto, B. S.; Silva, M. M. F. P.; Terribile, L. C.; Rangel, T. F. L. V. B.; Pinto, M. P.; Sousa, N. P. R.; Vieira, L. C. G.; Melo, A. S.; Marco JR, P.; Vieira, C. M.; Balmires, D.; Bastos, R. P.; Carvalho, P.; Ferreira, L. G.; Telles, M. P. C.; Rodrigues, F. M.; Silva, D. M.; Silva JR, N. J.; Soares, T. N. 2009a. Macroecologia, biogeografia e áreas prioritárias para conservação no cerrado. *Oecologia Brasiliensis* 13: 470-497.
- Diniz-Filho, J. A. F.; Nabout, J. C.; Bini, L. M.; Soares, T. N.; Campos Telles, M. P.; Marco, P.; Collevatti, R. G. 2009b. Niche modelling and landscape genetics of *Caryocar brasiliense* (Pequi tree: Caryocaraceae) in Brazilian Cerrado: an integrative approach for evaluating central peripheral population patterns. *Tree Genetics & Genomes* 5: 617-627.
- Diniz-Filho, J. A. F.; Nabout, J. C.; Telles, M. P. C.; Soares, T. N.; Rangel, T. F. L. V. B. 2009c. A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and Molecular Biology* 32: 203-211.
- Diniz-Filho, J. A. F.; Oliveira, G.; Bini, L. M.; Loyola, R. D.; Nabout, J. C.; Rangel, T. F. L. V. B. 2009d. Conservation biogeography and climate change in brazilian cerrado. *Natureza & Conservação* 7: 8-18.

- Diniz-Filho, J.; Ferro, V.; Santos, T.; Nabout, J.; Dobrovolski, R.; De Marco JR, P. 2010a. The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatix* (Lepidoptera, Arctiidae). *Revista Brasileira de Entomologia* 54: 339-349.
- Diniz-Filho, J.; Nabout, J.; Bini, L.; Loyola, R.; Rangel, T.; Nogues-Bravo, D.; Araújo, M. 2010b. Ensemble forecasting shifts in climatically suitable areas for *Tropidacris cristata* (Orthoptera: Acridoidea: Romaleidae). *Insect Conservation and Diversity*, 3: 213-221,
- Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7: 1-26.
- Elith, J.; Graham, C. H.; Anderson, R. P.; Dudik, M.; Ferrier, S.; Guisan, A.; Hijmans, R. J.; Huettmann, F.; Leathwick, J. R.; Lehmann, A.; Li, J.; Lohmann, L. G.; Loiselle, B. A.; Manion, G.; Moritz, C.; Nakamura, M.; Nakazawa, Y.; Overton, J. M.; Peterson, A. T.; Phillips, S. J.; Richardson, K. S.; Scachetti-Pereira, R.; Schapire, R. E.; Soberon, J.; Williams, S.; Wisz, M. S.; Zimmermann, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.
- Elith, J.; Leathwick, J. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13: 265-275.
- Elith, J.; Graham, 2009. C. H. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66-77.
- Elith, J.; Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics* 40: 677-97.
- Elith, J.; Kearney, M. ; Phillips, S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution*. doi: 10.1111/j.2041-210X.2010.00036.x.
- Elith, J.; Phillips, S. J.; Hastie, T.; Dudik, M.; Chee, Y.; Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17 : 43-57.
- Elton, C. 1927. *Animal Ecology*. London: Sidwig and Jackson.
- Engler, R.; Guisan, A.; Rechsteiner, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41: 263-274.
- Fawcett, T. 2003. ROC graphs: Notes and practical considerations for data mining researchers. HP Labs Tech Report HPL-2003-4.

- Ferrier, S. 1984. *The status of the Rufous Scrub-Bird Atrichornis rufescens: habitat, geographical variation and abundance*. Armidale, PhD Thesis - University of New England.
- Fielding, A. H.; Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474–499.
- Giannini, T. C.; Saraiva, A. M.; Alves dos Santos, I. 2010. Ecological niche modeling and geographical distribution of pollinator and plants: a case study of *Peponapis fervens* (Smith, 1879) (Eucerini: Apidae) and *Cucurbita* species (Cucurbitaceae). *Ecological Informatics* 5: 59–66.
- Giannini, T. C.; Lira-Saade, R.; Ayala, R.; Saraiva, A. M.; Alves dos Santos, I. 2011. Ecological niche similarities of *Peponapis* bees and non-domesticated *Cucurbita* species. *Ecological Modelling* doi:10.1016/j.ecolmodel.2011.03.031.
- Godsoe, W.; Strand, E.; Smith, C. I.; Yoder, J. B.; Esque, T. C.; Pellmyr, O. 2009. Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytologist* 183: 589–599.
- Graham, C. H.; Ferrier, S.; Huettman, F.; Moritz, C.; Peterson, A. T. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19: 497–503.
- Grinnell, J. 1904. The origin and distribution of the chestnut-backed chickadee. *Auk* 21: 375–377.
- Grinnell, J. 1917. The niche-relationship of the California thrasher. *Auk* 34: 427–433.
- Grinnell, J. 1924. Geography and evolution. *Ecology* 5: 225–229.
- Guisan, A.; Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993–1009.
- Guisan, A.; Graham, C. H.; Elith, J.; Huettmann, F.; The NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13: 332–340.
- Heikkinen, R. K.; Luoto, M.; Virkkala, R.; Pearson, R. G.; Körber, J. H. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* 16: 754–763.

- Hijmans, J. R.; 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.
- Hirzel, A.; Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157: 331–341.
- Hirzel, A. H.; Hausser, J.; Chessel, D.; Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83: 2027–2036.
- Hortal, J.; Jimenez-Valverde, A.; Gomez, J.F.; Lobo, J.M.; Baselga, A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117: 847–858.
- Hutchinson, G. E. 1944. Limnological studies in Connecticut. Part 7. A critical examination of the supposed relationship between phytoplankton periodicity and chemical changes in lake waters. *Ecology* 35: 3–26.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology* 22: 415–427.
- Hutchinson, G. E. 1978. *An introduction to population ecology*. New Haven: Yale University Press.
- Jiménez-Valverde, A.; Gómez, J. F.; Lobo, J. M.; Baselga, A.; Hortal, J. 2008a. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. *Annales Zoologici Fennici* 45: 200–210.
- Jiménez-Valverde, A.; Lobo, J. M.; Hortal, J. 2008b. Not as good as they seem: the importance of concepts in species distribution modeling. *Diversity and Distributions* 14: 885–890.
- Leathwick, J. R.; Austin, M. P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82: 2560–2573.
- Levine, R. S.; Peterson, A. T.; Yorita, K. L.; Carroll, D.; Damon, I. K.; Reynolds, M. G. 2007. Ecological niche and geographic distribution of human monkeypox in Africa. *Plos One* 2: e176.
- Liu, C.; Berry, P. M.; Dawson, T. P.; Pearson, R. G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385–393.
- Lobo, J. M. 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation* 17: 873–881.
- Lobo, J. M.; Jiménez-Valverde, A.; Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151.

- Lobo, J. M.; Tognelli, M. F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation* 19: 1–7.
- Margules, C. R.; Austin, M. P. 1991. *Nature Conservation: Cost Effective Biological Survey and Data Analysis*. Canberra: CSIRO.
- Mcpherson, J. M.; Jetz, W.; Rogers, D. J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41: 811-823.
- Metz, C. E. 1986. ROC methodology in radiologic imaging. *Investigational Radiology* 21:720-733.
- Muñoz, M. E. S.; Giovanni, R.; Siqueira, M. F.; Sutton, T.; Brewer, P.; Pereira, R. S.; Canhos, D. A. L.; Canhos, V. P. 2009. openModeller: a generic approach to species' potential distribution modeling. *Geoinformatica*. DOI 10.1007/s10707-009-0090-7.
- Nix, H.; McMahon, J.; Mackenzie, D. 1977. Potential areas of production and the future of pigeon pea and other grain legumes in Australia. In: Wallis, E. S.; Whiteman, P. C. (eds). *The potential for pigeon pea in Australia. Proceedings of Pigeon Pea (Cajanus cajan (L.) Millsp.)*. Queensland: Field Day University of Queensland, p. 1–12.
- Oberhauser, K.; Peterson, A. T. 2003. Modeling current and future potential wintering distributions of eastern North American monarch butterflies. *Proceedings of the National Academy of Sciences* 100: 14063-14068.
- Olson, D. M.; Dinerstein, E.; Wikramanayake, E. D.; Burgess, N. D.; Powell, G. V. N.; Underwood, E. C.; D'amico, J.A.; Itoua, I.; Strand, H. E.; Morrison, J. C.; Loucks, C. J.; Allnutt, T. F.; Ricketts, T. H.; Kura, Y.; Lamoreux, J. F.; Wettengel, W. W.; Hedao, P.; Kassem, K. R. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* 51: 933-938.
- Ortega-Huerta, M. A.; Peterson, A. T. 2004. Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Diversity and Distributions* 10: 39–54.
- Pearce, J. L.; Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* 128: 127–147.
- Pearson, R. G.; Thuiller, W.; Araújo, M. B.; Martinez-Meyer, E.; Brotons, L.; McClean, C.; Miles, L.; Segurado, P.; Dawson, T. C.; Lees, D. C. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33: 1704-1711.

- Pearson, R. G.; Raxworthy, C. J.; Nakamura, M.; Peterson, A. T. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102-117.
- Peduzzi, P.; Concato, J.; Kemper, R.; Holford, T. R.; Feinstein, A. R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49: 1373-1379.
- Peterson, A. T.; Ortega-Huerta, M. A.; Bartley, J.; Sánchez-Cordero, V. Soberón, J.; Buddemeier, R. H.; Stockwell, D. R. B. 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416: 626-629.
- Peterson, A. T.; Papes, M.; Kluza, D. A. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science* 51: 863-868.
- Peterson, A. T.; Robins, C. R. 2003. Using ecological-niche modeling to predict Barred Owl invasions with implications for Spotted Owl conservation. *Conservation Biology* 17: 1161-1165.
- Peterson, A. T. 2006. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3: 59-72.
- Peterson, A. T.; Papes, M.; Reynolds, M. G.; Perry, N. D.; Hanson, B.; Regnery, R. L.; Hutson, C. L.; Muizniek, B.; Damon, I. K.; Carroll, D. S. 2006a. Native-range ecology and invasive potential of *Cricetomys* in North America. *Journal of Mammalogy* 87: 427-432.
- Peterson, A. T.; Lash, R. R.; Carroll, D. S.; Johnson, K. M. 2006b. Geographic potential for outbreaks of Marburg hemorrhagic fever. *American Journal of Tropical Medicine and Hygiene* 75: 9-15.
- Peterson, A. T.; Papes, M.; Soberón, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213: 63-72.
- Peterson, A. T.; Barve, N.; Bini, L. M.; Diniz Filho, J. A. F.; Jimenez-Valverde, A.; Lira-Noriega, A.; Lobo, J.; Maher, S.; Marco JR., P.; Martinez-Meyer, E.; Nakazawa, Y.; Soberón, J. 2009. The climate envelope may not be empty. *PNAS* 106: E47.
- Phillips, S. J.; Anderson, R. P.; Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
- Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson *et al.* (2007). *Ecography* 31: 272-278.

- Polasky, S.; Solow, A. R. 2001. The value of information in reserve site selection. *Biodiversity and Conservation* 10: 1051–1058.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3: 349-361.
- Rubel, F.; Kottek, M. 2010. Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift* 19: 135-141.
- Rushton, S. P.; Ormerod, S. J.; Kerby, G. 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology* 41: 193-200.
- Sánchez-Cordero, V.; Martínez-Meyer, E. 2000. Museum specimen data predict crop damage by tropical rodents. *Proceedings of the National Academy of Sciences* 97: 7074-7077.
- Santana, F. S.; Siqueira, M. F.; Saraiva, A. M.; Correa, P. L. P. 2008. A reference business process for ecological niche modeling. *Ecological Informatics* 3: 75-86.
- Santana, F. S.; Saraiva, A. M. 2010. Challenges in ecological niche modeling. In: XVIIIth World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR). Symposium. Québec, p. 1-10.
- Saraiva, A. M.; Ruggiero, M.; Adams, L. D.; Sellers, E.; Correa, P. L. P. 2009a. Pollinators information network for the Americas – IABIN-PTN. In: e-biosphere Conference 09. Conference Abstracts. Londres, p.79-80.
- Saraiva, A. M.; Cartolano Junior, E. A.; Giovanni, R.; Giannini, T. C.; Correa, P. L. P. 2009b. Exchanging specimen interaction data using Darwin Core. In: Weitzman, A. L. (ed.). *Proceedings of TDWG*, Montpellier.
- Schoener, T. W. 1989. The ecological niche. In: Cherrett, J. M. (Ed.) *Ecological concepts*. Oxford: Blackwell.
- Siqueira, M. F.; Peterson, A. T. 2003. Consequences of global climate change for geographic distributions of cerrado tree species. *Biota Neotropica* 3: 1-14.
- Siqueira, M. F.; Durigan, G. 2007. Modelagem da distribuição geográfica de espécies lenhosas de cerrado no Estado de São Paulo. *Revista Brasileira de Botânica* 30: 239-249.
- Siqueira, M. F.; Durigan, G.; De Marco Jr., P.; Peterson, A. T. 2009. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *Journal for Nature Conservation* 17: 25-32.
- Soberón, J.; Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics* 2: 1-10.

- Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115–1123.
- Soberón, J.; Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. *PNAS* 106: 19644–19650.
- Soberón, J. 2010. Niche and area of distribution modeling: a population ecology perspective. *Ecography* 33: 159–167.
- Stockwell, D. R. B.; Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1–13.
- Terribile, L. C.; Diniz-Filho J. A. F.; Marco JR., P. 2010. How many studies are necessary to compare niche-based models for geographic distributions? Inductive reasoning may fail at the end. *Brazilian Journal of Biology* 70: 263–269.
- Thomas, C. D.; Cameron, A.; Green, R. E.; Bakkenes, M.; Beaumont, L. J.; Collingham, Y. C.; Erasmus, B. F. N.; de Siqueira, M. F.; Grainger, A.; Hannah, L.; Hughes, L.; Huntley, B.; van Jaarsveld, A. S.; Midgley, G. F.; Miles, L.; Ortega-Huerta, M. A.; Peterson, A. T.; Phillips, O. L.; Williams, S. E. 2004. Extinction risk from climate change. *Nature* 427: 145–148.
- Thuiller, W. 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9: 1353–1362.
- Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology* 10: 2020–2027.
- Thuiller, W.; Albert, C.; Araújo, M. B.; Berry, P. M.; Cabeza, M.; Guisan, A.; Hickler, T.; Midgley, G. F.; Paterson, J.; Schurr, F. M.; Sykes, M. T.; Zimmermann, N. E. 2008. Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics* 9: 137–152.
- Thuiller, W.; Lafourcade, L.; Engler, R.; Araujo, M. B. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373.
- Tilman, D. 1982. *Resource competition and community structure*. Princeton: Princeton University Press.
- Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E.; Steininger, M. 2003. Remote sensing for biodiversity and conservation. *Trends in Ecology & Evolution* 18: 306–314.
- Vanderwal, J.; Shoo, L. P.; Graham, C.; Williams, S. E. 2009. Selecting pseudoabsence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling* 220: 589–594.

- Verner, J.; Morrison, M. L.; Ralph, C. J. 1986. *Wildlife 2000: modelling habitat relationships of terrestrial vertebrates*. Madison: University of Wisconsin Press.
- Wiens, J. A.; Stralberg, D.; Jongsomjit, D.; Howell, C. A.; Snyder, M. A. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19729-19736.
- Wiley, E. O.; Mcnyset, K. M.; Peterson, A. T.; Robins, C. R.; Stewart, A. M. 2003. Niche modeling and geographic range predictions in the marine environment using a machine learning algorithm. *Oceanography* 16: 120–127.
- Wisz, M. S.; Hijmans, R. J.; Li, J.; Peterson, A. T.; Graham, C. H.; Guisan, A.; NCEAS. 2008. Predicting Species Distributions Working Group. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763-773.

CAPÍTULO 2

Ecological niche modeling and geographical distribution of pollinator and plants: a case study of *Peponapis fervens* (Smith, 1879) (Eucerini: Apidae) and *Cucurbita* species (Cucurbitaceae)

Tereza C. Giannini, Antonio M. Saraiva, Isabel Alves-dos-Santos

***Ecological Informatics* (2010) v. 5, p. 59–66.**

Peponapis fervens
Foto: T. C. Giannini

Ecological niche modeling and geographical distribution of pollinator and plants: a case study of *Peponapis fervens* (Smith, 1879) (Eucerini: Apidae) and *Cucurbita* species (Cucurbitaceae)

Tereza C. Giannini¹, Antonio M. Saraiva², Isabel Alves-dos-Santos¹

1. Instituto de Biociências da Universidade de São Paulo. Rua do Matão, Trav. 14, nº 321, Butantã, São Paulo, SP, Brasil, 05508-900

2. Escola Politécnica da Universidade de São Paulo. Av. Prof. Luciano Gualberto, nº 158, Butantã, São Paulo, SP, Brasil, 05508-900

Abstract

The bees of the *Peponapis* genus (Eucerini, Apidae) have a Neotropical distribution with the center of species diversity located in Mexico and are specialized in *Cucurbita* plants, which have many species of economic importance, such as squashes and pumpkins. *Peponapis fervens* is the only species of the genus known from southern South America. The *Cucurbita* species occurring in the same area as *P. fervens* include four domesticated species (*C. ficifolia*, *C. maxima maxima*, *C. moschata* and *C. pepo*) and one non-domesticated species (*C. maxima andreana*). It was suggested that *C. m. andreana* was the original pollen source to *P. fervens*, and this bee expanded its geographical range due to the domestication of *Cucurbita*. The potential geographical areas of these species were determined and compared using ecological niche modeling that was performed with the computational system openModeller and GARP with best subsets algorithm. The climatic variables obtained through modeling were compared using Cluster Analysis. Results show that the potential areas of domesticated species practically spread all over South America. The potential area of *P. fervens* includes the areas of *C. m. andreana* but reaches a larger area, where the domesticated species of *Cucurbita* also occur. The Cluster Analysis shows a high climatic similarity between *P. fervens* and *C. m. andreana*. Nevertheless, *P. fervens* presents the ability to occupy areas with wider ranges of climatic variables and to exploit resources provided by domesticated species.

Keywords: interacting species; crops; squash bee; openModeller; potential distribution.

Introduction

As previously pointed out by Kearns et al. (1998), pollination interactions are keystones in human-managed as well as natural ecosystems. The “pollinator crisis”, i.e. the reduction of the pollination service performed by animal pollinator agents, is widespread all over the world and is related to human activities, especially habitat loss (Buchmann and Nabhan, 1996). Also, the recent decline of honeybee populations (Stokstad 2007) points to a more important role of wild bees as providers of ecological services (Losey and Vaughan 2006, Winfree et al. 2007). Specialized relationships between plant and pollinator are of special concern because of the vulnerability of this narrow interaction (Biesmeijer et al. 2006, Girao et al. 2007, Memmott et al. 2007). Thus, it is necessary to increase knowledge related to the organisms involved in these interactions.

The bees of *Peponapis* occur only in the Americas and are specialized in *Cucurbita* species, being considered their effective pollinators (Fronk and Slater 1956, Michener and Lange 1958, Linsley 1960, Hurd and Linsley 1966, 1967, Hurd et al. 1971, 1974, Willis and Kevan 1995, Meléndez-Ramírez et al. 2002, Cane et al. 2004, Cane 2005, Michener 2007). They are solitary, but live in aggregation in the ground, usually in clay soils, near *Cucurbita* fields (Rozen and Ayala 1987). The males sleep in the closed flowers (Michener and Lange 1958, Hurd and Linsley 1964).

While honeybees, *Apis mellifera*, have been used in the pollination of *Cucurbita* (Delaplane and Mayer 2000), bees of the genus *Peponapis* present some behavioral and morphological adaptations to *Cucurbita* flowers that improve their efficiency (Hurd et al. 1971, Tepedino 1981, Ordway et al. 1987, Canto-Aguilar and Parra-Tabla 2000, Shuler et al. 2005).

Hurd et al. (1971, 1974) suggested that *Peponapis* followed the different centers of domestication of *Cucurbita* at different times throughout history, and that the extensive cultivation of *Cucurbita* in America may have contributed to expand its geographical distribution. Evidences to support this hypothesis were also reported recently by Bischoff et al. (2009). They suggested that the geographic range of the American native populations of *Peponapis pruinosa* naturally expanded following the expansion of the distribution of the wild buffalo gourd, *Cucurbita foetidissima*.

Three species of *Peponapis* are known from South America: *P. citrulina* Cockerell, which has a distribution encompassing Central America and northern South America; *P. melonis* (Friese), which is endemic to Ecuador; and *P. fervens*, which occurs predominantly in southern South America, including Argentina, Bolivia, Brazil, Paraguay and Uruguay. All of them are considered as occurring in mesophytic areas (Hurd and Linsley 1970).

The *Cucurbita* species occurring in the same area as *P. fervens* include four domesticated species (*C. ficifolia* Bouche, *C. m. maxima* Duchesne, *C. moschata* (Duch. ex Lam.) Duch. ex Poir and *C. pepo* L.). It has been suggested that these domesticated species are native from South America excepting *C. pepo* that is native from North America (Decker 1988, Heiser 1990, Sanjur et al. 2002, Teppner 2004, Smith 2006). At the same area occurs a non-domesticated species (*C. m. andreana* Naudin) that also is native from South America (Nee 1990). All of these *Cucurbita* species were considered by Hurd et al. (1971) to have a potential relationship with *P. fervens* except *C. ficifolia*, which they reported as occurring at higher altitudes. Hurd and Linsley (1967) suggested that *C. m. andreana* may be the original pollen source of *P. fervens*, and that this bee probably spread to its current geographical range along the pollen avenues established by the domestication of *Cucurbita*.

One of the methods that can be applied to obtain more information about the geographic distribution of these species is ecological niche modeling that considers the niche concept. The ecological niche of a species has been described as a multidimensional space where the distinct dimensions represent distinct variables, such as interactions with other species (biotic factors) and environmental conditions (abiotic factors) (Hutchinson 1957), which reflect the species' geographical distribution (Brown and Lomolino 2006). Ecological niche modeling is a computational technique that combines environmental variables with known species occurrences and produces a model of the species ecological niche that identifies the potentially appropriate regions for the occurrence of a species, creating a map of its potential distribution (Stockwell and Peterson 2002, Anderson et al. 2003, Soberon and Peterson 2005, Soberon 2007). It requires a database with occurrence points, layers with different environmental variables, and the use of Geographical Information System tools to analyze the maps (Stockwell and Peters 1999).

This approach allows different types of analyses, for example: to estimate the impact of climatic changes in biodiversity (Peterson et al. 2002, Siqueira and Peterson 2003); to discuss the conservation of endangered species (Gaubert et al. 2006); to predict occupation patterns of invasive species (Peterson and Vieglais 2001, Hinojosa-Diaz et al. 2005); and to analyze interactions between the geographical distribution of species and vegetation patterns (Phillips et al. 2006) or between species (Leathwick and Austin 2001, Anderson et al. 2002, Jimenez-Valverde et al. 2008).

This paper presents a case study of the geographical distribution of the specialized wild bee *Peponapis fervens* and its host plants (*Cucurbita* spp) using ecological niche modeling. The main purpose was to compare the patterns of ecological niche and geographical distribution of these species and the correspondences between them.

Material and Methods

The main sources of occurrence points consulted to obtain data on the species *Peponapis fervens* and *Cucurbita* were GBIF (Global Biodiversity Information Facility - <http://www.gbif.net/>), speciesLink (<http://splink.cria.org.br/>), Discover Life (<http://www.discoverlife.org/>), and CONABIO (Comisión Nacional para el Conocimiento y uso de la Biodiversidad - <http://www.conabio.gob.mx/>).

In addition, the literature and the Coleção Entomológica Padre Jesus Santiago Moure at the Universidade Federal do Parana were consulted for data on *P. fervens*. Dr. M. A. J. F. Ferreira of the Brazilian Agricultural Research Corporation (Empresa Brasileira de Pesquisa Agropecuária - Embrapa) was consulted about occurrence points of domesticated *Cucurbita* in Brazil and Dr. Michael Nee of the New York Botanical Garden about occurrence points of *Cucurbita* in Bolivia.

Missing geographical coordinates were obtained at the speciesLink website (<http://splink.cria.org.br/>) which lists the coordinates used by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística - IBGE), and the Taseo website (<http://www.taseo.com/>) was used for the other countries. The coordinates were converted to decimal degrees and checked on the speciesLink website using the following tools: conversor (converting coordinate systems), datacleaning (checking coordinates and detecting errors) and speciesmapper (visualization of coordinates on a map). The data sources consulted provided 170 occurrence points for

the species *P. fervens* and 851 for *Cucurbita* (Table 1). The species with the lowest number of occurrence points was *C. m. andreana*, which presented 20 points, followed by *C. pepo* with 34.

Table 1. Number of occurrence points and countries reported in the datasets analyzed.

Scientific Name	Number of occurrence points	Countries where the species were reported
<i>Peponapis fervens</i>	170	Argentina, Bolivia, Brazil, Paraguay, Uruguay
<i>Cucurbita ficifolia</i>	47	Argentina, Bolivia, Brazil, Colombia, Peru
<i>Cucurbita maxima andreana</i>	20	Argentina, Bolivia, Uruguay
<i>Cucurbita maxima maxima</i>	290	Argentina, Bolivia, Brazil, Chile, Paraguay, Peru
<i>Cucurbita moschata</i>	460	Argentina, Bolivia, Brazil, Colombia, Paraguay, Peru
<i>Cucurbita pepo</i>	34	Argentina, Bolivia, Brazil, Colombia, Peru

Modeling was performed with the computational system openModeller (version 1.0.7) described by Santana et al. (2008). OpenModeller has many algorithms available, including GARP (Genetic Algorithm for Rule-set Production), which was chosen due to its wide application to modeling of ecological niches and geographical distribution of species (Stockwell and Peters 1999, Stockwell et al. 2006). GARP produces models based on non-random associations between the environmental characteristics of occurrence points of a certain species versus the overall region of study. The main objective is to search for similar environmental conditions, resulting in a model that indicates the potential areas of occurrence.

GARP is based on definitions of limits related to environmental variables through selection, evaluation, testing, incorporation or rejection of rules in order to produce a rule-set that characterizes the species' ecological requirements (Stockwell and Peters 1999, Anderson et al. 2003). Four rules are used simultaneously (Stockwell et al. 2006): an atomic rule that uses a single value of climatic parameter in the precondition of the rule; the Bioclim and range rules that are used in the Bioclim program (Nix 1986) and produce an “envelope” that encloses the range of the environmental values of the data points – the potential distribution of species considers those points that fall within the environmental envelope. Finally, GARP considers a fourth family of rules, the logit rules, which are an adaptation of logistic regression models to rules.

Two types of errors are considered: the omission and commission errors that are calculated by the values of confusion matrix (Elith et al. 2006, Philips et al. 2006). The

omission error has been related to the presence of occurrence points that are located outside the potential area (Anderson et al. 2003). This error is usually related to mistakes in identifying species or geographic coordinates. The commission error refers to areas of absence that are incorrectly predicted as potential areas, and one of the factors that could cause this error is the presence of ecological barriers (Anderson et al. 2003). Fielding and Bell (1997) suggested that this error could occur due to insufficient ecologically-relevant processes specified in the model. Anderson et al. 2003 also suggested the commission error can be related to inadequate sampling and, in this case, it is not a 'true' error.

The specific implementation of the algorithm used was GARP with best subsets – desktop GARP implementation that uses the two types of errors to select a subset of best models. The commission error threshold was 50%, a default value used for GARP with best subsets in the openModeller system. Using this algorithm, twenty models are produced, and a composite prediction is created based on all of them that can reveal a more reasonable estimation of the species' potential distribution (Anderson et al. 2003).

The model accuracy and the area under the curve (AUC) of receiver operating characteristic (ROC) are also calculated by the values of confusion matrix. Two tests were performed: an internal test using all points of occurrence and an external test (or independent test). The data of the external test were divided in sets before running GARP, randomly and without reposition, into test and train data (Table 2). The limit proposed by Stockwell and Peterson (2002) was used, in which the train data must have at least 20 occurrence points. The proportion between the number of train points and test points was 70% and 30% of the total points, respectively. The average and standard deviation of AUC values obtained for each data set of each species were calculated (Araujo et al. 2005). Swets (1988) suggested that the results of AUC could be interpreted as excellent when the AUC was higher than 0.90. However, the results of AUC have been frequently discussed in the literature, and there is not yet a consensus among the authors (McPherson et al. 2004, Austin 2007, Lobo et al. 2008, Peterson et al. 2008). The external test was performed only with those species that presented a sufficient number of occurrence points (5 out of 6 species).

The environmental layers (Bioclim layers with resolution of 5 arc-min obtained from Worldclim <http://www.worldclim.org>) considered the following variables: 1. Annual Mean Temperature; 2. Mean Diurnal Range; 3. Isothermality; 4. Temperature

Seasonality; 5. Max Temperature of Warmest Period; 6. Min Temperature of Coldest Period; 7. Temperature Annual Range; 8. Mean Temperature of Wettest Quarter; 9. Mean Temperature of Driest Quarter; 10. Mean Temperature of Warmest Quarter; 11. Mean Temperature of Coldest Quarter; 12. Annual Precipitation; 13. Precipitation of Wettest Period; 14. Precipitation of Driest Period; 15. Precipitation Seasonality; 16. Precipitation of Wettest Quarter; 17. Precipitation of Driest Quarter; 18. Precipitation of Warmest Quarter; 19. Precipitation of Coldest Quarter.

The openModeller system can extract the values of environmental variables used for each occurrence point. These values were used to compare the climatic variables of species in order to analyze their ecological niche. A Cluster Analysis was performed using the averages of all environmental variables to estimate the similarity between pairs of species (Sneath and Sokal 1973). The unweighted pair-group average algorithm (UPGMA) was applied, where clusters are joined based on the average distance between all members in the two groups and the Euclidean distance measure was applied to convert distance to similarity. Dendrogram distortion was evaluated by the cophenetic correlation index that is obtained correlating the original similarity matrix with the matrix obtained from the dendrogram. A value higher than 0.75 is considered a good representation of the original data (McGarigal et al. 2000). Cluster analysis was performed with the software Past Program (<http://folk.uio.no/ohammer/past/>).

Species occurrence frequency was also analyzed in relation to the averages of some of the climatic variables used: temperatures of the warmest and coldest quarter and precipitation of the wettest and driest quarter of the year.

Results

The maps obtained through modeling (Figure 1) show that the potential occurrence area of *P. ferverns* extended north beyond the latitude of 20°S, including some areas of northeastern and central western Brazil, northern Argentina, and Bolivia (Figure 1A). It completely overlaps with the areas of *C. m. andreana* (Figure 1C) and presents a broader geographical distribution, encompassing the areas of domesticated species of *Cucurbita* (Figure 1B, D, E, F).

Figure 1. Potential areas of *Peponapis fervens* and *Cucurbita* species obtained through ecological niche modeling.

Regarding the statistics for each model (Table 2), the highest omission error was obtained for *C. pepo* (8.8%) and the highest values of commission errors were obtained for *C. moschata* (19.0%) and *C. m. maxima* (20.1%). These species also presented the highest percentage of cells predicted present (18.4% and 18.2%). All the AUC values obtained from the modeling with all occurrence points (internal test) exceeded 0.90, and the highest AUC was obtained for *C. m. andreana* (AUC=1.0). The external test shows that all models presented an average AUC higher than 0.90. This test was not performed with *C. m. andreana* due to insufficient data.

Table 2. Statistics results of internal and external test performed with ecological niche modeling (see Material and Methods).

Scientific name	Internal Test						External test					
	AUC	Accuracy (%)	Omission error (%)	Commission error (%)	% cells that predict presence	Number of sets	Test (30%)			Train (70%)		
							Number of points	AUC average	Standard deviation	Number of points	AUC average	Standard deviation
<i>P. fervens</i>	0.96	100	0	12.2	11.2	4	10	0.98	0.02	24	0.99	0.02
<i>C. ficifolia</i>	0.95	95.8	4.2	14.5	13.9	1	10	1.0	-	24	0.96	-
<i>C. maxima andreana</i>	1.0	100	0	1.2	1.4	-	-	-	-	-	-	-
<i>C. maxima maxima</i>	0.92	99.0	1.0	20.1	18.4	8	10	1.0	0	24	0.97	0.03
<i>C. moschata</i>	0.92	99.1	0.9	19.0	18.2	13	10	0.98	0.03	24	0.98	0.02
<i>C. pepo</i>	0.95	91.2	8.8	16.4	13.1	1	10	1.0	-	24	0.98	-

The histograms created with the frequency of individuals in different environmental conditions show that *P. fervens* has a close association with *C. m. andreana* (Figure 2). Both are present mainly between 22-28°C in the warmest quarter (Figure 2A) and between 10-22°C in the coldest quarter (Figure 2B). Also, they occur between 320-640mm of precipitation in the wettest quarter (Figure 2C). However, *P. fervens* also occurs with a lower frequency at temperature values between 14-20°C in the warmest quarter (Figure 2A), precipitation values above 720mm in the wettest quarter (Figure 2C) and 350mm in the driest quarter (Figure 2D).

Figure 2. Histograms of environmental variables for a three-month period (a quarter) for *Peponapis fervens*, *Cucurbita maxima andreana* and domesticated *Cucurbita*: A. Mean

temperature of the warmest quarter; B. Mean temperature of the coldest quarter; C. Precipitation of the wettest quarter; D. Precipitation of the driest quarter.

The average, maximum and minimum temperature and precipitation for each species analyzed are shown in Figure 3. *Peponapis fervens* and *C. m. andreana* presented similar values of mean temperature, but *P. fervens* can reach lower values (Figure 3A). Both also presented similar mean precipitation, but *P. fervens* also occurs at higher values (Figure 3B). The domesticated species of *Cucurbita* exhibit the highest variation between maximum and minimum values of temperature and precipitation. *C. m. maxima* and *C. moschata* presented the highest mean values of temperature (Figure 3A) and precipitation (Figure 3B) and also showed the largest difference in precipitations between the wettest and the driest quarters (averages approximately between 650 and 65mm). *C. ficifolia* presented the lowest values of temperature (Figure 3A). *C. ficifolia* and *C. pepo* presented the lowest values of precipitation within domesticated species (Figure 3B).

Figure 3. Averages, maximums and minimums of environmental variables for the three-month period (quarter) for all species analyzed: A. Temperature of the warmest and the coldest quarter; B. Precipitation of the wettest and the driest quarter.

The Cluster Analysis using Euclidean distance shows two clusters (Figure 4): one with *P. fervens* and *C. m. andreana*, and the other including the domesticated species. *C. m. maxima* and *C. moschata* present high climatic similarity between them and both with *C. ficifolia*. *C. pepo*, the only species introduced in South America, showed the lowest similarity within this cluster. The cophenetic correlation index showed a good fit for the cluster obtained.

Figure 4. Cluster Analysis dendrogram performed with paired group algorithm and Euclidean distance using the averages of environmental variables analyzed of *Peponapis fervens* and associated *Cucurbita* species.

Discussion

The potential areas obtained through modeling for *P. fervens* (Figure 1A) overlapped with the areas of *C. m. andreana* (Figure 1C). Both share common areas, but *P. fervens* presents a broader geographical distribution, including some occurrence and potential areas of domesticated species of *Cucurbita*.

The occurrence points obtained for *P. fervens* (Figure 1A) showed a distribution that extended north beyond the threshold of approximately 20°S originally suggested by Hurd and Linsley (1967). These points were obtained in the literature (Bouseman 1975, Gomes 1991, Batalha Filho et al 2007). The modeling resulted in potential areas that extended to approximately 15°S (Figure 1A), including some areas of northeastern and central western Brazil characterized by seasonally dry climates, under the biomes of subtropical savannas and xeric shrublands. This species had not been reported in the central west, but was collected recently in the state of Mato Grosso do Sul by A. Aguiar (personal communication, 2009). Still in this region, there is an extensive crop of

Cucurbita (indicated in Figure 1D and E), and additional surveys of *P. fervens* in these areas would provide important results.

Peponapis fervens and *C. m. andreana* were the species with the lowest percentage of commission error (Table 2). Nevertheless, the map of potential areas obtained for *C. m. andreana* clearly showed some overprediction. Maybe its points of occurrence are underestimated due to the lack of surveys or the difficulties of accessing this data, but on the other hand, this species has never been reported in southwestern Brazil, which was shown as a potential area in the map (Figure 1C). Only some hybrids of *C. m. maxima* and *C. m. andreana* have been found in low altitudes in Brazil (Cutler and Whitaker 1961).

Michael Nee (personal communication, 2009) found *C. m. andreana* in Bolivia only in the areas marked on the map as occurrence points (Figure 1C). But the potential areas of this species resulting from modeling emphasize mainly the northern Argentina, Paraguay, and Uruguay as potential areas, more than Bolivia itself.

C. m. andreana has been pointed out as the ancestor of *C. m. maxima*, and it was supposed to have originated in Argentina (Heiser 1979, Nee 1990). But recently Sanjurjo et al. (2002) suggested “humid lowland regions of Bolivia in addition to warmer temperate zones in South America from where *C. m. andreana* was originally described should possibly be considered as an area of origin for *C. m. maxima*.” Cutler and Whitaker (1961) also reported hybrids of *C. m. maxima* with *C. m. andreana* in lowlands in both Argentina and Bolivia.

Only two studies have mentioned the occurrence of *P. fervens* in Bolivia: Hurd and Linsley (1967), who reported three specimens from an unspecified location and Bouseman (1975) who reported two localities that were added to the modeling (Figure 1A). A survey in this country involving these species would be very interesting to provide new insights regarding their distributional range.

The domesticated species of *Cucurbita* have been cultivated for a long time under different conditions and are adapted to a broad variety of climatic parameters indicating their viability as important crops. Nowadays they present a widespread distribution and the potential areas shown in the maps could probably be used as areas of domestication. For example, although *C. m. maxima* (Figure 1D) and *C. moschata* (Figure 1E) showed

the highest values of commission error (Table 2), Ferreira et al. (2006) verified that both species occur throughout Brazil and are under-represented in current surveys.

With respect to *C. moschata*, its area of high landrace diversity has been identified as northern South America, particularly Colombia (Whitaker 1956, Cutler and Whitaker 1968), where they show basal characteristics (Wessel-Beaver 2000, Sanjur et al. 2002). Saade and Hernandez (1992) reported that the geographical distribution of the known archaeological remains of *C. moschata* indicates that it has been cultivated for more than 5000 to 6000 years, and its dissemination to other countries was very early. Dillehay et al. (2007) suggested an even earlier origin (9 to 10 thousand years ago), based on analysis of remains in an archaeological site on the western slopes of the northern Peruvian Andes. In both areas (Colombia and Peru), another species of *Peponapis* occurs: *P. citrulina* Cockerell.

Cucurbita ficifolia also shows a high commission error, and its distribution encompasses many potential areas of occurrence (Figure 1B). This species was thought to have originated in the Andes, maybe in Bolivia or Peru (Nee 1990, Saade and Hernandez 1992). Cutler and Whitaker (1961) suggested that this is the oldest cultivated species and clearly stands apart from the other cultivated species, maintaining some similarities with the wild ones. This finding agrees with Sanjur et al. (2002), who placed *C. ficifolia* as an outlier to other domesticated species of *Cucurbita* and considered it as being basal to the mesophytic species in the genus.

Hurd et al. (1971) considered that *C. ficifolia* was not related to *P. fervens* due to its distribution in higher altitudes. But although this species is usually mentioned in the literature as being cultivated at higher altitudes (Hurd et al. 1971, Woodson et. al 1978, Nee 1990, Sanjur et. al 2002), Saade et al. (1995) emphasized that it could be found in lowlands, also suggesting their domestication in widespread areas. The results concerning the climatic parameter used for modeling showed that this species occurs in areas with the lowest values of mean temperature. *C. ficifolia* might be a potential host plant to *P. fervens*, which also occur at lower values of temperature and in associated regions.

Cucurbita pepo presented the highest omission error (8.8% - Table 2), probably due to the low number of occurrence points, which could have affected the results. It also presented the lowest percentage of cells that predict the presence within the domesticated species (13.1% - Table 2), indicating that it is the domesticated species

with the smallest potential area of occurrence. Nevertheless, in Brazil it is considered one of the most important cultivated species, along with *C. m. maxima* and *C. moschata* (Ferreira et al. 2006), and it is also reported in all Brazilian territory (Ferreira 2007). The model (Figure 1F) shows no occurrence points in northern Brazil, but this is probably due to the lack of data from this region (Ferreira 2007).

Conclusion

Ecological niche modeling was applied to analyze and compare known and potential geographic distribution patterns of interacting species. This case study shows that *P. fervens* is closely associated with *C. maxima andreana*. The potential areas of *P. fervens* overlap with those of *C. maxima andreana*, but *P. fervens* shows a larger area of occurrence. This species occurs at broader climatic conditions, in which the domesticated cucurbits also occur, indicating its capacity to occupy niches other than those occupied by *C. maxima andreana*. Results show that ecological niche modeling can be an important tool to understand specialized interactions, such as those involving plant-pollinator species. Nevertheless, further investigation should be conducted to amplify the analyses, such as new surveys in areas with little available data and improvements in analyzing the results of modeling.

Acknowledgments

The authors are grateful to Dr. Maria A. J. F. Ferreira of EMBRAPA – CENARGEM, Dr. Michael Nee of the New York Botanical Garden, Dr. Gabriel Melo and Dr. Danuncia Urban of Universidade Federal do Parana, Dr. Antonio Aguiar of Museu de Zoologia, Dr. Roberto Shimizu of Instituto de Biociencias of Universidade de São Paulo and the anonymous referee for the careful revision of the manuscript. They are also grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 04/15801-0, 04/11012-0 and 03-08134-4) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – scholarship grant to the first author).

References

- Anderson, R. P., Peterson, A. T. and Gomez-Laverde, M. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* 98, 3–16.
- Anderson, R. P., Lew, D. and Peterson, A. T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162, 211-232.
- Araujo, M. B., Pearson, R. G., Thuiller, W. and Erhard, M. 2005. Validation of species-climate impact models under climate change. *Global Change Biology* 11, 1504-1513.
- Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modeling* 200, 1-19.
- Batalha Filho, H., Nunes, L. A., Pereira, D. G. and Waldschmidt, A. M. 2007. Inventário da fauna de abelhas (Hymn; Apoidea) em uma área de caatinga da região de Jequié, BA. *Bioscience Journal* 23, 24-9.
- Bischoff, I., Schroder, S. and Misof, B. 2009. Differentiation and range expansion of North American Squash Bee, *Peponapis pruinosa* (Apidae: Apiformes) populations assessed by geometric wing morphometry. *Annals of the Entomological Society of America* 102, 60-69.
- Biesmeijer, J. C., Roberts, S. P. M., Reemer, M., Ohlemuller, R., Edwards, M., Peeters, T., Schaffers, A. P., Potts, S. G., Kleukers, R., Thomas, C. D., Settele, J. and Kunin, W. E. 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* 313, 351-354.
- Bouseman, J. K. 1975. *Peponapis fervens* (Smith) in Bolivia (Hymenoptera: Apoidea). *Proceedings of the Entomological Society of Washington* 77, 508.
- Brown, J. H. and Lomolino, M. V. 2006. Biogeografia. 2a edição. Funpec. Ribeirão Preto. 692p.
- Buchmann, S. L. and Nabhan, G. P. 1996. The Forgotten Pollinators. Island. Washington, DC. 292p.

- Cane, J. et al. 2004. Importance of squash bees (*Peponapis* and *Xenoglossa*) as pollinators of domestic *Cucurbita* in the Americas: SPAS (Squash Pollinators of the Americas Survey). Case Studies on Conservation of Pollination Services as a Component of Agricultural Biological Diversity. <http://www.fao.org/>.
- Cane, J. 2005. Squash Pollinators of the Americas Survey (SPAS). USDA. Agricultural Research Service. <http://www.ars.usda.gov/>.
- Canto-Aguilar, A. and Parra-Tabla, V. 2000. Importance of conserving alternative pollinators: assessing the pollination efficiency of the squash bee, *Peponapis limitaris* in *Cucurbita moschata* (Cucurbitaceae). Journal of Insect Conservation 4, 203–210.
- Cutler, H. C. and Whitaker, T. W. 1961. History and distribution of the cultivated cucurbits in the Americas. American Antiquity 26, 469-485.
- Cutler, H. C. and Whitaker, T. W. 1968. A new species of *Cucurbita* from Ecuador. Annals of the Missouri Botanical Garden 55, 392-396.
- Decker, D. S. 1988. Origin(s), Evolution, and systematics of *Cucurbita pepo* (Cucurbitaceae). Economy Botanic 42, 4-15.
- Delaplane, K. S. and Mayer, D. F. 2000. Crop pollination by bees. CABI Publishing. Wallingford, UK and New York. 352p.
- Dillehay, T. D., Rossen, J., Andres, T. C. and Williams, D. E. 2007. Preceramic adoption of peanut, squash, and cotton in Northern Peru. Science 316, 1890.
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberon, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129-151.
- Ferreira, M. A. J. F., Melo, A. M. T., Carmo, C. A. S., Silva, D. J. H., Lopes, J. F., Queiroz, M. A., Moura, M. C. C. L. and Negrini, R. C. S. 2006. Mapeamento da distribuição geográfica e conservação dos parentes silvestres e variedades crioulas

- de *Cucurbita*. In: Coradin, L. (coord.) Parentes Silvestres das Espécies de Plantas Cultivadas. Ministério do Meio Ambiente.
- Ferreira, M. A. J. F. 2007. Abóboras, morangas e abobrinhas: estratégias para coleta, conservação e uso. <http://www.embrapa.br/>.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24, 38–49.
- Fronk, W. D. and Slater, J. A. 1956. Insect fauna of the cucurbit flowers. *Journal of Kansas Entomological Society* 29, 141-145.
- Gaubert, P., Papes, M. and Peterson, A. T. 2006. Natural history collections and the conservation of poorly known taxa: Ecological niche modeling in central African rainforest genets (*Genetta* spp.). *Biological Conservation* 130, 106 –117.
- Girao, L. C., Lopes, A. V., Tabarelli, M. and Bruna, E. M. 2007. Changes in tree reproductive traits reduce functional diversity in a fragmented Atlantic Forest landscape. *PLoS ONE* 9, 1-12.
- Gomes, M. F. F. 1991. Polinização entomófila na produção de sementes híbridas (*C. maxima* x *C. moschata*). Dissertation for Master's Degree. Universidade Federal de Viçosa. Minas Gerais, Brasil.
- Heiser, C. B. 1979. Origins of some cultivated New World plants. *Annual Review of Ecology and Systematics* 10, 309-326.
- Heiser, C. 1990. New perspectives on the origin and evolution of New World domesticated plants. *Economic Botany* 44, 111-116
- Hinojosa-Diaz, I. A., Yanez-Ordóñez, O., Chen, G., Peterson, A. T. and Engel, M. S. 2005. The North American invasion of the giant resin bee (Hymenoptera: Megachilidae). *Journal of Hymenoptera Research* 14, 69–77.
- Hurd, P. D. Jr. and Linsley, E. G. 1964. The squash and gourd bees – Genera *Peponapis* Robertson and *Xenoglossa* Smith – Inhabiting America North of Mexico (Hymenoptera: Apidae). *Hilgardia* 35, 375-477.

- Hurd, P. D. Jr. and Linsley, E. G. 1966. The Mexican squash and gourd bees of the genus *Peponapis* (Hymenoptera: Apoidea). *Annals of the Entomological Society of America* 59, 835-851.
- Hurd, P. D. Jr. and Linsley, E. G. 1967. South American squash and gourd bees of the genus *Peponapis*. *Annals of the Entomological Society of America* 60, 647-61.
- Hurd, P. D. Jr. and Linsley, E. G. 1970. A classification of the squash and gourd bee *Peponapis* and *Xenoglossa* (Hymenoptera: Apoidea). In: Belkin, J. N., Bohart, R. M., DeBach, P., Douthett, R. L., Jensen, D. D., Lange, W. H. and Schlinger, E. I. *University of California Publications in Entomology* 69, 1-39.
- Hurd, P. D. Jr., Linsley, E. G. and Whitaker, T. W. 1971. Squash and gourd bees (*Peponapis*, *Xenoglossa*) and the origin of the cultivated *Cucurbita*. *Evolution* 25, 218-234.
- Hurd, P. D. Jr., Linsley, E. G. and Michelbacher, A. E. 1974. Ecology of the squash and gourd bee, *Peponapis pruinosa*, on cultivated cucurbits in California (Hymenoptera: Apoidea). *Smithsonian Contributions to Zoology* 168, 1-17.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology* 22, 415-427.
- Jimenez-Valverde, A., Gómez, J. F., Lobo, J. M., Baselga, A. and Hortal, J. 2008. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. *Annales Zoologici Fennici* 45, 200-210.
- Kearns, C. A., Inouye, D. W. and Waser, N. M. 1998. Endangered mutualisms: the conservation of plant-pollinator interactions. *Annual Review of Ecology, Evolution, and Systematics* 29, 83-112.
- Leathwick, J. R. and Austin, M. P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82, 2560-2573.
- Linsley, E. G. 1960. Observations on some matinal bees at flowers of *Cucurbita*, *Ipomoea* and *Datura* in desert areas of New Mexico and southeastern Arizona. *Journal New York Entomological Society* 68, 13-20.
- Lobo, J. M., Jimenez-Valverde, A. and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17, 145-151.

- Losey, J. E. and Vaughan, M. 2006. The economic value of ecological services provided by insects. *BioScience* 56, 311-323.
- McGarigal, K., Cushman, S. and Stafford, S., 2000. Multivariate statistics for wildlife and ecology research. Springer-Verlag. New York, 283p.
- McPherson, J. M., Jetz, W. and Rogers, D. J. 2004. The effect's of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41, 811-823.
- Meléndez-Ramírez, V., Magaña-Rueda, S., Parra-Tabla, V., Ayala, R. and Navarro, J. 2002. Diversity of native bee visitors of cucurbit crops (Cucurbitaceae) in Yucatán, México. *Journal of Insect Conservation* 6, 135–147.
- Memmott, J., Craze, P. G., Waser, N. M. and Price, M. V. 2007. Global warming and the disruption of plant-pollinator interactions. *Ecology Letters* 10, 710-717.
- Michener, C. D. and Lange, R. B. 1958. Observations of the ethology of Neotropical Anthophorine bees (Hymenoptera: Apidae). *The University of Kansas Science Bulletin* 39, 69-96.
- Michener, C. D. 2007. *The bees of the world*. John Hopkins University. Baltimore, 913p.
- Nee, M. 1990. The domestication of *Cucurbita* (Cucurbitaceae). *Economic Botany* 44, 56-68.
- Nix, H.A. 1986. A biogeographic analysis of Australian elapid snakes. In: Longmore, R. (Ed.), *Atlas of Elapid Snakes of Australia*. Australian Flora and Fauna Series Number 7. Australian Government Publishing Service, Canberra, p. 4-15.
- Ordway, E., Buchmann, S. L., Kuehl, R. O. and Shipman, C. W. 1987. Pollen dispersal in *C. foetidissima* (Cucurbitaceae) by bees of the genera *Apis*, *Peponapis* and *Xenoglossa* (Hymenoptera: Apidae). *Journal Kansas Entomological Society* 60, 489-503.
- Peterson, A. T. and Vieglais, D.A. 2001. Predicting species invasions using ecological niche modeling. *BioScience* 51, 363–71.

- Peterson, A. T., Ortega-Huerta, M. A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H. and Stockwell, D.R. 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416, 626-9.
- Peterson, A. T., Papes, M. and Soberon, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213, 63-72.
- Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231–259.
- Rozen, J. G. and Ayala, R. 1987. Nesting biology of the squash bee *Peponapis utahensis* (Hymenoptera; Anthophoridae; Eucerini). *Journal of New York Entomology Society* 95, 28-33.
- Saade, R. L. and Hernandez, S. M. 1992. Cucurbits. In: Bermejo, J. E. and Leon, J. 1992 (Editors). *Neglected crops: 1492 from a different perspective*. Plant Production and Protection Series n. 26. FAO, Rome, Italy.
- Saade, R. L., Andrés, T. C. and Nee, M. 1995. *Cucurbita* L. In: Saade, R. L. *Estudios taxonómicos y ecogeográficos de las Cucurbitaceae latinoamericanas de importancia económica*. Instituto de Biología da UNAM e IPGRI. 255p.
- Sanjur, O. I., Piperno, D. R., Andres, T. C. and Wessel-Beaver, L. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *PNAS* 99, 535–540.
- Santana, F. S., Siqueira M. F., Saraiva, A. M. and Correa, P. L. P. 2008. A reference business process for ecological niche modeling. *Ecological Informatics* 3, 75-86.
- Shuler, R. E., Roulston, T. H. and Farris, G. E. 2005. Foraging practices influence wild pollinator populations on squash and pumpkin. *Journal of Economic Entomology* 98, 790-5.
- Siqueira, M. F. and Peterson, A. T. 2003. Consequences of global climate change for geographic distributions of cerrado tree species. *Biota Neotropica* 3, 1-14.
- Smith, B. D. 2006. Eastern North America as an independent center of plant domestication. *PNAS* 103, 12223–12228.

- Sneath, P. H. A. and Sokal, R. R. 1973. Numerical taxonomy. W. H. Freeman. San Francisco, 573p.
- Soberon, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2, 1-10.
- Soberon, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10, 1115–1123.
- Stockwell, D. and Peters, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13, 143-158.
- Stockwell, D. R. B. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148, 1–13.
- Stockwell, D. R. B., Beach, J. H., Stewart, A., Vorontsov, G., Vieglais, D. and Pereira, R. C. 2006. The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological Modelling* 195, 139-145.
- Stokstad, E. 2007. The case of the empty hives. *Science*, 316, 970–972.
- Swets, K. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Tepedino, V. J. 1981. The pollination efficiency of the squash bee (*Peponapis pruinosa*) and the honey bee (*Apis mellifera*) on summer squash (*Cucurbita pepo*). *Journal of the Kansas Entomological Society* 54, 359-377.
- Teppner, H. 2004. Notes on *Lagenaria* and *Cucurbita*. *Phyton* 44, 245-308.
- Wessel-Beaver, L. 2000. Evidence for the center of diversity of *Cucurbita moschata* in Colombia. *Cucurbit Genetics Cooperative Report* 23, 54-55.
- Whitaker, T. W. 1956. The origin of the cultivated *Cucurbita*. *The American Naturalist* 90, 171-176.
- Willis, D. S. and Kevan, P. G. 1995. Foraging dynamics of *Peponapis pruinosa* (Hymenoptera: Anthophoridae) on pumpkin (*Cucurbita pepo*) in southern Ontario. *Canadian Entomologist* 127, 167-175.

- Winfree, R., Williams, N. M., Dushoff, J. and Kremen, C. 2007. Native bees provide insurance against ongoing honey bee losses. *Ecology Letters* 10, 1105–1113.
- Woodson Jr., R. E., Schery, R. W. and Wunderlin, R. P. 1978. Flora of Panama. Part IX. Family 182. Cucurbitaceae. *Annals of the Missouri Botanical Garden* 65, 285-366.

Supplementary Material: Data sources

Peponapis

- Discover Life (<http://www.discoverlife.org/>)
 - American Museum of Natural History
- speciesLink (<http://splink.cria.org.br>)
 - Coleção da Pontifícia Universidade Católica de Rio Grande do Sul (PUCRS)
 - Coleção de Abelhas do Museu de Ciências e Tecnologia da PUCRS
 - Coleção Entomológica Paulo Nogueira Neto
- Coleção Entomológica Pe. Jesus Santiago Moure of Federal University of Paraná (UFPR) – personal visit
- Dr. Antonio Aguiar of Museu de Zoologia – personal communication
- Bouseman, J. K. 1975. *Peponapis fervens* (Smith) in Bolivia (Hymenoptera: Apoidea). *Proceedings of the Entomological Society of Washington* 77, 508.
- Gomes, M. F. F. 1991. Polinização entomófila na produção de sementes híbridas (*C. maxima* x *C. moschata*). Dissertation for Master's Degree. Universidade Federal de Viçosa. Minas Gerais, Brasil.
- Holmberg, E. L. 1903. *Delectus Hymenopterologicus Argentinus*. *Anales del Museo Nacional de Historia Natural de Buenos Aires* 2, 377-468.
- Hurd, P. D. Jr. and Linsley, E. G. 1967. South American squash and gourd bees of the genus *Peponapis*. *Annals of the Entomological Society of America* 60, 647-61.
- Jörgensen, P. 1909. Beobachtungen über Blumenbesuch, Biologie, Verbreitung usw. der Bienen von Mendoza. (Hym.). *Deutsche Entomologische Zeitschrift* 1909, 53-65.
- Krug, C. 2007. A comunidade de abelhas. (Hymenoptera - Apiformes) da Mata com. Araucária em Porto União-SC e abelhas visitantes florais da abóboreira (*Cucurbita*. L.) em Santa Catarina, com notas sobre *Peponapis fervens* (Eucerini, Apidae). Universidade do Extremo Sul Catarinense. Programa de Pós-Graduação em Ciências Ambientais. 120p.
- LaBerge, W. E. & Moure, J. S. 1962. Type specimens of American Eucerini bees deposited in the British Museum (Hymenoptera, Apoidea). *Boletim da Universidade Federal do Paraná, Zoologia* 1, 1-12.
- Schrottky, C. 1903. Enumération des hyménoptères connus jusqu'ici de la République Argentine, de l'Uruguay et du Paraguay. *Anales de la Sociedad Científica Argentina* 55, 176-186.
- Weiss, G. & Melo, G. A. R. 2007. Comportamento forrageiro de *Peponapis fervens* Smith (Hymenoptera, Apidae, Eucerini) em flores de *Cucurbita* (Cucurbitaceae). *Anais do VIII Congresso de Ecologia do Brasil*.

Cucurbita

- Instituto Nacional de Pesquisas da Amazônia Manaus - Amazonas (accessed through <http://angelim.inpa.gov.br>)
- Herbário do Jardim Botânico do Rio de Janeiro (accessed through http://jabot.jbrj.gov.br/frmfiltroespecimes_pub.php)
- The New York Botanical Garden - Virtual Herbarium (accessed through <http://sciweb.nybg.org/science2/vii2.asp>)
- CONABIO (<http://www.conabio.gob.mx/remib>)
 - University of Aarhus Herbarium Jutlandicum
 - Bailey Hortorium Herbarium (BH) Cornell University
 - Universidad Nacional de Córdoba Herbario (CORD)
 - Herbário HUA Centro de Investigaciones en Ciencias Exactas y Naturales, Universidad de Antioquia
 - Herbario Nacional de Bolivia (LPB)
 - Missouri Botanical Garden
 - Pontificia Universidad Católica del Ecuador Herbario (QCA)
 - Museo Ecuatoriano de Ciencias Naturales Herbario Nacional del Ecuador (QCNE)

- Herbario SI Instituto de Botánica Darwinion
- The New York Botanical Garden
- Herbario USZ Universidad Autónoma Gabriel René Moreno Herbario del Oriente Boliviano
- GBIF (<http://www.gbif.org/>)
 - Bioversity International
 - Herbarium of the University of Aarhus
 - Herbier de la Guyane
 - Missouri Botanical Garden
 - US National Plant Germplasm System
- speciesLink (<http://splink.cria.org.br>)
 - ESA Herbario da Escola Superior de Agricultura Luiz de Queiroz Escola Superior de Agricultura Luiz de Queiroz Piracicaba - São
 - Herbario Dárdano de Andrade Lima (Instituto Agronômico de Pernambuco)
 - Herbario Dimitri Sucre Benjamin Jardim Botânico do Rio de Janeiro, JBRJ
 - Herbario Dom Bento Pickel Instituto Florestal São Paulo - São Paulo
 - Herbario Instituto Nacional de Pesquisas da Amazônia Manaus - Amazonas
 - The New York Botanical Garden, NYBG
- Discover Life (at <http://www.discoverlife.org/>)
 - Missouri Botanical Garden
- Herbario SPF – Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo – personal visit
- Herbario do Estado Maria E.P.K. Fidalgo - Instituto Botânico – personal visit
- Dr. Maria A. J. F. Ferreira of EMBRAPA – CENARGEM (Empresa Brasileira de Pesquisa Agropecuária) – personal communication
- Dr. Michael Nee of the New York Botanical Garden – personal communication
- Ashworth, L.; Galetto, L. 2001. Pollinators and Reproductive Success of the Wild Cucurbit *Cucurbita maxima* ssp. *andreana* (Cucurbitaceae). Plant Biology 3, 398-404.

CAPÍTULO 3

Ecological niche similarities of *Peponapis* bees and non-domesticated
Cucurbita species

Tereza C. Giannini, Rafael Lira-Saade, Ricardo Ayala, Antonio M. Saraiva,
Isabel Alves-dos-Santos

***Ecological Modelling* (2011) v. 222, p. 2011-2018.**

Peponapis* em flor de *Cucurbita
Foto: T'ai Roulston

Ecological niche similarities of *Peponapis* bees and non-domesticated *Cucurbita* species

Tereza C. Giannini¹, Rafael Lira-Saade², Ricardo Ayala³, Antonio M. Saraiva⁴, Isabel Alves-dos-Santos¹

1. Instituto de Biociências da Universidade de São Paulo. Rua do Matão, Trav. 14, nº 321, Butantã, São Paulo, SP, Brasil, 05508-900.
2. Facultad de Estudios Superiores Iztacala. Universidad Nacional Autónoma de México. Av. de los Barrios 1, Tlanepantla, México, México, 54090.
3. Estación de Biología Chamela, Instituto de Biología. Universidad Nacional Autónoma de México. Apartado Postal 21, San Patricio, Jalisco, México, 48980.
4. Escola Politécnica da Universidade de São Paulo. Av. Prof. Luciano Gualberto, nº 158, Butantã, São Paulo, SP, Brasil, 05508-900.

Abstract

Peponapis bees are considered specialized pollinators of *Cucurbita* flowers, a genus that presents several species of economic value (squashes and pumpkins). Both genera originated in the Americas, and their diversity dispersion center is in Mexico. Ten species of *Peponapis* and ten species of *Cucurbita* (only non-domesticated species) were analyzed considering the similarity of their ecological niche characteristics with respect to climatic conditions of their occurrence areas (abiotic variables) and interactions between species (biotic variables). The similarity of climatic conditions (temperature and precipitation) was estimated through cluster analyses. The areas of potential occurrence of the most similar species were obtained through ecological niche modeling and summed with geographic information system tools. Three main clusters were obtained: one with species that shared potential occurrence areas mainly in deserts (*P. pruinosa*, *P. timberlakei*, *C. digitata*, *C. palmata*, *C. foetidissima*), another in moist forests (*P. limitaris*, *P. atrata*, *C. lundelliana*, *C. o. martinezii*) and a third mainly in dry forests (*C. a. sororia*, *C. radicans*, *C. pedatifolia*, *P. azteca*, *P. smithi*, *P. crassidentata*, *P. utahensis*). Some species with similar ecological niche presented potential shared areas that are also similar to their geographical distribution, like those occurring predominantly on deserts. However, some clustered species presented larger geographical areas, such as *P. pruinosa* and *C. foetidissima* suggesting other drivers

than climatic conditions to shape their distributions. The domestication of *Cucurbita* and also the natural history of both genera were considered also as important factors.

Keywords: interspecific interaction, ecological niche modeling, specialized species, pollinators

Introduction

Multiple environmental variables, such as climatic conditions, play a role in the requirements of species with respect to their fundamental niche and affect their geographical distribution (Hutchinson 1957). Although suitable conditions may exist in many regions, other biological (e.g., interactions with other species) and historical processes can prevent a species from existing throughout its full potential geographical distribution, which reduces the fundamental niche to a smaller portion of the potential range, i.e., the realized niche (Hutchinson 1957). Climatic variables are considered as conditions, or *scenopoetic* variables, and there is supposedly no competition for them; however, there may be competition for biotic variables, such as resources (*bionomic* variables) (Hutchinson 1978), especially between species that co-occur in the same areas (Waser 1983, Mitchell et al. 2009a).

Ecological niche modeling has been used as a computational tool to combine occurrence data with environmental variables based on a correlative approach to build a representation of the ecological requirements of a species (Anderson et al. 2003). Several algorithms have been applied to create such models, and their computational results can be projected onto a map showing areas that are similar to those where the species is known to occur and that represents the potential areas of occurrence (Stockwell and Peters 1999; Phillips et al. 2006). Although ecological niche has been described as a multidimensional space where the multiple dimensions represent distinct abiotic and biotic factors (Hutchinson 1957), few studies have included species interactions (biotic factors) in the modeling processes. In most examples in the literature, the layers provided as inputs to algorithms are abiotic in nature; i.e., algorithms are generally not provided with information about biotic variables because such information is generally unavailable or complex to interpret (Soberón and Peterson 2005). Nevertheless, biotic and historical processes can be added into the interpretation

of modeling to analyze the distribution of interacting species (such as Anderson et al. 2002, Jiménez-Valverde et al. 2008, Godsoe et al. 2009).

Pollination is a classic ecological interaction in which plants provide floral visitors with rewards such as nectar, pollen or oil, and pollinating animals facilitate plant reproduction by dispersing pollen grains. Interactions involving specialized plant-pollinator species can offer an interesting case study for discussion of niche similarity and patterns of geographical distribution. A plant is said to present specialized pollination if it is successfully pollinated only by a subset of pollinators, and such a plant is also said to occupy pollination niches (Fenster et. al 2004). Mitchell et al. (2009b) summarized the complexities of generalized and specialized systems as follows: the shifting costs and benefits of plant-pollinator interactions may also play an important role in determining whether plant-pollinator interactions are more generalized or specialized. If a plant species has many different visitor taxa that provide similar pollination services, and if costs of the interaction are comparable, the net benefits to plants should also be similar, and there is little incentive for plants to specialize in attracting a particular group of pollinators. On the other hand, if some floral visitors are more effective in terms of the quantity or quality of pollen transfer, selection should favor traits promoting these effective pollinators. Such selection that favors specialization of particular pollinator species or functional groups might explain why some plants have traits that appear to restrict their pollinators. However, Thompson (1994) suggested that the specialization of interactions is an evolutionarily dynamic state that changes in short periods of time. Such interactions are intrinsically related to the geographical distribution (Thompson 2005) by which the different populations are shaped and can evolve to include different species.

This particular kind of mutualism may be substantially altered if flowering species compete for the services of shared pollinators; such competition provides an counterpoint to mutualistic and competitive interactions (Mitchell et al. 2009a). Plant-pollinator interactions are often viewed as tightly coevolved relationships, but despite the potential for mutual benefits, these interactions also present inherent conflicts that may vary spatially and temporally and need not involve tight coevolution (Waser 2001, Mitchell et al. 2009b). Although co-existing species can exhibit some differences in their ecological requirements to avoid or minimize competition (Waser 1983), this is not a rule because in some cases, traits of these species have been found to be more similar

to one another than expected by chance (Chase and Leibold 2003). At present, interspecific competition is assumed to be either niche-based or neutral (Ruokolainen et al. 2009). Under the first hypothesis, the community presents general and predictable processes along geographical gradients (Law and Morton 1996), and in the second case, the communities are structured by neutral drift (Hubbell 2001).

Among specialized plant-pollinator species, the case of the *Peponapis* genus is well known. These bees are specialized for *Cucurbita* flowers and dependent on the pollen grains of these flowers to feed their larvae (Fronk and Slater 1956, Michener and Lange 1958, Linsley 1960, Hurd and Linsley 1964, 1966, Hurd et al. 1971). The center of dispersion for these two genera is located in Mexico, and some species also occur in other countries of North and Central America. Other *Cucurbita* species have been cultivated in America for thousands of years, and this has probably contributed to the expansion of the distribution of *Peponapis* bees (Hurd et al. 1971).

Because of the importance of pollination as a key ecosystem service (Daily 1997), it is necessary to improve our knowledge about pollinators, especially because of the increasing threats to and decline of these organisms (Buchmann and Nabhan 1996, Kearns et al. 1998, Potts et al. 2010). This study examines the similarity of climatic requirements of North and Central American species of *Peponapis* and non-domesticated *Cucurbita* and analyzes the potential areas of occurrence that are shared by similar species.

Material and Methods

The sites of occurrence of *Peponapis* and non-domesticated *Cucurbita* species were obtained from internet biodiversity portals and specialized literature. The coordinates where the surveys were conducted were used whenever a clear specification about geographical coordinates was not available in the original sources (see Supplementary Material 1). The coordinates were converted to decimal degrees, checked and corrected to build the final database (Table 1).

Table 1. Number of occurrence points of the examined species obtained from the consulted sources

Species*	Number of points	Countries**
1. <i>Peponapis apiculata</i> (Cresson, 1878)	12	Mexico, (<i>Guatemala, El Salvador</i>), Costa Rica, Panama
2. <i>Peponapis atrata</i> (Smith)	8	Mexico (<i>Guatemala</i>)
3. <i>Peponapis azteca</i> Hurd & Linsley, 1966	77	Mexico (<i>Guatemala</i>)
4. <i>Peponapis crassidentata</i> (Cockerell, 1949)	68	USA, México, El Salvador, (<i>Honduras, Nicaragua</i>), Costa Rica
5. <i>Peponapis limitaris</i> (Cockerell, 1906)	35	México, (<i>Guatemala, El Salvador, Honduras, Nicaragua</i>), Costa Rica, Panama
6. <i>Peponapis michelbacherorum</i> Hurd & Linsley, 1964	17	USA, Mexico
7. <i>Peponapis pruinosa</i> (Say, 1837)	617	Canada, USA, Mexico
8. <i>Peponapis smithi</i> Hurd & Linsley, 1966	56	Mexico
9. <i>Peponapis timberlakei</i> Hurd & Linsley, 1964	37	USA, Mexico
10. <i>Peponapis utahensis</i> (Cockerell, 1905)	92	USA, Mexico, Guatemala, El Salvador, (<i>Honduras, Nicaragua</i>), Costa Rica
Total	1019	
11. <i>Cucurbita argyrosperma sororia</i> (L. H. Bailey) Merrick & Bates	39	Mexico, Nicaragua (<i>Panama</i>)
12. <i>Cucurbita cordata</i> S. Watson	56	Mexico
13. <i>Cucurbita digitata</i> A. Gray	69	USA, Mexico
14. <i>Cucurbita foetidissima</i> H.B.K.	203	USA, Mexico
15. <i>Cucurbita lundelliana</i> L. H. Bailey	47	Mexico, Guatemala, Belize, Nicaragua
16. <i>Cucurbita okeechobeensis</i> subsp. <i>martinezii</i> (L. H. Bailey) Walters & Decker-Walters 1993	34	Mexico
17. <i>Cucurbita okeechobeensis</i> subsp. <i>okeechobeensis</i> (J. K. Small) L. H. Bailey	11	USA
18. <i>Cucurbita palmata</i> S. Watson	123	USA, Mexico
19. <i>Cucurbita pedatifolia</i> L. H. Bailey	22	Mexico
20. <i>Cucurbita radicans</i> Naudin	47	Mexico
Total	651	

* *Peponapis* according to Hurd and Linsley, 1970, and *Cucurbita* according to Lira-Saade et al., 1995.

**The countries emphasized in italics are reported as having these species, but it was not possible to obtain any geographical coordinates from the data sources (see Supplementary Material 1 for details about data sources).

Ecological niche modeling was performed for each species separately using the openModeler computational system, version 1.0.9 (Santana et al. 2008; Muñoz et al. 2009). This system has many algorithms, including GARP (Genetic Algorithm for Rule-set Production). Modeling using GARP is based on common operations of genetic algorithms (selection, crossover and mutation) that are applied to rules that define the limits of environmental variables by considering the species occurrence points (Stockwell and Peters 1999). The specific algorithm used was GARP with best subsets (Desktop GARP implementation) (Stockwell et al. 2006), which produces twenty models. A composite prediction is created based on all of the models to generate a more reasonable estimation of the species' potential distribution (Anderson et al. 2003). This algorithm has been widely used by several authors and applied on different species for ecological niche modeling (Anderson et al. 2002, Peterson and Holt 2003, Oberhauser and Peterson 2003, Schussman et al. 2006, Levine et al. 2007, Sweeney et al. 2007, DeVaney et al. 2009, Siqueira et al. 2009) and proved to be a reliable method for modeling based on the use of sets of data originating from biological collections (Stockwell et al. 2006), presence only data (Tsoar et al. 2007), small sample of occurrence points (Stockwell and Peterson 2002, Wisz et al. 2008) and with ordered (i.e. non-categorical) data (Elith and Graham 2009).

Environmental layers with five arc-minutes of resolution were obtained from the Worldclim website (Hijmans et al. 2005), and an Americas mask was used to limit the model to this continent. The environmental layers considered the following variables: 1) annual mean temperature; 2) mean diurnal range; 3) isothermality; 4) temperature seasonality; 5) maximum temperature of the warmest period; 6) minimum temperature of the coldest period; 7) temperature annual range; 8) mean temperature of the wettest quarter; 9) mean temperature of the driest quarter; 10) mean temperature of the warmest quarter; 11) mean temperature of the coldest quarter; 12) annual precipitation; 13) precipitation of the wettest period; 14) precipitation of the driest period; 15) precipitation seasonality; 16) precipitation of the wettest quarter; 17) precipitation of the driest quarter; 18) precipitation of the warmest quarter; 19) precipitation of coldest quarter.

The AUC values (area under the curve) were used to determine the quality of results. Two tests based on AUC were performed: an internal test using all the points of occurrence obtained, and an external test (Araújo et al. 2005). The data were divided

into partitions, randomly and without replacement, and each partition was divided into test data and training data. The proportion of training and test points was 90% and 10%, respectively. Swets (1988) suggested that the results of the average AUC could be interpreted as follows: excellent, when above 0.90, good between 0.90-0.80; fair between 0.80-0.70; poor, between 0.70-0.60; and fail, between 0.60-0.50. The AUC of the external test has been considered as the most reasonable estimate for the model's quality (Fielding and Bell 1997). However, methods to estimate the quality of the resulting models have been discussed in the literature, and there is not yet a consensus (Peterson et al. 2008). The final models that were used to build the maps were obtained using all occurrence points.

The openModeller system can extract the values of environmental variables used for each occurrence point, and these values were used to perform a cluster analysis. The averages of all parameters, the Euclidian distance and a UPGMA algorithm were applied using Past Program (Hammer and Harper 2005) (free access on <http://folk.uio.no/ohammer/past/>). The cophenetic correlation index was calculated, for which a value higher than 0.75 is considered a good representation of the original data (McGarigal et al. 2000). Some climatic parameters of clustered species were also analyzed through histograms.

To analyze the shared potential areas, the maps of clustered species were summed, i.e. the potential areas of all species per cluster were represented in the resulting map. Therefore, these shared potential areas correspond to the regions where virtually all of the clustered species can be found together. The WWF (World Wildlife Fund) database of Terrestrial Ecoregions (Olson et al. 2001) was used to characterize the shared potential areas following the suggestion of Soberón and Peterson (2005). All of these procedures were performed using ArcGIS (9.3) (ESRI Inc.).

Results

Cluster analyses performed with the averages of climatic parameters resulted in three main clusters (Figure 1). Cluster A was composed of *P. pruinosa*, *P. timberlakei*, *C. digitata*, *C. palmata*, and *C. foetidissima*. Cluster B included *P. limitaris*, *P. atrata*, *C. lundelliana*, and *C. o. martinezii*, and Cluster C, *C. a. sororia*, *C. radicans*, *C. pedatifolia*, *P. azteca*, *P. smithi*, *P. crassidentata*, and *P. utahensis*.

A fourth cluster (D) presented three species with low similarity: *P. michelbacherorum*, *C. o. okeechobeensis* and *C. cordata*. *Peponapis apiculata* was also isolated from Cluster A and showed little similarity with the other clusters.

The cophenetic correlation index obtained was 0.87, which represents a good fit for the clusters obtained.

Figure 1. Cluster Analyses performed with *Peponapis* and *Cucurbita* analyzed species using the averages of the environmental layers.

The shared potential areas of all species in the same cluster coincided with certain ecoregions. Areas of Cluster A mainly included the Mojave, Sonoran and Chihuahan deserts. Species of Cluster B were mainly associated with the Veracruz, Pelen-Veracruz and Yucatan moist forest regions. They also presented a potential area of occurrence in Honduras; however, no occurrence point was obtained for this area. Cluster C species were mainly associated with the Sinaloan, Bajio and Balsas dry forests. The species of Cluster D presented no shared potential occurrence sites (Figure 2).

Figure 2. Shared potential areas obtained through the sum of potential areas of clustered *Peponapis* and *Cucurbita* species.

The quality of all models obtained through ecological niche modeling was considered suitable presenting AUC values higher than 0.91 (Table 2).

Table 2. AUC values of external tests performed with 10% of the occurrence points

Species	AUC	Species	AUC
<i>Peponapis apiculata</i>	1.0	<i>Cucurbita argyrosperma sororia</i>	0.96
<i>Peponapis atrata</i>	0.98	<i>Cucurbita cordata</i>	0.91
<i>Peponapis azteca</i>	0.96	<i>Cucurbita digitata</i>	0.91
<i>Peponapis crassidentata</i>	0.95	<i>Cucurbita foetidissima</i>	0.97
<i>Peponapis limitaris</i>	0.94	<i>Cucurbita lundelliana</i>	1.0
<i>Peponapis michelbacherorum</i>	1.0	<i>Cucurbita okechobeensis martinezii</i>	0.99
<i>Peponapis pruinosa</i>	0.92	<i>Cucurbita okechobeensis okechobeensis</i>	1.0
<i>Peponapis smithi</i>	0.99	<i>Cucurbita palmata</i>	0.91
<i>Peponapis timberlakei</i>	0.99	<i>Cucurbita pedatifolia</i>	1.0
<i>Peponapis utahensis</i>	0.94	<i>Cucurbita radicans</i>	0.99

The variation of temperature (Figure 3) is higher in the occurrence areas of Cluster A species, especially for *P. pruinosa*. The areas of these species also presented the lowest values of precipitation. Lower variations of temperature and the highest values of precipitation were obtained for areas of Cluster B species. The areas of occurrence related to Cluster C also presented lower values of precipitation that mainly fell within the driest quarter.

Figure 3. Average, maximum and minimum temperature of the warmest and coldest quarter and precipitation of the wettest and driest quarter for *Peponapis* and *Cucurbita* species of each cluster.

Discussion

The areas of potential co-occurrence found for the most ecologically similar species are, in some cases, smaller than the total area occupied by each species, as suggested by their occurrence points. For example, *P. pruinosa* presents a wide area of distribution that extends to the east coast of the USA and to the south of Mexico, *C. foetidissima* also presents occurrence points in south Mexico, *C. a. sororia* extends to the west coast of Mexico and Central America and the distribution of *P. limitaris* also extends to Central America. Other examples are *P. crassidentata* and *P. utahensis*, which present wide ranges of occurrences extending to Costa Rica, where they have been collected on the same flowers of *Cucurbita* by Wille (1985). However, all of these species of *Peponapis* occur in areas that present high variations of temperature and, especially, of precipitation, which could amplify their ecological niche under other environmental conditions. Moreover, they also forage on domesticated cucurbits (Hurd et al. 1971), which could increase their access to more abundant resources. These characteristics have probably contributed to their occupation of a wider geographical area.

On the other hand, the areas of potential co-occurrence cover important extensions of the geographical distribution of each species. These areas coincide with three important ecoregions of North America: deserts, moist and dry forests. This relationship is not surprising because the ecoregions have been described as relatively homogenous regions with respect to their ecological systems (Wright et al. 1998), and they often integrate vegetation, topographic and climatic patterns (Johnson 2000).

Deserts of the USA and Mexico (Mojave, Sonoran and Chihuahuan) present a topography that is characterized by extensive plains with low mountains. They have long, hot summers, and although winters are moderate, the regions are subject to occasional frosts (Bailey 1995). Annual precipitation at sites in the lower Sonoran and Mojave deserts is minimal and occurs with the highest between-year unpredictability of any North American desert (Minckley et al. 2000). Three species of this cluster, *P. timberlakei*, *C. palmata* and *C. digitata*, presented smaller occurrence areas with low variation in precipitation. Additionally, the two species of *Cucurbita* are visited only by *P. timberlakei* and *P. pruinosa*, both from the same cluster.

The climate of the moist forests located on the Gulf of Mexico coast is tropical humid. The Veracruz moist forest (in the north of Veracruz and south of Tamaulipas state) presents precipitation during seven months of the year with mild temperature oscillations. Pelen-Veracruz forest (situated in the state of Tabasco) consists of a matrix of wetlands, riparian habitats and moist forests, whose soil is considered as one of the most productive in the country and highly desirable for agriculture. The forests on the Yucatán Peninsula are characterized by little topographic relief (WWF 2001), but they are drier than the others; nevertheless, the fauna is similar to that of the previous moist forests cited. All species of this cluster presented distributional areas that extend to Central America, except for *C. o. martinezzi*. This cluster also presented a potential area of occurrence in Honduras, which has no associated occurrence site. Within this area, only *P. limitaris* was registered on Honduras, but no occurrence point was found. This region is characterized by mixed ecoregions (Central America pine oak, moist and dry forests), and a survey of this area could offer additional information about the occurrence of these species. However, *P. atrata* and *C. o. martinezzi* seem not to occur beyond Guatemala. Hurd et al. (1971) suggested that *P. atrata* occurs mainly in the uplands of Mexico and is probably an oligoletic species of *C. ficifolia*, a domesticated cucurbit that is also related to areas with higher altitudes. Although modeling has suggested other areas where this species could find appropriate climatic conditions, its restriction due to resource utilization probably participates in the reduction of its distributional area.

Dry forests of Mexico presented the smallest potential area of co-occurrence and the highest number of species. The Sinaloan forests constitute the northernmost dry forests of Mexico. This region has been isolated from other warm environments, although it

shares some species with the Sonoran Desert. Dry forests of Sinaloa are considered as a type of thorny scrub belonging to the arid ecological zones of Mexico instead of as true dry forests. The Bajío dry forest presents a topography consisting of valleys interspersed with steep slopes and canyons, and the climate is tropical subhumid with a long dry season that can last up to eight months. Balsas Dry is situated along the Balsas River. The climate is also tropical subhumid with a severe dry season that can last up to eight months (WWF 2001). *Cucurbita pedatifolia* and *C. radicans* occurred only within their potential area of co-occurrence, especially in the Bajío dry forest. *Peponapis azteca* and *P. smithi* species also presented a sympatric occurrence in the Bajío forest, and the distribution of *P. smithi* extends to northern areas and *P. azteca* to southern ones. On the other hand, *Peponapis crassidentata* and *P. utahensis* presented wider and sympatric occurrence areas reaching as far as Costa Rica. All of these species are also able to exploit the resources of domesticated *Cucurbita* (Hurd et al. 1971).

Peponapis apiculata was related to two clusters (B and C) but was isolated from them, showing little similarity. This species extends to Panama, as does *P. limitaris*, and is also able to visit the domesticated species (Hurd and Linsley 1970). It occurs on uplands and was collected on *C. ficifolia* in some localities of Guatemala (Hurd and Linsley 1966).

The species of Cluster D exhibited no apparent relationship: *P. michelbacherorum* was reported only on the west coast of Sonora State (Mexico) and visits only the domesticated species of *Cucurbita* (Hurd et al. 1971). The *Cucurbita* species of this cluster also have a restricted geographical distribution. *Cucurbita cordata* is endemic to Baja California (Mexico – west coast), and *C. o. okeechobeensis* is endemic to Florida (USA – east coast). The population of this last species probably represents a peripheral isolation of a taxon that was more widely distributed in eastern Mesoamerica north to the Gulf Coast and east to Florida (Walters and Decker-Walters 1993, Andres and Nabhan 1988). It so far has not been possible to identify the pollinator of *C. o. okeechobeensis*, and it has been considered as an endangered species (U.S. Fish and Wildlife Service 1999).

In summary, our results highlighted three main areas on North and Central America where *Peponapis* and *Cucurbita* species with high similarity of ecological requirements can be found. These areas probably also present high probability of interaction and resource partitioning between species. However, the geographical distribution of the

ecologically similar species was not totally equivalent to these shared areas, showing larger or smaller ranges. On a similar analyses, *P. fervens*, one species that occurs on southern of South America, presented the highest ecological similarity with the only non-domesticated *Cucurbita* of this area (*C. maxima andreana* Naudin) but its distribution was found to be more spread, probably following the distribution of domesticated *Cucurbita* that also occurs on this region (Giannini et al. 2010).

The correspondence between the geographical distributions of interactive species has been suggested by other works (Bohan et al. 2000, Anderson et al. 2002, Winder et al. 2005, Bell et al. 2010). For example, two congeneric fig wasps analyzed by Warren et al. (2010) showed similar abiotic requirements and geographical distribution, in spite of their physiological and behavioral divergences and strong competition for their host. Analyzing another case of obligate mutualism, Godsoe et al. (2009) suggested that non climatic factors, such as the biotic interaction between *Yuca brevifolia* and its pollinators, are also responsible for divergences on the distribution of different host varieties. However, the distribution of *Peponapis* and *Cucurbita* is a complex issue, because in spite of being two specialized genera their distribution has been influenced also by the spreading of cucurbit crops.

Hortal et al. (2010) suggested four main drives to be considered as shaping the distribution of species. They stated that abiotic and biogeographic factors must act mainly on larger scales while biotic and occupancy dynamic factors act on smaller. On the intermediary scales these four factors are intermingled and play different roles depending on the natural history of the species involved. It is likely that agriculture and also biogeographic factors (their phylogeny and probable co-evolution) are shaping the distributions of *Peponapis* and *Cucurbita* and had been also important besides climatic features and interactions. Nevertheless, our results indicate overall tendencies of these specialized genera and more accurate analysis would require surveys on a local scale, where the interactions between species (biotic factors) and the processes affecting populations and their individuals (occupancy dynamic) could be better observed (Hortal et al. 2010, Soberón 2010).

Jiménez-Valverde et al. (2008) emphasized that species occurrence models are a first step in a large-scale study of geographical range. Their output could be used to detect suitable locations and drive future surveys and are a complement rather than a replacement for expert knowledge. However, our approach suggests a method to

analyze the similarity of climatic conditions required by interactive species. Additionally, we hope that it might also present implications for integrating biotic interactions into niche modeling.

Acknowledgments

The authors are grateful to Professor Paulo Inácio de Knecht López de Prado from the Department of Ecology of the University of São Paulo for suggestions in analyzing the data, to Dr. James Cane from the United States Department of Agriculture for commentaries and suggestions on earlier versions, and two anonymous referees for suggestions on manuscript. We also thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 04/15801-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – for a scholarship grant to the first author).

References

- Anderson RP, Gomez-Laverde M, Peterson AT. Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecol Biogeogr* 2002;11:131-41.
- Anderson RP, Lew D, Peterson AT. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol Model* 2003;162:211-32.
- Andres T, Nabhan GP. Taxonomic rank and rarity of *Cucurbita okeechobeensis*. *Cucurbit Genetics Cooperative Report* 1988;11:83-5.
- Araújo MB, Pearson RG, Thuiller W, Erhard M. Validation of species-climate impact models under climate change. *Glob Change Biol* 2005;11:1504-13.
- Bailey RG. Description of the Ecoregions of the United States. Available at <http://www.fs.fed.us/land/ecosysmgmt/index.html>. 1995.
- Bell JR, King RA, Bohan DA, Symondson WOC. Spatial co-occurrence networks predict the feeding histories of polyphagous arthropod predators at field scales. *Ecography* 2010;33:64-72.
- Bohan DA, Bohan AC, Glen DM, Symondson WOC, Wiltshire CW, Hughes L. Spatial dynamics of predation by carabid beetles on slugs. *J Anim Ecol*; 2000;69:367-79.

- Buchmann SL, Nabhan GP. The Forgotten Pollinators. Washington: Island; 1996.
- Chase JM, Leibold MA. Ecological niches: linking classical and contemporary approaches. The University of Chicago Press; 2003.
- Daily GC. Nature's services: societal dependence on natural ecosystems. Washington: Island; 1997.
- DeVaney SC, McNyset KM, Williams, JB, Peterson AT, Wiley EO. A tale of four “carp”: invasion potential and ecological niche modeling. PLoS ONE 2009;4:1-10.
- Elith J, Graham CH. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography 2009;32:66-77.
- Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD. Pollination syndromes and floral specialization. Annu Rev of Ecol Syst 2004;35:375-403.
- Fielding AH, Bell JF. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 1997;24:38-49.
- Fronk WD, Slater JA. Insect fauna of the cucurbit flowers. J Kansas Entomol Soc 1956;29:141-5.
- Giannini TC, Saraiva AM, Alves dos Santos I. Ecological niche modeling and geographical distribution of pollinator and plants: a case study of *Peponapis fervens* (Smith, 1879) (Eucerini: Apidae) and *Cucurbita* species (Cucurbitaceae). Ecol Inform 2010;5:59-66.
- Godsoe W, Strand E, Smith CI, Yoder JB, Esque TC, Pellmyr O. Divergence in an obligate mutualism is not explained by divergent climatic factors. New Phytol 2009;183:589-99.
- Hammer O, Harper DAT. Paleontological Data Analysis. Oxford: Blackwell; 2005.
- Hijmans JR, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 2005;25:1965-78.
- Hortal J, Roura-Pascual N, Sanders NJ, Rahbek C. Understanding (insect) species distributions across spatial scales. Ecography 2010;33:51-3.
- Hubbell SP. The unified neutral theory of biodiversity and biogeography. Princeton: Princeton University Press; 2001.

- Hurd Jr PD, Linsley EG. The squash and gourd bees - genera *Peponapis* and *Xenoglossa* - inhabiting America North of Mexico. *Hilgardia* 1964;35:375-477.
- Hurd Jr PD, Linsley EG. The Mexican squash and gourd bees of the genus *Peponapis*. *Ann Entomol Soc America* 1966;59:835-51.
- Hurd Jr PD, Linsley EG. A classification of the squash and gourd bee *Peponapis* and *Xenoglossa* (Hymenoptera: Apoidea). In: Belkin JN, Bohart RM, DeBach P, Douthett RL, Jensen DD, Lange WH, Schlinger EI, editors. *University of California Publications in Entomology* vol. 69; 1970. p. 1-39.
- Hurd Jr PD, Linsley EG, Whitaker TW. Squash and gourd bees (*Peponapis*, *Xenoglossa*) and the origin of the cultivated *Cucurbita*. *Evolution* 1971;25:218-34.
- Hutchinson GE. Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology* 1957; 22:415-27.
- Hutchinson GE. *An introduction to population ecology*. New Haven: Yale University Press; 1978.
- Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A, Hortal J. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. *Ann Zool Fennici* 2008;45:200-10.
- Johnson RK. Spatial congruence between ecoregions and littoral macroinvertebrate assemblages. *JNABS* 2000;19:475-86.
- Kearns CA, Inouye DW, Waser NM. Endangered mutualisms: the conservation of plant-pollinator interactions. *Ann Rev Ecol Evol Syst* 1998;29:83-112.
- Law R, Morton RD. Permanence and the assembly of ecological communities. *Ecology* 1996;77:762-775.
- Levine RS, Peterson AT, Yorita KL, Carroll D, Damon IK, Reynolds MG. Ecological niche and geographic distribution of human monkeypox in Africa. *PLoS ONE* 2007;1:1-7.
- Linsley EG. Observations on some matinal bees at flowers of *Cucurbita*, *Ipomoea* and *Datura* in desert areas of New Mexico and southeastern Arizona. *J N Y Entomol Soc* 1960;68:13-20.

- Lira-Saade R, Andres TC, Nee M. *Cucurbita*. In: Lira-Saade, R, editor. Estudios taxonómicos y ecogeográficos de las Cucurbitaceae Latinoamericanas de importancia económica: *Cucurbita*, *Sechium*, *Sicana* y *Cyclanthera*. Systematic and Ecogeographic Studies on Crop Gene Pools. Rome: 9th International Plant Genetic Resources Institute; 1995. p. 1-115.
- McGarigal K, Cushman S, Stafford S. Multivariate statistics for wildlife and ecology research. New York: Springer-Verlag; 2000.
- Michener CD, Lange RB. Observations of the ethology of Neotropical Anthophorine bees (Hymenoptera: Apidae). Univ Kans Sci Bull 1958;39:69-96.
- Minckley RL, Cane JH, Kervin L. Origins and ecological consequences of pollen specialization among desert bees. Proc R Soc Lond B 2000;267:265-71.
- Mitchell RJ, Flanagan RJ, Brown BJ, Waser NM, Karron JD. New frontiers in competition for pollination. Ann Bot 2009a;103:1403-13.
- Mitchell RJ, Irwin RE, Flanagan RJ, Karron JD. Ecology and evolution of plant-pollinator interactions. Ann Bot 2009b;103:1355-63.
- Muñoz MES, Giovanni R, Siqueira MF, Sutton T, Brewer P, Pereira RS, et al. 2009. openModeller: a generic approach to species' potential distribution modeling. Geoinformatica DOI 10.1007/s10707-009-0090-7.
- Oberhauser K, Peterson AT. Modeling current and future potential wintering distributions of eastern North American monarch butterflies. PNAS 2003;100:14063-8.
- Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, et al. Terrestrial ecoregions of the world: a new map of life on earth. BioScience 2001;51:933-8.
- Peterson AT, Holt RD. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. Ecol Lett 2003;6:774-82.
- Peterson AT, Papes M, Soberón J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 2008;213:63-72.
- Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecol Model 2006;190:231-59.

- Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE. Global pollinator declines: trends, impacts and drivers. *TREE* 2010;25:345-53.
- Ruokolainen L, Ranta E, Kaitala V, Fowler MS. When can we distinguish between neutral and non-neutral processes in community dynamics under ecological drift? *Ecol Lett* 2009;12: 909-19.
- Santana FS, Siqueira MF, Saraiva AM, Correa PLP. A reference business process for ecological niche modeling. *Ecol Inform* 2008;3:75-86.
- Schussman H, Geiger E, Mau-Crimmins T, Ward J. Spread and current potential distribution of an alien grass, *Eragrostis lehmanniana* Nees, in the southwestern USA: comparing historical data and ecological niche models. *Diversity Distrib* 2006;12:582-92.
- Siqueira MF, Durigan G, Marco Jr P, Peterson AT. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *J Nat Cons* 2009;17:25-32.
- Soberón J. Niche and area of distribution modeling: a population ecology perspective. *Ecography* 2010;33:159-67.
- Soberón J, Peterson AT. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiv Inf* 2005;2:1-10.
- Stockwell D, Peters D. The GARP modeling system: problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci* 1999;13:143-58.
- Stockwell DRB, Peterson AT. Effects of sample size on accuracy of species distribution models. *Ecol Model* 2002;148:1-13.
- Stockwell DRB, Beach JH, Stewart A, Vorontsov G, Vieglais D, Pereira RC. The use of the GARP genetic algorithm and internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecol Model* 2006;195:139-45.
- Sweeney AW, Beebe NW, Cooper RD. Analysis of environmental factors influencing the range of anopheline mosquitoes in northern Australia using a genetic algorithm and data mining methods. *Ecol Model* 2007;20:375-86.
- Swets KA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-93.
- Thompson JN. The coevolutionary process. University Chicago Press; 1994.

- Thompson JN. The geographic mosaic of coevolution. University Chicago Press; 2005.
- Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R. A comparative evaluation of presence only methods for modelling species distribution. *Diversity Distrib* 2007;13:397-405.
- U.S. Fish and Wildlife Service. South Florida multi-species recovery plan. U.S. Fish and Wildlife Service, Atlanta, on <http://www.fws.gov/>; 1999.
- Walters TW, Decker-Walters DS. Systematics of the endangered okeechobee gourd (*Cucurbita okeechobeensis*: Cucurbitaceae). *Sys Bot* 1993;18:175-87.
- Warren M, Robertson MP, Greeff JM. A comparative approach to understanding factors limiting abundance patterns and distributions in a fig tree-fig wasp mutualism. *Ecography* 2010;33:148-58.
- Waser NM. Competition for pollination and floral character differences among sympatric plant species: a review of evidence. In: Jones CE, Little RJ, editors. *Handbook of pollination biology*. New York: Scientific and Academic Editions; 1983. p. 277-93.
- Waser NM. Pollinator behavior and plant speciation: looking beyond the “ethological isolation” paradigm. In: Chittka L, Thomson JD, editors. *Cognitive Ecology of Pollination: Animal Behaviour and Floral Evolution*. Cambridge University Press; 2001. p. 318-35.
- Wille A. Las abejas *Peponapis* y *Xenoglossa* en Costa Rica y su importancia en la polinización de las Cucúrbitas domésticas. *Rev Biol Trop* 1985;33:17-24.
- Winder L, Alexander CJ, Holland JM, Symondson WOC, Perry JN, Woolley, C. Predatory activity and spatial pattern: the response of generalist carabids to their aphid prey. *J Anim Ecol* 2005;74:443-54.
- Wisn MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, NCEAS Predicting Species Distributions Working Group. Effects of sample size on the performance of species distribution models. *Diversity Distrib* 2008;14:763-73
- Wright RG, Murray MP, Merrill T. Ecoregions as a level of ecological analysis. *Biol Cons* 1998;86:207-13.
- WWF (World Wildlife Fund). Terrestrial Ecoregions. Available on <http://www.worldwildlife.org/>; 2001

Supplementary Material. Data sources***Peponapis***

- Discover Life (at <http://www.discoverlife.org/>)
 - American Museum of Natural History
 - Kansas Natural History Museum
 - Sam Droege, U.S. Geological Survey database
 - U.S. Geological Survey database
 - University of Kansas Insect Collection
- The Global Biodiversity Information Facility (at <http://www.gbif.org>)
 - Instituto Nacional de Biodiversidad (Costa Rica)
 - Ohio State University Insect Collection
 - University of Kansas Insect Collection
 - US Department of Agriculture -ARS
 - USDA-ARS Bee Biology and Systematics Laboratory
- Illinois Natural History Survey - INHS Insect Collection (at http://ctap.inhs.uiuc.edu/Insect/search_inhs.asp)
- Bouseman JK. The squash and gourd bees *Peponapis azteca* Hurd and Linley and *Xenoglossa gabii crawfordi* Cockerell in Chihuahua Mexico. Proc Entomol Soc Wash 1977;79:488-9.
- Hurd Jr PD, Linsley EG. The Mexican squash and gourd bees of the genus *Peponapis*. Ann Entomol Soc America 1966;59:835-51.
- Hurd Jr PD, Linsley EG. The squash and gourd bees - genera *Peponapis* and *Xenoglossa* - inhabiting America North of Mexico. Hilgardia 1964;35:375-477.
- Lutz FE, Cockerell TDA. Notes on the distribution and bibliography of North American bees of the families Apidae, Meliponidae, Bombidae, Euglossidae and Anthophoridae. Bull Am Mus Nat Hist 1920;42:491-641.
- Michener CD. Bees of Panama. Bull Am Mus Nat Hist 1954;104:1-176.
- Ordway E, Buchmann SL, Kuehl RO, Shipman CW. Pollen dispersal in *Cucurbita foetidissima* (Cucurbitaceae) by bees of the genera *Apis*, *Peponapis* and *Xenoglossa* (Hymenoptera, Apidae, Anthophoridae). J Kansas Entomol Soc 1987;60:489-503.
- Tepedino VJ. The pollination efficiency of the squash bee (*Peponapis pruinosa*) and the honey bee (*Apis mellifera*) on summer squash (*Cucurbita pepo*). J Kansas Entomol Soc 1981;54:359-77.
- Vergara CH. Diversity and abundance of wild bees in a mixed fruit orchard in Central Mexico. In: Kevan P, Imperatriz Fonseca VL, editors. Pollinating Bees. Brasília: Ministry of Environment; 2002. p.169-76.
- Willis DS, Kevan PG. Foraging dynamics of *Peponapis pruinosa* (Hymenoptera: Anthophoridae) on pumpkin (*Cucurbita pepo*) in southern Ontario. Can Entomol 1995;127:167-75.

Cucurbita

- Comisión nacional para el conocimiento y uso de la biodiversidad (at <http://www.conabio.gob.mx/>)
 - Centro de Investigación Científica de Yucatán (CICY)
 - Escuela Nacional de Ciencias Biológicas (ENCB)
 - Herbario de Arizona
 - Herbario de la Universidad de Texas - Austin
 - Centro de Investigaciones Biológicas del Noroeste (CIBNOR)
 - Herbario IEB del Instituto de Ecología, A.C., México
 - Herbario XAL del Instituto de Ecología, A.C., México
 - Lundell Herbarium (LL) Universidad de Texas - Austin
 - Missouri Botanical Garden
 - New York Botanical Garden

- Dunn-Palmer Herbarium (UMO) University of Missouri
 - Universidad de Sonora (USON)
- The Global Biodiversity Information Facility (at <http://www.gbif.org>)
 - Arizona State University, International Institute for Species Exploration
 - Berkeley Natural History Museums
 - Bioversity International
 - Comisión nacional para el conocimiento y uso de la biodiversidad
 - Herbario del Instituto de Ecología. Mexico
 - Herbario Nacional de México
 - Missouri Botanical Garden
 - Missouri Botanical Garden
 - The New York Botanical Garden
 - University of California, Davis (UCD)
 - Instituto de Biología, Universidad Nacional Autónoma de México (UNIBIO, IBUNAM)
 - Universidad Nacional Autónoma de México - IBiología
 - University of Alabama Biodiversity and Systematics
 - US National Plant Germplasm System
 - US Department of Agriculture PLANTS
 - Utah State University
 - Utah Valley State College (UVSC)
- The Southwest Environmental Information Network (at <http://swbiodiversity.org/seinet/index.php>)
 - Arizona State University Vascular Plant Herbarium
 - California State University. Chico
 - Rancho Santa Ana Botanic Garden
 - San Diego Natural History Museum
 - University of California Riverside
 - University and Jepson Herbaria
 - University of Arizona Herbarium
 - University of California. Irvine
- Arizona-Sonora Desert Museum (at http://www.desertmuseum.org/programs/succulents_table1.html)
- Jepson Herbaria (at <http://ucjeps.berkeley.edu/interchange.html>)
- Missouri Botanical Garden database (at <http://www.discoverlife.org/>)
- The New York Botanical Garden - Virtual Herbarium (at <http://sciweb.nybg.org/science2/vii2.asp>)
- The New York Botanical Garden (at <http://sweetgum.nybg.org/vh/specimen.php?irn=849022>)
- University Herbaria (at http://cms.huh.harvard.edu/databases/specimen_index.html)
- University of Arizona Herbarium (at <http://ag.arizona.edu/herbarium/>)
- University of New Mexico (at <http://abies.nmsu.edu/INRAM/biodiversity/index2.php?Method=simple>)
- USDA (at <http://plants.usda.gov/>)
- Hurd Jr PD, Linsley EG. The squash and gourd bees - genera *Peponapis* and *Xenoglossa* - inhabiting America North of Mexico. *Hilgardia* 1964;35:375-477.
- Hurd Jr PD, Linsley EG. The Mexican squash and gourd bees of the genus *Peponapis*. *Ann Entomol Soc America* 1966;59:835-51.
- Ordway E, Buchmann SL, Kuehl RO, Shipman CW. Pollen dispersal in *Cucurbita foetidissima* (Cucurbitaceae) by bees of the genera *Apis*, *Peponapis* and *Xenoglossa* (Hymenoptera, Apidae, Anthophoridae). *J Kansas Entomol Soc* 1987;60:489-503.
- Walters TW, Decker-Walters DS. Systematics of the endangered okeechobee gourd (*Cucurbita okeechobeensis*: Cucurbitaceae). *Syst Botany* 1993;18:175-87.

CAPÍTULO 4

Ecological Niche Modeling and Principal Component Analysis of *Krameria*
Loefl. (Krameriaceae)

Tereza C. Giannini, Adriana Takahasi, Maria C. M. P. Medeiros, Antonio M.
Saraiva, Isabel Alves-dos-Santos

***Journal of Arid Environment* (2011) v. 75, p. 870-872**
(short communication)

Krameria lanceolata
Foto: University of Texas

**Ecological Niche Modeling and Principal Component Analysis of *Krameria* Loeffl.
(Krameriaceae)**

Tereza C. Giannini¹, Adriana Takahasi¹, Maria C. M. P. Medeiros², Antonio M.
Saraiva³, Isabel Alves-dos-Santos¹

1. Departamento de Ecologia do Instituto de Biociências da Universidade de São Paulo.
Rua do Matão, Trav. 14, 321, Cidade Universitária, São Paulo, SP, Brasil, 05508-900

2. Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo.
Rua do Matão, 277, Cidade Universitária, São Paulo, SP, Brasil, 05508-900

3. Escola Politécnica da Universidade de São Paulo. Av. Prof. Luciano Gualberto, 158,
Cidade Universitária, São Paulo, SP, Brasil, 05508-970

Abstract

Krameria plants are found in arid regions of the Americas and present a floral system that attracts oil-collecting bees. Niche modeling and multivariate tools were applied to examine ecological and geographical aspects of the 18 species of this genus, using occurrence data obtained from herbaria and literature. Niche modeling showed the potential areas of occurrence for each species and the analysis of climatic variables suggested that North American species occur mostly in deserted or xeric ecoregions with monthly precipitation below 140mm and large temperature ranges. South American species are mainly found in deserted ecoregions and subtropical savannas where monthly precipitation often exceeds 150mm and temperature ranges are smaller. Principal Component Analysis (PCA) performed with values of temperature and precipitation showed that the distribution limits of *Krameria* species are primarily associated with maximum and minimum temperatures. Modeling of *Krameria* species proved to be a useful tool for analyzing the influence of the ecological niche variables in the geographical distribution of species, providing new information to guide future investigations.

Keywords: geographical distribution, climatic variables, neotropical flora, oil plants

Krameria Loefl. (Krameriaceae) is a genus comprised of 18 species of herbs and shrubs that present root-hemiparasitism. The genus is largely restricted to arid regions of the neotropics with Mexico and Brazil harboring eleven and five species respectively (Simpson 1989). Traditional taxonomy has suggested one radiation event giving rise to a single North American and a single South American clade (Simpson 1989). However recent molecular evidence suggested a more complicated evolutionary history involving at least two independent intercontinental dispersal or vicariance events leading to at least two new lineages (Simpson et al. 2004).

The entire genus exhibits an obligatory relationship with *Centris* bees (Centridini, Apidae) to pollination. Specific pollinators are attracted by oils produced by specialized floral glands (Vogel 1974, Simpson 1989). The single seeded fruits are spiny and animal dispersed (Simpson 1989).

In this study we have attempted to consolidate distributional information of the genus in order to examine the role that climate might play in controlling biogeographic patterns within *Krameria*. We created a matrix of over 2,742 occurrences of *Krameria* gleaned from both the botanical collections record and literature (Appendix 1 electronic version only). We used Principal Component Analysis (PCA) to discern climatic variables potentially important in the geographic organization of *Krameria* and utilized the entire matrix in a niche model to estimate the area of potential occupancy of each species (Stockwell and Peters 1999) applying climatic and environmental variables (maximum and minimum temperatures of 12 months, precipitation of 12 months and altitude) extracted from publicly available databases (Appendix 2 electronic version only). Finally, we discuss our results in light of current understanding of species distributions.

The first two principal components obtained with the PCA explained 57 and 19% of the variation respectively, totaling 75.76%. The climatic variables most associated with the distribution of species along axis 1 of the PCA diagram were minimum temperature in November, maximum temperature in November, and the minimum temperature in May (loading values approximately 0.49 for each). Axis 2 was a contrast between maximum temperature in May (loading value = 0.67) and precipitation in November (loading value = -0.60) (Figure 1).

Figure 1. Ordination diagram of PCA for the points of occurrence of 18 species of *Krameria* and six climatic variables. Group A = *Krameria* species that occur in South America (except Andes); group B = *Krameria* species that occur in North America and Andes. Where **a**: *K. argentea*; **b**: *K. bahiana*; **c**: *K. cistoidea*; **d**: *K. cytisoides*; **e**: *K. erecta*; **f**: *K. grandiflora*; **g**: *K. grayi*; **h**: *K. lanceolata*; **j**: *K. lappacea*; **m**: *K. paucifolia*; **n**: *K. revoluta*; **o**: *K. sonoreae*; **p**: *K. secundiflora*; **r**: *K. ramosissima*; **s**: *K. spartioides*; **v**: *K. ixine*; **x**: *K. pauciflora*; **w**: *K. tomentosa*.

North American species (*K. erecta*, *K. grayi*, *K. paucifolia*, *K. ramosissima*, *K. sonoreae*, *K. cytisoides*, *K. lanceolata* and *K. secundiflora*) were mainly associated with the positive side of axis 1 while the ‘Andean’ species (*K. cistoidea* and *K. lappacea*) polarized the negative side of this axis. Both these species (North American and Andean ones) polarized the negative side of axis 2. North American species suffer the highest annual temperature fluctuations. Minimum temperatures in the coldest months (November to February) can reach negative values and the warmest months (May to July) can reach a stifling 38°C. In contrast, Andean species presented the lowest annual temperature range (Figure 2).

Excepting the Andean ones, South American species (*K. argentea*, *K. bahiana*, *K. grandiflora*, *K. spartioides* and *K. tomentosa*) were associated with the positive side of axis 2 and occur in areas where maximum temperatures obtained for the coolest months (May to July) can reach 30°C, defining more moderate annual temperature fluctuations

from those experienced by their North American counterparts. Furthermore, South American species occur in more seasonal habitats with greater annual precipitation values (Figure 2).

Figure 2. Variables of species occurrence areas: latitude of occurrence points; ecoregion (1. Desert and xeric, 2. Dry Forest, 3. Montane grassland and shrubland, 4. Subtropical Coniferous Forests, 5. Temperate Coniferous Forests, 6. Mediterranean Forests, 7. Temperate Savannas, 8. Subtropical Savannas – according Olson et al. 2001), average altitude and ranges of precipitation and temperature.

Most of the models showed potential areas that were congruent with the occurrence points obtained through the botanical collections record and literature, suggesting that the environmental variables considered during modeling are, in fact, important in defining the distributional limits of *Krameria* species (Appendix 3 electronic version only).

For example, the North American species *K. lanceolata* (Appendix 3.4 electronic version only) presented a clear disjunction within the temperate broadleaf forest of southeastern U.S., a pattern already signaled by Simpson (1989). This same pattern has been well-documented for other plant populations and is considered to be an artifact of Pleistocene glacial cycles (Zomlefer et al. 2006). A similar distribution of some species of *Centris* bees in the same region was also reported (Simpson 1989, Moure et al. 2008), suggesting that current distributions may be due to historical events and ecological interactions with specific pollinators.

Simpson (1989) also suggested a disjunctive population of *K. cytisoides* on southern Puebla and Oaxaca (Appendix 3.1 electronic version only), a notion fully supported by

our model. The relatively higher Trans-Mexican Volcanic Belt pine-oak forests of central-eastern Mexico (according Olson et al. 2001) correspond to a small area of non-occurrence in the model. This relatively young, east-west running mountain belt may effectively divide two *K. cytisoides* populations, representing a modern day vicariance event for this, and other similarly distributed plant populations.

On the other hand, several species presented wider modeled distributions not corroborated by the botanical evidence. The list includes both North American (*K. erecta* and *K. secundiflora* - respectively Appendix 3.2 and 3.8 electronic version only) and South American (*K. grandiflora*, *K. spartioides*, *K. tomentosa* and *K. lappacea* - respectively Appendix 3.14, 3.15, 3.16 and 3.18 electronic version only) species. Hutchinson (1957) suggested that the realized niche of a given species may be constrained by a variety of historical events, dispersal limitations, and ecological interactions. Indeed this may be the case for *Krameria*, especially considering their specific pollinator interactions and hemiparasitic habit that may cause them to be highly sensitive to other environmental variables not considered in the model. Little is known of how soil type, luminosity, host, and vegetative as well as animal communities affecting *Krameria* species distributions.

For example, *K. lappacea* (Appendix 3.18 electronic version only) showed a wide modeled distribution extending into south-central Argentina, for which there are no available points of occurrence in the literature. Moure et al. (2008) reported the presence of *Centris* bees only as far south as northwest Argentina to the states of Entre Ríos and La Pampa. In this case, a more extensive distribution reaching central Argentina may be inhibited by the absence of specific pollinators rather than climatic factors involved in modeling.

Alternatively, what we consider to be model overpredictions for these species may simply be the lack of botanical collections in difficult-to-access areas (Anderson et al. 2003), especially for the South American species. For example, *K. grandiflora* (Appendix 3.14 electronic version only) presented a pattern of points of occurrence and potential areas similar to that found by Simpson (1989), but included areas of deserted ecoregions of northeastern Brazil in addition to subtropical savannas from central Brazil to Paraguay. The modeling also showed a potential area of occurrence covering the upper Amazon of northern Bolivia and Peru where collections of the species have been severely limited. Similarly, the distribution model of *K. tomentosa* (Appendix 3.16

electronic version only) suggested a potential range spanning northern Argentina to northern Amazonian Peru. Much of this area, however, remains comparatively poorly collected. As these Amazonian areas present few available data on biodiversity, it is difficult to be sure about the presence or absence of both species, which would justify a survey in these areas.

The modeling of *Krameria* species distribution has contributed basic information concerning this poorly known genus, especially South American species. The congruence obtained between most of the modeled distributions and the botanical collections suggested that the environmental and climatic variables analyzed are indeed important in determining biogeographic patterns of *Krameria*. Temperature played a key role in the genus' geographic organization, suggesting further studies concerning the potential impact of climatic change on this group as well as their specialist pollinators (Mommott et al. 2007).

On the other hand, some species with fairly restricted distributions in the botanical record were found to have extensive modeled distributions, suggesting sensitivity to other factors that shape their biogeographic patterns. For example, in South American subtropical savanna there is high species richness of *Centris* bees (Alves dos Santos 2009) and in this area, species of Malpighiaceae have been reported. Some of these species are oil producing and also depends on *Centris* bees to pollination, probably competing with *Krameria* for the same pollinators. Analyses of pollinator interactions and its role in the geographic limits of these species could be important parameters for a more refined modeling in the future.

Acknowledgments

We are grateful to Prof. Dr. José Rubens Pirani from the Department of Botany of the University of São Paulo whose course of Biogeography of Vascular Plants inspired us to conduct this study. We are grateful also to Prof. Sergio Tadeu Meirelles, Jaime Rissi Passarini Jr., two anonymous referees and to The State of São Paulo Research Foundation (FAPESP 04/15801-0, 04/11012-0 and 03/08134-4) and the Coordination for the Improvement of Higher Education Personnel (CAPES - scholarship granted to the first author).

References

- Alves dos Santos, I., 2009. Bees of the Brazilian Savanna, in: Del-Claro, K, Oliveira P.S., Rico-Gray, V. Tropical Biology and Conservation Management - Savannah Ecosystems. Encyclopedia of Life Support Systems (EOLSS), vol. 10. Oxford, UK.
- Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162, 211-232.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology* 22, 415–427.
- Memmott, J.; Craze, P. G.; Waser, N. M.; Price, M. V., 2007. Global warming and the disruption of plant–pollinator interactions. *Ecology Letters* 10, 710–717.
- Moure, J.S., Melo, G.A.R., Vivallo, F., 2008. Centridini Cockerell & Cockerell, 1901, in: Moure, J.S., Urban, D., Melo, G.A.R. (Eds), *Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region - online version*. Available at <http://www.moure.cria.org.br/catalogue>. Accessed Nov/04/2009.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* 51, 933-938.
- Simpson, B.B., 1989. Krameriaceae. *Flora Neotropica*. Monograph 49, 1-109.
- Simpson, B.B., Weeks, A., Helfgott, D.M., Larkin, L.L., 2004. Species relationships in *Krameria* (Krameriaceae) based on ITS sequences and morphology: implications for character utility and biogeography. *Systematic Botany* 29, 97-108.
- Stockwell, D., Peters, D., 1999. The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13, 143-158.
- Vogel, S., 1974. Ölblumen und ölsammelnde Bienen. *Tropische und Subtropische Pflanzenwelt* 7, 285-547.

Zomlefer, W.B., Whitten, W.M., Williams, N.H., Judd, W.S., 2006. Infrageneric phylogeny of *Schoenocaulon* (Liliales: Melanthiaceae) with clarification of cryptic species based on ITS sequence data and geographical distribution. American Journal of Botany 93, 1178-1192.

Appendix 1. Data Sources

Accessed through CONABIO data portal (<http://www.conabio.gob.mx/remib> 30/10/2007)

- Agentes Bioactivos de Plantas Desérticas de Latinoamérica
- Árboles de la Península de Yucatán
- Herbario de Geo. B. Hinton
- Herbario de la Escuela Nacional de Ciencias Biológicas
- Herbario de la Universidad Autónoma de Baja California
- Herbario de la Universidad de Arizona
- Herbario de la Universidad de Sonora
- Herbario de la Universidad de Texas – Austin
- Herbario del Centro de Investigación Científica de Yucatán
- Herbario IEB del Instituto de Ecología
- Herbario XAL del Instituto de Ecología
- w3TROPICOS

Accessed through CRIA data portal (<http://splink.cria.org.br/> 20/10/2007)

- Herbário Antônio Nonato Marques - Empresa Baiana de Desenvolvimento Agrícola
- Herbário da Escola Superior de Agronomia Luis de Queiros da Universidade de São Paulo
- Herbario da Universidade Estadual de Feira de Santana Universidade Estadual de Feira de Santana
- Herbário Dimitri Sucre Benjamin - Jardim Botânico do Rio de Janeiro
- Herbário do Instituto Nacional de Pesquisas da Amazônia
- Herbário Lauro Pires Xavier - Universidade Federal da Paraíba
- JBRJ_RB - Herbário Dimitri Sucre Benjamin Jardim Botânico do Rio de Janeiro
- JPB - Herbário Lauro Pires Xavier - Universidade Federal da Paraíba
- Missouri Botanical Garden
- Museu Botânico Municipal de Curitiba
- The New York Botanical Garden

Accessed through GBIF data portal (<http://data.gbif.org/> 31/10/2007)

- Arizona State University Vascular Plant Herbarium
- Australian National Herbarium (CANB)
- Biodiversidad de Costa Rica
- Bishop Museum Natural History Specimen Data
- Database Schema for UC Davis [Herbarium Labels]
- Fairchild Tropical Botanic Garden Virtual Herbarium Darwin Core format Harvard University Herbaria
- herbario
- Herbario de la Universidad de Sevilla, SEV-Historico
- Herbario del Instituto de Ecología, A.C., México (IE-XAL)
- Herbarium Universitat Ulm
- Herbarium WU
- Herbier de la Guyane
- Missouri Botanical Garden
- Museu Botanico Municipal
- NatureServe Network Species Occurrence Data
- Phanerogamie
- SANT herbarium vascular plant collection
- SERNEC - University of North Carolina at Chapel Hill - Plants
- SysTax
- The Deaver Herbarium, Northern Arizona University
- United States National Plant Germplasm System Collection
- University and Jepson Herbaria DiGIR provider

- USDA PLANTS Database
- USU-UTC Specimen Database
- Utah Valley State College Herbarium
- Vascular Plant Type Specimens

Accessed through New Mexico Biodiversity Collections Consortium (<http://nmbiodiversity.org/> 28/10/2007)

- Center for Natural History Collections Herbarium
- Natural History Collection
- Museum of Southwestern Biology Herbarium
- Gila Center for Natural History

Accessed through Southwest Environmental Information Network (<http://swbiodiversity.org/seinet/> 30/10/2007)

- ARIZ - University of Arizona Herbarium
- ASC - Deaver Herbarium (Northern Arizona University)
- ASU - Arizona State University Vascular Plant Herbarium
- COCHISE - The Cochise County Herbarium
- CS - Colorado State University Herbarium
- DES - Desert Botanical Garden Herbarium Collection
- ENMU Natural History Collection Herbarium
- GCNP - Grand Canyon National Park
- MABA - Madrean Archipelago Biodiversity Assessment Observations
- NAVA - Navajo Nation Herbarium
- NMC - NMSU Center for Natural History Collections Herbarium
- NMCR - NMSU Center for Natural History Collections Range Science
- NMSU Center for Natural History Collections Range Science
- NY - New York Botanical Garden
- SNM - Gila Center for Natural History (WNMU) Herbarium
- TES - US Forest Service Southwestern Region - TEUI Herbarium
- UCR - University of California, Riverside Plant Herbarium
- UNM - University of New Mexico Herbarium
- USON - Herbario de la Universidad de Sonora
- UTC - Intermountain Herbarium

Other data portal

- Utah State University Herbarium database (accessed through Discover Life <http://discoverlife.org/> 29/10/2007)
- The Kansas State University Herbarium (accessed through Humboldt State University Herbarium <http://sorrel.humboldt.edu/~herb/links.html> 30/10/2007)

Brazilian Herbaria (São Paulo State) (personal visit)

- Herbário do Instituto de Biociências da Universidade de São Paulo
- Herbário do Instituto de Botânica de São Paulo

Academic literature

- Simpson, B.B., 1989. Krameriaceae. Flora Neotropica. Monograph 49, 1-109.

Appendix 2. Material and Methods

To create the database on the geographic distribution of *Krameria*, the species points of occurrence were searched on Internet websites that are specialized in biodiversity (Appendix 1. Data sources). The occurrence points presented in Simpson (1989) were also included in the data matrix, with one modification: Simpson et al. (2004) separated *K. grayi* into two species (*K. grayi* and *K. sonorae*) and because of this, their occurrence points were not considered.

The coordinates were used to perform the ecological niche modeling through GARP algorithm with best subsets (Desktop GARP implementation) that is found in the computer system openModeller (version 1.0.9.) (Santana et al. 2008, Muñoz et al. 2009). GARP is an algorithm that produces logical models based on nonrandom associations between points of occurrence of species and sets of georeferenced environmental layers. This algorithm has been widely used by several authors for ecological niche modeling (Stockwell and Peterson 2002, Anderson et al. 2003) and proved to be a reliable method for modeling based on the use of sets of data originating from biological collections (Stockwell et al. 2006).

The accuracy of the model and the area under the receiver operating characteristic curve (AUC) have been applied to estimate the quality of the models (Phillips et al. 2006). Two tests based on AUC were performed: an internal test using all the points of occurrence obtained, and an external test (Araújo et al. 2005) (Table 1). The data were divided into partitions, randomly and without replacement, and each partition was divided into test data and training data. The limit proposed by Stockwell and Peterson (2002) was used, which states that the training data must contain at least 20 points of occurrence. The proportion of number of training points and number of test points was 70% and 30%, respectively. The average and standard deviation of AUC values obtained for each data partition of each species were calculated. Swets (1988) suggested that the results of the mean AUC could be interpreted as follows: excellent, when above 0.90, good between 0.90-0.80; fair between 0.80-0.70; poor, between 0.70-0.60; and fail, between 0.60-0.50.

For all models, 37 layers of environmental variables (5 minutes of arc resolution) obtained from Worldclim (<http://www.worldclim.org>) were used: maximum and minimum temperatures for 12 months, precipitation for 12 months and altitude (Hijmans et al. 2005).

As a resulting product of modeling through openModeller, a matrix with the environmental variables attributed to each occurrence point is obtained. These values were used to analyze the climatic parameters and to conduct the PCA (Principal Component Analysis). To implement the PCA, the original matrix was randomly reduced to 10 points for each species and their data were centralized and standardized. As several variables had high correlation coefficients (values above 0.90), some were excluded from the analysis by selecting only one among the higher correlated pairs. Thus, a PCA matrix was obtained with 176 points of occurrence and six environmental variables: precipitation, maximum and minimum temperatures of the months of November and May. In order to obtain a more easily interpretable configuration of the variables on the PCA axes, the rotation of the data *a posteriori* (Varimax) was performed, using the software SPSS v.13.0 (SPSS for Windows, Chicago, Illinois, USA).

Table 1. Accuracy and AUC resulted from internal and external tests of modeling

Species	Internal test			External test						
	Total number of points	Accuracy (%)	AUC	Number of sets	Number of points	AUC average	Standard deviation	Number of points	AUC average	Standard deviation
1. <i>K. cytoides</i> Cav.	134	98.5	0.99	3	12	1	0	30	0.980	0.010
2. <i>K. erecta</i> Willd. ex Schult.	762	96.4	0.97	18	12	1	0	30	0.997	0.006
3. <i>K. grayi</i> Rose & Painter	512	96.7	0.98	12	12	1	0	30	0.997	0.007
4. <i>K. lanceolata</i> Torr.	467	97.9	0.98	11	12	0.999	0.003	30	0.997	0.006
5. <i>K. pauciflora</i> ex DC.	81	92.7	0.99	1	12	1	-	30	1	-
6. <i>K. paucifolia</i> Rose	7	71.4	1	-	-	-	-	-	-	-
7. <i>K. ramosissima</i> (A.Gray) S.Watson	66	95.5	1	1	12	1	-	30	1	-
8. <i>K. secundiflora</i> ex DC.	47	93.6	0.97	1	12	1	-	30	1	-
9. <i>K. sonorae</i> Britton	19	94.7	1	-	-	-	-	-	-	-
10. <i>K. ixine</i> L.	173	98.8	0.99	4	12	1	0	30	0.97	0
11. <i>K. revoluta</i> O. Berg	83	97.5	1	1	12	1	-	30	0.97	-
12. <i>K. argentea</i> Mart. ex Spreng.	45	97.8	1	1	12	1	-	30	1	-
13. <i>K. bahiana</i> B. Simpson	17	100	1	-	-	-	-	-	-	-
14. <i>K. grandiflora</i> A. St.-Hil.	37	94.4	0.97	1	11	1	-	26	0.97	-
15. <i>K. spartioides</i> Klotzsch ex O.Berg	27	96.3	1	-	-	-	-	-	-	-
16. <i>K. tomentosa</i> A. St.-Hil.	172	98.3	0.96	4	12	1	0	30	0.980	0.010
17. <i>K. cistoidea</i> Hook. & Arn.	21	100	1	-	-	-	-	-	-	-
18. <i>K. lappacea</i> (Dombey) H.M.Burdet & B.B.Simpson	72	100	0.98	1	12	0.98	-	30	0.97	-

References

- Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162, 211-232.
- Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. *Global Change Biology* 11, 1504-1513.
- Hijmans, J.R., 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.
- Muñoz, M.E.S., Giovanni, R., Siqueira, M.F., Sutton, T., Brewer, P., Pereira, R.S., Canhos, D.A.L., Canhos, V.P., 2009. openModeller: a generic approach to species' potential distribution modeling. *Geoinformatica*. DOI 10.1007/s10707-009-0090-7.
- Philips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231-259.
- Santana, F.S., Siqueira, M.F., Saraiva, A.M., Correa, P.L.P., 2008. A reference business process for ecological niche modeling. *Ecological Informatics* 3, 75-86.
- Simpson, B.B., 1989. *Krameriaceae*. *Flora Neotropica*. Monograph 49, 1-109.
- Simpson, B.B., Weeks, A., Helfgott, D.M., Larkin, L.L., 2004. Species relationships in *Krameria* (*Krameriaceae*) based on ITS sequences and morphology: implications for character utility and biogeography. *Systematic Botany* 29, 97-108.
- Stockwell, D.R.B., Beach, J.H., Stewart, A., Vorontsov, G., Vieglais, D., Pereira, R.C., 2006. The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological Modelling* 195, 139-145.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148, 1-13.

Appendix 3. Areas of potential occurrence of *Krameria* species resulting from ecological niche modeling

CAPÍTULO 5

Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants

Tereza C. Giannini; Daniel S. Chapman; Antonio M. Saraiva; Isabel Alves-dos-Santos; Jacobus C. Biesmeijer

Ecography (submetido)

Bombus bohemicus
Foto: J. K. Lindsey

Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants

Tereza C. Giannini¹; Daniel S. Chapman²; Antonio M. Saraiva³; I. Alves-dos-Santos¹;
Jacobus C. Biesmeijer⁴

1. Bioscience Institute. University of Sao Paulo. R. do Matao, 321. 05508-900. Sao Paulo, Brazil.

2. Centre for Ecology and Hydrology. Bush Estate, Edinburgh EH10 4HT, UK.

3. School of Engineering. University of Sao Paulo. Av. Prof. Luciano Gualberto, 380. 05508-970. Sao Paulo, Brazil.

4. NCB Naturalis PO Box 9517, 2300 RA Leiden, Netherlands.

Abstract

Biotic interactions have been considered as an important feature in species distribution modeling, but little is known about the effect of including different types of interactions or performing different strategies of integrating biotic interactions in species distribution modelling. This study compares different methods for including interspecific interactions in species distribution models for bees, their cleptoparasites, and the plants they pollinate. Host-parasite interactions among bumble bees (genus *Bombus*: generalist pollinators and brood parasites) and specialist plant-pollinator interactions between *Centris* bees and *Krameria* flowers were used as case studies. We used 7 different modelling algorithms available in the BIOMOD R package. Adding biotic information to present day predictions of potential occurrence significantly improved the cross-validated area under the receiver operating characteristic curve (AUC), a measure often applied to estimate model accuracy. Different species and types of interaction showed different AUC results in line with data quality, level of biological linkage and interdependence of each interaction. Furthermore, the species that presented the best improvement of AUC was projected under future climate scenarios. The results showed marked differences when using abiotic data only, and when including biotic interactions of various types. The results show that choosing the correct biotic

information, based on biological and ecological knowledge, is critical to improve the accuracy of species distribution models, and also to forecast distribution change.

Keywords

Species distribution modelling, species interactions, host-parasite, plant-pollinator, climate change, BIOMOD, *Bombus*, *Centris*, *Krameria*

Introduction

Species distribution modelling has been used as a computational tool with multiple objectives, including predicting the impacts of global change, especially global warming (Franklin 2009). This technique combines occurrence data with environmental variables based on a correlative approach to build a representation of a species' ecological requirements (Anderson et al. 2003). Several algorithms have been applied to create such models, and their computational results can show areas that are similar to those where the species is known to occur and thus may represent its potential areas of occurrence (Stockwell and Peters 1999, Phillips et al. 2006). Future scenarios can be projected assuming that current climatic requirements of a species remain unchanged under climate change (Thuiller et al 2005).

Recently, the inclusion of biotic information as input in species distribution models (SDM) has been advocated as a way of representing effects of interspecific interactions (e.g. competition, predation, parasitism, mutualism) on species distributions. For example, plant–insect interactions (Jiménez-Valverde et al. 2008) and specialized plant–pollinator interactions (Godsoe et al. 2009, Giannini et al. 2010, Warren et al. 2010) have been modelled. In most cases incorporating biotic interactions significantly improved the models of the focal species' present day distribution and altered the projections of its future distribution (Araujo and Luoto 2007, Meier et al. 2010, Van der Putten et al. 2010).

Once the biotic information has been selected it can be incorporated in the modelling exercise in different ways. First, the actual points of occurrence of the interacting species might be added directly as a predictor variable for the focal species. This might make sense when the biotic component has an obligate relationship with the focal

species. Alternatively, the estimated occurrence or habitat suitability of the interacting species (i.e. its SDM prediction) can be used as input layer in the focal species' models. To our knowledge, a comparative analysis considering both options has not been performed.

The potential value of biotic information to species distribution models should depend on the relationship between the focal species and the biotic component. The occurrence of an obligate parasite, for example, can be taken as good evidence for the presence of its host. While the presence of a single food plant species of a generalist forager merely points to the potential occurrence of that forager species. It is even possible that information on a non-interacting species can improve the distribution model if both species happen to be associated with similar environmental features (e.g. soil, climate, microhabitat). However, previous studies have only compared abiotic SDMs to biotic models that incorporate known interacting species (Heikkinen et al. 2007, Pellissier et al. 2010). As such, it is not clear whether the inclusion of interacting species improves the model because it actually represents interspecific interactions or whether the improvement simply arises through the inclusion of an additional spatially structured predictor.

In this paper, we tested to what extent the modeling methods, trophic position, and level of ecological dependence actually change the SDM of the focal species. We chose two different groups of interacting organisms: one group of host-parasite pairs and one group of specialized plant-pollinator species. The host-parasite group was comprised of species of bumblebees and their conspecific brood parasites (both *Bombus* spp.) that are common throughout Europe and Asia and tend to have a broad floral diet (Benton 2006). The specialized plant-pollinator group was comprised of species of *Krameria* plants (Krameriaceae) and *Centris* bee pollinators - two genera with a Neotropical distribution. *Krameria* plants are pollinated only by *Centris* bees that collect floral oil to raise their offspring (Vogel 1974, Simpson 1989).

The main objectives of this work were 1) to compare the fit of species distribution models with and without biotic data included in different ways; 2) to compare the effect of including biotic data on interacting species pertaining to different trophic positions or ecological dependence; 3) to analyze to what extent projections under climate change depend on modelling methods and the way biotic interactions are included.

Material and Methods

The selected species of *Bombus*, representing host-parasite pairings, were: 1) *B. terrestris* (host) and *B. vestalis* (parasite), and 2) *B. lucorum*, *B. cryptarum* and *B. magnus* (all hosts) and *B. bohemicus* (parasite of all three hosts). Analyzes were focused on Great Britain for which good distribution records were available.

The selected species of bees and plants representing pollinator-plant interactions were: 1) *Krameria erecta* (plant); *Centris atripes* and *C. rhodopus* (pollinating bees) all occurring in western areas of Mexico and USA, and 2) *K. tomentosa* (plant) and *C. fuscata* and *C. tarsata* (pollinating bees) occurring in northeastern and central areas of Brazil.

To compensate for variation in data sampling density between different areas, we created a subsample of our original dataset by overlaying a 10x10 km grid, sampling only one species presence point from each cell (Godsoe et al. 2009). This procedure was performed with ArcGIS 9.3 (ESRI Inc.). The details about data sources used to perform the models can be found as Supplementary Material.

The BIOMOD package (Thuiller 2003) for R 2.11.1 (The R Foundation for Statistical Computing) was applied to model present day and future species distributions. We created pseudo-absence points for each Neotropical species using the BIOMOD “circles” strategy. With this strategy, a circle is drawn individually around each presence point with a radius of 5° and pseudo-absence points selected from this region. The total number of pseudo-absences per species was set so that the number of occurrences and pseudo-absences summed to 10000 (Van der Wal et al. 2009, Lobo and Tognelli 2011). For Britain, where biological recording of bumblebees is good, we selected occurrences reported since the year 1990 and considered the cells with no reported occurrence as true absence points. The original databases were randomly split in calibration (80%) and evaluation subsets (20%) (Fielding and Bell 1997).

Using BIOMOD, we applied 7 model classes: artificial neural networks (ANN) (Ripley 1996), classification tree analysis (CTA) (Breiman et al. 1984), generalised linear models (GLM) (McCullagh and Nelder 1989), generalised boosted models (GBM) (Ridgeway 1999), flexible discriminant analysis (FDA) (Hastie et al. 1994), multivariate adaptive regression splines (MARS) (Friedman 1991) and random forest (RF) (Breiman 2001). We used the cross-validated area under the receiver operating characteristic

curve (AUC) to evaluate the models. Swets (1988) suggested that the results of AUC could be interpreted as follows: excellent, when above 0.90; good, between 0.90-0.80; fair, between 0.80-0.70; poor, between 0.70-0.60; and fail, between 0.60-0.50.

We used 19 bioclimatic variables with a grid resolution of 5 arc-minutes to model the species' geographical distributions. These represent gradients in average temperature and precipitation, over diurnal, seasonal and annual timescales (Hijmans et al. 2005). The same variables were used for predicted future climate conditions in 2050, assuming a moderate climatic change scenario (Special Report on Emission Scenarios B2) (Ramirez and Jarvis 2008) and the HADCM3 model (Gordon et al. 2000). In modelling the present day species distributions, we also used land use information for the year 2000 (Global Land Cover 2000 database) that was converted to 10km resolution. Because many of the environmental variables were highly correlated we applied principal component analysis (PCA) using R 2.11.1 (The R Foundation for Statistical Computing) (Powney et al. 2010). We identified six independent axes of variables representing 80% of the total environmental variation in Britain, 97% in North America and 92% in South America.

We added the information about biotic interaction as a supplementary layer to the six PCA variables following different strategies. First, we performed modelling with only the abiotic variables to obtain the probability of occurrence of each species. Second, we created new models including the biotic information in two different ways: (A) We included the actual presence-absence of the biotic component as a supplementary explanatory variable (predictor) to modelling the focal species; (B) We included the occurrence probability of interacting species from abiotic SDMs as a supplementary predictor. Where the modeled species had more than one interacting species (e.g. *B. bohemicus* with its hosts and *Krameria* with its pollinators), we fitted individual models for each interacting species, as well as models in which all interacting species were included individually, and models based on the richness of the interacting species. Where models were based on occurrence probabilities, a richness index was derived as the sum of probabilities for the interacting species.

As well as modeling effects of interacting species, we repeated the modeling exercise specifying non-relevant interactions in the SDMs, i.e. we included interactions between species that do not strongly interact via specialist host-parasite or plant-pollinator relationships. In this way, we aimed to test whether the AUC variation was due to

increased model complexity or statistical artifacts of spatially structured predictors. It was not possible to apply this procedure to *Krameria* species because the different interacting species groups occupied distinct geographic regions. Thus, we considered non-relevant interactions between parasitic *Bombus* species and hosts they did not use, and interactions between regionally co-occurring *Centris* bees. Differences in the resulting AUC evaluations of the abiotic and different biotic models were tested using paired *t* tests.

We also analyzed the effect of including biotic data on model projections with future climate conditions. Future land use scenarios were not available to us, and so for this exercise we developed present day models based on just the climate and interspecific interactions. For brevity, we only forecasted the distribution of the species for which biotic interactions had the largest effect on the SDMs for present day conditions (*B. bohemicus*). To obtain future projections, we applied the ensemble forecasting strategy to these models (Araújo and New 2007, Thuiller et al. 2009). With BIOMOD ensemble forecasting, the final model is a consensual result of all models used (in our case, the same seven models described before), weighted on the basis of their predictive accuracy on the test data. Using this function, one could choose also different kinds of measures of accuracy and we have chosen AUC to calculate the weights.

Results

Including relevant interactions in the SDMs for *Bombus* species nearly always significantly increased their fit, as measured by the AUC score (Fig. 1A). Furthermore, in every case the effect on the parasite species was stronger than that on the host. For the parasite with multiple hosts (*B. bohemicus*), there was little difference between models that used all three hosts individually, used host richness or only used the most widespread host (*B. lucorum*), but inclusion of the rarer hosts resulted only in slight model improvement. Results from models including non-relevant interactions between species not engaged in host-parasite relationships were mixed (Fig. 1B). Unexpected significant improvements in AUC were obtained for three of the eight models, while significant decreases in AUC were obtained in two cases. Indeed, including a non-host of *B. vestalis* (*B. lucorum*) gave a similar model improvement seen when the correct host was included.

On the other hand, the inclusion of biotic interactions in models for the Neotropical plant and pollinator species resulted in almost no significant AUC variation (Figure 1C-D). Significant improvement was observed only in one case when including relevant interactions (*K. erecta* including *C. atripes*) and when including non-relevant species (*C. fuscata* with *C. tarsata*).

Figure 1. Summary of the effects of including (A) expected actual and (B) non-expected hypothetical interactions between host and brood parasite *Bombus* bumblebees in species distribution models for Great Britain. Interactions were modelled by including the presence/absence of the other species as a predictor alongside six abiotic gradients. Arrows point towards the modelled species and are labelled with the change in average cross-validated AUC caused by including the biotic interaction over the seven modelling algorithms. Bold labels show statistically significant differences ($P < 0.05$ in paired t-tests). For *B. bohemicus*, which has three host species, host richness and all species were also included as a predictor. (C and D) Specialist plant-pollinator interactions between *Centris* bees and *Krameria* flowers. Here

species interactions are included by using the abiotic-only distribution model scores for the interactor species as a predictor variable. For both *Krameria* species, richness and both pollinators were included as a predictor. Because of geographical separation between the two species groups, it was not possible to test for unexpected interactions other than between co-occurring pollinators.

The best models for each of the species showed high fits (cross-validated AUC range 0.74-0.98). The resulted average of each variable importance to each model (7 models quoted) was converted to percentage and compared (Figure 2A and B). This measure is a function of the correlation scores between the original prediction and the prediction made with a permuted variable (Thuiller et al. 2009a). The results to host-parasite species showed that the biotic component was the most important variable in the SDMs of focal species and thus more important than any single climate or environmental variable (i.e. PCA axis) (Figure 2A). Considering plant-pollinator (Figure 2B), the average variable importance of interacting species was also higher than PC1 scores, but the results showed higher variation considering PC1 and interacting species.

Figure 2. Variable importance of each principal component (abiotic features) and interacting(s) species (biotic feature). Top: Host-parasite; Bottom: Plant-Pollinator.

Model projections of distributions in 2050 with moderate climate change were sensitive to the inclusion of interspecific interactions. For *B. bohemicus* (the species with the highest improvement in AUC for biotic models), abiotic models and biotic models with appropriate and inappropriate interactions all give similar representations of the present-day distribution (Figure 3a-d). However, projections under climate change differ markedly. The abiotic model and the biotic model based on a non-existent interaction both predict that *B. bohemicus* will retreat into its present-day range core in Scotland (northwest Britain) (Figure 3e and h). By contrast, both biotic models suggest the species will maintain or expand its Scottish distribution and also expand into Wales (southwest Britain) (Figure 3f-g).

Figure 3. Projections of *Bombus bohemicus* occurrence areas. Top: present climatic conditions using a) abiotic information only; b) *B. lucorum*; c) richness of its three host species (*B. lucorum*, *B. magnus* and *B. cryptarum*); d) non-relevant interacting species (*B. terrestris*). Bottom: 2050 climatic conditions using e) abiotic information only; f) *B. lucorum*; g) richness of its three host species (*B. lucorum*, *B. magnus* and *B. cryptarum*); h) non-relevant interacting species (*B. terrestris*). White areas in the future projection maps show the regions where one or more of the climatic PCA gradients are outside the present day range, and on which projections may be unreliable.

Discussion

We found that including interspecific interactions of known relevance to the focal species, resulted in large increases in model fit compared to abiotic models for host-parasite species pairs, but had little effect for plant-pollinator species. This could be due to better data quality in Great Britain than in the neotropics. Alternatively, the interactions among *Bombus* species might be more specialized than those between *Krameria* and *Centris*. *Krameria* plants depend more on the *Centris* bees (their only pollinator) than the bees depend on *Krameria* oil rewards, since these bees can obtain their resource from other oil plant producers, such as Malpighiaceae species (Simpson 1989). *Centris* is also a widespread genus and other congeneric species could potentially pollinate the *Krameria* species analyzed here. Thus, our results suggest that stronger and more specialized biotic interactions are likely to be more important than weak and generalized interactions in determining regional-scale distribution patterns.

Our results revealed an improvement in *Bombus*' model accuracy when including known strong interspecific interactions in species distribution models. However, in some cases information on non-relevant species also improved the model. Indeed, when modeling *B. vestalis* the non-host *B. lucorum* gave the same increase in model fit as its true host *B. terrestris*. This might be due to unknown indirect interactions between the species, but is most likely caused by similarity in their relationship with climate and environmental conditions. Therefore, our results show that previous knowledge about species interactions is necessary to establish which biotic information should be added to distribution models and also to correctly interpret the results (i.e. what is the underlying ecological or biological mechanism).

For *Bombus*, the higher trophic level (parasite) always benefited more from including interactions than lower level (host). This may reflect a mainly bottom-up control of host-parasite communities, as suggested by Vazquez et al. (2007) who analyzed trophic networks. One would certainly expect hosts to be able to exist in areas unsuitable for their parasites, but not the other way around. Recent reviews also emphasize that higher trophic levels are often disproportionately affected by drivers such as climate change, competition from invasive species, and habitat modification (Tylianakis et al. 2008, Gilman et al 2010).

We also demonstrated that SDM projected distributions under climate change are sensitive to the inclusion of biotic interactions, as illustrated by our *B. bohemicus* projection to 2050. Analyzing trophic networks, Chen et al. (2008) showed that host species occupying prominent network positions have many interacting partners and the associated parasite species are buffered against fluctuations in host availability; if an interacting partner goes extinct, the parasite can use alternative hosts to continue its life-cycle and survive. Our analysis featured a parasite with few hosts (three species) and so one would expect changes in the distribution of these hosts to be very important in its response to ongoing global change. In any case, our results suggest that a failure to include biotic interactions, or their mis-specification, causes great uncertainty in SDM-based climate change projections. This brings into question earlier studies that predict biodiversity impacts of global change by modeling single-species range shifts without regard for interspecific interactions (Thuiller et al. 2005, Rebelo et al. 2010).

Other works have already reviewed the importance of considering biotic interactions when analyzing impacts of climatic change (Brooker et al. 2007, Hegland et al. 2009, Van der Putten et al. 2010). The analyses addressed here highlight the importance of considering different methods to include this information, and also the importance of reliable knowledge about focal species biology. Our analyzes also support the utility of the “community modules” approach of Gilman et al. (2010) whereby entire food webs are simplified into tractable sub-webs of strongly interacting species. In our study, these modules represent defined host-parasite and plant-pollinator linkages. Ferrier and Guisan (2006) also emphasized that predicting the effects of climate change is dependent upon identifying those interactions between species that are most vulnerable to changing climate and are key determinants of the structure and function of a community.

In this study we considered specialized, presumably quite strong interactions. However, weak interactions dominate communities, and it is not clear whether a species’ response to climate change would be more influenced by a small number of strong interactions *versus* a large number of weak ones. Such weak interactions might be better represented by multivariate distribution model approaches that represent interspecific associations or shared responses to environmental gradients and predict the distributions of multiple species (Baselga and Araújo 2009; Elith and Leathwick 2007). According to Tylianakis et al. (2008) the presence of weak interactions has long been known to promote web

stability and can promote the stability of metacommunities at landscape scales. On the other hand, specialized pollinators, especially ones with small diet breadths, are vulnerable to disruption in the plant community (Memmott et al 2007). As pointed out by Gilman et al. (2010), the key question is not the resulting effects of climate change on individual species, but the stability of the system as a whole.

Conclusions

In conclusion, we found that the distributions of species that have strong and specialized interactions with a focal species are useful predictors in species distribution models. However, including non-relevant interactions can also yield a model improvement, showing that positive correlations between species distribution patterns are not necessarily indicative of interspecific interactions. Furthermore, we also showed how projections of climate change impacts on species distributions are very sensitive to the specification of biotic interactions. Thus, we suggest that a better understanding of species interactions, and how species assemblages face global changes as an integrated system, is needed to understand their resilience to these impacts.

Acknowledgments

The authors are grateful to Brazilian Federal Agency for Support and Evaluation of Postgraduate Education (CAPES 3030-10-5).

References

- Anderson, R. P. et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. – *Ecol. Model.* 162: 211-32.
- Araújo, M. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – *Global Ecol. Biogeogr.* 16: 743-753.
- Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – *TREE* 22: 42-47.
- Baselga, A. and Araújo, M. B. 2009. Individualistic vs community modelling of species distributions under climate change. – *Ecography* 32: 55-65.

- Benton, T. 2006. Bumblebees: the natural history and identification of the species found in Britain. – Collins.
- Breiman, L. 2001. Random forests. – Mach. Learn. 45: 5-32.
- Breiman, L. et al. 1984. Classification and regression trees. – Chapman and Hall.
- Brooker, R. W. et al. 2007. Modelling species' range shifts in a changing climate: the impacts of biotic interactions, dispersal distance and the rate of climate change. – J. Theor. Biol. 245: 59-65
- Chen, H. W. et al. 2008. Network position of hosts in food webs and their parasite diversity. – Oikos 117: 1847-1855.
- Elith, J. & Leathwick, J. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. – Divers. Distrib. 13: 265-275.
- Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393-404.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38-49.
- Franklin, J. 2009. Mapping species distribution. Spatial inference and prediction. – Cambridge.
- Friedman, J. 1991. Multivariate adaptive regression splines. – Ann. Stat. 19: 1-141.
- Giannini, T. C. et al. 2010. Ecological niche modeling and geographical distribution of pollinator and plants: a case study of *Peponapis fervens* (Smith, 1879) (Eucerini: Apidae) and *Cucurbita* species (Cucurbitaceae). – Ecol. Inform. 5: 59-66.
- Gilman, S. E. et al. 2010. A framework for community interactions under climate change. – TREE 25: 325-331.
- Global Land Cover 2000 database. European Commission, Joint Research Centre, 2003. <http://bioval.jrc.ec.europa.eu/glc2000/glc2000.php>.
- Godsoe, W. et al. 2009. Divergence in an obligate mutualism is not explained by divergent climatic factors. – New Phytol. 183: 589-599.

- Gordon, C. et al. 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. – *Clim. Dynam.* 16: 147-168.
- Hastie, T. et al. 1994. Flexible Discriminant Analysis by Optimal Scoring. – *J. Am. Stat. Assoc.* 89: 1255-1270.
- Hegland, S. J. et al. How does climate warming affect plant-pollinator interactions? – *Ecol. Lett.* 12: 184-195.
- Heikkinen, R. K. et al. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. – *Global Ecol. Biogeogr.* 16: 754-763.
- Hijmans, J. R. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965-78.
- Jiménez-Valverde, A. et al. 2008. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. – *Ann. Zool. Fenn.* 45: 200-210.
- Lobo, J. M.; Tognelli, M. F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. – *J. Nat. Conserv.* 19: 1–7.
- McCullagh, P. and Nelder, J. A. 1989. Generalized linear models. – Chapman and Hall.
- Meier, E. S. et al. 2010. Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica* L. – *J. Biogeogr.* doi:10.1111/j.1365-2699.2010.02405.x
- Memmott, J. et al. 2007. Global warming and the disruption of plant-pollinator interactions. – *Ecol. Lett.* 10: 710-717.
- Pellissier, L. et al. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. – *Ecography* 33: 1004-1014.
- Phillips, S. J. et al. 2006. Maximum entropy modelling of species geographic distributions. – *Ecol. Model.* 190: 231-59.
- Powney, G. D. et al. 2010. Synchrony of butterfly populations across species' geographic ranges. – *Oikos* 119: 1690–1696.

- Ramirez, J. and Jarvis, A. 2008. High resolution statistically downscaled future climate surfaces. International Centre for tropical Agriculture, CIAT. Available on <http://gisweb.ciat.cgiar.org/GCMPPage>
- Rebelo H et al. 2010. Predicted impact of climate change on European bats in relation to their biogeographic patterns. – *Glob. Change Biol.* 16: 561-576.
- Ridgeway, G. 1999. The state of boosting. – *Comput. Sci. Stat.* 31: 172-181.
- Ripley, B. D. 1996. Pattern recognition and neural networks. – Cambridge Univ. Press.
- Simpson, B. B. 1989. Krameriaceae. – *Flora Neotropica. Monograph* 49: 1-109.
- Stockwell, D. and Peters, D. 1999. The GARP modeling system: problems and solutions to automated spatial prediction. – *Int. J. Geogr. Inf. Sci.* 13: 143-58.
- Swets, K. A. 1988. Measuring the accuracy of diagnostic systems. – *Science* 240: 1285-1293.
- Thuiller, W. 2003. BIOMOD: optimizing predictions of species distributions and projecting potential future shifts under global change. – *Glob. Change Biol.* 9: 1353-1362.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – *PNAS* 102: 8245–8250.
- Thuiller, W. et al. 2009. BIOMOD: a platform for ensemble forecasting of species distributions. – *Ecography* 32: 369-373.
- Thuiller, W. et al. 2009a. ModOperating Manual for BIOMOD. Laboratoire d'Écologie Alpine and Université Joseph Fourier.
- Tylianakis, J. M. et al. 2008. Global change and species interactions in terrestrial ecosystems. – *Ecol. Lett.* 11: 1351-1363.
- Van der Putten, W. H. et al. 2010. Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. – *Philos. T. Roy. Soc. B* 365: 2025-2034.
- Van der Wal, J. et al. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? – *Ecol. Model.* 220: 589–594.

- Vazquez, D. P. et al. 2007. Species abundance and asymmetric interaction strength in ecological networks. – *Oikos* 116: 1120-1127.
- Vogel, S. 1974. Ölblumen und ölsammelnde Bienen. – *Tropische und Subtropische Pflanzenwelt* 7: 285-547.
- Warren, M. et al. 2010. A comparative approach to understanding factors limiting abundance patterns and distributions in a fig tree - fig wasp mutualism. – *Ecography* 33: 148-158.

1. Supplementary Material. Data Sources

Bombus

Data provider

- National Biodiversity Network (<http://www.searchnbn.net/>) - 10km squares
 - Anglesey Abbey wildlife species data held by The National Trust.
 - Bedfordshire and Luton Biodiversity Recording and Monitoring Centre
 - Bees, Wasps and Ants Recording Society
 - Bristol Regional Environmental Records Centre
 - BTCV Scotland
 - Countryside Council for Wales
 - Derbyshire Wildlife Trust
 - Devon incidental species records 1950-2002
 - Dorset Environmental Records Centre
 - East Ayrshire Countryside Ranger Service
 - EcoRecord
 - Environmental Records Information Centre North East
 - Greater Manchester Ecology Unit
 - Greenspace Information for Greater London
 - Hatfield Forest species data held by The National Trust.
 - Highland Biological Recording Group
 - Ickworth species data held by The National Trust.
 - Invertebrate Site Register - England.
 - Leicestershire and Rutland Environmental Records Centre
 - Lothian Wildlife Information Centre Secret Garden Survey
 - Merseyside BioBank Chris Felton's notebooks
 - Merseyside BioBank Steve White's Dataset
 - Merseyside Environmental Advisory Service Dataset
 - National Trust for Scotland (staff)
 - Norfolk Biodiversity Information Service
 - Records for Fife Nature Records Centre
 - Royal Horticultural Society
 - Scottish Wildlife Trust
 - Shire Group of Internal Drainage Boards
 - Shropshire Ecological Data Network
 - Somerset Environmental Records Centre
 - Staffordshire Ecological Record
 - Suffolk Biological Records Centre
 - Surrey Biodiversity Information Centre
 - Sutton Hoo species data held by The National Trust.
 - SWT Scottish Borders Local Wildlife Site Survey data 1996-2000 - species information
 - Thames Valley Environmental Records Centre
 - Tullie House Museum
 - Welsh Peatland Invertebrate Survey (WPIS)
 - West Wales Biodiversity Information Centre
 - Wicken Fen nature reserve species data held by The National Trust
 - Wiltshire and Swindon Biological Records Centre
 - Worcestershire Biological Records Centre

Krameria

Data providers

- Berkeley and Jepson Herbaria (<http://ucjeps.berkeley.edu/interchange.html>)
- Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (<http://www.conabio.gob.mx/remib>)
 - Herbario de Geo. B. Hinton

- Herbario de la Universidad Autónoma de Baja California
 - Herbario de la Universidad de Arizona
 - Herbario de la Universidad de Sonora
 - Herbario de la Universidad de Texas - Austin
 - Herbario IEB del Instituto de Ecología
 - w3TROPICOS
 - Centro de Referência em Informação Ambiental (<http://splink.cria.org.br/>)
 - Herbário - IPA Dárdano de Andrade Lima
 - Herbário Antônio Nonato Marques
 - Herbário da Escola Superior de Agricultura Luiz de Queiroz
 - Herbário da Universidade de Brasília
 - Herbário da Universidade Estadual de Feira de Santana
 - Herbário da Universidade Estadual de Londrina
 - Herbário da Universidade Federal de Sergipe, Departamento de Biologia
 - Herbário do Departamento de Botânica
 - Herbário do Instituto Agrônomo de Campinas
 - Herbário do Instituto do Meio Ambiente do Estado de Alagoas
 - Herbário do Museu Botânico Municipal
 - Herbário Dr. Roberto Miguel Klein, FURB
 - Herbário Lauro Pires Xavier
 - Herbário Prisco Bezerra - EAC
 - Herbário Sérgio Tavares
 - Herbário Universidade Federal do Rio Grande do Norte
 - Herbário Vale do São Francisco
 - Missouri Botanical Garden - Brazilian records
 - Smithsonian Department of Botany - Brazilian records
 - Global Biodiversity Information Facility (<http://data.gbif.org/>)
 - Arizona State University Vascular Plant Herbarium
 - Tropicos Specimen Database
 - Australian National Herbarium (CANB)
 - Harvard University Herbaria
 - University and Jepson Herbaria DiGIR provider
 - United States National Plant Germplasm System Collection
 - USU-UTC Specimen Database
 - Herbario del Instituto de Ecología, A.C., México (IE-BAJIO)
 - California State University, Chico
 - NMNH Botany Collections
 - New Mexico Biodiversity Collections Consortium database
 - UA Herbarium
 - Botany Vascular Plant Collection
 - Herbarium of The New York Botanical Garden
 - Consortium of California Herbaria
 - EURISCO, The European Genetic Resources Search
 - Royal Botanic Gardens, Kew
 - Herbario del CIBNOR
 - Herbarium de Geo. B. Hinton, México
 - Herbario XAL del Instituto de Ecología, A.C., México (IE-XAL)
 - Herbario de la Universidad de Arizona, EUA
 - Utah Valley State College Herbarium
 - University of Alberta Museums, Vascular Plant Herbarium
 - Institute of Natural Resource Analysis and Management (<http://abies.nmsu.edu/INRAM/>)
 - Southwest Environmental Information Network (<http://swbiodiversity.org/seinet/>)
- Brazilian Herbaria (São Paulo State) (personal visit)
- Herbário do Instituto de Biociências da Universidade de São Paulo

- Herbário do Instituto de Botânica de São Paulo

Academic literature

- Aguiar CML, Zanella FCV, Martins CF, Carvalho CAL. 2003. Plantas visitadas por *Centris* spp. (Hymenoptera: Apidae) na caatinga para obtenção de recursos florais. *Neotropical Entomology* 32 (2):247-259.
- Simpson BB. 1989. Krameriaceae. *Flora Neotropica*. Monograph 49: 1-109.

Centris

Data providers

- Centro de Referência em Informação Ambiental (<http://splink.cria.org.br/>)
 - Coleção Entomológica do Depto. de Sistemática e Ecologia – Universidade Federal da Paraíba
 - Coleção Entomológica Paulo Nogueira-Neto – Instituto de Biociências da Universidade de São Paulo
 - Coleção Entomológica Pe. Jesus Santiago Moure (Hymenoptera) – Universidade Federal do Paraná
 - Laboratório de Ecologia e Biogeografia de Insetos da Caatinga - Universidade Federal de Campina Grande
 - Plebeia - Coleção Entomológica de Polinizadores – Universidade Federal de Pernambuco
- Global Biodiversity Information Facility (<http://data.gbif.org/>)
 - Ohio State University Insect Collection
 - Unidad de Informática para la Biodiversidad (UNIBIO) Instituto de Biología, Universidad Nacional Autónoma de México (IBUNAM)
 - United States Department of Agriculture (USDA) and the Agricultural Research Service (ARS) Bee Biology and Systematics Laboratory
 - University of Kansas Biodiversity Research Center
 - Yale University Peabody Museum

Academic literature

- Aguiar CML, Zanella FCV, Martins CF, Carvalho CAL. 2003. Plantas visitadas por *Centris* spp. (Hymenoptera: Apidae) na caatinga para obtenção de recursos florais. *Neotropical Entomology* 32 (2):247-259.
- Antonini Y, Martins RP 2003. The Flowering-Visiting Bees at the Ecological Station of the Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. *Neotropical Entomology* 32: 565-575.
- Araujo VA, Antonini Y, Araujo APA. 2006. Diversity of bees and their floral resources at altitudinal areas in the Southern Espinhaço Range, Minas Gerais, Brazil. *Neotropical Entomology* 35: 30-40
- Azevedo AA. 2002. Composição de Faunas de Abelhas (Hymenoptera, Apoidea) e da Flora Associada em Áreas de Cerrado de Minas Gerais, Brasil. Universidade Federal de Viçosa. Master thesis.
- Batalha-Filho H, Nunes LA, Pereira DG, Waldschmidt AM. 2007. Inventário da fauna de abelhas (Hymenoptera, Apoidea) em uma área de caatinga da região de Jequié, BA. *Biosci. J.* 23: 24-29.
- Boaventura MC. 1998. Sazonalidade e estrutura de uma comunidade de abelhas silvestres (Hymenoptera: Apoidea) numa área de cerrado do jardim botânico de Brasília, Universidade Nacional de Brasília. Master thesis.
- Boff SV. 2008. Flora de capões e Hymenoptera (Apoidea: abelhas e vespas) visitantes de flores no Pantanal de Miranda-Abobral. Universidade Federal de Mato Grosso do Sul. Master thesis.
- Bortoli C, Laroca S. 1997. Melissocenologia no Terceiro Planalto Paranaense. Abundância relativa das abelhas silvestres (Apoidea) de um biótopo urbano de Guarapuava (PR, Brasil). *Acta Biol. Paranaense* 26: 51-86.

- Carvalho CAL. 1993. Abelhas (Hymenoptera, Apoidea) no município de Cruz das Almas-Bahia: levantamento, identificação e material coletado em plantas de importância econômica. Universidade Federal da Bahia. Master thesis.
- Carvalho CAL, Marques OM. 1995. Abelhas (Hymenoptera, Apoidea) em Cruz das Almas-Bahia. *Insecta* 4: 18-25.
- Carvalho CAL, Santos FM, Silva RF, Souza BA. 2007. Phenology of bees (Hymenoptera: Apoidea) in a transition area between the Cerrado and the Amazon region in Brazil. *Sociobiology* 50: 1177-1190.
- Cure JR, Bastos-Filho GS, Oliveira MJF, Silveira FA. 1993. Levantamento de abelhas silvestres na zona da mata de Minas Gerais - pastagem na região de Viçosa (Hymenoptera – Apoidea). *Revista Ceres* 40: 131-161.
- Freitas RIP. 1998. Abelhas silvestres (Hymenoptera: Apoidea) e a floração de plantas em áreas de cerrado recém-queimadas no distrito federal. Universidade Nacional de Brasília. Master thesis.
- Lima MFC. 2004. Comunidade de abelhas, nidificação de abelhas solitárias em cavidades preexistentes (Hymenoptera: Apoidea) e interação abelha-planta na reserva biológica de Guaribas, Mamanguape, Paraíba, Brasil. Universidade Federal da Paraíba. Doctoral thesis.
- Martins CF. 1990. Estrutura da comunidade de abelhas (Apoidea) na caatinga (Casa Nova BA) e na Chapada Diamantina (Lençóis BA). Universidade de São Paulo. Doctoral thesis.
- Martins CF. 1994. Comunidade de abelhas (Hymenoptera, Apoidea) da caatinga e do cerrado com elementos de campo rupestre do Estado da Bahia, Brasil. Caatinga arbórea aberta/ Cerrado gramíneo-lenhoso com elementos de campo rupestre. *Revista Nordestina de Biologia* 9: 225-257.
- Martins CF. 1995. Flora apícola e nichos tróficos de abelhas (Hymenoptera, Apoidea) na Chapada Diamantina (Lençóis, BA, Brasil). *Revista Nordestina de Biologia* 10: 119-140.
- Michelette ER. 2005. Distribuição local das abelhas. In: Carvalho CM, Vilar JC (ed). Parque Nacional Serra de Itabaiana Levantamento da Biot. Aracaju, Ibama, p. 113-119.
- Milet-Pinheiro P, Schlindwein C. 2008. Comunidade de abelhas (Hymenoptera, Apoidea) e plantas em uma área do Agreste pernambucano, Brasil. *Revista Brasileira de Entomologia* 52 (4): 625-636.
- Mouga DMDS. 2004. As comunidades de abelhas (Hymenoptera, Apoidea) em Mata Atlântica na região nordeste do Estado de Santa Catarina, Brasil. Universidade de São Paulo. Doctoral thesis.
- Parizotto DR. 2007. Biocenótica das abelhas silvestres (Hymenoptera, Apoidea, Anthophila) do Parque Estadual do Caxambú, Município de Castro, Paraná, Sul do Brasil. Universidade Federal do Paraná. Master thesis.
- Santos FM, Carvalho CAL, Silva RF. 2004. Diversidade de abelhas (Hymenoptera: Apoidea) em uma área de transição Cerrado-Amazônia. *Acta Amazônica* 34: 319 – 328.
- Santos MSS. 2003. Relação abelha-flor em comunidades de restinga tropical: sistemas polinizadores de flores melitófilas especializadas. Universidade Federal da Bahia. Master thesis.
- Silveira FA. 1989. Abelhas silvestres (Hymenoptera, Apoidea) e suas fontes de alimento no cerrado da estação florestal de experimentação de Paraopeba - Minas Gerais. Universidade Federal de Viçosa. Doctoral thesis.
- Truylio B, Harter-Marques B. 2007. A comunidade de abelhas (Hymenoptera, Apoidea) em áreas florestais do Parque Estadual de Itapuã (Viamão, RS): diversidade, abundância relativa e atividade sazonal. *Iheringia* 97: 392-399.
- Viana BF. 1999. A Comunidade de Abelhas (Hymenoptera: Apoidea) das Dunas Interiores do Rio São Francisco, Bahia, Brasil. *Neotropical Entomology* 28: 635-645.
- Zanella FCV. 2002. Sistemática, filogenia e distribuição geográfica das espécies sul-americanas de *Centris* (*Paracentris*) Cameron, 1903 e de *Centris* (*Penthemisia*) Moure,

1950, incluindo uma análise filogenética do "grupo Centris" sensu Ayala, 1998 (Hymenoptera, Apoidea, Centridini). Revista Brasileira de Entomologia 46: 435-488.

- Zanella FCV, Martins CF. 2005. Abelhas (Hymenoptera, Apoidea, Apiformes) da área do Curimataú, Paraíba. In: Araújo FS, Rodal MJN, Barbosa MV (ed) Análise das variações da biodiversidade do bioma Caatinga: suporte a estratégias regionais de conservação. Brasília, Ministério do Meio Ambiente, p. 75-134.

Occurrences

Species	Original dataset*	Final dataset**
1. <i>Bombus terrestris</i> (Linnaeus, 1758)	1551	947
2. <i>Bombus vestalis</i> (Geoffroy, 1785)	740	415
3. <i>Bombus lucorum</i> (Linnaeus, 1761)	2626	1441
4. <i>Bombus cryptarum</i> (Fabricius, 1775)	93	73
5. <i>Bombus magnus</i> Vogt, 1911	633	265
6. <i>Bombus bohemicus</i> Seidl, 1837	891	370
7. <i>Krameria erecta</i> Willd. ex Schult.	764	413
8. <i>Centris atripes</i> Mocsáry, 1899	719	122
9. <i>Centris rhodopus</i> Cockerell, 1897	988	116
10. <i>Krameria tomentosa</i> A. St.-Hil	327	214
11. <i>Centris fuscata</i> Lepeletier, 1841	504	67
12. <i>Centris tarsata</i> Smith, 1874	1078	81

*total number of occurrence points

**number of 10kmx10km occurrence cells (see details on Material and Methods)

Distribuição geográfica de abelhas e plantas associadas através de modelagem computacional

Conclusão

- Em relação às espécies de *Peponapis* e *Cucurbita* do sul da América do Sul, demonstrou-se que as áreas de ocorrência da espécie *P. fervens* apresentam maior similaridade climática com as áreas da única espécie de *Cucurbita* não domesticada presente na região (*C. maxima andreana*). No entanto, *P. fervens* exibe uma tolerância maior a variações climáticas e uma área de ocorrência mais ampla, na qual as espécies domesticadas de *Cucurbita* (abóbora) também ocorrem. É provável que a expansão de *P. fervens* a outras regiões tenha sido favorecida pela presença de recursos florais disponíveis nas áreas de cultivo das abóboreiras.
- Em relação às espécies de *Peponapis* e *Cucurbita* das Américas do Norte e Central, foi demonstrada a correspondência entre três grupos de espécies que ocorrem em áreas climaticamente similares e três das ecorregiões importantes nesses subcontinentes. Algumas zonas dessas três ecorregiões apresentaram a maior probabilidade de co-ocorrência dessas espécies, constituindo regiões potenciais de maior interação e partição de recursos. Porém, as áreas de distribuição geográfica dessas espécies não são totalmente equivalentes, sendo que algumas delas ocorrem em áreas maiores e outras, menores. Apesar da influência do clima, outros fatores participam da distribuição espacial das espécies, no caso, suas interações, história evolutiva e presença dos cultivos de abóboreiras.
- Em relação às espécies do gênero *Krameria*, foi demonstrada a equivalência entre as áreas de ocorrência potencial representadas pelos modelos e o que já foi registrado sobre sua distribuição. As variáveis climáticas analisadas mostraram-se importantes na determinação dos padrões de ocorrência de *Krameria*, especialmente a temperatura. No entanto, algumas espécies que foram previamente registradas na literatura como sendo de distribuição restrita resultaram em distribuições potenciais mais amplas, sugerindo outros fatores determinantes em suas áreas de ocupação além do clima. Provavelmente, a dependência das abelhas do gênero *Centris* para polinização é um dos fatores que atua em suas distribuições. Além disso, algumas

áreas especialmente na América do Sul, talvez tenham sido sub-amostradas, o que acarretaria na ausência de informações acerca da presença das espécies nessas áreas.

- Demonstrou-se que a inclusão de informações bióticas no processo de modelagem é útil para determinar com maior acurácia, a distribuição geográfica de uma espécie focal. No entanto, em alguns casos, a adição de dados suplementares não relevantes sobre espécies pode também aumentar a acurácia dos modelos, exigindo um conhecimento preciso acerca de quais interações devem ser realmente consideradas. A inclusão de informações sobre as espécies com as quais a espécie focal interage também é importante na construção de cenários futuros de distribuição baseados em mudanças de clima, uma vez que altera o resultado dos modelos.
- A modelagem computacional foi uma ferramenta auxiliar importante no estudo apresentado aqui e apresenta grande potencial para ser aplicado em outras análises. Esse estudo utilizou principalmente, dados disponíveis em coleções, museus e herbários que são fundamentais para a aplicação dessa ferramenta. Portanto, deve-se ressaltar a necessidade de um maior número de coletas, especialmente em algumas áreas do Brasil, que ainda estão sub amostradas. Destaca-se também, a importância fundamental das coleções biológicas, que devem ser conservadas e modernizadas, e seus acervos, conferidos, digitalizados e compartilhados para fomentar novas pesquisas. Os dados e sua disponibilização devem seguir os padrões internacionais de trocas de informações sugeridas pelos órgãos responsáveis, para viabilizar as análises com maior precisão e eficiência.

Distribuição geográfica de abelhas e plantas associadas através de modelagem computacional

Resumo

As abelhas e plantas apresentam diferentes graus de especialização em suas interações. Parceiros mais especialistas frequentemente apresentam uma história evolutiva mútua e sobreposição nas áreas de ocorrência. No entanto, a estrutura espacial dos ambientes nos quais esses grupos se distribuem é caracterizada por padrões complexos e dinâmicos. Para analisar a influência dos fatores que atuam na distribuição de espécies de abelhas e plantas associadas foram utilizadas a modelagem de distribuição de espécies, análise multivariada e ferramentas de sistemas de informações geográficas. Os resultados indicaram que a distribuição de gêneros estritamente associados, como é o caso de *Peponapis* e *Cucurbita*, é influenciada pelo clima das áreas de ocupação, bem como provavelmente, por sua história evolutiva e pela domesticação das abóboras (*Cucurbita*). Já os gêneros *Krameria* e *Centris* apresentam um padrão mais intrincado de distribuição, uma vez que a interação entre ambos é mais complexa. *Centris* é um grupo diverso que utiliza recursos florais de outras famílias botânicas além de *Krameriaceae*, o que provavelmente influencia seus padrões de distribuição. No entanto, os resultados obtidos para *Krameria* demonstraram de maneira geral, a influência de características climáticas na distribuição do grupo e uma provável dependência maior de *Centris* em alguns casos. Finalmente, foi também demonstrada a importância da inclusão de dados bióticos no processo da modelagem de distribuição, resultando no aumento da acurácia dos modelos e na alteração da projeção da distribuição para o futuro, considerando-se um cenário de mudança climática. Os resultados foram mais conspícuos quando foram consideradas interações mais estreitas entre espécies de abelhas parasitas e hospedeiras do gênero *Bombus*, do que entre *Centris* e *Krameria*. As técnicas utilizadas, em especial a modelagem de distribuição, representaram uma importante contribuição para a análise efetuada. No entanto, embora crescentemente utilizada, a modelagem de distribuição demanda técnicas e testes mais robustos para avaliar a acurácia dos modelos gerados. Além disso, um desafio adicional a ser vencido consiste no aumento e melhoria da qualidade dos pontos de ocorrência das espécies, principalmente no Brasil. Faz-se necessário um esforço adicional de coleta, especialmente em algumas áreas específicas, bem como, a conservação e digitalização dos dados das coleções biológicas. Porém, as técnicas utilizadas mostraram um grande potencial a ser explorado em outras análises, envolvendo questões biológicas diferentes, ou outros grupos taxonômicos e camadas de dados.

Geographical distribution of associated bees and plants through computational modeling

Abstract

Bees and plants present different degrees of specialism in their interactions. More specialized partners generally present a mutual evolutionary history and overlap with their occurrence areas. Nevertheless, the spatial structure of environments occupied by them is characterized by complex and dynamic patterns. Species distribution modelling, multivariate analyses and geographical system information tools were used in order to analyze the influence of different factors that act in the geographical distribution of associated bees and plants. Results showed that the geographical distribution of close associated genera, such as *Peponapis* and *Cucurbita* are influenced by the climate of occurrence areas, and also, by their evolutionary history and cucurbits domestication (squashes and pumpkins). On the other hand, *Centris* and *Krameria* genera presented a more intricate distribution pattern, since their interaction is more complex. *Centris* is a diverse group that uses other floral resources than those provided by the Krameriaceae family, which has probably influenced its distribution, also. However, the results obtained for *Krameria* showed the influence of climate in its distribution and a stronger relationship with *Centris* in some cases. Finally, the importance of including biotic data in the species distribution modelling process was also demonstrated, resulting in a general increase in the models accuracy and also altering future scenarios projection, considering climate changes. Stronger interaction, such as the host-parasite bee species of *Bombus* showed more conspicuous results than those found for *Krameria* and *Centris*. The techniques, especially distribution modelling, made an important contribution to the analyses. However, in spite of being increasingly used, distribution modelling demands more robust tests and techniques to evaluate the accuracy of final models. Besides, an additional challenge to be achieved consists in the increase and improvement of species occurrence data, mainly in Brazil. An additional survey effort is necessary, especially in specific areas, as well as the conservation and data digitalization of biological collections. However, the techniques used here showed a great potential to be further explored in other analyses, involving different biological issues, other taxonomic groups and other data layers.