



INCT-Herbário Virtual
da Flora e dos Fungos

Boletim de notícias

NÚMERO 1 | janeiro de 2014



Editorial

A partir de hoje vamos enviar periodicamente notícias sobre o INCT-Herbário Virtual da Flora e dos Fungos e projetos associados (Reflora e Sisbiota), de modo a deixá-los mais inteirados das ações e avanços do Instituto. Também disponibilizamos o espaço para incluir notícias que vocês nos enviarem para divulgação e desde já agradecemos aos que puderem mandar informações sobre atividades do pesquisador e/ ou herbário participante da rede INCT-Herbário Virtual. As notícias também continuarão sendo veiculadas na página (<http://inct.florabrasil.net>), juntamente com anúncios de cursos, visitas e demais atividades do INCT-HVFF.

Este 1º Boletim inclui textos da Dra. Ma.Regina Barbosa sobre o *Ciência sem Fronteiras*, e do colega Ricardo Braga-Neto, sobre o sistema *BioGeo*.

Leonor Costa Maia

Comitê Gestor INCT – HVFF



Ciência sem Fronteiras: Bolsistas do INCT- Herbário Virtual da Flora e dos Fungos do Brasil no Jardim Botânico de Nova York

Por Maria Regina Barbosa/UFPA

O Jardim Botânico de Nova York (NYBG) tem uma longa história de colaboração com o Brasil no contexto da Botânica. Ao longo dos anos foi construída uma sólida parceria entre pesquisadores do NYBG e pesquisadores brasileiros para estudo, principalmente, da flora do Cerrado (Projeto Plantas do Planalto do Brasil), da Amazônia (Projeto Flora Amazônica), e da Floresta Atlântica (Projeto Mata Atlântica Nordeste). Recentemente, o NYBG iniciou uma parceria com o INCT- HVFF

disponibilizando, através da plataforma speciesLink, mais de 330 mil registros e 120 mil imagens de plantas coletadas no Brasil desde o século XVIII, que fazem parte da coleção do William and Lynda Steere Herbarium. Estas informações e imagens contribuíram de forma significativa, junto com os dados dos demais herbários disponibilizados pelo INCT HVFF, para a elaboração e consolidação da Lista de Espécies da Flora do Brasil, da qual participaram também pesquisadores do NYBG.

Anualmente, botânicos do mundo todo vem ao NYBG principalmente para consultar suas coleções, herbário e biblioteca, e também, mais recentemente, para realizar estudos moleculares no Pfizer Laboratory através do Cullman Program for Molecular Systematics. Até meados de outubro de 2013, cerca de 270 botânicos, 24% dos quais brasileiros, visitaram o NYBG com estas finalidades. Além desses visitantes, que passam normalmente pequenas temporadas, estão



atualmente desenvolvendo pesquisas no NYBG 10 brasileiros, dois alunos de doutorado pleno, seis de doutorado sanduíche, e dois pesquisadores pós-doc. Desses, três vieram com bolsa do Programa Ciência Sem Fronteiras/CNPq, vinculada ao INCT-herbário Virtual, por um período de seis meses cada: a Dra. Maria Regina Barbosa, da Universidade Federal da Paraíba, para realizar estágio pós-doutoral, e as alunas de doutorado Gêssica Costa, da UFPE e Maria José Reis da Rocha, da UFMG, ambas com bolsa de doutorado sanduíche.

dois pesquisadores do NYBG foram contemplados com bolsas de Pesquisador Visitante Especial, do Programa Ciência sem Fronteiras-Capes, o Dr. Wayt Thomas, vinculado ao Programa de Pós-Graduação em Biologia Vegetal da UFPE, e o Dr. Douglas Daly, vinculado à Faculdade de Ciências Agrônômicas da UNESP, Botucatu.

Biogeografia da Flora e dos Fungos do Brasil

Por Ricardo Braga-Neto/CRIA

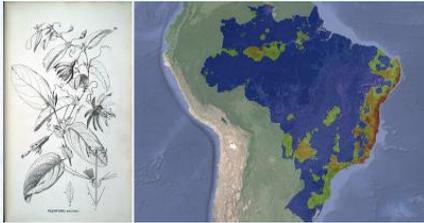
Obs: Notícia divulgada no Blog do CRIA.



Da esquerda para a direita: Douglas Daly, Marcelo Reginato, Felipe Saiter, Ana Paula Caetano, Fernanda dos Santos, Catarina Medeiros, Maria Regina Barbosa e Wayt Thomas, nos jardins do NYBG, out/2013.

Fortalecendo mais ainda a cooperação com o Brasil, atualmente

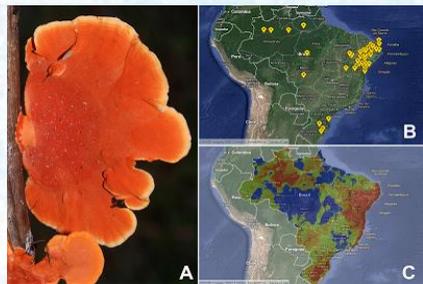
Compreender mais a distribuição geográfica de espécies é fundamental para promover a conservação da biodiversidade. Visando ampliar o conhecimento sobre a biogeografia de plantas e fungos do Brasil, está sendo desenvolvido um sistema para modelar a distribuição potencial das espécies, contando com a participação ativa de especialistas.



*Projeção de consenso entre os modelos de nicho ecológico para **Passiflora mucronata** (Bernacci & Giovanni 2013), indicando a área potencial de distribuição da espécie no Brasil. Prancha do [Herbario Virtual Flora Brasiliensis](#).*

Visando expandir o conhecimento sobre a biogeografia das espécies de plantas e fungos do Brasil, o INCT-Herbário Virtual da Flora e dos Fungos (INCT-HVFF) está desenvolvendo em parceria com o Centro de Referência em Informação Ambiental (CRIA) um sistema para modelar a distribuição potencial de espécies, contando com a participação ativa de especialistas. O sistema, [Biogeografia da Flora e dos Fungos do Brasil](#) (Biogeo), está sendo desenvolvido no âmbito do Sistema Nacional de Pesquisa em Biodiversidade (SISBIOTA Brasil), utilizando a rede speciesLink como base de dados da ocorrência das espécies.

O sistema visa contribuir para ampliar a compreensão das necessidades ambientais das espécies, investigar diversas questões envolvendo pesquisa e conservação, indicar espécies com maior carência de dados e orientar novas coletas. O sistema abre a perspectiva para a comunidade botânica e micológica construírem um banco de dados sobre plantas e fungos que no futuro deverá conter pelo menos um modelo de distribuição potencial para cada espécie.



*Modelos gerados para **Pycnoporus sanguineus**, um fungo amplamente distribuído em regiões tropicais e subtropicais do mundo (Braga-Neto 2013). **A)** Corpo de frutificação: produz os esporos que são dispersos pelo vento. **B)** Pontos de ocorrência: Dos 594 registros disponíveis na rede speciesLink, apenas 7,9% foram incluídos na modelagem. **C)** Modelo de consenso: Ainda que o número de pontos*



seja suficiente para incluir todos os algoritmos, o modelo pode ser considerado preliminar porque existem grandes lacunas de registros com coordenadas geográficas, influenciando a qualidade final do modelo.

Como funciona o Biogeo?

A interface do sistema possui uma seção aberta, onde todos os modelos publicados podem ser visualizados, e uma seção reservada aos supervisores cadastrados, os quais são responsáveis pelo processo de modelagem das espécies. Atualmente o sistema tem 55 supervisores cadastrados e cerca de 700 espécies com modelos gerados, incluindo angiospermas, samambaias e fungos macroscópicos. Todos esses modelos podem ser vistos a partir do menu de navegação na barra superior (clicando em Taxonomia pode-se visualizar as opções disponíveis, realçadas com fundo branco) ou buscados pelo nome científico no canto superior direito.

O Biogeo busca padronizar a geração de modelos e compartilhar os resultados de forma que os

experimentos possam ser facilmente reproduzidos e verificados por qualquer pessoa. O sistema permite que vários modelos sejam gerados ao longo do tempo para a mesma espécie, porém somente um deles é exibido como referência. O modelo de referência é sempre o último modelo aprovado para a espécie, pois se espera que cada novo modelo seja melhor que os anteriores. Os modelos gerados ficam disponíveis para serem avaliados pelos especialistas, que podem aprovar ou descartar o resultado. O sistema utiliza dados de ocorrência de espécies disponíveis na [rede speciesLink](#), a Lista de Espécies da Flora do Brasil 2012 como base taxonômica e variáveis ambientais bioclimáticas do [WorldClim](#) que afetam a distribuição de grande parte das espécies vegetais:

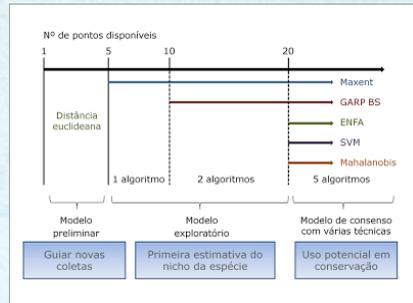
- Altitude (modelo digital de elevação)
- Variação média de temperatura no dia (BIO 2)
- Temperatura máxima no mês mais quente (BIO 5)



- Temperatura mínima no mês mais frio (BIO 6)
- Precipitação no trimestre mais úmido (BIO 16)
- Precipitação no trimestre mais seco (BIO 17)
- Precipitação no trimestre mais quente (BIO 18)
- Precipitação no trimestre mais frio (BIO 19)

O funcionamento do Biogeo é baseado na ferramenta [openModeller](#), que inclui 5 algoritmos que podem ser usados para a modelagem, dependendo do número de pontos de ocorrência. Para espécies com **menos de 5 pontos** gera-se apenas um modelo de dissimilaridade ambiental pelo cálculo da distância euclídeana ao ponto de ocorrência mais próximo. De **5 a 9 pontos** gera-se um modelo com o algoritmo Maxent; de **10 a 19 pontos** são gerados dois modelos, um com o Maxent e outro com o GARP Best Subsets (GARP BS). A partir de **20 pontos** são utilizados 5 algoritmos: Maxent, GARP BS, Distância Mahalanobis, ENFA e Máquina

Vetores de Suporte de classe única (SVM).



A estratégia de modelagem de nicho ecológico depende do número de pontos de ocorrência disponíveis para cada espécie. Menos de 5 pontos permite gerar modelos preliminares, com base em um algoritmo, que podem ser úteis para guiar novas coletas. Entre 5 e 19 pontos os modelos servem como uma estimativa preliminar do nicho da espécie, sendo gerados com um ou dois algoritmos. Modelos gerados a partir de 20 pontos permitem a inclusão de todos os algoritmos, tendendo a ser mais robustos.

O trabalho é realizado separadamente para cada espécie e basicamente envolve: (1) a seleção dos nomes a serem utilizados na busca de registros de ocorrência; (2) a seleção dos registros de ocorrência a serem utilizados no procedimento de modelagem; (3) a avaliação do modelo gerado. Embora a maior parte dos modelos originais sejam contínuos



(i.e., prevê a ocorrência potencial da espécie em uma gama de valores **entre 0 e 1**), os modelos podem ser transformados em modelos binários (i.e., prevê a ocorrência potencial da espécie em valores **iguais a 0 ou 1**), com base em um limiar de corte, visando facilitar a geração de um modelo de consenso que exibe apenas os locais onde há concordância entre os algoritmos utilizados. Para mostrar o funcionamento escolhemos a espécie *Parkia pendula* (Willd.) Benth. ex Walp. como exemplo.



Exsicata de **Parkia pendula** coletada em 1979 (Estação Experimental de Silvicultura Tropical, entre Manaus e Boa Vista). Amostra depositada no herbário do Jardim Botânico de Nova Iorque (NY 1190430) e, por possuir coordenadas geográficas do local de coleta, foi incluída na geração dos modelos de nicho ecológico.

Explorando o Biogeo

O primeiro passo para gerar os modelos de *Parkia pendula* envolveu a seleção dos nomes a serem utilizados na busca de registros de ocorrência, considerando a possibilidade de haver sinônimos na Lista de Espécies da Flora do Brasil. Em seguida o supervisor selecionou 50 pontos de ocorrência com a ajuda de filtros automáticos de qualidade que evitam a inclusão de registros com problemas de georreferenciamento ou identificação. Os pontos de ocorrência selecionados provêm de 18 herbários integrados à rede *speciesLink*.

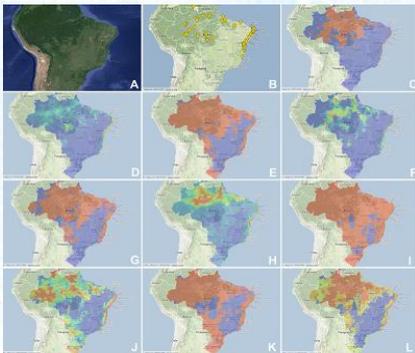


Registros de **Parkia pendula** utilizados no procedimento de modelagem (Lima 2013).

Em seguida, os modelos foram projetados no espaço geográfico para visualização. Como existem mais de 20 pontos disponíveis para esta



espécie, foram gerados modelos com base nos 5 algoritmos. No momento, a geração e a projeção dos modelos estão restritas ao território brasileiro porque a ampliação da cobertura geográfica depende de uma série de alterações não triviais, como conhecer os *datums* utilizados em outros países, ampliar a base de dados de estados e municípios, além de outros detalhes.



*Projeções dos modelos de nicho ecológico gerados para **Parkia pendula** (Leguminosae – Mimosoideae) (Lima 2013). A) Extensão geográfica abordada nos modelos. B) Pontos de ocorrência. C) SVM one-class. D) Maxent original. E) Maxent binário. F) Distância Mahalanobis original. G) Distância Mahalanobis binário. H) ENFA original. I) ENFA binário. J) GARP BS original. K) GARP BS binário. L) Modelo de consenso. As áreas em vermelho em C, E, G, I e K indicam modelos binários (presença ou ausência). Os modelos D, F, H e J são contínuos e as cores quentes indicam maior probabilidade de ocorrência.*

As áreas em vermelho no modelo de consenso L indicam concordância entre os 5 algoritmos, em laranja 4 e em amarelo 3.

Como avaliar os modelos gerados?

Os modelos gerados NÃO representam a distribuição real da espécie, mas sim a distribuição potencial em áreas ambientalmente adequadas à espécie de acordo com os pontos e variáveis ambientais utilizados no procedimento de modelagem. Vários motivos históricos podem ter contribuído para que a espécie não tenha ocupado todas as áreas potencialmente adequadas. Assim, a projeção do modelo é frequentemente maior que a distribuição real.

O procedimento de modelagem visa gerar resultados melhores e mais precisos ao longo do tempo à medida que mais registros de ocorrência são disponibilizados. Contudo, dois tipos de erro podem ocorrer num modelo (algumas vezes simultaneamente em diferentes regiões):



1. **Omissão:** quando o modelo não prevê áreas que são ambientalmente adequadas para a espécie
2. **Sobreprevisão:** quando o modelo erra ao prever determinadas áreas como sendo adequadas

Erros de omissão são mais fáceis de detectar ao confrontar o modelo com pontos de ocorrência. Dois indicadores são calculados para este fim: a *omissão interna*, que é calculada com os mesmos pontos utilizados na geração do modelo, e a *omissão externa*, que é calculada com pontos que não foram utilizados na geração do modelo. Normalmente sugere-se que a primeira não ultrapasse 5% e a segunda 20%. Para modelos contínuos a omissão é calculada com base no limiar mínimo (acima de zero) de presença.

Algoritmo	AUC (prom. crescente)	AUC omissão média
Consenso Maxent, GARP BS, ENFA, Mahalanobis, SVM	0.86 0.87 0.87 0.87 0.87 0.86 0.86 0.86 0.86 0.86 0.86	0.86 4.0%
Distância Mahalanobis	0.59 0.69 0.75 0.89 0.89 0.89 0.89 0.89 0.89 0.89 0.89	0.87 10.0%
ENFA	0.77 0.83 0.86 0.87 0.84 0.84 0.84 0.84 0.84 0.84 0.84	0.92 2.0%
GARP BS (june=10)	0.76 0.78 0.84 0.86 0.84 0.84 0.85 0.85 0.86 0.86 0.86	0.91 10.0%
Maxent	0.87 0.77 0.77 0.87 0.87 0.86 0.86 0.86 0.86 0.86 0.86	0.91 2.0%
SVM one-class (k=0.85)	0.69 0.69 0.69 0.79 0.79 0.89 0.89 0.89 0.89 0.89 0.89	0.83 32.0%

Os valores de AUC são calculados com validação cruzada do tipo 10-fold. Valores de AUC >0.9 são considerados como tendo

muito boa, >0.8 boa e >0.7 razoável capacidade de discriminação.

Modelos gerados com pelo menos 20 pontos de ocorrência podem ser avaliados por meio da AUC (Area under the Curve). Modelos com 5 a 19 pontos de ocorrência são testados com o procedimento de leave-one-out, seguido do cálculo da probabilidade associada ao número de acertos e modelos com menos de 5 pontos não são testados.



Condições ambientais nos pontos de ocorrência indicando para cada variável a amplitude de variação e a distribuição de frequência dos valores.

Além disso, os modelos aprovados agora incluem um perfil ambiental, disponível na página de detalhes do procedimento de modelagem. Procedimentos com um mínimo de 10 pontos incluem histogramas das variáveis ambientais, mostrando a



amplitude de variação do nicho ecológico da espécie, modelado com base nos pontos de ocorrência.

Como participar?

O cadastro de usuários está aberto a todos os especialistas interessados. Não é preciso ser um profundo conhecedor de modelagem para utilizar o sistema, pois se aplica um procedimento padronizado com algoritmos e variáveis ambientais previamente escolhidos. Por outro lado, é importante conhecer a taxonomia, a nomenclatura e a distribuição geográfica das espécies em questão.

Para participar, entre em contato pelo email abaixo, informando nome, instituição e grupo(s) taxonômico(s) que deseja monitorar:

biogeo@inct.florabrasil.net

Saiba mais!

[Biogeografia da Flora e Fungos do Brasil](#)

[Perguntas & Respostas \(FAQ do Biogeo\)](#)

[Twitter do Biogeo](#)

Modelagem de nicho ecológico em geral

Peterson A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, H., Nakamura, M., Araújo, M.B., 2011. *Ecological Niches and Geographical Distributions*. Princeton University Press, Princeton.

GARP Best Subsets : 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.

Distância Mahalanobis : Farber, O., Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, 160: 115–130.

ENFA : 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.

Maxent : Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190: 231–259.

SVM one-class : Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A.J., Williamson, R.C., 2001. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13: 1443–1471.



INCT-Herbário Virtual
da Flora e dos Fungos

www.inct.florabrasil.net
inct@florabrasil.net