

Modelagem da distribuição geográfica potencial de cinco espécies de Metzgeria no Brasil, visando à conservação.

Felipe Sodré Mendes Barros (felipe.b4rros@gmail.com);
Marinez Ferreira de Siqueira (marinez@jbrj.gov.br);
Denise Pinheiro da Costa (dcosta@jbrj.gov.br);

Rua Maria Eugenia 210/201,
CEP: 22261-080 Rio de Janeiro-RJ

Instituto de Pesquisa Ecologica Jardim Botânico
Rua Pacheco Leão, 915
Jardim Botânico - CEP 22460-030.

Abstract: The family *Metzgeriaceae* is distributed from the tropics to temperate regions, having its center of dispersion the Neotropical areas. From all species recognized in the world, approximately 43% lies in this region, where 19% of which is found in Brazil (Giulietti *et al.*, 2005). The Mata Atlantica ecosystem is considered the center of diversity of the family in the country. The southeast and southern concentrate a large part of the diversity of this genus containing six species endemic to the country, five of which are our object of study: *M. bahiensis*, *M. brasiliensis*, *M. cratoneura*, *M. psilocraspeda*, *M. subaneura*. Considering the knowledge of the species of that family, this project aims to a phytogeographic analysis based on modeling of potential distribution for the five species listed above, establishing distribution patterns, pointing out gaps or disjunctions and characterize the conservation status according to IUCN criteria. The algorithm used in modeling was the MaxEnt (Phillips *et al.*, 2006), based on nine environmental layers (climate and topography) with spatial resolution of five minutes (≈ 10 km) to reconcile with the georeferencing accuracy of the records (based most of all in localities and municipalities). Regarding partitions, were made 10 partition for the species with 15 or more points of occurrence using the bootstrap method, and JackKnife technique for the species with less than 15 points of occurrence. The statistical validation of the models generated showed quite satisfactory results and allowed us to identify areas of possible co-occurrence and information gaps.

Palavras-Chave: *MaxEnt*, *Metzgeriaceae*, *environmental suitability*, *geographical modeling*, modelagem preditiva.

1-Introdução

A família Metzgeriaceae distribui-se da região dos trópicos para as regiões temperadas, tendo como centro de dispersão e de diversidade o Neotrópico (Kuawahara apud Costa, 2008). Das espécies reconhecidas para o mundo, aproximadamente 43% encontram-se em tal região, sendo 17 endêmicas.

Costa, em 2008, publicou o tratamento taxonômico das espécies da família Metzgeriaceae para o Neotrópico como uma contribuição a série de monografias da flora Neotropical, sendo a primeira ferramenta elaborada na América Tropical para identificar as espécies dessa importante família de hepáticas.

O Brasil, com 8,5 milhões Km² de extensão ocupa parte significativa da zona Neotropical. Este caracteriza-se, não apenas por sua significativa diversidade biológica (estimada em 1,8 milhões de espécies, segundo Lewinsohn & Prado, 2005), como também pelo alto grau de endemismo, sendo considerado centro de diversidade para espécies da Mata Atlântica.

Costa comprovou que as espécies de Metzgeria no Brasil habitam preferencialmente os ecossistemas costeiros de Mata Atlântica (desde o nordeste até o sul do país), sendo esta, por tanto, considerada o centro de diversidade genética de Metzgeria no Brasil.

Com relação à sua distribuição nos outros biomas brasileiros, a distribuição de tal gênero é praticamente ausente na Amazônia (área pouco conhecida do ponto de vista da briologia), e cerrado, não ocorrendo na caatinga e pantanal.

Desta distribuição, as regiões sudeste e sul do país concentram grande diversidade deste gênero, sendo as áreas de Mata Atlântica, consideradas centros de diversidade genética, contendo seis espécies endêmicas para o país, das quais cinco são nosso objeto de estudo: *M. bahiensis*, *M. brasiliensis*, *M. cratoneura*, *M. psilocraspeda*, *M. subaneur*.

Entretanto, Costa a partir de sua pesquisa realizada, com coletas nos diferentes biomas brasileiros, descarta a possibilidade de o alto número de táxons para a região ser devido à falta de coletas em outras regiões do Brasil.

Considerando-se o atual grau de conhecimento das espécies da família Metzgeriaceae no Neotrópico, este projeto propõe integrar informações disponíveis sobre a ocorrência de táxons pré-selecionados e ferramentas de modelagem para cinco espécies de Metzgeria. As espécies foram selecionadas com base na disponibilidade de dados já trabalhados no Neotrópico (Costa, 2008) e por sua importância na composição florística dos Biomas brasileiros.

Dessa forma, temos como objetivo realizar uma análise fitogeográfica, com base na distribuição potencial de cinco táxons endêmicos do Brasil, estabelecendo os principais padrões de distribuição, estabelecendo lacunas ou disjunções, caracterizar o status de conservação desses táxons no país segundo os critérios da IUCN e contribuir com a revisão taxonômica da família no Brasil.

O entendimento dos padrões de distribuição espacial das espécies é fundamental para a conservação da diversidade biológica. Atualmente vêm sendo desenvolvidas e utilizadas várias técnicas de modelagem da distribuição geográfica de espécies com os mais variados objetivos. Algumas dessas técnicas envolvem modelagem baseada em análise ambiental, nas quais os algoritmos procuram por condições ambientais semelhantes àquelas onde as espécies foram encontradas, resultando em áreas potenciais onde as espécies foram encontradas, resultando em áreas potenciais onde as condições ambientais seriam propícias ao desenvolvimento dessas espécies possibilitando, não só o enriquecimento das análises ambientais resultantes, mas também, determinar áreas para futuras coletas, pesquisas e prioritárias para conservação.

2-Metodologia

A metodologia de modelagem preditiva procura por condições ambientais semelhantes àquelas onde as espécies foram encontradas, estabelecendo padrões de adequabilidade ambiental, através da aplicação de algoritmos. Portanto, para o desenvolvimento da modelagem geográfica, são necessários, basicamente, três tipos de dados:

1. Os dados bióticos, que são os pontos de existência das espécies georreferenciados;
2. Os dados abióticos, que são os dados ambientais, na maioria das vezes, climáticos e topográficos;

Assim sendo, nossos dados bióticos (tabela 1) dizem respeito aos pontos de ocorrência (registros) georeferenciados de cada espécie. Já os dados abióticos (tabela 2) utilizados estavam relacionados a dados climáticos, como de precipitação, temperatura, evapotranspiração, e dados topográficos de altimetria. Tendo em vista a quantidade de pontos de registros e a precisão destes, realizou-se uma amostragem dos dados abióticos os adequando à precisão dos dados bióticos utilizados (em sua maioria georeferenciado pelo município) estabelecendo a esses a resolução espacial de aproximadamente 10 km (5 Minutos).

Tabela 1: Relação das espécies estudadas e a quantidade de registros georeferenciados.

Espécies	N. Registros
M. bahiensis	6
M. brasiliensis	32
M. cratoneura	14
M. psilocraspeda	28
M. subaneura	9

Tabela 2: Fonte e resolução espacial dos dados abióticos utilizados na modelagem.

Dados abióticos utilizados	Fonte dos dados	Resolução espacial
BIO1 = Annual Mean Temperature	World Clim	5'
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))	World Clim	5'
BIO5 = Max Temperature of Warmest Month	World Clim	5'
BIO6 = Min Temperature of Coldest Month	World Clim	5'
BIO12 = Annual Precipitation	World Clim	5'
BIO13 = Precipitation of Wettest Month	World Clim	5'
BIO14 = Precipitation of Driest Month	World Clim	5'
Potential Evapo-Transpiration	CGIAR-CSI	5'
Altimetria	SRTM	5'

Como se pode perceber, na modelagem preditiva, um item essencial, é o algoritmo que possibilita a interpolação entre os pontos de existência e as camadas ambientais, de forma a estabelecer parâmetros de adequabilidade ambiental a partir da probabilidade. Para o presente trabalho foram realizados alguns testes prévios para averiguação dos melhores algoritmos a serem utilizados, e o que obteve melhor performance for o *MaxEnt*, além de ser um algoritmo de referencia nas bibliografias de tal temática.

Pode-se dizer que tal algoritmo baseia-se no princípio da máxima entropia (a melhor aproximação para uma distribuição de probabilidades desconhecida é aquela que satisfaça qualquer restrição à distribuição). Dessa forma, estima a probabilidade de ocorrência da espécie encontrando a distribuição de probabilidade da máxima entropia (que é a distribuição mais espalhada, mais próxima da distribuição uniforme), submetidas a um conjunto de restrições que representam a verdade, que é a informação presente (pontos de ocorrência) da distribuição alvo.

Tal algoritmo foi utilizado na modelagem de todas as espécies de estudo, apenas variando as formas de partição dos dados, de acordo com a quantidade de registros de ocorrência. Assim, ficou estabelecido (a partir de testes e, principalmente, de revisão bibliográfica) que para as espécies com registros de ocorrência (N) >15, teriam como método de partição *bootstrap*, na qual 70% dos dados seriam utilizados para treino e os 30% restantes para teste, com reposição em 10 partições diferentes. Já os dados com N<15, foram utilizados o método de partição sem reposição chamada *Jackknife*, onde N-1 eram utilizados para treino e 1 para teste. Ao estabelecermos o número de replicações igual ao N total, garantimos que todos os registros de ocorrência fossem utilizados uma vez para teste.

A tabela 3 resume os métodos de partição utilizados para cada espécie estudada, bem como seu N total, o N utilizado para treino e teste.

Tabela 3: Número de registros de treino e teste resultantes da partição dos dados através de métodos de bootstrap e jackknife.

Espécie	N total	N. Treino	N. Teste	Método Partição	N. Partições
<i>M. Bahiensis</i>	6	5	1	Jackknife	6
<i>M. Brasiliensis</i>	32	70%	30%	Bootstrap	10
<i>M. Cratoneura</i>	14	13	1	Jackknife	14
<i>M. Psilocraspeda</i>	28	70%	30%	Bootstrap	10
<i>M. subaneura</i>	9	8	1	Jackknife	9

3-Resultados e discussão

Os resultados das modelagens podem ser vistos nos mapas a seguir. Em todos os resultados apresentados foram utilizados como limite de corte (*threshold*) 10 percentil de treino. Ou seja, um limite de corte que assume a omissão de 10% de todos os registros utilizados para treino, com menores valores de adequabilidade ambiental.

O primeiro resultado apresentado é o da espécie *M. bahiensis*. De uma forma geral, pode-se perceber que esta espécie possui os registros de ocorrência bem espaçados, prevalecendo as regiões sul/sudeste do país, nas áreas do bioma de mata atlântica. Como resultado da modelagem, estabelecemos três níveis de adequabilidade ambiental: baixo, médio e alto.

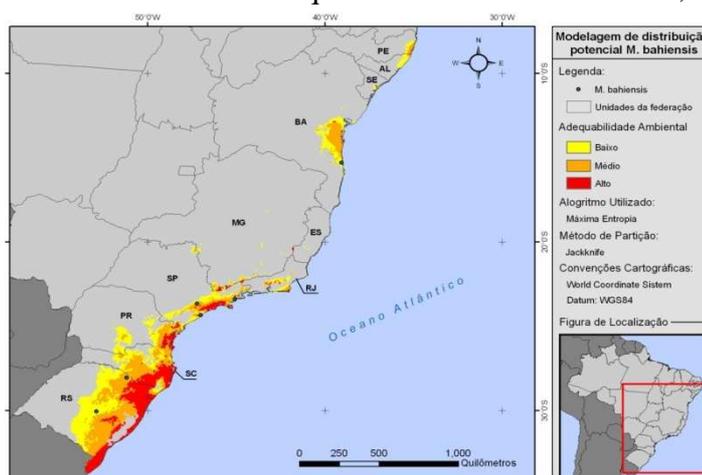


Figura 1: Mapa de adequabilidade ambiental para a ocorrência de *M. bahiensis*.

Taxa de acerto = 0,83; *p-value* = 0,000

Analisando o resultado apresentado, vemos que há algumas áreas de alta adequabilidade ambiental sem registros de ocorrência da espécie por perto, o que poderia indicar áreas potenciais para novas coletas e estudos futuros.

A respeito da modelagem em si, a partir dos dados probabilísticos apresentados, evidencia-se que a taxa de acerto foi alta e a probabilidade de o resultado ter sido dado ao acaso (*p-value*), foi baixa, configurando um resultado satisfatório do modelo gerado.

O resultado obtido para a espécie *M. brasiliensis* evidencia um nicho mais restrito ao litoral, mas também com grande ocorrência longitudinal. Apenas a área de alta adequabilidade ambiental, em Pernambuco, não apresenta ponto de registro incidente, nem próximo.

Com relação aos dados sobre a qualidade do dado gerado, por se tratar de um modelo gerado através do método de partição *bootstrap*, evidenciamos duas informações diferentes: a *AUC*, que é calculada independentemente do limite de corte escolhido, e a probabilidade binomial relativa ao limite de corte escolhido. Assim, podemos dizer que o modelo gerado também teve um desempenho satisfatório, uma vez que a *AUC* foi próximo a um, ou seja, diferente do aleatório, e com desvio padrão baixo, e a probabilidade binomial foi 0,000 também indicando a baixa probabilidade deste modelo ter sido gerado ao acaso.

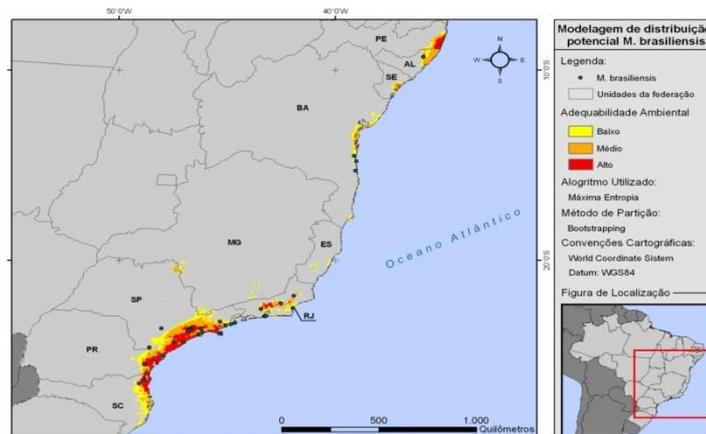


Figura 2: Mapa de adequabilidade ambiental para a ocorrência de *M. brasiliensis*.
 $AUC = 0.9900 \pm 0.0039$; $Binomial\ Probability = 0,000$

O modelo gerado para a espécie *M. cratoneura* também apresentou bons índices, com alta taxa de acerto e p-value baixa. Esta espécie apresenta-se mais restrita à região sul/sudeste do Brasil, com pontos de alta adequabilidade ambiental próximos aos litoral, sendo que alguns deles sem ocorrência da espécie, mas com alguns pontos bem próximos.

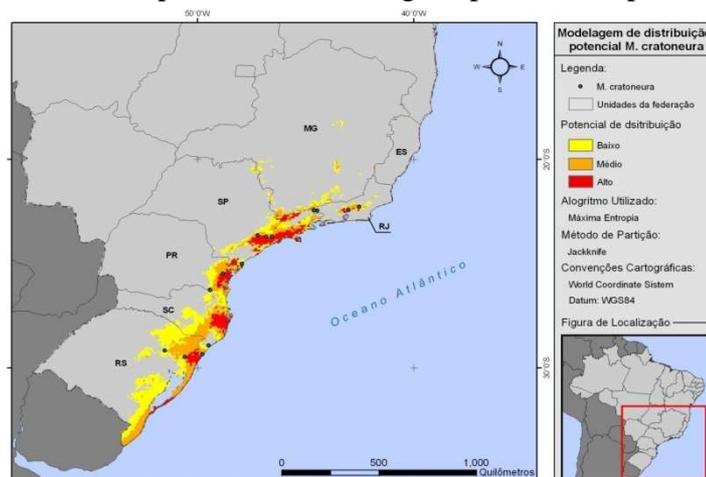


Figura 3: Mapa de adequabilidade ambiental para a ocorrência de *M. cratoneura*.
 Taxa de acerto = 0,78; $p\text{-value} = 0,000$

A partir dos pontos de existência da espécie *psilocraspeda* pode-se deduzir que se trata de uma espécie com nicho mais amplo, não apenas devido ao seu maior espaçamento entre os pontos de ocorrência, mas também pelo fato desses não se limitarem ao litoral. O interessante da modelagem resultante é que apesar desta distribuição mais ampla, as áreas de alta adequabilidade ambiental se restringiram ao litoral sul/sudeste, com alguns pontos isolados como, por exemplo, em Roraima. O modelo gerado apresentou taxa de AUC excelente, e baixa probabilidade deste resultado ter sido gerado ao acaso.

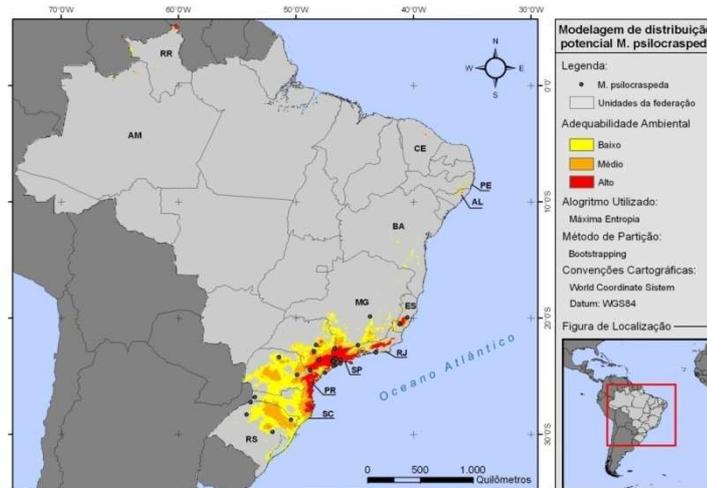


Figura 4: Mapa de adequabilidade ambiental para a ocorrência de *M. psilocraspeda*.
 $AUC = 0.9692 \pm 0.0129$; $Binomial Probability = 0.0053$

Por ultimo temos a modelagem preditiva para a espécie *subaneura*. Esta espécie aparentemente possui uma maior distribuição pelo Brasil tendo registros de ocorrência, inclusive, no Estado do Acre. Porém, esta espécie foi uma das três, de nosso estudo, que apresenta poucos pontos de presença, o que dificulta nossa análise, uma vez que possivelmente estamos manuseando dados com lacunas de informação quanto ao nicho de distribuição.

Assim mesmo, o modelo foi gerado e apresentou uma boa taxa de acerto. Porém, o p-value calculado, não foi tão baixo, sendo um pouquinho acima do limite estabelecido a ser considerado bom (0,05). O resultado, considerado insatisfatório, pode ser visto no resultado da modelagem com as áreas de adequabilidade ambiental muito abrangente.

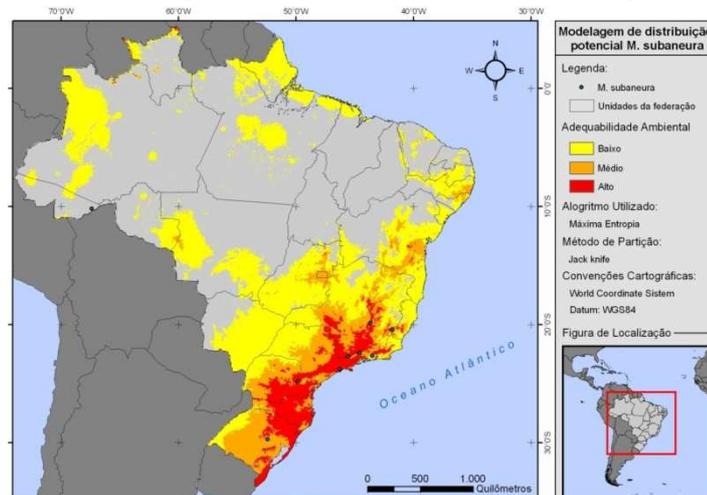


Figura 5: Mapa de adequabilidade ambiental para a ocorrência de *M. subaneura*.
 Taxa de acerto = 0,89; $p\text{-value} = 0,053$

A baixo evidenciamos uma tabela com os dados mais importantes que resumem a qualidade dos modelos gerados, com as taxas de qualidade do modelogerado (p(binomial), para os modelos utilizando bootstrap; (p-value), para os modelos gerados a partir do Jackknife).

Tabela 4: Estatísticas de validação dos resultados obtidos na modelagem.

Species	#Training samples	AUC (test points)	AUC test Standard Deviation	10 percentile training presence logistic threshold	p (binomial)	Taxa de acerto*	p (value)*
M._bahiensis	5	N/A	N/A	0.4342	N/A	0,83	0,000
M._brasiliensis	21	0.99	0.0039	0.365	0	N/A	N/A
M._cratoneura	13	N/A	N/A	0.3848	N/A	0,78	0,000
M._psilocraspeda	19	0.9692	0.0129	0.2759	0.0053	N/A	N/A
M._subaneura	8	N/A	N/A	0.0824	N/A	0,89	0,053

*Estatísticas resultantes da técnica para amostras pequenas baseado em Pearson (2007).

Dessa forma, de todos os modelos gerados, apenas o modelo da espécie subaneura foi considerado insatisfatório uma vez que, apesar da boa taxa de acerto, o seu p-valeu, ficou mais alto do o valor estabelecido como satisfatório. Tal resultado nos induz a rever os dados utilizados para tal espécie, buscando possíveis erros, bem como um maior estudo sobre a espécie e, se possível, buscar novas coletas na tentativa de aumentar e melhorar a qualidade dos dados de distribuição desta espécie.

4-Conclusões

A modelagem geográfica preditiva se mostrou um importante instrumento de análise, tendo bons resultado mesmo em espécies com poucos pontos de ocorrência. Pode-se observar algumas áreas de alta adequabilidade ambiental e sem nenhum ponto de ocorrência da espécie modelada, podendo assim, a modelagem ser usada para a identificação de novas áreas de coleta e áreas pertinentes à reintrodução da espécie.

Como mencionado, apenas um resultado não foi satisfatório, o que nos obriga uma maior atenção sobre a espécie *M.subaneura*. Por mais insatisfatório que a modelagem possa ter se apresentado, esta pode estar evidenciando uma lacuna de conhecimento da espécie. Assim, há que se ter um estudo maior sobre a espécie, buscando maiores informações de nicho e novas coletas, de forma a apurar os dados atuais.

Desta observação, nos deparamos com uma reflexão importante a ser feita a respeito dos dados utilizados na modelagem. Mais importante do que a quantidade de pontos de ocorrência das espécies, e a resolução espacial dos *rasters* ambientais, é a qualidade do georreferenciamento dos dados de ocorrência. Este dado é de fundamental importância, uma vez que, além de ditar a resolução espacial dos dados abióticos utilizados, conduzirá a resultados mais precisos.

Para os resultados positivos, temos em perspectiva a continuidade da análise a partir de um refinamento na modelagem gerada, estabelecendo novo recorte de análise que seja mais próximo da distribuição reconhecida/esperada para as espécies (bioma Mata Atlântica), identificando disjunções entre a distribuição conhecida e potencial das espécies, visando avançar na discussão quanto aos critérios da IUCN e apontar lacunas de conhecimento das espécies analisadas.

Agradecimentos

Agradecemos à Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) pelo apoio e auxílio prestado na execução deste projeto, ao Instituto de Pesquisas Jardim Botânico do Rio de Janeiro por ter aceito e abrigado nosso trabalho.

Aproveito para agradecer também às Drs. Denise Pinheiro da Costa e Marinez Ferreira de Siqueira, pelos ensinamentos, orientação e paciência ao longo do trabalho desenvolvido.

Referências Bibliográficas

- Costa, D.P. **Metzgeriaceae (Hepaticae) Flora Neotropica Monograph**; Rio de Janeiro, 2008. 169p.
- Giulietti, A.M., Raymond, M.H., Queiroz, L.P., Wanderley, M.G.L. & Van den Berg, C. **Biodiversity and Conservation of Plants in Brazil**. Cons. Biol. 2005 p. 632-639.
- Lewinsohn, T.M. & Prado, P.I.. **How many species are there in Brazil?** Cons. Biol. 2005 p. 619-624.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. **Maximum entropy modeling of species geographic distributions**. Ecological Modelling 2006 p.231-259.