

25. Meio Ambiente

Separando a influência da variação climática e da distância geográfica na estrutura genética da castanheira-do-Brasil na Amazônia brasileira: A importância do delineamento amostral

Garuti, Giovana¹; giovana.lgaruti@gmail.com; Costa, Patricia²; Thomas, Evert³; Wadt, Lúcia Helena²; Martins, Karina (orientadora)¹; karimartins@yahoo.com

¹Universidade Federal de São Carlos, ²Empresa Brasileira de Pesquisa Agropecuária,

³Bioversity International

Resumo

A genômica da paisagem estuda a distribuição geográfica da variação genética em escala genômica. O grande desafio é separar os efeitos do isolamento por distância (*isolation by distance* – IBD) e do isolamento por ambiente (*isolation by environment* – IBE) na diferenciação genética. Um delineamento amostral inadequado dificulta a separação dos efeitos da distância geográfica e dos gradientes ambientais na estrutura genética e a identificação de regiões genômicas potencialmente envolvidas na adaptação local. O objetivo deste estudo é determinar o delineamento amostral mais adequado para estudar a genômica da paisagem da castanheira-do-Brasil (*Bertholletia excelsa*) em toda a sua distribuição na Amazônia brasileira. O banco de dados inclui 51 potenciais pontos de amostragem. Estes foram comparados com dados de 489 localizações obtidas de um trabalho de modelagem ecológica. Para todas as 540 localizações foi feito o download de 19 variáveis bioclimáticas na plataforma Worldclim e excluídas as variáveis altamente correlacionadas ($|r| \geq 0,70$). Seis variáveis foram mantidas para análise das distâncias climáticas. Para cada variável, as distâncias geográficas e climáticas foram plotadas em gráficos de dispersão e histogramas. Todas as análises foram realizadas no ambiente computacional R. Nossos resultados mostram que o delineamento amostral de apenas 44 pontos amostrais é adequado para separar IBD e IBE, uma vez que toda a gama de possíveis distâncias geográficas e climáticas entre pares de populações estão bem representadas. Além do mais, os 44 pontos amostrais representam a ampla variação climática ao longo da área de distribuição da *B. excelsa* na Amazônia brasileira e serão amostrados para análises genéticas.

Palavras-chave: *Bertholletia excelsa*, Genômica da Paisagem, Isolamento por distância.

Introdução

A genômica da paisagem é uma abordagem que busca estudar como as características geográficas e ambientais atuam na variação genética entre indivíduos ou populações. Segundo Holderegger e Wagner (2008), a genômica da paisagem é uma integração da genética de populações com a ecologia da paisagem, cujo o foco é o estudo do fluxo gênico e da variação genética adaptativa entre os ambientes.

A disponibilidade de sequências genômicas proporciona examinar efeitos históricos da população, ocorrências de migração e seleção natural (Sork et al., 2013). Adicionalmente, a identificação dos genes que expressam os fenótipos relacionados à adaptação local das populações contribui para o conhecimento da base genética da adaptação e da evolução por seleção natural (Rieseberg & Bourke, 2001; Sork et al., 2013).

Como as diferenças adaptativas são impulsionadas pela seleção em ambientes diferentes ecologicamente (Räsänen & Hendry, 2008), a genômica da paisagem vale-se da coleta de amostras genéticas ao longo de gradientes ambientais e da identificação de marcadores moleculares que estão estatisticamente relacionados a determinado ambiente (Holderegger & Wagner, 2008). Compreender como o fluxo gênico atua com os gradientes ambientais possibilita um estudo da evolução na

adaptação de espécies ou populações durante períodos de rápida mudança ambiental (Sexton, Hangartner & Hoffmann, 2013).

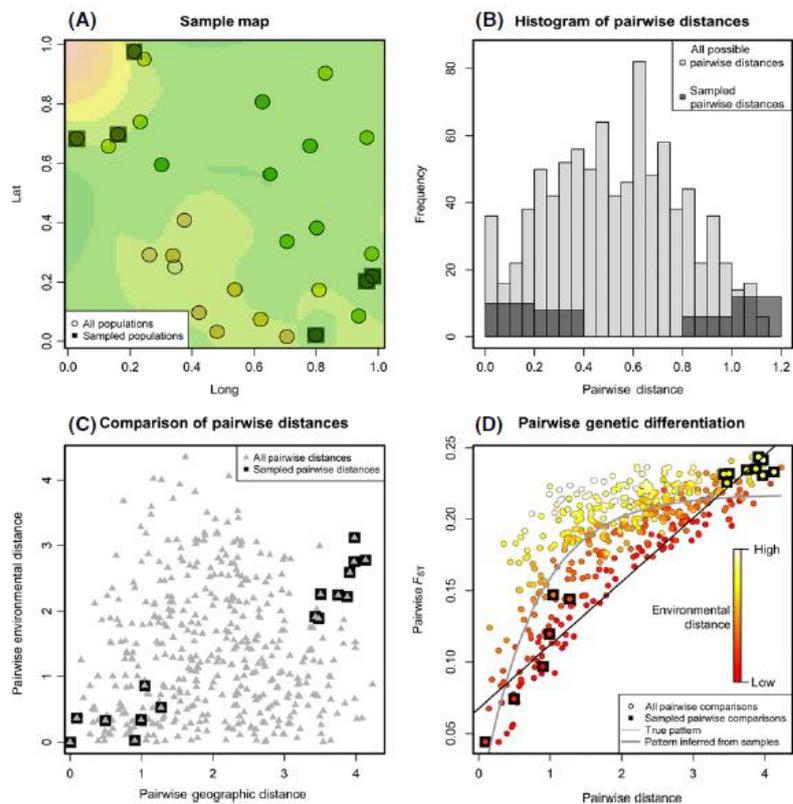
O grande desafio dos estudos em genômica da paisagem é separar a interação e relação do 'isolamento por distância' (IBD – *Isolation by Distance*) e do 'isolamento por ambiente' (IBE – *Isolation by Environment*) na diferenciação genética espacial (Meirmans, 2012; Meirmans, 2015; Wang & Bradburd, 2014), o que dificulta a avaliação dos efeitos da seleção natural, da história geográfica e da migração no estudo evolutivo da espécie. O isolamento por ambiente pode ser definido como um modelo em que a diferenciação genética aumenta com as diferenças ambientais, independente da distância geográfica. Já o isolamento por distância é um modelo em que a diferenciação genética aumenta com a distância geográfica. (Wang & Bradburg, 2014).

A divergência genética entre populações pode ser resultado de fluxo gênico reduzido associado a fatores geográficos e ambientais. Compreender os efeitos das contribuições de IBE e IBD sobre a redução do fluxo gênico entre populações podem fornecer informações de como as paisagens moldam essa divergência genética. O grande desafio de separar os efeitos de IBE e IBD se deve ao fato de que as populações que estão distantes entre si também ocupam diferentes ambientes (Wang, Glor & Losos, 2013).

Nos últimos anos, diversos autores têm argumentado que, devido à correlação entre os modelos IBD e IBE na distribuição da variação genética espacial, é essencial que haja um delineamento amostral, que permita separar os efeitos relativos de distância geográfica e distância ambiental na variação genética (de Mita et al. 2013; Lotterhos & Whitlock, 2015; Meirmans, 2012; Meirmans, 2015; Wang & Bradburg, 2014). Para isso, Wang & Bradburg (2014) argumentam que amostras de diferentes ambientes devem ser delineadas utilizando populações que estejam geograficamente próximas e distantes, sendo que as populações que

ocupam ambientes semelhantes também devem ser amostradas. Desse modo, a associação entre IBE e IBD será reduzida. Eles também ilustram como um delineamento amostral inadequado pode impossibilitar que as questões referentes à genômica da paisagem sejam adequadamente respondidas. A Figura 1, extraída do trabalho de Wang & Bradburg (2014), usa simulação para ilustrar um delineamento amostral inadequado.

Figura 1. Ilustração de um delineamento amostral inadequado de uma paisagem hipotética apresentada por Wang e Bradburg (2014). Em (A) é mostrada a localização espacial de todas as populações presentes no gradiente ambiental, com destaque para populações amostradas (*sampled populations*). Em (B) o histograma mostra as distâncias geográficas entre todos os pares de populações representadas em (A). O gráfico de dispersão em (C) mostra a comparação das distâncias geográficas e ambientais entre todas as populações, com destaque para as populações amostradas (*sampled pairwise distances*). Em (D) são plotadas as curvas de diferenciação genética entre pares de distância ambiental (escala de cor) e distância geográfica (eixo X). Fonte: Wang & Bradburg (2014). Reprodução autorizada por John Wiley and Sons (*license number 4391380583796*)



Nota-se que no plano de amostragem em uma paisagem hipotética proposto por Wang e Bradburd (2014), as amostras coletadas não representam toda a variação ambiental da paisagem (Figura 1A) e nem toda a gama de distâncias geográficas possíveis entre pares de populações (Figura 1B), ou seja, as distâncias intermediárias entre pares de populações não estão representadas nessa amostragem. Na Figura 1C pode-se também perceber como a amostragem parcial das populações resulta em distâncias geográficas e ambientais correlacionadas. As comparações de distâncias geográficas curtas e longas entre pares de amostras coletadas coincidem com as distâncias ambientais. Nesse delineamento não houve amostras com distâncias geográficas intermediárias entre as populações. Dessa forma, os pares de populações com distâncias geográficas curtas apresentam sempre distâncias ambientais curtas e o mesmo ocorre com as distâncias geográficas grandes. O efeito dessa correlação entre as distâncias geográficas e ambientais resultantes do delineamento amostral ruim poderá gerar inferências erradas sobre o efeito das distâncias geográficas e ambientais (IBD e IBE respectivamente) na estrutura genética (Figura 1D), ou seja, a partir desse delineamento amostral inadequado as curvas são mal estimadas.

Segundo Sork et al. (2013),

espécies arbóreas são de interesse à conservação devido à importância econômica e ecológica. As árvores armazenam uma enorme quantidade de carbono, influenciando diretamente o ciclo global do carbono e a taxa de mudança climática. Por isso, estudar a genômica da paisagem de espécies arbóreas proporciona informações importantes para a conservação da floresta diante das rápidas mudanças ambientais que vêm sendo causadas pelo homem. A *Bertholletia excelsa* Bonpl. (Lecythidaceae), conhecida como a castanheira-do-Brasil ou castanheira-da-Amazônia, é abundante na região amazônica e suas amêndoas são exploradas por diversas comunidades em florestas nativas no curto prazo e com baixo custo (Tonini Costa & Kaminski, 2008), para comercialização tanto para consumo *in natura* como para uso na indústria de cosméticos. A espécie é considerada ameaçada devido ao desmatamento e classificada como espécie vulnerável de acordo com a Lista Vermelha das Espécies da Flora Ameaçadas de Extinção (IUCN, 2016). No Brasil, a derrubada da árvore é proibida por lei (IN-IBDF 1/1980).

A distribuição geográfica da *B. excelsa* é ampla, apresentando assim diferenças no gradiente de pluviosidade e sazonalidade por toda a Amazônia. (Nepstad et al. 2004). Tendo em vista que a seleção natural é o principal fator

causador da adaptação local das populações a ambientes, acredita-se que a distribuição da variação genética entre as populações de castanheira na Amazônia brasileira seja fortemente explicada pela seleção natural. Estudos sobre a história demográfica da *B. excelsa* e a distribuição de sua variabilidade genética vêm sendo realizados (Sujii, Martins, Wadt, Azevedo & Solferini, 2015; Thomas, Caicedo, Loo & Kindt, 2014; Thomas, Caicedo, Mcmichael, Corvera & Loo, 2015), mas não há ainda estudos que utilizaram uma amostragem ampla para avaliar a genômica da paisagem dessa espécie. O conhecimento dos padrões geográficos da variação genética adaptativa e sua associação com gradientes climáticos pode trazer informações sobre a habilidade de populações locais tolerarem mudanças climáticas e na definição mais eficiente de estratégias e áreas prioritárias para conservação da espécie. O presente trabalho é a primeira parte de um projeto que visa estudar a genômica da paisagem em *B. excelsa* ao longo da Amazônia brasileira.

Objetivos

O objetivo desta pesquisa foi definir um delineamento amostral adequado para que as análises estatísticas em genômica da paisagem permitam separar os efeitos das distâncias geográficas e climáticas na variação genômica de *Bertholletia excelsa* na Amazônia brasileira.

Materiais e Métodos

O banco de dados do projeto incluiu 51 localizações de potenciais pontos de coleta que foram indicados pelos pesquisadores participantes do projeto em todos os estados da Amazônia brasileira. Estas localizações foram comparadas com o banco de dados de 489 pontos de ocorrência de *B. excelsa* apresentados em Thomas et al. (2014) e disponibilizados para este estudo. Para todas as 540 localizações foi realizado o download das 19 variáveis bioclimáticas no banco de dados do WorldClim – Global Climate Data (<http://www.worldclim.org/>).

Inicialmente foi calculada a correlação linear simples entre pares de variáveis climáticas nos 51 locais para exclusão de variáveis altamente correlacionadas ($|r| \geq 0,70$) e escolha de apenas seis variáveis para as análises subsequentes. Na escolha, optou-se por manter números equivalentes de variáveis relacionadas à temperatura e à precipitação. Calculou-se também a correlação entre os dados climáticos e latitude e longitude, com o objetivo de averiguar se há uma tendência latitudinal ou longitudinal nos gradientes climáticos.

Para cada variável climática selecionada, foram calculadas distâncias climáticas entre pares de populações e plotadas em gráficos de dispersão e histogramas seguindo o procedimento apresentando em Wang e Bradburd (2014)

como ilustrado na Figura 1B e 1C. Os gráficos foram observados para comparar as distâncias geográficas e climáticas de todos 489 pontos e das 51 populações.

Após a análise da eficiência amostral das 51 localizações potenciais foi simulada a exclusão de pontos amostrais potencialmente redundantes e a inclusão de novos locais, que não haviam sido previstos quando da elaboração do projeto. Além disso, houve a necessidade de excluir locais potenciais devido a dificuldades logísticas e financeiras que ocorreram durante a execução do projeto. Após a confirmação dos locais viáveis para amostragem, 44 pontos de coleta foram definidos. Todas as análises descritas anteriormente foram refeitas para esse novo conjunto de dados.

Todas as análises foram realizadas no ambiente computacional R (R core Team, 2016) utilizando os pacotes 'dismo' (Hijmans, Phillips, Leathwick & Elith, 2017), 'raster' (Hijmans, 2017) 'maptools' (Bivand & Lewin-Koh, 2017) e dependências. Os gráficos e mapas também foram realizados no R com uso do pacote 'ggplot2' (Wickham, 2009).

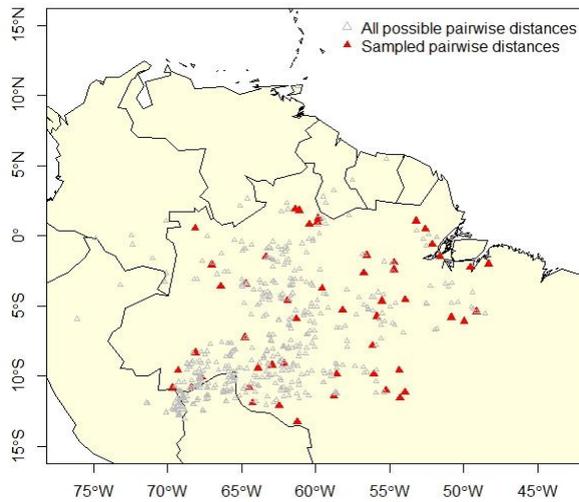
Resultados e Discussão

Os 51 pontos potenciais de coleta estão amplamente distribuídos ao longo da

área de ocorrência da espécie no Brasil (Figura 2A). Esses pontos compreendem regiões que inclusive não estão bem representadas no banco de dados das 489 localizações (centro-oeste do Mato Grosso e leste do Pará). O histograma das distâncias (km) entre pares de populações (Figura 2B) confirma a boa distribuição das amostras ao longo da área de ocorrência. Tanto as distâncias curtas entre populações como longas estão bem representadas. A zona em roxo na Figura 2B mostra o quanto das 51 amostras potenciais representam as 489 populações conhecidas, através da sobreposição das distâncias geográficas dos potenciais pontos e dos 489 pontos conhecidos.

Nota-se que as distâncias geográficas tanto das amostras potenciais quanto as das populações conhecidas possuem um comportamento parecido. Somente os pontos com uma distância menor (de 200 a 500 km) entre as populações possuem uma quantidade menor com relação às 489 populações conhecidas. Além disso, nota-se também uma maior quantidade de amostras com 1000 km de distância.

A) Mapa de simulación de amostragem



B) Histograma de distâncias em pares

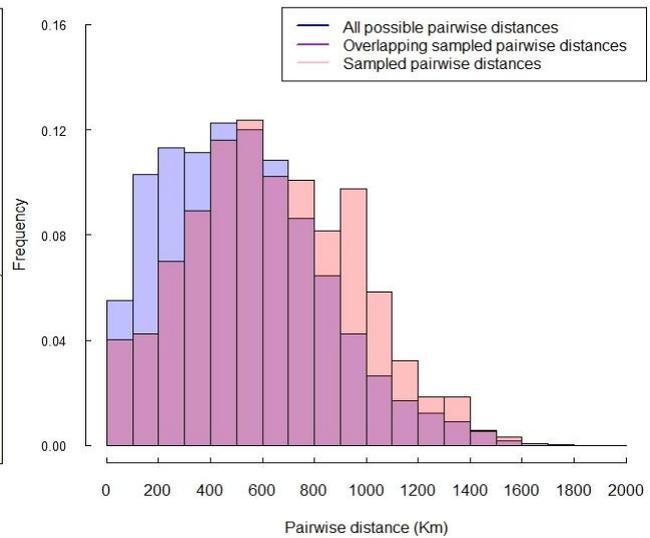
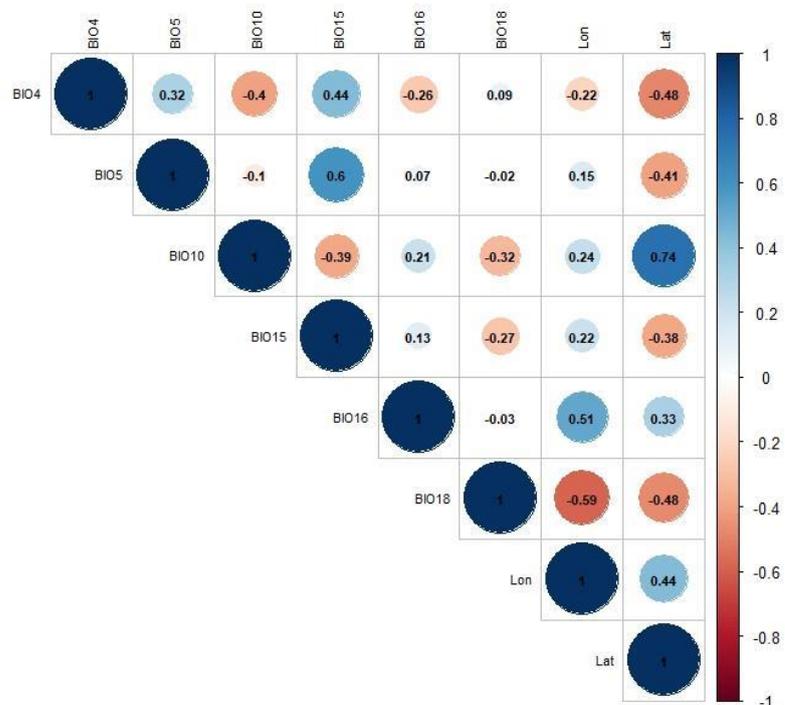


Figura 2. Simulação de delineamento amostral das 51 potenciais localizações. (A) Mapa de simulação de amostragem com os 51 potenciais pontos de amostragem em vermelho e as 489 populações conhecidas na área de ocorrência em cinza. (B) Histograma das distâncias geográficas em pares de populações de todas as 540 populações, na qual em azul as distâncias entre pares das 489 populações conhecidas, em rosa estão as distâncias entre pares das 51 localizações dos possíveis pontos de coleta e em roxo as amostras que se sobrepõem.

Figura 3. Correlação linear entre as variáveis climáticas e os dados de Longitude e Latitude dos locais de amostragem BIO1: Temperatura Sazonal; BIO5: Temperatura Média no Mês Mais Quente; BIO10: Temperatura Média do Trimestre Mais Quente; BIO15: Precipitação Sazonal; BIO16: Precipitação do Trimestre Mais Úmido; BIO18: Precipitação do Trimestre Mais Quente; Lon: Longitude; e Lat: Latitude.



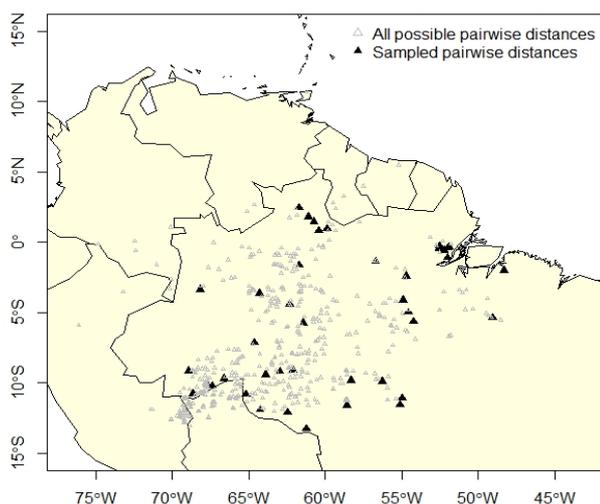
A análise de correlação entre variáveis climáticas nos 51 potenciais

locais de coleta resultou na manutenção de apenas seis variáveis bioclimáticas para esse trabalho, sendo três relacionadas à temperatura e três relacionadas à precipitação. As variáveis escolhidas foram: Sazonalidade da Temperatura; Máxima Temperatura do mês mais quente; Temperatura Média do Trimestre Mais Quente; Sazonalidade da Precipitação; Precipitação do Trimestre Mais Úmido; e Precipitação do Trimestre Mais Quente.

A escolha das variáveis climáticas se manteve a mesma após a diminuição do número de amostras de 51 para 44 localizações (Figura 3).

Com relação aos dados de longitude e latitude das 44 amostras, podemos observar apenas uma correlação de $|r| = 0,74$ entre a variável climática BIO10 (Temperatura Média do Trimestre Mais Quente) e a latitude, mostrando que esta variável varia em função da latitude.

A) Mapa de amostragem



B) Histograma de distância entre pares

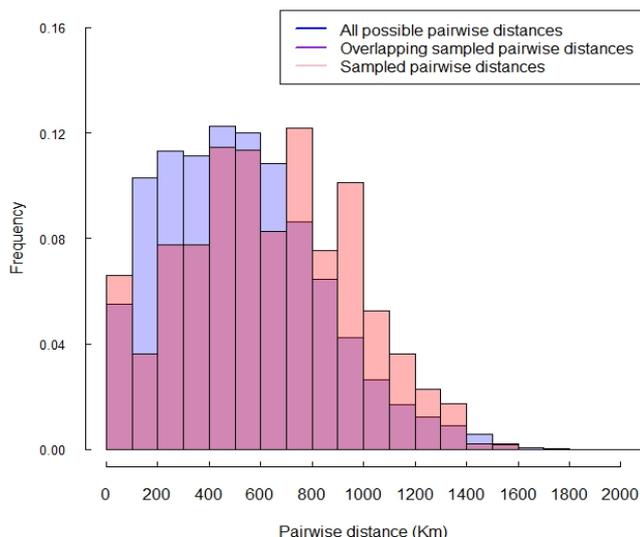


Figura 4. Delineamento amostral dos 44 pontos de coleta. (A) Mapa de amostragem com os 44 pontos de amostragem em preto e as 489 populações conhecidas na área de ocorrência em cinza. (B) Histograma das distâncias geográficas em pares de populações de todas as 533 populações, na qual em azul as distâncias entre pares das 489 populações conhecidas, em rosa estão as distâncias entre pares das 44 populações amostradas e em roxo as amostras que se sobrepõem.

É possível afirmar que as amostras representam todo o gradiente climático da área de ocorrência e que o delineamento amostral com as 44 populações a serem

amostradas é eficiente (Figura 5). Observa-se que não há um comportamento de correlação linear entre as distâncias ambientais de cada variável

com as distâncias geográficas entre as populações.

Na comparação de distâncias ambientais e geográficas entre pares de populações proposta por Wang e Bradburg (2014) na Figura 1C, a correlação existente entre distâncias geográficas e ambientais é nítida. A falta da amostragem de populações com distâncias geográficas

intermediárias ocasiona a baixa representatividade do gradiente ambiental ao longo da área, diferentemente da amostragem proposta para as 44 populações do presente trabalho, a qual conseguiu alcançar uma boa representatividade do gradiente de ambiental da área de ocorrência da *B. excelsa*. (Figura 5)

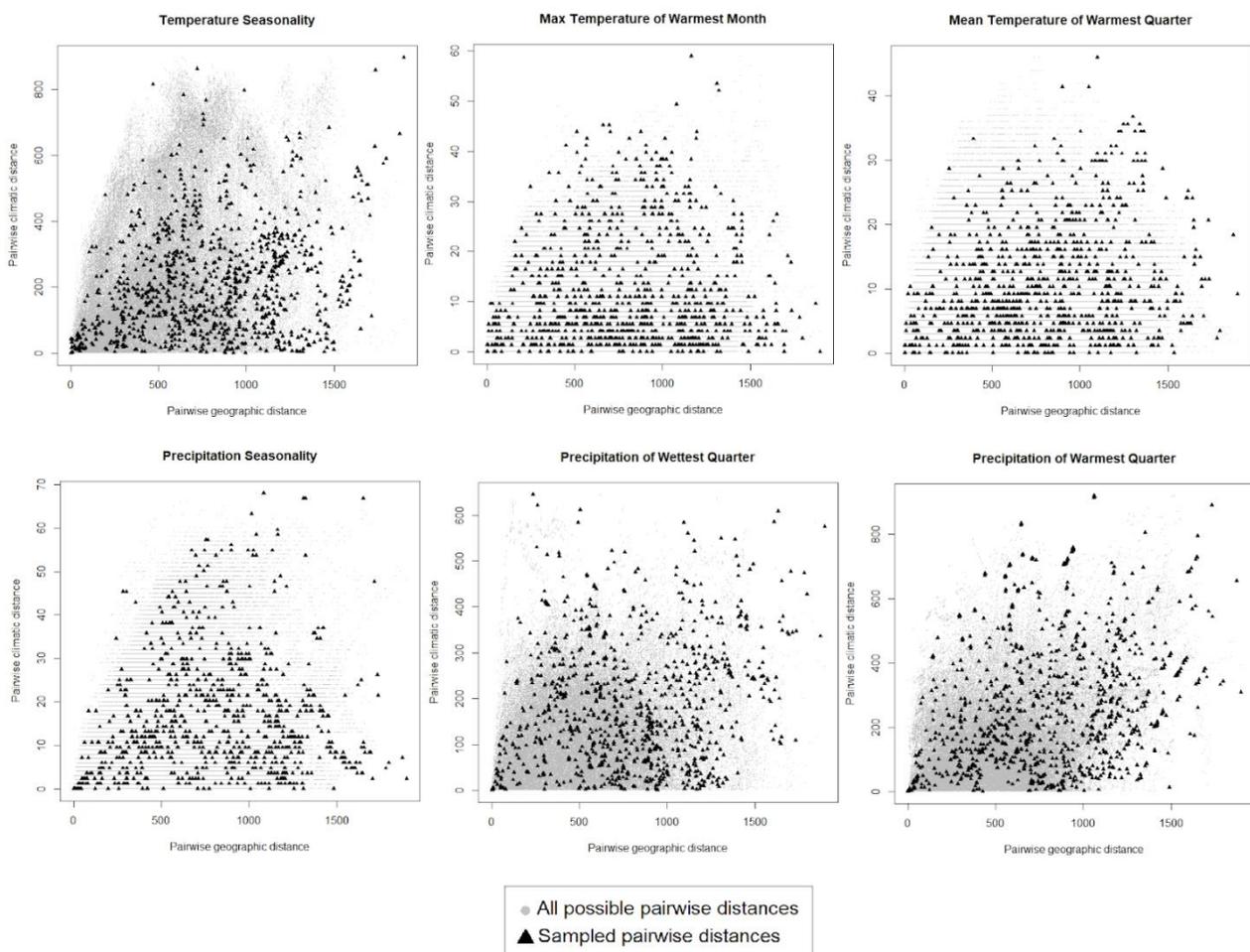


Figura 5. Comparação das distâncias de cada variável climática e distâncias geográficas em pares de populações. Da direita para esquerda respectivamente: Sazonalidade da Temperatura; Máxima Temperatura do mês mais quente; Temperatura Média do Trimestre Mais Quente; Sazonalidade da Precipitação; Precipitação do Trimestre Mais Úmido; Precipitação do Trimestre Mais Quente.

Conclusões

Os objetivos do trabalho foram devidamente alcançados, uma vez que foram definidos os locais de amostragem da *B. excelsa* na Amazônia brasileira, de modo que estes representassem todo o gradiente das variáveis climáticas. E, principalmente, que o delineamento amostral definido permitiu que os efeitos de correlação entre Isolamento por Distância e Isolamento por Ambiente fossem reduzidos. Com base nos resultados desse estudo, as expedições de coleta estão sendo realizadas para o estudo da genômica da paisagem da castanheira-do-brasil.

Bibliografía

Bivand, R., Lewin-Koh, N. (2017). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.9-2. <https://CRAN.R-project.org/package=maptools>

De Mita, S., Thuillet, A., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, 22(5), 1383-1399. doi: 10.1111/mec.12182

Dixo, M., Metzger, J., Morgante, J., & Zamudio, K. (2009). Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biological Conservation*, 142(8), 1560-1569. doi: 10.1016/j.biocon.2008.11.016

Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J. (2017). *dismo: Species Distribution Modeling*. R package version 1.1-4. <https://CRAN.R-project.org/package=dismo>

Hijmans, R. J. (2017). *raster: Geographic Data Analysis and Modeling*. R package version 2.6-7. <https://CRAN.R-project.org/package=raster>

Holderegger, R., & Wagner, H. (2008). *Landscape Genetics*. *Bioscience*, 58(3), 199-207. doi: 10.1641/b580306

Lotterhos, K., & Whitlock, M. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031-1046. doi: 10.1111/mec.13100

Meirmans, P. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839-2846. doi: 10.1111/j.1365-294x.2012.05578.x

Meirmans, P. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, 24(13), 3223-3231. doi: 10.1111/mec.13243

Nepstad, D., Lefebvre, P., Lopes da Silva, U., Tomasella, J., Schlesinger, P., & Solorzano, L. et al. (2004). Amazon drought and its implications for forest flammability and tree growth: a basin-wide analysis. *Global Change Biology*, 10(5), 704-717. doi: 10.1111/j.1529-8817.2003.00772.x

Pádua, J. A. R. (2011) *Genética da Paisagem: Áreas prioritárias para manejo e a conservação de Eremanthus erythropappus (DC.) Macleish no estado de Minas Gerais*. Dissertação (Mestrado em Engenharia Florestal). Universidade Federal de Lavras, UFLA, Lavras, Minas Gerais, Brasil.

- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rieseberg, L., & Burke, J. (2001). The Biological Reality of Species: Gene Flow, Selection, and Collective Evolution. *Taxon*, 50(1), 47. doi: 10.2307/1224511
- Räsänen, K., & Hendry, A. (2008). Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, 11(6), 624-636. doi: 10.1111/j.1461-0248.2008.01176.x
- Sexton, J., Hangartner, S., & Hoffmann, A. (2013). Genetic isolation by environment or distance: Which pattern of gene flow is most common?. *Evolution*, 68(1), 1-15. doi: 10.1111/evo.12258
- Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., & Neale, D. (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, 9(4), 901-911. doi: 10.1007/s11295-013-0596-x
- Sujii, P., Martins, K., Wadt, L., Azevedo, V., & Solferini, V. (2015). Genetic structure of *Bertholletia excelsa* populations from the Amazon at different spatial scales. *Conservation Genetics*, 16(4), 955-964. doi: 10.1007/s10592-015-0714-4
- The IUCN Red List of Threatened Species. Version 2016-3. <www.iucnredlist.org>. Downloaded on 08 March 2017.
- Thomas, E., Alcázar Caicedo, C., Loo, J., Kindt, R. (2014). The distribution of the Brazil nut (*Bertholletia excelsa*) through time: from range contraction in glacial refugia, over human-mediated expansion, to anthropogenic climate change. *Boletim do Museu Paraense Emílio Goeldi. Ciências Naturais*, 9(2): 267-291.
- Thomas, E., Alcázar Caicedo, C., McMichael, C., Corvera, R., & Loo, J. (2015). Uncovering spatial patterns in the natural and human history of Brazil nut (*Bertholletia excelsa*) across the Amazon Basin. *Journal Of Biogeography*, 42(8), 1367-1382. doi: 10.1111/jbi.12540
- Tonini, H., Costa, P., & Kaminski, P. (2008). Estrutura e Produção de duas populações nativas de Castanheira-do-brasil (*Bertholletia excelsa* O. Berg) em Roraima. *Floresta*, 38(3). doi: 10.5380/uf.v38i3.12410
- Wang, I., & Bradburd, G. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649-5662. doi: 10.1111/mec.12938
- Wang, I., Glor, R. E., Losos, J. B. (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, 16, 175-182. doi:10.1111/ele.12025
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- WORLDCLIM, s. d. Disponível em: <www.worldclim.org/current>. Acesso em: 7 Março 2017.

Agradecimientos

Contribuíram para indicação de locais de coleta diversos pesquisadores das unidades da EMBRAPA nos estados de AC, AM, AP, PA, RO, RR e MT; e da Universidade do Oeste do Pará (UFOPA).

Agradecemos especialmente aos colegas Ricardo Scoles, Susan Aragon, Raquel Rodrigues Santos, Lucieta Guerreiro Martorano, Kátia Emídio da Silva, Marcelino Guedes Carneiro, Aisy Baldoni, Eulalia Sobreira e Tatiana de Campos.

Financiamento

Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Projeto EcogenCast: Ecología e genética da castanheira (*Bertholletia excelsa* Bonpl.) como subsídio à conservação e uso sustentável da espécie. Processo SEG: 02.13.05.017.00.00 (auxílio financeiro). Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), bolsa PIBIC para GG.