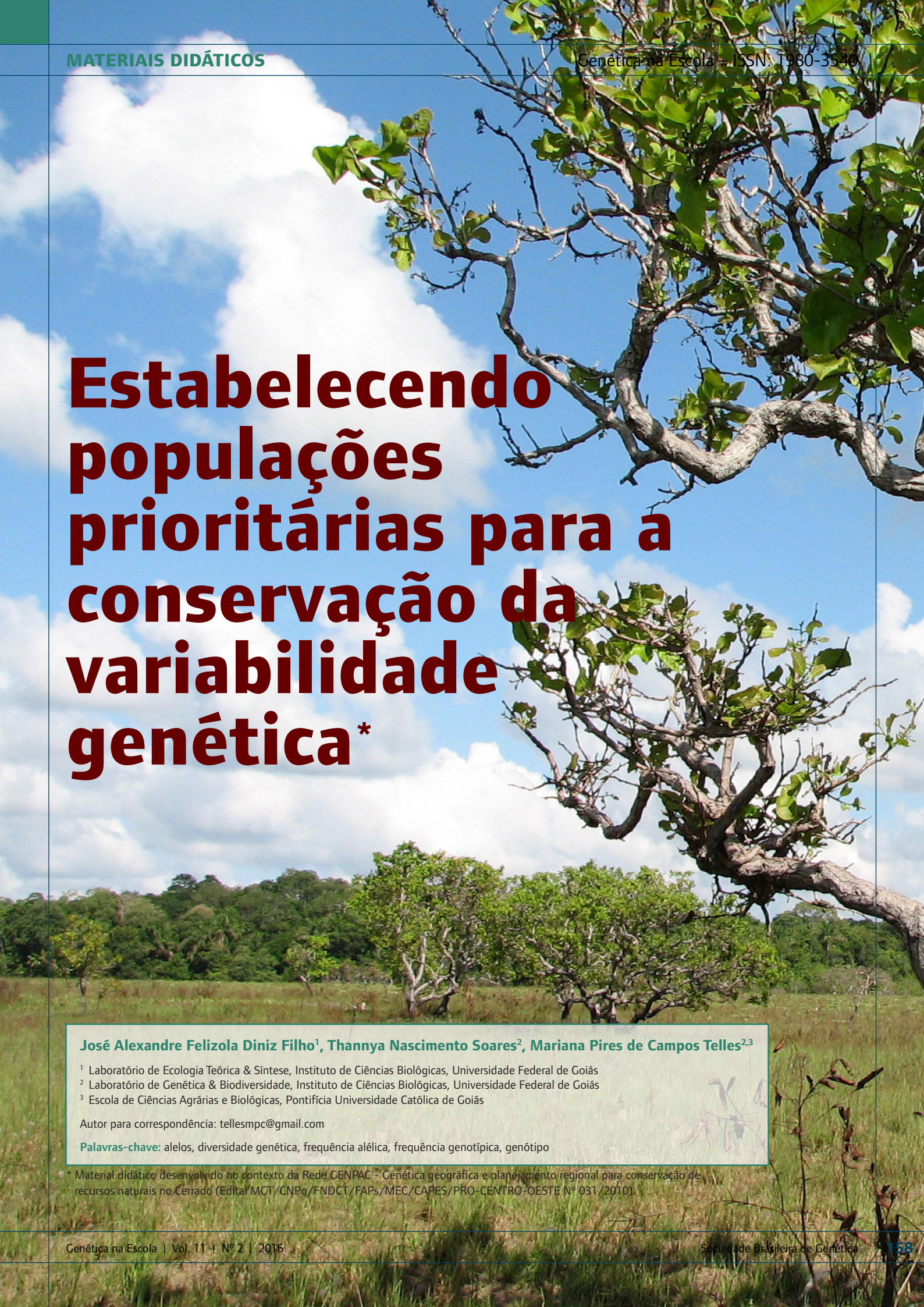
Genética na Escola - ISSN: 1980-3540



Estabelecendo populações prioritárias para a conservação da variabilidade genética*

José Alexandre Felizola Diniz Filho[1], Thannya Nascimento Soares[2], Mariana Pires de Campos Telles[2,3]

[1] Laboratório de Ecologia Teórica & Síntese, Instituto de Ciências Biológicas, Universidade Federal de Goiás
[2] Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás
[3] Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás

Autor para correspondência: tellesmpc@gmail.com

Palavras-chave: alelos, diversidade genética, frequência alélica, frequência genotípica, genótipo

* Material didático desenvolvido no contexto da Rede GENPAC - Genética geográfica e planejamento regional para conservação de recursos naturais no Cerrado (Edital MCT/CNPq/FNDCT/FAPs/MEC/CAPES/PRO-CENTRO-OESTE Nº 031/2010).

Genética na Escola – ISSN: 1980-3540

Nesta proposta apresentamos, em um contexto de Genética da conservação, um método simples para o estabelecimento de áreas prioritárias para conservação da variabilidade genética, possibilitando aos estudantes de graduação ou pós-graduação perceber a importância de utilizar métodos matemáticos e computacionais para delinear estratégias de conservação da biodiversidade. O exemplo utilizado para o exercício diz respeito à utilização de métodos para o estabelecimento de populações prioritárias para a conservação, de uma espécie hipotética de plantas no Cerrado, levando-se em consideração também o nível de antropização em diferentes regiões do bioma. Para estudantes mais avançados ou em cursos nos quais seja possível uma integração entre Genética e Bioestatística, foi proposto um *script* na plataforma computacional "R" para realizar os procedimentos apresentados.

FUNÇÃO PEDAGÓGICA

Colocar o estudante em contato com novos aspectos conceituais e metodológicos na área da **Biologia da Conservação**, utilizando dados genéticos obtidos por marcadores moleculares, a fim de estabelecer estratégias de conservação da biodiversidade. Mais especificamente, é importante que o estudante compreenda que: 1) existem métodos objetivos para estabelecer prioridades de conservação utilizando procedimentos matemáticos e estatísticos, ao invés de usar estratégias oportunistas ou simplesmente baseadas em uma percepção subjetiva da importância da biodiversidade; 2) esses métodos podem ser utilizados para analisar a biodiversidade em diferentes níveis da hierarquia biológica, desde o nível de genótipos (alelos) até o nível de espécies ou características da paisagem e do ecossistema, permitindo construir "redes" ou "conjuntos" de unidades espaciais (populações, locais, regiões etc) que permitam representar ou conservar a diversidade biológica com maior eficiência e o menor custo possível.

PROBLEMA PROPOSTO

Em função do crescente impacto das atividades humanas e da consequente perda de diversidade em diversos níveis da hierarquia biológica, passou a ser necessário estabelecer prioridades para a conservação. De modo geral, o objetivo é encontrar soluções que permitam conservar a **biodiversidade** no nível de alelos nas populações para tentar garantir sua persistência com o menor custo possível, sem restringir o desenvolvimento de atividades humanas (tais como agricultura, criação de empreendimentos ou expansão urbana) que são consideradas igualmente importantes para o bem-estar ou qualidade de vida da espécie humana.

Conservar a biodiversidade é importante por uma série de razões, indo desde questões éticas, morais e filosóficas, até razões mais operacionais sob o ponto de vista mais "utilitarista", com as quais a conservação (de **recursos genéticos**, por exemplo) pode ser importante para a persistência da própria espécie humana, fornecendo produtos naturais que podem ser importantes para a indústria farmacêutica ou alimentícia, ou provendo diversos serviços ecossistêmicos.

A variabilidade em nível genético-molecular, frequentemente observada ou avaliada em nível de populações, é importante já que populações com elevada diversidade genética têm mais chances de persistir diante de mudanças no ambiente. Ao mesmo tempo, essa variabilidade genética pode estar ligada a funções biológicas da espécie que podem, por exemplo, ser importantes em um contexto de melhoramento genético. Assim, nesse contexto, surgiu, a partir da década de 1980, uma nova área dentro da Genética, a chamada **Genética da Conservação**.

Assumindo que é importante conservar a variabilidade genética, mas ao mesmo tempo, essa variabilidade não é constante ao longo da distribuição geográfica da espécie (um fenômeno conhecido dos geneticistas desde o início do século XX), podem ser utilizados procedimentos de **planejamento sistemático em conservação** (SCP, do inglês "*Systematic Conservation Planning*") (MARGULES; PRESSEY, 2000) para estabelecer qual o menor conjunto de populações necessário para conservar (ou representar) a variabilidade genética conhecida para uma dada espécie. Essas populações passariam a ser prioritárias tanto para estratégias de conservação "*in situ*", orientando a criação de reservas biológicas em áreas prioritárias, quanto para estratégias de conservação "*ex-situ*", direcionando a coleta de material biológico para montagem de **coleções de germoplasma** (DINIZ-FILHO *et al.*, 2012).

Imagine-se que, para a região do Cerrado no Brasil Central (Figura 1), foram amostradas 10 populações de uma espécie de planta com potencial de utilização (por exemplo: frutíferas e medicinais), e que essa amostragem é suficiente para conhecer a maior parte da variabilidade genética da espécie. Os indivíduos dessas 10 populações foram avaliados para um **marcador microssatélite** (um lócus), exibindo um total de 8 alelos (A1 a A8). A presença ou ausência de cada um dos alelos em cada população está representada como valores de 1 (um) ou 0 (zero), respectivamente, como descrito na Tabela 1. O valor de 1 é colocado na tabela se, pelo menos, um indivíduo da população possui o alelo (ou seja, se a frequência do alelo é diferente de zero na população). Não importa, em um

Biologia da Conservação: "É uma ciência multidisciplinar que foi desenvolvida como resposta à crise com a qual a diversidade biológica se confronta atualmente". Tem como um dos objetivos principais entender os efeitos da atividade humana sobre as espécies e desenvolver abordagens práticas para prevenir a extinção de espécies (SOULÉ, 1985).

Genética da conservação: "O uso das análises genéticas moleculares para elucidar aspectos da biologia das espécies relevantes para a conservação" (FRANKHAM *et al.*, 2008).

Planejamento sistemático em conservação – área da Biologia da Conservação que procura sistematizar os procedimentos para estabelecer prioridades para conservação, usualmente em um contexto geográfico. O objetivo é estabelecer quais as regiões ou locais prioritários para tal sistematização a partir de dados de biodiversidade existentes nesses locais, em geral maximizando a complementaridade entre esses locais (ou seja, evitando ao máximo a repetição de características) (MARGULES; PRESSEY, 2000).

Coleções de germoplasma: As coleções são compostas por conjuntos de indivíduos representativos da variabilidade genética da espécie que será o objeto da conservação (BORÉM, 2009).

Biodiversidade: "Diversidade biológica significa a variabilidade de organismos vivos de todas as origens, compreendendo, dentre outros, os ecossistemas terrestres, marinhos e outros ecossistemas aquáticos e os complexos ecológicos de que fazem parte; compreendendo ainda a diversidade dentro de espécies, entre espécies e de ecossistemas." (MMA/CDB, 2000).

Marcadores microssatélites – marcadores genéticos de regiões microssatélites do genoma, amplamente utilizados para estudos genético-populacionais.

Recursos genéticos: "material genético de valor real ou potencial" (MMA/CDB, 2000).

primeiro momento, se esses alelos se encontram em regiões neutras ou adaptativas do genoma, uma vez que a metodologia pode ser utilizada para qualquer região do genoma que apresente um marcador molecular associado. O problema, objetivamente, é encontrar o menor número de populações que forme um subconjunto que contenha todos os alelos pelo menos uma vez.

ORIENTAÇÃO PARA O PROFESSOR

O professor deve inicialmente apresentar o contexto teórico da Genética da Conservação aos estudantes (ver, por exemplo, FRANKHAM *et al.*, 2008) e colocar o problema concreto apresentado neste trabalho (i.e., qual o menor subconjunto de populações necessário para representar toda a variabilidade genética, representada aqui pelo número de alelos da população), utilizando os dados da Tabela 1 e da Figura 1. É importante que, em um primeiro momento, o professor certifique-se de que os estudantes compreendam esse contexto e, mais importante, que entendam o significado do conjunto de dados da Figura 1 e Tabela 1 (ou seja, 1 significa que o alelo está presente, e 0, está ausente).

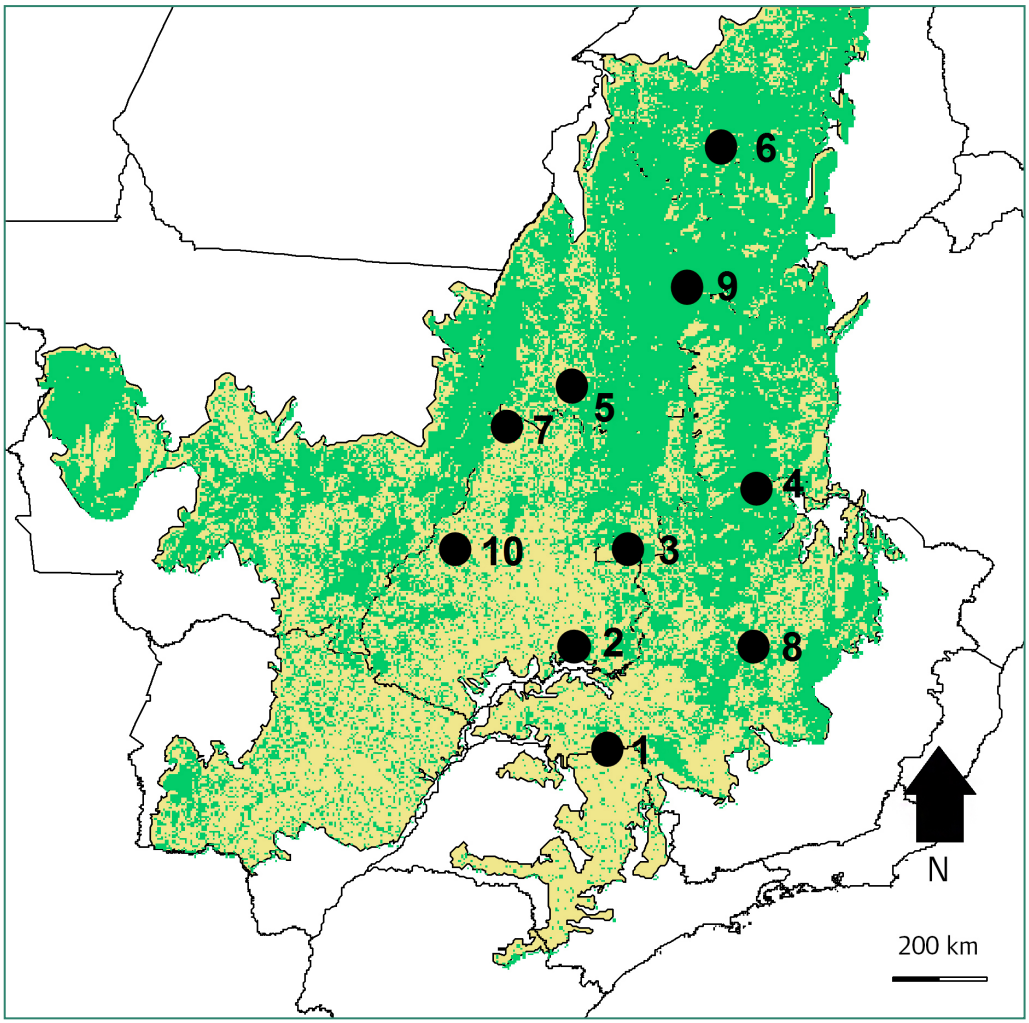


Figura 1.
Mapa base da região do Cerrado, mostrando a localização de 10 populações de uma espécie hipotética de planta para as quais foram obtidos os dados genéticos da Tabela 1. A região em verde indica remanescentes de vegetação natural de Cerrado (com menor nível de antropização).

População	Alelo 1	Alelo 2	Alelo 3	Alelo 4	Alelo 5	Alelo 6	Alelo 7	Alelo 8
1	0	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0
3	1	1	1	0	1	0	0	1
4	0	0	1	0	0	0	1	1
5	1	1	1	0	1	0	0	0
6	1	1	1	1	0	0	0	0
7	0	1	1	1	1	0	0	0
8	0	0	1	1	0	0	0	0
9	0	0	0	1	1	1	0	1
10	0	0	0	1	1	1	0	1

Tabela 1.
Dados de presença (1) e ausência (0) de alelos de um lócus de marcador microssatélite em cada uma das 10 populações indicadas da Figura 1.

Após apresentar o problema, sugere-se que o professor não apresente de imediato o modo de resolver o problema. É importante que haja um tempo para que os estudantes tentem encontrar as soluções, inclusive porque existem soluções alternativas e maneiras diferentes de resolver tal problema. O professor pode sugerir diversas soluções (ver abaixo) e deixar que os estudantes tentem executar algumas delas, avaliando a sua viabilidade.

Em seguida, o professor pode apresentar algumas ideias gerais sobre como resolver o problema. Por exemplo, será que alguma das populações possui todos os alelos? Em caso positivo, o problema já está resolvido (ou seja, basta assegurar a persistência dessa população que todos os alelos estão representados). Esse não é o caso da Tabela 1, mas será que existem combinações de duas-em-duas populações, ou de três-em-três populações que permitam representar todos os alelos? É possível tentar achar todas as combinações possíveis de populações, ou usar um método computacional para tentar, por exemplo, 1.000.000 de combinações, e tentar encontrar a solução usando uma estratégia de "força bruta" (ver DINIZ-FILHO *et al.*, 2016).

Existem métodos específicos que procuram essas soluções usando métodos de inteligência artificial. Essas maneiras de resolver o problema são extremamente eficientes, mas implicam em um grande esforço computacional e, dependendo do tamanho do problema (do tamanho da matriz de dados) podem levar muito tempo. Por outro lado, pode ser utilizado um algoritmo bem mais simples, apelidado de "guloso" ("*greedy*", em inglês). Esse algoritmo é rápido e fácil de ser compreendido, embora em situações complexas não garantam que uma solução com o menor número de populações seja encontrada. De qualquer forma, para entender a lógica de definir prioridade para conservação da variabilidade genética de modo heurístico e para fins didáticos, esse algoritmo será apresentado na próxima seção.

SOLUÇÃO SEQUENCIAL DO PROBLEMA

Após dar oportunidade aos estudantes para discutirem e fazerem propostas de como solucionar o problema, segue a aplicação do algoritmo "guloso". O primeiro passo é calcular quantos alelos estão presentes em cada uma das populações, somando-se as colunas da Tabela 1, e estimando o "número de alelos" em cada uma delas (Painel 1). Assim, pode ser verificado, por exemplo, que a população 1 possui apenas 1 alelo, enquanto a população 3 possui 5 alelos. Como o objetivo é representar todos os alelos com o menor número de populações, assume-se que a população 3 é a primeira a ser preservada (garantindo, portanto, que os 5 alelos, A1, A2, A3, A5 e A8, estejam representados). Se esses 5 alelos da população 3 já estão representados, podem ser excluídas as colunas correspondentes a eles, restando, portanto, os alelos A4, A6 e A7.

Repetindo o procedimento, e possível verificar que as populações 9 e 10 possuem alelos que não são encontrados na população 3. Assim, a população 10 pode ser escolhida, considerando, por exemplo, que ela se encontra um pouco mais próxima, geograficamente, da população 3, que foi escolhida inicialmente. Com isso, está sendo formada uma "rede" contendo duas populações (a população 3 e a população 10) que apresentam, conjuntamente, todos os alelos exceto o alelo 7. Esse alelo 7 só existe na população 4, que passa a ser, portanto, a última população a ser incluída na "rede" de populações prioritárias. Com esse procedimento, é possível concluir que 3 populações (3, 10 e 4) são suficientes para representar, pelo menos uma vez, todos os 8 alelos (Figura 2), e se o objetivo for, por exemplo, montar uma coleção de germoplasma para conservar a variabilidade genética avaliada por esses marcadores moleculares, não é necessário coletar amostras nas 10 populações, bastam essas 3. Assim, torna-se possível minimizar o esforço de coleta e o custo para maximizar o esforço de conservação.

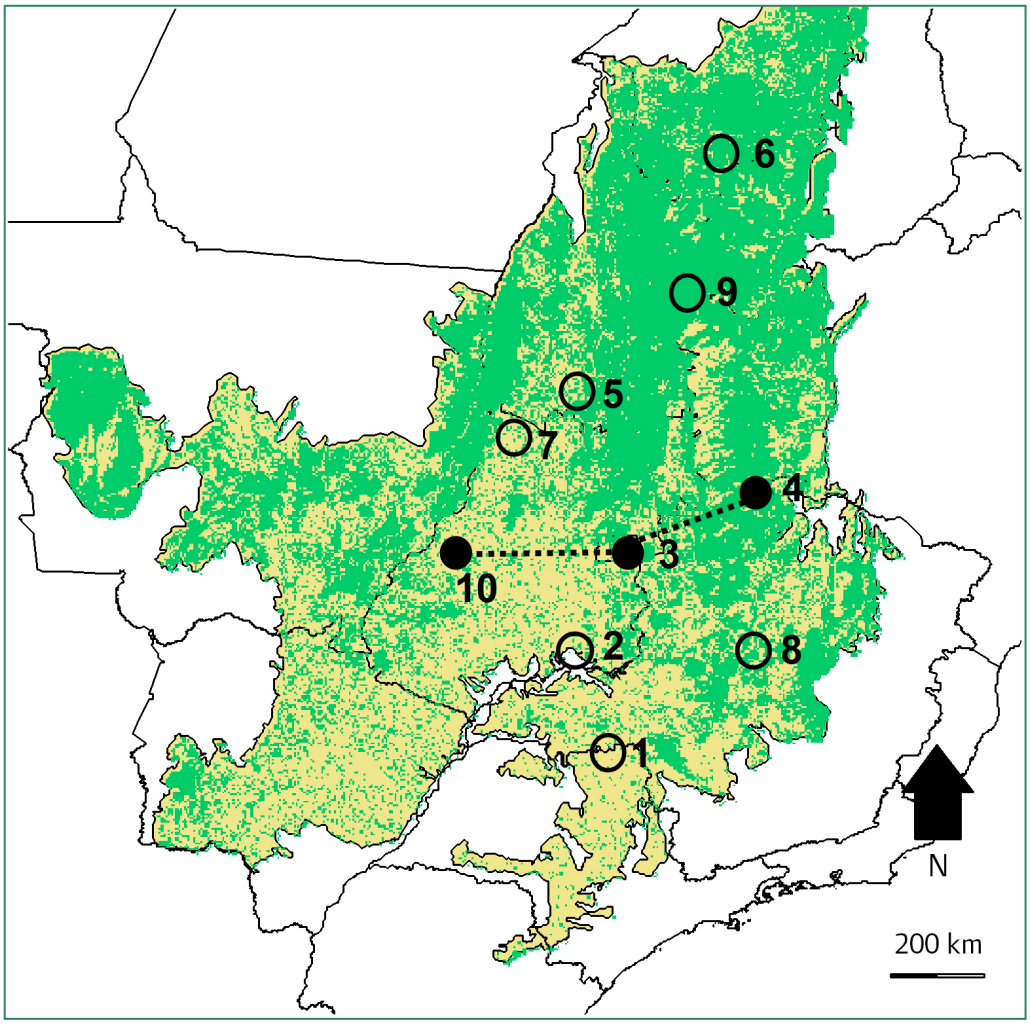


Figura 2.
Representação da rede de populações prioritárias para a conservação da variabilidade genética (círculos preenchidos) obtida pelo algoritmo *"greedy"*, na solução 1, a partir dos dados descritos no Painel 1.

Como sugestão, o professor pode pedir para os estudantes que comecem a resolver o problema a partir da Tabela 1, esclarecendo o algoritmo acima de forma simplificada, adotando os passos a seguir:

a) Calcular o número de alelos encontrados em cada população (soma considerando as colunas de cada população, que será o número total de alelos);

b) Selecionar como prioritária a população com maior número de alelos;

c) Observar quais alelos são encontrados na população prioritária e os exclua da Tabela 1 (rasurando ou excluindo as colunas dos alelos representados, por exemplo);

d) Recalcular o número total de alelos após a exclusão dos alelos presentes na população prioritária, voltando ao passo "a" e repetindo os passos acima até que todos os alelos estejam representados (ou seja, todas as colunas estejam rasuradas ou excluídas).

A fim de estabelecer a rede de populações prioritárias para a conservação, os estudantes devem anotar, em cada passo, qual a população selecionada e quantos alelos foram encontrados nessa população (ou seja, colunas riscadas ou excluídas), correspondendo aos passos "b" e "c".

TORNANDO O PROBLEMA MAIS COMPLEXO

No mapa da Figura 1 é possível ver que uma região em verde, que representa remanescentes de vegetação natural do Cerrado e as áreas em marrom claro, que já apresentam um elevado nível de alteração das áreas de vegetação natural. Com isso, pode ser questionado se é possível, ou se é economicamente viável, manter as populações da espécie de interesse em regiões com elevado nível de antropização. Assim, o problema apresentado pode se tornar um pouco mais complexo e passar a incorporar mais uma variável, que é o nível de alteração das áreas naturais de vegetação pela ocupação humana no Cerrado. Isso pode ser feito de forma quantitativa e usando variáveis que expressem numericamente essa ocupação humana mas, por simplificação, pode-se assumir que algumas das populações estão em regiões (marrons) com alto nível de ocupação humana e que não seria viável manter populações naturais nessas regiões pelo risco que elas têm em perder os alelos pela deriva genética, por exemplo.

A população que possui maior número de alelos (população 3, com 5 alelos), encontra-se em uma região com elevada ocupação humana, próxima ao Distrito Federal. Com isso, a população 3 não será considerada como a população inicial no algoritmo, em função desse fator indesejável. Uma possibilidade então seria escolher, por exemplo, populações que apresentem 4 alelos (que seriam as populações 5, 6, 7 e 9). Dessas populações que apresentam 4 alelos, apenas a população 6, mais ao norte do Cerrado, encontra-se em uma região mais bem preservada (verde).

Para dar início ao algoritmo, a população 6 será escolhida e os passos apresentados anteriormente devem ser repetidos (Painel 2). A população 9 poderia ter sido escolhida também, mas já que ela se encontra mais próxima à região antropizada, optou-se pela população 6. Como os alelos A1, A2, A3 e A4 estão representados na população 6, essas 4 colunas devem ser excluídas, faltando apenas os alelos de A5, A6, A7 e A8. O número total de alelos deve ser recalculado para esse novo conjunto, exibindo o número de alelos que não estão representados na população anterior (a população 6). Neste caso pode ser observado mais uma vez um "empate", ou seja, tanto a população 9 quanto a 10 apresentam 3 alelos (os alelos A5, A6 e A8).

Seguindo com a utilização do mesmo critério em relação ao nível de ocupação humana na área, percebe-se que a população 10, selecionada na seção anterior, à oeste do bioma Cerrado, encontra-se em uma região com elevado nível de ocupação humana. Com isso, a população 9 pode ser escolhida e, após isso, pode ser verificado que falta representar apenas o alelo A7, que só está presente na população 4.

A nova "rede" de populações, agora considerando tanto o menor conjunto possível para representar todos os alelos, quanto as regiões com menor nível de ocupação humana no Cerrado, é formada pelas populações 6,

9 e 4 (Figura 3). Note-se que a população 4 está presente nas duas soluções, o que é esperado já que ela é a única que apresenta o alelo 7. Assim, a população 4 deve ser considerada "insubstituível" nesse contexto de conservação, pois deve estar presente em qualquer solução que busque representar todos os alelos.

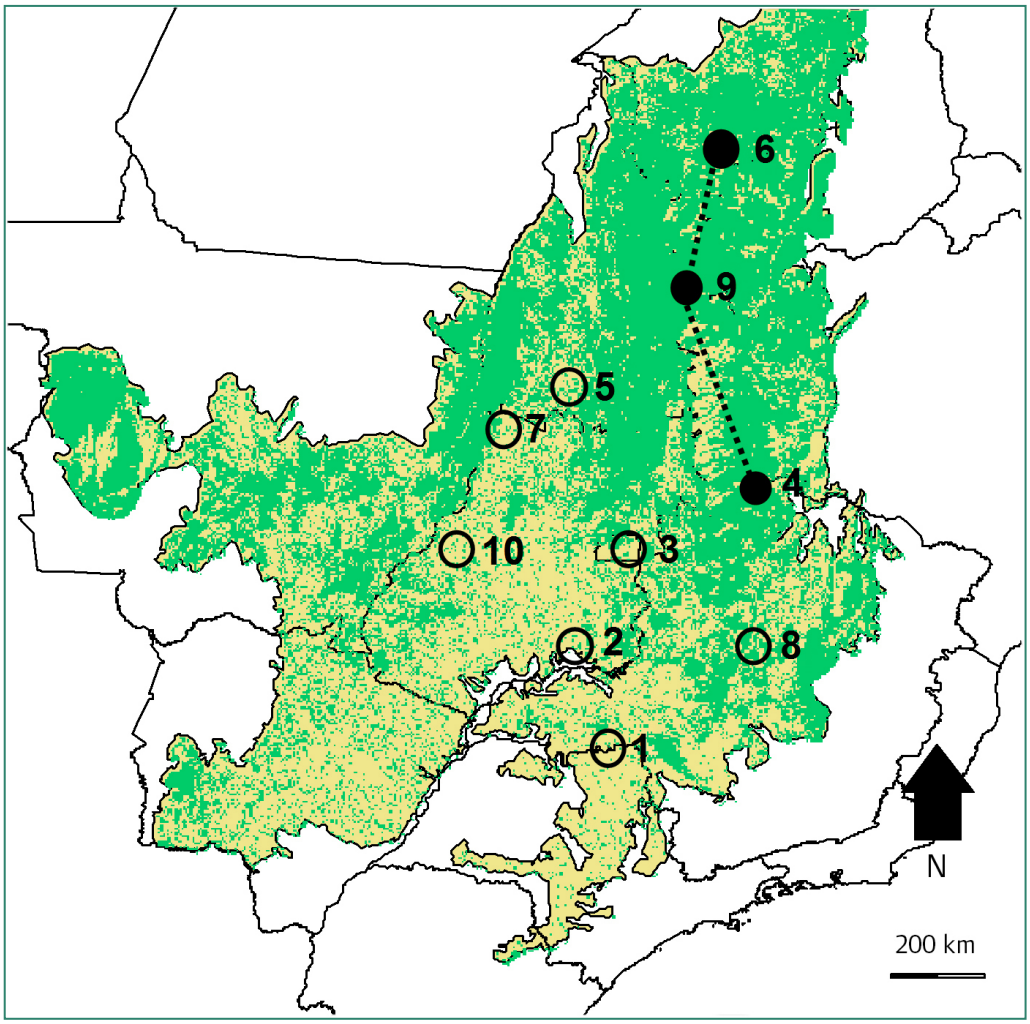


Figura 3.
Representação da rede de populações prioritárias para a conservação da variabilidade genética (círculos preenchidos) obtida pelo algoritmo "greedy", na solução 2, a partir dos dados descritos no Painel 2, no qual a escolha leva em consideração a região (com maior ou menor nível de antropização) onde a população se encontra.

As soluções 1 (populações 3, 10, 4) e 2 (populações 6, 9, 4) são equivalentes no sentido de representar todos os alelos com o menor número de populações possível, mas é importante destacar que a segunda solução é melhor, uma vez que as populações representadas estão situadas em regiões de menor impacto da ocupação humana. Isso pode ser interessante, sob diversos aspectos, se o objetivo é criar reservas ou parques naturais, que irão persistir durante muitos anos e manter a viabilidade das populações em longo prazo. Por outro lado, outras soluções podem ser possíveis. Se o objetivo é coletar material para montar uma coleção de germoplasma, pode ser mais viável diminuir os custos financeiros de viagem para a coleta e não se preocupar se as populações persistirão de forma "natural" e, neste caso, a primeira solução poderia ser mais interessante.

Para conjuntos de dados reais, com um número maior de populações e alelos, torna-se mais complexo aplicar o algoritmo "greedy" e é preciso utilizar programas computacionais. No Box 1 está apresentado um "script" em linguagem R (R Core Team,2015; ver www.r-project.org), que é uma plataforma

computacional livre ("freeware"), que permite a realização de diversas análises estatísticas. Pode ser uma boa oportunidade para realizar atividades integradas entre disciplinas de Bioestatística (que podem já utilizar a plataforma R) e Genética.

Para utilizar o "script", a matriz de dados da Tabela 1 deve ser colocada em um arquivo texto (ASC), denominada "exemplo_gen_esc.txt" (esse é o arquivo de entrada, que pode ser alterado). O objeto RESULTADOS no final do "script" contém o número de alelos preservados em cada passo e o conjunto mínimo de populações na solução. Em seguida, encontra-se o "script" para gerar soluções aleatórias com o mesmo número mínimo de populações encontradas no algoritmo "greedy", para o mesmo conjunto de dados, gerando ao final o histograma da Figura 4 (ver próxima seção) e a probabilidade de encontrar o número de alelos escolhendo populações ao acaso.

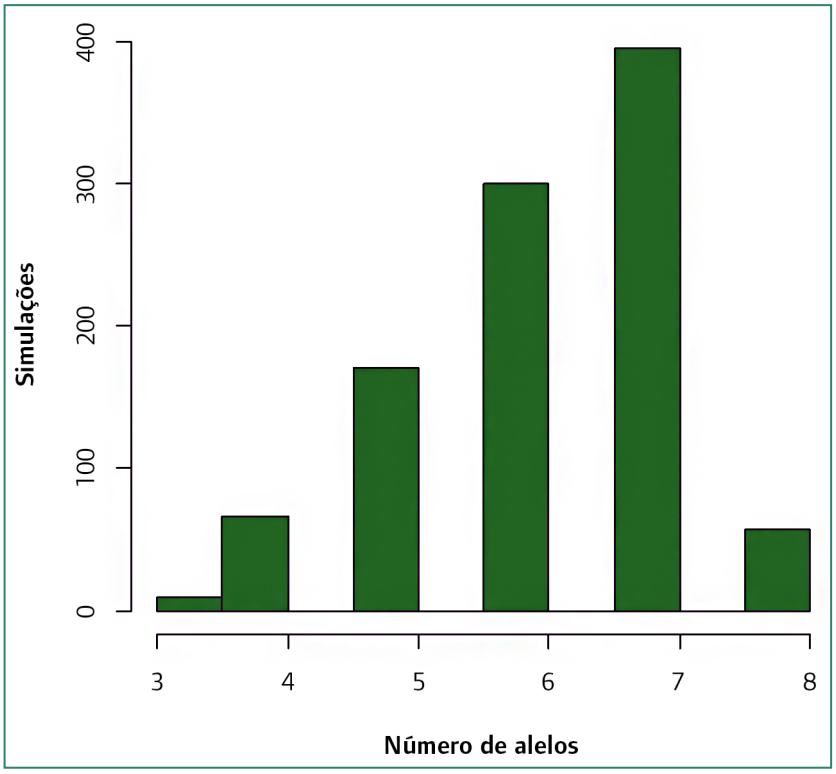
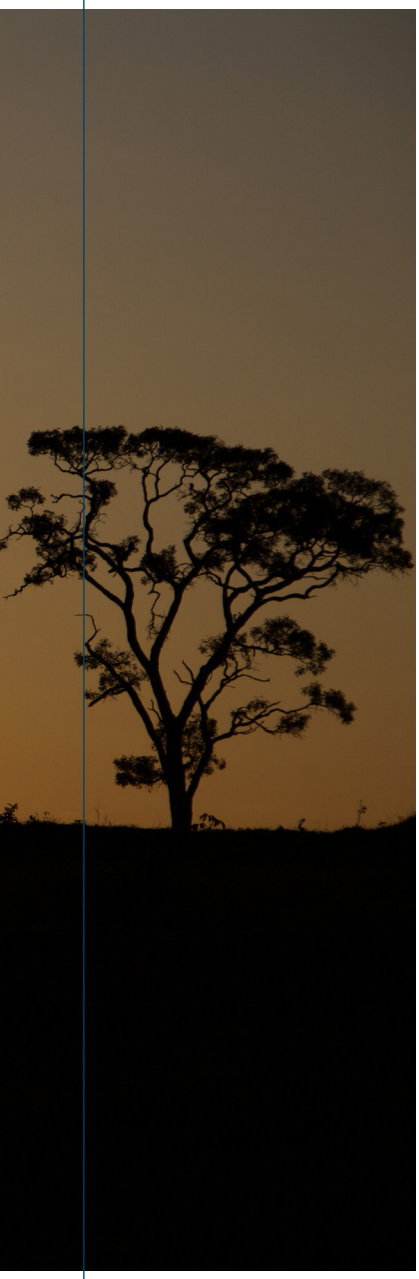


Figura 4.
Histograma com o número de alelos representados em 1000 soluções que amostram, ao acaso, 3 das 10 populações, mostrando que a chance de encontrar ao acaso um conjunto de populações que contenha os 8 alelos, para o conjunto de dados hipotéticos deste exemplo, é pequena (em torno de 5%).

AVALIANDO A EFICIÊNCIA DA ESCOLHA DE POPULAÇÕES

Uma possibilidade de trabalho adicional, ligado às soluções alternativas, é tentar avaliar a "eficiência" do algoritmo utilizado. Uma vez que os estudantes encontraram que 3 é o menor número de populações necessárias para representar os 8 alelos pelo menos uma vez, a pergunta que se pode fazer é: qual a chance de, selecionando ao acaso 3 populações, esse subconjunto represente os 8 alelos?

Os estudantes podem ser distribuídos em grupos, ou cada um deles pode tentar algumas soluções, anotando os resultados. Por exemplo, se as populações 7, 10 e 5, forem "preservadas", estarão sendo representados 6 alelos, ao passo que, se as populações 2, 3 e 8 forem selecionadas, estarão representados 7 alelos. Ao final, pode ser construído um gráfico com a distribuição, em um histograma, do número de alelos representado ao acaso e cada solução. Se o número de soluções em que o número de 8 alelos foi encontrado (incluindo-se a solução do algoritmo) for dividido pelo número de tentativas, é possível responder à questão a respeito da probabilidade de que 3 populações, tomadas ao acaso, representem toda a variabilidade genética avaliada por este marcador molecu-



lar. Quando utilizado um programa computacional, realizando 1000 soluções ao acaso com 3 populações, é possível verificar que em nenhuma delas podem ser encontrados os 8 alelos, de modo que a chance de, por acaso, encontrar a solução "correta" é em torno de 5% (Figura 4).

Um problema relacionado a esse, em termos de eficiência, seria avaliar um sistema já implementado de unidades de conservação, ou achar uma solução, considerando que já existe uma coleção de germoplasma que contenha parte da variabilidade genética existente na espécie (i.e., alguns alelos). Por exemplo, supondo-se que a população 5 se encontre em uma reserva ou parque nacional, ou que um pesquisador já tenha montado uma coleção de germoplasma a partir dessa mesma população. Desse modo, pode-se assumir que os alelos presentes já estão preservados. Com isso, o algoritmo pode ser iniciado com a população 5 e, em seguida, pode ser encontrado qual o menor subconjunto de populações que será necessário para complementar, da melhor forma possível, a população 5.

CONSIDERAÇÕES FINAIS

Espera-se que o problema apresentado, bem como as interpretações associadas a cada uma das soluções, mostre de forma clara ao estudante que existem métodos objetivos capazes de estabelecer prioridades de conservação da biodiversidade. Certamente, o professor deve chamar a atenção dos estudantes para o problema apresentado que é bastante simples e possui objetivo puramente didático. Na prática, em geral existe uma matriz bem maior de dados (com mais populações, lócuss e alelos) sendo necessário considerar diversos fatores adicionais, além do simples efeito da ocupação humana indicada por remanescentes de vegetação. Os dados genéticos podem, além disso, ser a frequência alélica (e não a "presença" do alelo), sendo também importante avaliar outras propriedades da população, tais como as proporções esperadas pelo equilíbrio de Hardy-Weinberg ou diversidade genética (heterozigose) (ver SCHLOTTFELDT *et al.*, 2015). O aumento da complexidade e do "tamanho" do problema gera a necessidade de utilizar procedimentos computacionais mais sofisticados a fim de encontrar as soluções para o problema, que podem ser inclusive combinadas e apresentadas como "cenários" alternativos para tomadores de decisão que implementarão, na prática, essas soluções, quando for o caso. De qualquer modo, apesar da simplicidade do exemplo apresentado, acredita-se que ele seja importante para destacar as bases conceituais do problema proposto e principais estratégias para solucioná-los.

REFERÊNCIAS

BORÉM, A.; MIRANDA, G. V. *Melhoramento de plantas*. 5.ed. Viçosa: Editora UFV, (2009).

DINIZ-FILHO, J. A. F.; DINIZ, J. V. B. P. L.; TELLES, M. P. C. Exhaustive search for conservation networks of populations representing genetic diversity. *Genetics and Molecular Research* v.15, n.1, Genetics and Molecular Research.15017525, (2016).

DINIZ-FILHO, J. A. F.; MELO D. B.; DE OLIVEIRA, G.; COLLEVATTI, R. G., et al. Planning for optimal conservation geographical genetic variability within species. *Conservation Genetics* v.13, p. 1085-1093, 2012.

FRANKHAM, R.; BALLOU, J. D.; BRISCOE, D. A. *Fundamentos de genética da conservação*. Ribeirão Preto: Sociedade Brasileira de Genética, 2008.

MARGULES, C. R ; PRESSEY, R. L. Systematic conservation planning. *Nature* v. 405, p. 243-253, 2000.

MINISTÉRIO DO MEIO AMBIENTE – MMA A Convenção sobre Diversidade Biológica – CDB. Série Biodiversidade n. 1, 2000. Disponível em: http://www.mma.gov.br/estruturas/sbf_dpg/_arquivos/cdbport.pdf. Acesso: 28-03/2016.

R DEVELOPMENT CORE TEAM. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL https://www.R-project.org/.

SCHLOTTFELDT, S.; MEMT, W.; CARVALHO, A. C. P. L. F, ;SOARES, T. N. ; TELLES, M. P. C. ; LOYOLA, R. D.; DINIZ-FILHO, J. A. F. Multi-objective optimization for plant germplasm collection conservation of genetic resources based on molecular variability. *Tree Genetics Genome* v.11, p.16, 2015.

Genética na Escola – ISSN: 1980-3540

SOLUÇÃO DO PROBLEMA

População	Alelo 1	Alelo 2	Alelo 3	Alelo 4	Alelo 5	Alelo 6	Alelo 7	Alelo 8	Total
1	0	1	0	0	0	0	0	0	1
2	0	1	0	0	0	0	0	0	1
3	1	1	1	0	1	0	0	1	5
4	0	0	1	0	0	0	1	1	3
5	1	1	1	0	1	0	0	0	4
6	1	1	1	1	0	0	0	0	4
7	0	1	1	1	1	0	0	0	4
8	0	0	1	1	0	0	0	0	2
9	0	0	0	1	1	1	0	1	4
10	0	0	0	1	1	1	0	1	4

1

População	Alelo 4	Alelo 6	Alelo 7	Total
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	1	1
5	0	0	0	0
6	1	0	0	1
7	1	0	0	1
8	1	0	0	1
9	1	1	0	2
10	1	1	0	2

2

População	Alelo 7	Total
1	0	0
2	0	0
3	0	0
4	1	1
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0

Painel 1.
Sequência da busca de uma solução do algoritmo *"greedy"* a partir da Tabela 1. A população com maior riqueza alélica (população 3 – ver última coluna) é selecionada, e os alelos que são encontrados nela (alelos 1, 2, 3, 5 e 8) são excluídos, repetindo-se o procedimento até que todos os alelos sejam representados (subconjunto de populações 3, 10 e 4).

População	Alelo 1	Alelo 2	Alelo 3	Alelo 4	Alelo 5	Alelo 6	Alelo 7	Alelo 8	Total
1	0	1	0	0	0	0	0	0	1
2	0	1	0	0	0	0	0	0	1
3	1	1	1	0	1	0	0	1	5
4	0	0	1	0	0	0	1	1	3
5	1	1	1	0	1	0	0	0	4
6	1	1	1	1	0	0	0	0	4
7	0	1	1	1	1	0	0	0	4
8	0	0	1	1	0	0	0	0	2
9	0	0	0	1	1	1	0	1	4
10	0	0	0	1	1	1	0	1	4

1

População	Alelo 5	Alelo 6	Alelo 7	Alelo 8	Total
1	0	0	0	0	0
2	0	0	0	0	0
3	1	0	0	1	2
4	0	0	1	1	2
5	1	0	0	0	1
6	0	0	0	0	0
7	1	0	0	0	1
8	0	0	0	0	0
9	1	1	0	1	3
10	1	1	0	1	3

2

População	Alelo 7	Total
1	0	0
2	0	0
3	0	0
4	1	1
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0

Painel 2.
Algoritmo *"greedy"* no qual a ocorrência da população em uma região com elevado nível de antropização (região marrom na Fig. 1) passa a ser um fator de restrição na escolha das populações. Ao invés de selecionar a população 3, inicia-se o algoritmo com a população 6 (pode-se escolher também a população 9), que é a segunda com maior número de alelos e que, ao mesmo tempo, se encontra em uma região com muitos remanescentes de vegetação. As demais populações com 4 alelos (populações 5, 7 e 10) encontram-se em regiões muito antropizadas (marrom).

BOX 1 – ESTABELECENDO REDES DE POPULAÇÕES USANDO A PLATAFORMA "R"

```
data <-as.matrix(read.table("exemplo_gen_esc.txt", h=T))
#GREEDY
pa <-as.matrix(read.table("exemplo_gen_esc.txt",h=T))
npop <-nrow(pa)
nall <-numeric()
poppr <-numeric()
run <-numeric()
rundel <-numeric()

paorig <-pa

for(i in 1:(npop)){
        nall[i] <-ncol(pa)
        richness <-apply(pa,1,sum)#check...
        range <-apply(pa,2,sum)#check...
        maxrich <-max(richness)
        popmax <-which(richness==maxrich)

        if(length(popmax)==1){
                del=popmax} else
                {del <-sample(popmax,1)}

        alldel <-which(pa[del,]==1)

        pa <-as.matrix(pa[,-alldel])
        rundel[i]=del
        run[i] <-i
        poppr[i] <-nall[i]-ncol(pa)
}

resul <-cbind(run,nall,poppr,rundel)
colnames(resul)<-c("Run","No. Alel","No.Alel.Repres.","Pop.Prioritaria")

r <-which(resul[,3]>0)
repr <-resul[1:length(r),3]/ncol(paorig)

RESULTADOS <-resul[r,]
allcumul <-cumsum((resul[,3])/ncol(paorig))
popsel <-resul[r,4]
minpop <-length(popsel)

#GERANDO SOLUÇÕES AO ACASO PARA O NUMERO DE POPULACOES NA SOLUÇÃO
sel <-seq(nrow(data))
k <- minpop
nsim <-1000

ALR <-numeric()
for(i in 1:nsim){
        spop <-sample(sel,k,replace=F)
        redata <-data[spop,]
        ALR[i] <-sum(ifelse(apply(redata,2,sum)>=1,1,0))
}
sum(ifelse(ALR==8,1,0)) / nsim
hist(ALR,xlab="Número de alelos",ylab="Simulações",col="green",main="")
```