

Proposta escolhida

Frente à devolutiva do monitor responsável pela avaliação das minhas propostas, aplicabilidade da função e familiaridade com o tema, decidi prosseguir com a **Proposta A: Suficiência amostral em trabalhos com diversidade genética**.

Contextualização

Tamanho amostral adequado é essencial em qualquer estudo empírico cujo objetivo seja fazer inferências sobre a população a partir de uma amostragem¹⁾. Em geral, avaliar a suficiência amostral é uma prática frequente em trabalhos de ecologia. No entanto, em outras áreas, como genética de populações, tal abordagem ainda é dificilmente incorporada às análises de dados. Alguns trabalhos tem buscado estabelecer generalizações a respeito do tamanho amostral adequado para estimativa de diversidade genética das populações (e.g. Hale *et al.* 2012)²⁾. Apesar desse esforço, é sabido que essa suficiência amostral pode variar com o tipo de marcador molecular, nível de polimorfismo dos *loci* e história natural das espécies. Sendo assim, a função `sample.suff` desenvolvida tem como objetivo permitir aos pesquisadores interessados em genética de populações avaliar se, dado o conjunto de *loci* microssatélites genotipado, seu esforço amostral foi suficiente para inferir a diversidade genética das populações ou grupos de indivíduos diplóides sob estudo.

Alterações e Justificativas

Ao longo da elaboração do trabalho final, algumas alterações foram necessárias com relação à proposta inicial. Abaixo, encontra-se uma relação dessas alterações seguida da respectiva justificativa.

Proposta	Alteração	Justificativa
Incluir riqueza de alelos privados (pA) como uma das opções de <code>gen.div</code>	Opção removida	pA não depende apenas dos indivíduos amostrados na população, mas também do número de populações amostradas. Logo, não é uma estimativa de diversidade genética intrínseca à população. Portanto, essa opção foi removida
Usar pacotes PopGenReport e adegenet	Usar apenas pacote adegenet	O pacote adegenet permitiu realizar todas as tarefas previstas na função. Portanto, não houve necessidade de usar outro pacote para análises de diversidade genética
Possibilidade de incluir argumento <code>mol.marker</code> ³⁾	Opção não incluída	Conforme previsto na proposta, o argumento <code>mol.marker</code> apenas seria incluído na função caso houvesse tempo disponível. No entanto, a função com os argumentos propostos consumiu bastante tempo para elaboração, tornando inviável a inclusão do argumento <code>mol.marker</code> no prazo previsto ⁴⁾

Proposta	Alteração	Justificativa
No objeto dados, organizar cada <i>locus</i> em duas colunas, uma para cada alelo	Organizar dados de forma que haja uma coluna por <i>locus</i> , com alelos separados por barra (/)	Alteração fez-se necessária para adequada transformação do data frame em objeto da classe <i>genind</i> , visto que esta é a classe de objetos utilizada pelo pacote <i>adeqenet</i> para as análises genéticas
Dentro do 1º ciclo for, estimar diversidade genética para cenários simulados apenas	Além do proposto, no primeiro ciclo for também foram incluídas as estimativas de diversidade genética para as populações observadas (ou seja, fora do contexto das simulações)	Tal procedimento foi feito para reduzir a redundância do código e não ter que criar um outro ciclo for também com contador de i até o número de populações
Gerar subamostragens das populações, usando <code>sample</code> com <code>size=j</code> e <code>replace=TRUE</code>	Usar <code>sample</code> com <code>size=j</code> e <code>replace=FALSE</code>	Evitar que um mesmo indivíduo possa ser amostrado nas populações simuladas, o que enviesaria a estimativa de diversidade genética
Plotar gráfico com linhas correspondentes aos intervalos de confiança	Ao invés de linhas, plotar área sombreada no gráfico, com cor com transparência, correspondente ao intervalo de confiança das estimativas de diversidade genética	Alteração deixou gráfico menos poluído e esteticamente mais bonito
Colocar legendas com cores das populações e o que cada linha indica	Na legenda são mostradas as cores da linha (média) e da área sombreada (intervalo de confiança) e a população correspondente	A interpretação da linha (média da estimativa de diversidade genética) e da área sombreada (intervalos de confiança) encontra-se explicitamente explicada na seção VALUE do help da função

1)

Fonte: https://en.wikipedia.org/wiki/Sample_size_determination

2)

Hale ML, Burg TM, Steeves TE (2012) Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. PLoS ONE 7(9): e45170

3)

daria ao usuário a opção de escolher entre marcadores SNPs e microssatélites

4)

fazer isso demandaria o uso de outra abordagem de análise, bastante diferente da utilizada para marcadores microssatélites

From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2018:alunos:trabalho_final:azevedosilva.m:func



Last update: **2020/08/12 06:04**