

# Kelly



Doutoranda em Genética pelo Instituto de Biociências da Universidade de São Paulo. Orientada pelo prof. Diogo Meyer. Atua na área de genética de populações humanas com ênfase na história evolutiva, demográfica e seletiva das populações nativo-americanas.

## Exercícios

01F 02F 03F 04 05F 06F 07F 08F 09F

## Proposta principal

Criar uma função em R para estimar a Heterozigose Haplótipo Específica - HSH (Haplotype Specific Heterozygosity). A HSH é a heterozigose (diversidade gênica) de um determinado locus adjacente a um dado haplótipo ou a um outro locus específico. O índice HSH é calculado através da estimativa das frequências alélicas de um determinado locus em um dado haplótipo/locus específico. As frequências alélicas são então aplicadas na fórmula  $HSH = 1 - \sum_i p_i^2$ , onde  $k$  corresponde ao número de alelos de um determinado locus e  $p$  a frequência alélica (Malkki et al., 2005). Os resultados serão apresentados em uma tabela com a distribuição sumária da HSH de cada locus em função do haplótipo/locus específico. Esse índice pode ser usado como ferramenta em diversos tipos de estudos como, por exemplo, detecção de alelos vantajosos que foram selecionados positivamente ou detecção de alelos informativos (preditores) de um haplótipo ou locus que tenha associação com doenças. Portanto para ter certeza que os índices de HSH encontrados não são frutos do acaso, proponho colocar uma opção nessa função que permita embaralhar os dados e simular HSH permitindo a comparação dos dados reais com os simulados.

Malkki M, Single R, Carrington M, Thomson G, Petersdorf E. MHC Microsatellite Diversity and Linkage Disequilibrium Among Common HLA-A, HLA-B, DRB1 Haplotypes: Implications for Unrelated Donor Hematopoietic Transplantation and Disease Association Studies. *Tissue Antigens* 2005; 66: 114-124.

## Comentário

A proposta parece boa mas fico incapacitado de avaliar sua viabilidade por não ter informação sobre qual o formato e conteúdo do input na função. Falta definir melhor o objeto de saída, bem como argumentos que serão utilizados.

— [Alexandre Adalardo de Oliveira](#) 2010/03/31 13:14

# Plano B

Fazer uma função para a estimativa do Desequilíbrio de Ligação (associação não aleatória de alelos de diferentes loci).

## Página de ajuda

hsh

package:nenhum

R Documentation

Índice da Heterozigose Haplótipo/Alelo Específica (HSH)

### Description:

Estima o índice de heterozigose de um determinado locus em função de um alelo de outro locus ou haplótipo de interesse. Essa função tem como opção contrastar o hsh observado a partir dos dados com hshs empíricos através da opção `simular`.

### Usage:

```
hsh (x,y,simular=FALSE)
```

### Arguments:

`x`: Vetor numérico. Frequências alélicas de um locus em função de um alelo de outro locus ou haplótipo de interesse.

`y`: vetor numérico. Conjunto de frequências alélicas reais ou distribuição alélica aleatória, a partir da qual os dados de frequência serão embaralhados e reamostrados se o argumento `simular` for igual a "TRUE".

`simular`: simula a partir de `y` o hsh 10.000 vezes.

### Details:

A função primeiro normaliza os valores de `x` em função do tamanho do vetor. Em outras palavras, divide-se as frequências alélicas pelo tamanho do vetor que corresponde ao número de alelos desse locus, observados no alelo em questão de um outro locus ou haplótipo de interesse.

Em seguida, é calculada a estimativa de hsh, que corresponde a um menos a soma do quadrado das frequências alélicas normalizadas. Ou seja, a soma as

probabilidades de sortear dois alelos na amostra e que eles sejam idênticos ( $p^2$ ). Essa soma representa a homozigose e, portanto, ao subtrair de 1 (100%) obtemos a estimativa de heterozigose.

Se o usuário optar por simular os dados, o hsh será calculada 10.000 vezes a partir do conjunto de dados informado pelo usuário (que pode ser desde uma distribuição real de dados observada em outro locus ou ainda uma distribuição de frequências alélicas ao acaso). Esses dados serão embaralhados e então sorteados  $n$  valores. O  $n$  corresponde ao tamanho do vetor real ( $x$ ). A distribuição das 10.000 estimativas de hsh são plotadas em um gráfico (histograma). Em seguida é acrescentada uma linha que corresponde ao valor de hsh observado (valor real).

Para finalizar, a função retorna um pvalue que corresponde à proporção de hsh empírico que são  $<$  que o hsh observado.

Value:

frequências alélicas normalizadas: as frequências de um locus normalizadas em função do alelo de outro locus ou haplótipo de interesse.

hsh: estimativa de heterozigose haplótipo/alelo específica relativo aos dados observados

gráfico: histograma com a distribuição empírica dos valores de hsh gerado a partir dos dados simulados. Esse gráfico será apresentado apenas se o argumento `simular` for igual a `TRUE`

pvalue: proporção de hsh empírico  $<$  que hsh observado.

Warning:

O arquivo de entrada dos dados para a função hsh foi pensado com base no arquivo de saída de programas que estimam a fase haplotípica (Ex. Arlequin, Phase, Pypop). Esses programas apresentam a estimativa geral das frequências alélicas. Por esse motivo a função hsh normaliza as frequências alélicas em função do haplótipo/alelo de interesse e o usuário não precisa entrar com o dado normalizado.

Author(s):

Kelly Nunes

knunes\_bio@yahoo.com.br

References:

Malkki M, Single R, Carrington M, Thomson G, Petersdorf E. MHC Microsatellite Diversity and Linkage Disequilibrium Among Common HLA-A, HLA-

B, DRB1 Haplotypes: Implications for Unrelated Donor Hematopoietic Transplantation and Disease Association Studies. *Tissue Antigens* 2005; 66: 114-124.

See Also:

Examples:

```
hsh(c(0.5,0.3,0.1))
```

```
hsh(c(0.1,0.03,0.05,0.2), c(0,0.01,0.03,0.05,0.07,0.1,0.2,0.5),  
simular=TRUE)
```

Usando dados de uma tabela como input

Nome da tabela = teste

```
|Alelos|Locus3_A|Locus3_B|  
|Locus1_A|0.0377|0|  
|Locus1_B|0.0566|0.1320|  
|Locus1_C|0.0566|0|  
|Locus1_D|0.3018|0.4151|  
|Locus2_A|0.2075|0.3396|  
|Locus2_B|.24528|0.2075|
```

Interesse= Estimar hsh do Locus 1 em função do alelo A do Locus 3. Para simular hsh, os dados das frequências do Locus 1 nos alelos A e B do Locus 3 serão embaralhados e reamostrados.

```
hsh((teste[1:4,2]), (c(teste[1:4,2],teste[1:4,3])), simular=TRUE)
```

## Código da Função

```
hsh<- function(x,y,simular=FALSE)  
{  
  n=length(x)  
  freq.norm<- rep(0,n)  
  hsh.real<- rep(0,n)  
  for (i in 1:n)  
  {  
    freq.norm <- (x/sum(x))  
  }  
  cat("\n\n", "Frequências alélicas normalizadas")  
  print(freq.norm)  
  for (i in 1:n)  
  {  
    pi2<- freq.norm^2  
    hsh.real<- 1-(sum(pi2))
```

```
    }
cat("\n\n", "HSH")
print (hsh.real)
if (similar=="TRUE")
{
y
sample(y,n,replace=T)-> resample
n.resample <- length(resample)
freq.norm.resample<- rep(0,n.resample)
hsh.resample.vector<- rep(0, 10000)
hsh.resample <- 0
for (j in 1:10000)    {
    sample(y,n, replace=T)-> resample
    for (i in 1:n.resample)
    {
    freq.norm.resample<- (resample/sum(resample))
    pi2.resample<- freq.norm.resample^2
    hsh.resample<- 1-(sum(pi2.resample))
    hsh.resample.vector[j]<- hsh.resample
    }}
table(hsh.resample.vector>=hsh.real)-> p
pvalue <- p[[1]]/(p[[1]]+p[[2]])
cat ("\n\n", "pvalue")
print(pvalue)
hist(hsh.resample.vector, xlim=0:1, nclass=20, main= "HSH Distribution",
xlab="HSH", ylab= "Frequency")
abline(v=hsh.real, lwd=3, col= "red")
}}
```

## Arquivos da Função

[Função hsh](#) [Help hsh](#) [Paso a paso da função hsh](#)

From:

<http://labtrop.ib.usp.br/> - **Laboratório de Ecologia de Florestas Tropicais**

Permanent link:

[http://labtrop.ib.usp.br/doku.php?id=cursos:ecor:05\\_curso\\_antigo:r2010:alunos:trabalho\\_final:knunes\\_bio:start](http://labtrop.ib.usp.br/doku.php?id=cursos:ecor:05_curso_antigo:r2010:alunos:trabalho_final:knunes_bio:start)

Last update: **2020/07/27 18:46**