

MELHORAMENTO GENÉTICO VEGETAL

ANÁLISE DOS COEFICIENTES DE ENDOGAMIA E DE PARENTESCO PARA QUALQUER NÍVEL DE PLOIDIA USANDO O PACOTE ESTATÍSTICO R ⁽¹⁾

LUIZ ALEXANDRE PETERNELLI ^(2*); FÁBIO MEDEIROS FERREIRA ⁽³⁾; RODRIGO BARROS ROCHA ⁽⁴⁾;
WILLIAN SILVA BARROS ⁽⁵⁾, MÁRCIO HENRIQUE PEREIRA BARBOSA ⁽⁶⁾

RESUMO

São poucos os softwares disponíveis para a análise de parentesco e, até a presente data, nenhum pode realizar a análise de parentesco para organismos poliplóides, com número par (2k) de cromossomos. Implementou-se junto ao programa R uma rotina capaz de executar a análise de parentesco para qualquer nível 2k de ploidia, número de indivíduos (ou populações) e número de gerações envolvidas na árvore genealógica. A função principal, `calc.rxy()` calcula o coeficiente de endogamia (F_X) de cada indivíduo; coeficientes de parentesco (r_{XY}) entre dois indivíduos quaisquer; e coeficiente de parentesco médio, variância, mínimo e máximo, para cada indivíduo. Ela pode mostrar a matriz completa de parentesco ou uma submatriz pré-definida de indivíduos selecionados. Adicionalmente, foram incluídas funções que servem para identificar erros de redundância dos dados originais (`checar.nomes()`) e para identificar os pares de indivíduos com r_{XY} superior a certo limite especificado pelo pesquisador (`corte.rxy()`). Essas funções podem ser usadas na análise de pedigrees com um número elevado de indivíduos e poderão ser utilizadas pela comunidade científica livremente, sem restrições à plataforma operacional utilizada. Destaca-se a vantagem em se poder trabalhar com qualquer nível 2k de ploidia, mesmo quando existe a possibilidade de certo indivíduo ser oriundo de autofecundação, aspecto comum em várias espécies vegetais.

Palavras-chave: genética de populações, poliplóides, análise de pedigree.

ABSTRACT

ANALYSIS OF THE INBREEDING AND COANCESTRY COEFFICIENTS FOR ANY PLOIDY LEVEL USING THE STATISTICAL PACKAGE R

There are few softwares available to analyze relatedness among individuals and, to date, none can perform this analysis for polyploid organisms, with even number (2k) of chromosomes. This work implements within package able to execute an analysis of relatedness for any 2k-ploidy level, number of individuals (or populations) and number of generations in the pedigree. The main function, `calc.rxy()`, calculates the inbreeding coefficient (F_X) of each individual; coancestry coefficients (r_{XY}) between any two individuals; and average, variance, minimum and maximum of coancestry coefficient for each selected individual. A complete matrix of coancestry or any sub-matrix for selected individuals can be output. In

⁽¹⁾ Recebido para publicação em 1 de novembro de 2006 e aceito em 14 de maio de 2009.

⁽²⁾ Departamento de Estatística, Universidade Federal de Viçosa, 36570-000 Viçosa (MG). E-mail: peterNELLI@ufv.br.

(*) Autor correspondente. Bolsista de produtividade em Pesquisa do CNPq.

⁽³⁾ Instituto de Ciências Exatas e Tecnologia, Universidade Federal do Amazonas, 69100-000 Itacoatiara (AM). Email: ferreirafmt@ufam.edu.br

⁽⁴⁾ EMBRAPA - Centro de Pesquisa Agroflorestal de Rondônia, BR 364, km 5,5 Caixa Postal 406, 78900-970 Porto Velho (RO), Brasil. E-mail: rodrigo@cpafro.embrapa.br.

⁽⁵⁾ Instituto de Física e Matemática/Departamento de Matemática e Estatística, Universidade Federal de Pelotas. E-mail: willian.barros@ufpel.edu.br

⁽⁶⁾ Departamento de Fitotecnia, Universidade Federal de Viçosa, 36570-000 Viçosa (MG). E-mail: barbosa@ufv.br. Bolsista de produtividade em Pesquisa do CNPq.

addition, functions are included to identify redundant information in the original data set (`checar.nomes()`) and to identify pairs of individuals with r_{XY} higher than an upper bound specified by the researcher (`corte.rxy()`). These functions can be used to analyze a pedigree with a large number of individuals and can be implemented by the scientific community freely, regardless of the platform in use. The advantage of allowing to work with any 2k-ploidy level is highlighted, even when some individual is originated from selfing, a common scenario for many plant species.

Key words: population genetics, polyploidy, pedigree analysis.

1. INTRODUÇÃO

WRIGHT (1922) foi o primeiro autor a propor um coeficiente realmente capaz de descrever o grau de endogamia de um indivíduo a partir de sua genealogia. O coeficiente de endogamia de Wright expressa a correlação entre os valores gaméticos que formam a progênie derivada de uma população. Anos depois, o coeficiente de Wright foi generalizado por outros autores (HALDANE e MOSHINSKY, 1939; COTTERMAN, 1941; MALÉCOT, 1969). Conhecido como coeficiente f foi definido como a probabilidade de que dois alelos de um mesmo gene presentes em um mesmo indivíduo sejam idênticos por descendência. Dentro deste contexto, o coeficiente de parentesco r_{XY} entre dois indivíduos (X e Y) foi definido como a probabilidade de que um alelo, tomado ao acaso do indivíduo X seja idêntico por descendência ao alelo do indivíduo Y (FALCONER e MACKAY, 1996; LYNCH e WALSH, 1998). Inicialmente, tais considerações foram postuladas a organismos diplóides. Mais tarde, em um contexto mais generalizado, KEMPTHORNE (1973) estendeu a definição probabilística a organismos poliplóides, com número par (2k) de cromossomos.

O uso dos coeficientes de parentesco e de endogamia tem sido freqüente em análises genéticas das populações humanas e no gerenciamento de bancos de germoplasma associado aos programas de melhoramento de plantas e animais.

Uma característica inerente à estimação dos coeficientes de parentesco e de endogamia é a dependência sobre a informação do pedigree. Por exemplo, um coeficiente de endogamia não é uma característica intrínseca do indivíduo e sim uma informação genealógica dele mesmo (KHANG, 1989). Assim, a estimação do coeficiente f ou r_{XY} torna-se extremamente trabalhosa quando se dispõe de uma genealogia complexa, isto é, com várias gerações, grande número de indivíduos e sobreposição de gerações. Certamente, o uso de *softwares* para a estimação dos coeficientes de parentesco e endogamia veio a reduzir esforços em situações complexas.

KHANG (1989) relata uma série de métodos processuais de programação para a estimação dos coeficientes a partir da informação genealógica. No entanto, existem poucos softwares capacitados à

análise de parentesco e até a presente data não foi constatado nenhum dirigido a organismos poliplóides.

Sabe-se que várias espécies vegetais de importância agrônômica e comercial possuem ploidia elevada. No melhoramento vegetal, a análise do parentesco entre dois genótipos tem auxiliado muitos melhoristas na identificação de grupos parentais promissores e combinações a serem exploradas em programas de melhoramento, como é o caso da batata (MENDOZA e HAYNES, 1974), do trigo (KIM e WARD, 1997), da cana-de-açúcar (DEREN, 1995) e do milho (LÜBBERSTEDT et al., 2000). As estimativas de coeficiente de parentesco têm sido apenas indicadores preliminares das combinações mais divergentes, entretanto, têm sido úteis nos programas de melhoramento em estágios iniciais, cujo acúmulo de informações fenotípicas e moleculares das populações ainda são escassas.

LIMA et al. (2002) utilizaram o procedimento PROC INBREED incluído no programa SAS (SAS INSTITUTE, 1989) na análise de parentesco em cana-de-açúcar. Neste trabalho, os autores adotaram algumas pressuposições, como r_{XY} igual a zero para ancestrais distantes e coeficientes de endogamia também iguais a zero, em virtude da característica heterozigota dos genótipos da cultura da cana-de-açúcar, além de considerar a segregação bivalente da espécie. O PROC INBREED possui dois modos de operação. Um deles executa a análise de parentesco pressupondo que todos os indivíduos pertencem a uma mesma geração. O outro divide toda a genealogia em gerações não sobrepostas e analisa cada geração separadamente, assumindo que os pais dos indivíduos da atual geração são definidos na geração anterior.

O pacote estatístico R ou, simplesmente, programa R, obtido em <http://www.r-project.org> é um software livre com uma linguagem dirigida à análise gráfica e estatística, semelhante à linguagem de computação S. A linguagem R, assim como a S, é designada como uma verdadeira linguagem de computação, permitindo aos usuários adicionar novas funções e implementar novas rotinas. Na verdade, o sistema R é um ambiente dentro do quais técnicas estatísticas podem ser implementadas.

Somada as vantagens atribuídas ao programa R, entende-se que a criação de uma nova rotina destinada à análise de genealogias, em especial, de organismos poliplóides, vem a contribuir com aqueles pesquisadores que desejam fazer uso de informações de parentesco em suas investigações. Portanto, o objetivo deste trabalho foi o de desenvolver funções para obtenção de estimativas de parentesco para quaisquer genealogias considerando: o nível de ploidia, número de indivíduos e número de gerações envolvidas na árvore genealógica do organismo sob estudo. Além disso, propõe-se apresentar a aplicabilidade da rotina desenvolvida quanto à estimação dos coeficientes de parentesco, por intermédio de um exemplo conhecido na literatura.

2. MATERIAL E MÉTODOS

Todas as funções desenvolvidas foram escritas em linguagem R de programação em um computador IBM e sistema operacional Windows e podem ser executadas no ambiente R, disponibilizado em <http://www.r-project.org>. O R pode ser instalado nos sistemas Windows, Linux ou Mac. Esta função poderá ser utilizada livremente pela comunidade científica, independentemente da plataforma operacional utilizada. Será disponibilizado um serviço de recebimento e envio de arquivos para ser utilizado com os dados de entrada e de saída da análise sob requisição, respectivamente, no site www.det.ufv.br/~peternelli/projeto.rxy. Este trabalho foi realizado no Departamento de Estatística da Universidade Federal de Viçosa, em conjunto com o Laboratório de Bioinformática do Departamento de Biologia da mesma universidade.

Algoritmo

A função principal, `calc.rxy()` realiza o cálculo do coeficiente de parentesco r_{XY} entre dois indivíduos quaisquer (X e Y), coeficiente de endogamia F_X (ou F_Y) e algumas estatísticas relacionadas ao r_{XY} , como média geral, média por indivíduo, variância geral, variância por indivíduo etc.

A função `calc.rxy()` é dividida em 5 etapas, detalhadas a seguir:

Etapa 1: Entrada de dados e informação da ploidia;

Etapa 2: Leitura dos dados e codificação dos indivíduos em números;

Etapa 3: Reorganização do arquivo gerado na Etapa 2;

Etapa 4: Cálculos efetuados a partir do arquivo gerado na Etapa 3;

Etapa 5: Apresentação de resultados especificados pelo usuário.

Adicionalmente, duas funções foram incluídas no pacote. Uma serve para identificar erros de redundância dos dados originais (`checar.nomes()`) e a outra para identificar os pares de indivíduos com r_{XY} superior a certo limite especificado pelo pesquisador (`corte.rxy()`).

Detalhamento das etapas da função

O arquivo de dados original (na Etapa 1) deve conter três colunas na seguinte ordem e os respectivos nomes: filho, pai e mãe; ou quaisquer outros nomes, mas sabendo que a primeira coluna se refere ao filho dos respectivos pais apresentados na segunda e terceira colunas. As informações podem ser salvas em arquivos de texto (extensão `.txt`) ou planilha eletrônica (extensão `.csv`). O usuário ainda poderá verificar se existe alguma duplicação nos nomes dos clones (primeira coluna) no arquivo de dados no formato csv e corrigir tal arquivo antes de iniciar a função `calc.rxy()`. Para isso basta digitar no console do R o comando `checar.nomes()`. Se houver duplicação de nomes de clones no conjunto de dados aparecerá a informação da linha onde a duplicação ocorreu. O usuário deverá fazer as devidas correções para novamente submeter o conjunto de dados corrigido à função `calc.rxy()`.

Se um indivíduo tiver pais desconhecidos, essa informação poderá aparecer como `"?"`, `"na"`, `"ns"`, `","`, ou qualquer outra nomenclatura, desde que a nomenclatura utilizada não apareça na primeira coluna deste arquivo. Qualquer código (informação) presente na segunda ou na terceira coluna e que não apareça na primeira coluna do arquivo será transformado em zero (0) na codificação efetuada na etapa 2, indicando que o indivíduo é desconhecido.

Ressalta-se ainda o cuidado máximo quanto ao nome dos indivíduos. Por exemplo, a função vai considerar desiguais os indivíduos XXX e XXx. A mistura de letras maiúsculas e minúsculas é suficiente para defini-los como indivíduos diferentes. Não pode, também, haver espaços no nome de qualquer indivíduo, uma vez que a função considera como informações diferentes os dados separados por espaço.

Uma exigência do programa é que na primeira linha do arquivo de dados deve aparecer, obrigatoriamente, um indivíduo que não tenha a informação de pai e mãe, ou seja, que apresente pai e mãe desconhecidos. Tal situação tem sido corriqueira no pedigree de qualquer organismo. A ordem de entrada das informações para os demais indivíduos é irrelevante, uma vez que o arquivo de dados será reorganizado de maneira apropriada na Etapa 3.

A construção de um arquivo de dados é exemplificada na Tabela 1. Ao iniciar a função, o usuário terá as seguintes opções de entrada de dados: 1 - Criar e analisar dados; 2 - Analisar dados já existentes com extensão *.txt; 3 - Analisar dados já existentes com extensão *.csv; e 0 - Sair da função.

Tabela 1. Informações genealógicas adaptadas de KEMPTHORNE (1973), página 105

Filho	Pai*	Mãe
A	?	?
X	?	?
Y	?	?
B	?	A
C	X	A
D	B	Y
E	D	C
F	E	A

*O símbolo “?” indica indivíduos desconhecidos.

Uma vez escolhidas as opções 1, 2 ou 3 o usuário terá, posteriormente, a opção de mudar a pasta de trabalho e localizar os arquivos de dados para a análise. Ao indicar a pasta de trabalho, o usuário poderá usar os recursos “copiar” e “colar” para indicar o caminho completo onde está localizado o arquivo com os dados, com o cuidado de usar barras separadoras de diretórios da forma “/” (barras inclinadas para a direita). Alternativamente, a sessão do R poderia ser direcionada para o diretório onde os dados estiverem localizados, facilitando o processo.

Na etapa 2, é feita a leitura e recodificação dos dados em números, mas sem perder sua identidade. Essa recodificação ocorrerá mesmo que o arquivo original seja apresentado com números, uma vez que toda informação do arquivo original é lida como caractere e não como número.

Na etapa 3, as linhas do arquivo gerado na etapa 2 são rearranjadas de forma que o código fornecido ao pai e à mãe sempre aparecerão previamente ao código do(s) seu(s) filho(s). Portanto, o arquivo de dados original pode ser organizado sem preocupação quanto à ordem temporal dos indivíduos, ou seja, as informações de um indivíduo podem aparecer na linha i , enquanto as informações referentes aos seus pais podem aparecer em qualquer outra linha i' (para todo $i' \neq i$).

Suponha que o arquivo original codificado na etapa 2 tenha a seguinte formação, onde 0 (zero) indica pais desconhecidos.

Filho	Pai	Mãe
1	0	0
2	1	1
3	4	5
4	2	1
5	0	0

Observamos que os filhos codificados “4” e “5” aparecem posteriormente à linha onde são definidos como pai e mãe. Para poder usar o processo recursivo no cálculo de F_X e r_{XY} , na Etapa 3 esse arquivo teria suas linhas rearranjadas da seguinte forma:

Filho	Pai	Mãe
1	0	0
2	1	1
4	2	1
5	0	0
3	4	5

Na Etapa 4, são efetuados os cálculos a partir do arquivo gerado na etapa 3. Estes cálculos são realizados com base nas expressões generalizadas desenvolvidas por KEMPTHORNE (1973), para o caso de um loco no qual existem $2k$ cromossomos, os quais segregam como cromossomos, dadas por:

$$F_X = \frac{1}{2k-1} \left[kr_{A,B} + \frac{(k-1)}{2} (F_A + F_B) \right]$$

em que $k = 1, 2, 3$ e 4 para organismo diplóide, tetraplóide, hexaplóide e octaplóide, respectivamente; e

$$r_{XX} = \frac{1}{2k} [1 + (2k-1)F_X] \text{ e } r_{XY} = r_{AB,Y} = \frac{1}{2} [r_{AY} + r_{BY}]$$

em que X é filho de A e B e Y é um indivíduo qualquer com pais desconhecidos.

A função `calc.rxy()` calcula r_{XY} de maneira recursiva, a partir do primeiro indivíduo do arquivo oriundo da etapa 3, e vai organizando os resultados em uma matriz de dimensão $n \times n$, onde n é o número de indivíduos listados na primeira coluna do arquivo original, ou seja, é o número de filhos.

Na Etapa 5, ocorrem os cálculos e as apresentações de resultados especificados pelo usuário, como r_{XY} para certos indivíduos, médias, valores máximos e mínimos etc.

Estimação de Parâmetros Genéticos

A função `calc.rxy()` permite, através da informação genealógica, estimar os seguintes parâmetros: coeficientes de endogamia (F_X) de cada

indivíduo; coeficientes de parentesco (r_{XY}) entre dois indivíduos quaisquer (mostrando a matriz completa de parentesco ou submatrizes pré-definidas de indivíduos selecionados); coeficiente de parentesco médio, variância, mínimo e máximo, para cada indivíduo. Essas informações são exportadas para arquivos de texto (extensão *.txt) ou planilha eletrônica (extensão *.csv) na pasta onde o R está sendo executado, ou de onde o arquivo de dados original foi lido, ou em uma pasta a ser definida pelo usuário, para qualquer uso posterior.

3. RESULTADOS E DISCUSSÃO

Após o desenvolvimento das funções, buscou-se verificar se a mesma seria verdadeiramente apropriada à estimação do coeficiente de parentesco (r_{XY}) para qualquer ploidia e grau de parentesco entre os indivíduos. O uso dessas funções e interpretação de suas saídas é mais bem apresentado na forma de um exemplo de aplicação. Este exemplo será apresentado a seguir.

Considerando a genealogia apresentada por KEMPTHORNE (1973), página 105, de um organismo tetraplóide, podemos realizar a análise completa a partir do arquivo exemplo1.txt cujos dados estão apresentados na tabela 1.

Todos os arquivos gerados a partir desta análise estão apresentados na página: www.dpi.ufv.br/projeto.rxy. Aqui serão apresentados apenas alguns resultados que merecem maiores discussões.

A tabela 2 apresenta os resultados obtidos para o coeficiente de parentesco (r_{XY}) entre todos os indivíduos apresentados na tabela 1.

No caso de pedigrees com elevado número de indivíduos, caso o pesquisador esteja interessado em conhecer o r_{XY} de alguns pares de indivíduos, poderá selecioná-los no momento em que for questionado para tal processo. Considere que o interesse seja nos indivíduos C, D, E e F. Uma nova saída ("rxy.par.a.par.txt" e "rxy.par.a.par.csv") será criada na pasta de trabalho, cujos resultados estão presentes na tabela 3.

A função interna descritivas.vetor() possibilitou fornecer os resultados das estatísticas descritivas dos r_{XY} para os indivíduos C, D, E e F (Tabela 4). Os resultados estão apresentados nos arquivos "descri.indiv.selec.txt" e "descri.indiv.selec.csv", sendo enviados para a pasta de trabalho.

Para selecionar os pares de indivíduos com r_{XY} superior a um valor qualquer pré-definido (α) foi desenvolvida uma função, corte.rxy(), na qual basta informar o argumento α de interesse.

Tabela 2. Matriz de coeficientes de parentesco obtidos para todos os indivíduos listados na primeira coluna da Tabela 1

	A	X	Y	B	C	D	E	F
A	0,2500	0,0000	0,0000	0,1250	0,1250	0,0625	0,0938	0,1719
X	-	0,2500	0,0000	0,0000	0,1250	0,0000	0,0625	0,0313
Y	-	-	0,2500	0,0000	0,0000	0,1250	0,0625	0,0313
B	-	-	-	0,2500	0,0625	0,1250	0,0938	0,1094
C	-	-	-	-	0,2500	0,0313	0,1406	0,1328
D	-	-	-	-	-	0,2500	0,1406	0,1016
E	-	-	-	-	-	-	0,2656	0,1797
F	-	-	-	-	-	-	-	0,2995

Tabela 3. Coeficientes de parentesco (r_{XY}) para os indivíduos C, D, E e F

	indivíduo X	indivíduo Y	r_{XY}
1	C	D	0,0313
2	C	E	0,1406
3	C	F	0,1328
4	D	E	0,1406
5	D	F	0,1016
6	E	F	0,1797

Tabela 4. Estatísticas descritivas relativas aos coeficientes de parentesco dos indivíduos C, D, E e F

Indivíduo	Média	Variância	Máximo	Mínimo
C	0,0882	0,0032	0,1406	0,0000
D	0,0837	0,0029	0,1406	0,0000
E	0,1105	0,0020	0,1797	0,0625
F	0,1083	0,0036	0,1797	0,0313

Tabela 5. Pares de indivíduos cujo coeficiente de parentesco é maior que 0,10

	indivíduo X selecionado	indivíduo Y selecionado	r_{XY} .selecionados
1	A	B	0,1250
2	A	C	0,1250
3	A	F	0,1719
4	X	C	0,1250
5	Y	D	0,1250
6	B	D	0,1250
7	B	F	0,1094
8	C	E	0,1406
9	C	F	0,1328
10	D	E	0,1406
11	D	F	0,1016
12	E	F	0,1797

Na tabela 5 nota-se um exemplo para $r_{xy} = 0,10$, que pode ser comparado com os valores da tabela 2. Essas informações são enviadas para a pasta de trabalho com os nomes “rxy.selecionados.par.a.par.txt” e “rxy.selecionados.par.a.par.csv”.

Esse arquivo poderia ser usado na identificação de cruzamentos indesejáveis, conforme comentado por PETERNELLI e BARBOSA (2004) e BARBOSA et al. (2005), sendo esta informação útil em muitos programas de melhoramento genético. A lista de nomes dos indivíduos apresentados na Tabela 5 são disponibilizados no arquivo gerado “nome_individuos_selecionados.txt” e “nome_individuos_selecionados.csv”, na pasta de trabalho.

As funções utilizadas e maiores instruções para seu uso serão apresentadas, oportunamente, no *site* supracitado. As funções foram testadas em vários conjuntos de dados de pedigrees hipotéticos ou reais, obtidos do programa de melhoramento de cana-de-açúcar da Universidade Federal de Viçosa. Procurou-se incluir diferentes níveis de ploidia, além da inclusão de indivíduos oriundos de autofecundação. A função proposta mostrou-se adequada para resolução de genealogias compostas por indivíduos provenientes de autofecundação, aspecto comum em várias espécies vegetais.

Para o caso de organismos diplóides, a consistência dos resultados foi checada a partir de softwares alternativos, além do cálculo manual em pedigrees menores. Para o caso de organismos poliplóides, as verificações foram realizadas manualmente, em pedigrees pequenos e com variados níveis de complexidade.

4. CONCLUSÕES

As funções desenvolvidas permitem uma análise de parentesco rápida e detalhada para indivíduos com nível 2k de ploidia, facilitando esse tipo de estudo no caso de pedigrees com número elevado de indivíduos. Devido à sua execução no software livre R, há garantia de que estas funções poderão ser utilizadas pela comunidade científica livremente, sem restrições à plataforma operacional utilizada. Destaca-se a vantagem em se poder trabalhar com qualquer nível 2k de ploidia, mesmo quando existe a possibilidade de certo indivíduo ser oriundo de autofecundação.

AGRADECIMENTOS

Ao programa de Melhoramento de Cana-de-açúcar da Universidade Federal de Viçosa, pelo fornecimento de dados genealógicos, para teste e aplicação das funções aqui desenvolvidas. À FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais e ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelos recursos financeiros necessários para o desenvolvimento desta pesquisa.

REFERÊNCIAS

- BARBOSA, M.H.P.; RESENDE, M.D.V.; SILVEIRA, L.C.I.; PETERNELLI, L.A. Estratégias de melhoramento genético da cana-de-açúcar em universidades. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS; Seleção recorrente no melhoramento de plantas, 2005, Lavras. *Anais...* Lavras: UFLA, 2005. p.43-58.
- COTTERMAN, C.W. Relatives and Human Genetic Analysis. *Scientific Monthly*, v.53, p.227-234, 1941.
- DEREN, C.W. Genetic base of U.S. Mainland sugarcane, *Crop Science*, v.35, p.1195-1199, 1995.
- FALCONER, D.S.; MACKAY, T.F.C. *Introduction to quantitative genetics*. 4.ed. Essex: Longman, 1996. 464p.
- HALDANE, J.B.S.; MOSHINSKY, P. Inbreeding in Mendelian populations with special reference to human cousin marriage. *Annals of Eugenics*, v.9, p.321-340, 1939.
- KEMPTHORNE, O. *An introduction to genetic statistics*. Ames: Iowa State University Press, 1973. 545p.
- KHANG, J.V.T. A FORTRAN subroutine to compute inbreeding and kinship coefficients according to the number of ancestral generations. *Computer Applications in the Biosciences*, v.5, p.199-204, 1989.
- KIM, H.S.; WARD, R.W. Genetic diversity in Eastern U.S. soft winter wheat (*Triticum aestivum* L. em Thell.) based on RFLPs

and coefficient of parentage. **Theoretical and Applied Genetics**, v.94, p.472-479, 1997.

LÜBBERSTEDT, T.; MELCHINGERA, A.E.; DUŠLEA, C.; VUYLSTEKEB, M.; KUIPERB, M. Relationships among early European maize inbreds: IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. **Crop Science**, v.40, p.783-791, 2000.

LIMA, M.L.A.; GARCIA, A.A.F.; OLIVEIRA, K.M.; MATSUOKA, S.; ARIZONO, H.; DESOUSA Jr, C.L.; DESOUSA, A.P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). **Theoretical and Applied Genetics**, v.104, p.30-38, 2002.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980p.

MALÉCOT, G. **Les Mathématiques de L'Herédité**. Masson: Masson, 1948. 63p.

MENDOZA, H.A.; HAYNES, F.L. Genetic basis of heterosis for yield in the autotetraploid potato. **Theoretical and Applied Genetics**, v.45, p.21-25, 1974.

PETERNELLI, L.A.; BARBOSA, M.H.P. Desafios na condução e planejamento de experimentos de avaliação de clones de cana-de-açúcar. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS: DESAFIOS E PERSPECTIVAS DA INTERAÇÃO BIOMETRIA E O MELHORAMENTO DE PLANTAS, 3., 2004, Lavras. **Anais...** Lavras: UFLA, 2004. p.17-27.

SAS Institute. **SAS/STAT User's Guide**. version 6, 4th ed. Cary: USA, 1989. v.2, 846 p.

WRIGHT, S. Coefficients of inbreeding and relationship. **American Naturalist**, v.56, p.330-338, 1922.