

ESTUDIO DE INDICADORES DE VARIABILIDAD GENÉTICA MEDIANTE RFLP EN LA CAÑA DE AZÚCAR

M.T. Cornide, O. Coto, E. Canales, M. Ramos Leal, H. Leonard, A. Veitía y A. Sigarroa.*

Bioplantas, Centro Nacional de Investigaciones Científicas, Avenida 25 y 158, Playa,
*Dpto. Genética, Facultad de Biología, Universidad de la Habana, Calle 25 y J, Vedado,
Ciudad de la Habana, Cuba.

Recibido: 3 de abril de 1996.

RESUMEN. Se estimó la similitud genética de un grupo de 21 genotipos de caña de azúcar con diferente grado de parentesco mediante cinco indicadores, cuatro de los cuales se basaron en los patrones de RFLP de 11 combinaciones enzima-sondas. La proporción de variantes comunes entre los genotipos y los testigos no emparentados osciló entre 0,18-0,52, evidenciándose esta fuente de sesgo por sobreestimación de la similitud genética en la caña de azúcar. Los indicadores ajustados dieron estimados negativos en algunas combinaciones por lo que controlaron parcialmente este sesgo. No hubo correlación entre el coeficiente de consanguinidad y los indicadores estudiados. Estos resultados indican la presencia de otras fuentes de sesgos debidas a la contribución desigual de los genotipos parentales a los de la progenie. Se caracterizaron tres grupos de genotipos por sus patrones de 9 sistemas RFLP.

ABSTRACT. RFLP analysis was performed on 21 sugarcane genotypes with different degrees of genetic relatedness. Genetic similarity was estimated using four indexes based on RFLP patterns and the coefficient of coancestry. The average proportion of shared molecular marker variants between non-related genotypes oscillated between 0.18-0.52, evidencing the presence of this type of bias. The absence of correlation between the coefficient of coancestry and the RFLP genetic distance estimates in sugarcane was not unexpected as the former has limited uses in many crops due to the violation of some of the genetic assumptions for its calculation. A cluster analysis was performed on pairwise dissimilarity values between all genotypes resulting in the aggregation of genotypes into three groups. A synthetic picture of the molecular characteristics of these groups by 9 RFLP patterns was obtained performing iteratively four factor analysis of correspondences taking rare bands as inactive variables.

INTRODUCCION

El conocimiento de las relaciones genéticas entre variedades y especies permite elegir progenitores, así como evaluar y caracterizar los recursos genéticos disponibles para hacer un uso eficiente de la variabilidad genética.¹ El coeficiente de Malécot se ha empleado tradicionalmente para elegir progenitores y establecer las distancias genéticas mínimas para la protección varietal. Sin embargo, su estimación precisa está limitada por la naturaleza genética de muchos cultivos, la complejidad de los esquemas de cruzamiento o el desconocimiento de las genealogías.

Los marcadores moleculares, ofrecen estimaciones directas del polimorfismo del ADN a través del análisis de RFLP y han sido usados exitosamente en muchos cultivos como vía de estimar la similitud genética entre varios genotipos tomando en cuenta una amplia cobertura de sus genomas.²⁻⁵ Las relaciones genéticas así establecidas, se basan en la proporción de variantes moleculares comunes por lo que pueden estar sobreestimadas por la presencia de variantes genéticas iguales en estado, pero no idénticas por descendencia (ibd).

La taxonomía, genética y mejoramiento de la caña de azúcar son extremadamente difíciles por su complejidad genética,⁶ el lento progreso del mejoramiento en este cultivo se ha relacionado con la estrechez de la base genética de los progenitores modernos.⁷

El presente trabajo tuvo como objetivo comparar la eficiencia de cinco indicadores para la estimación de las relaciones genéticas en la caña de azúcar.

MATERIALES Y METODOS

Para los fines del trabajo, se escogieron veintidós genotipos de caña de azúcar con diferente grado de parentesco (f_{AB} 0-25) (Tabla I), procedentes del Banco de Germoplasma

del Instituto Nacional de Investigaciones de la Caña de Azúcar en Matanzas, los cuales fueron analizados por RFLP, obtenidos con 11 combinaciones (CES) de las enzimas EcoRV, Hind III y BamHI con las sondas heterólogas E10y P11 (ADN cloroplástico arroz), CSU 70 (ADN citoplasmático maíz) y GL 26 (ADN ribosomal maíz), a fin de limitar el sesgo de contribución parental desigual por selección. Las bandas diferenciales se evaluaron de forma binaria. El grado de similitud genética fue calculado para cada una de las combinaciones posibles entre pares de genotipos mediante las proporciones comunes de bandas (SG) y de patrones o variantes de RFLP considerando cada CES como un locus RFLP (S_{AB}).⁸

Se llevó a cabo un estudio comparativo de estos indicadores y dos ajustados para el sesgo de la variación común entre genotipos no-emparentados (S_{AB}): (f_{AB}^M) propuesto por Bernardo⁹ y S'_{AB} , como la diferencia entre S_{AB} y S_{AB} . Los coeficientes de consanguinidad se estimaron por genealogía.¹⁰⁻¹²

Los estimados de similitud genética así obtenidos fueron comparados para cada una de las combinaciones mediante el coeficiente de correlación de rangos de Spearman (r_s) y por la frecuencia relativa de genotipos con agrupamiento coincidente según los análisis multivariados.

Las relaciones genéticas entre el material estudiado fueron analizadas empleando las distancias respectivas mediante análisis de componentes principales y de conglomerados con el algoritmo de las medias ponderadas, por presentar el coeficiente cofenético más adecuado. Los valores negativos de f_{AB}^M y S'_{AB} se consideraron cero.

Los patrones de bandas diferenciales fueron analizados mediante análisis factorial de correspondencia simple (FAC)¹¹ para caracterizar los patrones y determinar los RFLP de mayor peso en la diferenciación.

TABLA I
Genotipos empleados

Nombre	Código	Nombre	Código
POJ 2878	(1)	Fiji	(12)
PSA 96	(2)	Morada	(13)
POJ 213	(3)	CP 29-116	(14)
POJ 2725	(4)	CP 44-101	(15)
BH (10)12	(5)	Co 244	(16)
B 42231	(6)	Co 312	(17)
C 431-62	(7)	US 1694	(18)
B 37161	(8)	C 45-52	(19)
C 157-74	(9)	B (30) L7	(20)
Co 421	(10)	Kassoer	(21)
CP 11-65	(11)		

El cálculo de los indicadores genéticos se realizó mediante el paquete MAT-GEN¹³ escrito en GW-BASIC, el cual permite estimar relaciones genéticas entre pares de genotipos empleando caracteres de forma binaria, incluyendo en el análisis varios tipos de indicadores y controlando el sesgo por presencia de variación común entre genotipos no emparentados. Los restantes procesamientos estadísticos se efectuaron con los paquetes STATITCF V. 4.0y NTSys.

RESULTADOS Y DISCUSION

Las CES empleadas mostraron un alto polimorfismo: se encontraron 81 variantes de RFLP, 21 de éstas representadas por genotipos únicos, en número de 2a2 0p orC ES y un promedio de 7,4 por cada uno.

En las tablas II y III, se presentan los estimados de similitud genética de las 210 combinaciones diferentes entre pares de genotipos estudiados.

TABLA II
Estimados de similitud basados en la proporción de bandas (SG) y variantes (S_{AB}) comunes de RFLP entre cada par de los genotipos estudiados

SG	S _{AB}																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1		64	455	5	554	5	364	5	274	5	455	52	71	85	55	55	55	53	64	54	5
2	78		55	55	73	55	45	45	36	45	36	45	27	9	55	55	45	45	27	45	36
3	85	77		82	64	55	64	64	55	36	27	36	18	9	36	36	36	36	27	27	27
4	84	76	9064			45	64	64	55	36	27	36	18	9	36	36	36	36	27	27	27
5	75	91	79	81		55	55	55	45	45	36	45	27	9	55	55	45	45	27	45	36
6	75	66	78	84	73		45	55	27	36	36	36	36	036		36	36	36	27	27	27
7	76	67	82	81	68	83		55	64	27	18	27	9	9	27	27	27	27	27	18	18
8	766	3	758	1	678	4	804		5	27	27	27	18	92	73	63	63	62	71	81	8
9	667	4	716	9	766	7	816	7		45	273	61	81	83	63	63	63	61	82	72	7
10	71	84	67	65	81	66	63	62	75		45	64	36	18	64	55	64	64	18	45	45
11	68	79	60	54	75	51	56	58	67	78		55	36	9	55	64	55	55	36	45	45
12	69	75	61	58	74	67	60	67	65	76	83		36	18	73	64	73	73	27	55	55
13	48	68	44	36	69	43	37	45	55	71	79	77		27	36	36	36	36	9	27	27
14	48	56	40	39	49	43	41	36	45	58	54	58	54		27	18	18	18	18	18	18
15	61	83	63	60	82	64	56	54	69	82	76	84	73	67		73	73	73	27	64	55
16	70	83	69	70	86	68	62	64	78	85	74	75	67	56	90		73	73	27	64	55
17	72	82	74	73	82	65	56	58	71	81	68	69	60	54	83	89		82	27	55	55
18	74	80	75	71	83	66	60	59	75	82	71	70	61	55	84	90	96		36	55	55
19	64	79	66	58	76	55	60	52	74	73	70	66	58	67	79	77	77	80		18	27
20	67	88	68	61	78	59	60	49	76	80	75	77	67	64	86	79	81	80	80		64
21	62	82	67	62	77	59	55	49	67	76	71	72	65	62	88	78	85	84	77	91	

1 ...2 1 Genotipos de caña de azúcar.

SG Coeficientes de similitud genética (X100) (Triangular inferior).⁸

TABLA III
Estimados de similitud genética ajustados para los sesgos por δ_{AB} y f_{AB}

S'_{AB}	f^M_{AB}																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1		326		22	24	3	-2	12	-19	-2	3	20	18	-19	9	12	12	12536			
2	17		20	20	53	17	11	10	-7	-5	-16	2	-21	-36	6	9	-9	-9	-11	1	-13
3	3	11		70	42	24	44	44	29	-13	-22	-6	-27	-27	-20	-17	-17	-17	-4	-22	-19
4	13	11	43		42	8	44	44	29	-13	-22	-6	-27	-27	-20	-17	-17	-17	-4	-22	-19
5	14	31	26	26		25	32	32	17	6	-4	11	-10	-25	17	19	3	3	-2	11	-2
6	2	10	14	5	16		15	28	-16	-16	-10	-9	-1	-43	-24	-20	-20	-20	-6	-25	-22
7	-1	6	29	29	22	9		35	47	-19	-28	-13	-32	-20	-26	-23	-23	-23	2	-28	-25
8	7	6	29	29	21	18	24		20	-20	-15	-14	-19	-20	-28	-9	-9	-9	2	-29	-26
9	-12	-4	19	19	11	-10	32	13		8	-16	0-21		-10	-14	-11	-11	-11	-12	-16	-14
10	-1	-3	-7	-7	3	-9	-12	-12	5		1	34	-6	-22	25	9	27	27	-25	1	3
11	2	-9	-13	-13	-3	-6	-18	-9	-100			22	-1	-30	12	31	14	14768			
12	12	1	-4	-4	7	-5	-8	-9	-1	19	13		0-16		48	32	49	49	-5	22	24
13	-11	-13	-17	-17	-7	-1	-22	-13	-14	-4	-1	0		3	-13	-10	-10	-10	-24	-15	-13
14	-13	-24	-19	-19	-18	-30	-15	-15	-7	-15	-21	-11	-2		-15	-26	-26	-26	-2	-17	-14
15	5	3	-11	-11	-9	-12	-15	-16	-8	12	6	25	-7	-9		41	41	41	-17	29	14
16	6	5	-9	-9	11	-11	-14	-5	-6	5	17	17	-6	-17	19		43	43	-15	31	17
17	6	-5	-9	-9	1	-11	-14	-5	-6	14	8	26	-6	-17	19	21		62	-15	14	17
18	6	-5	-9	-9	1	-11	-14	-5	-6	14	8	26	-6	-17	19	21	30		014		33
19	3	-7	-3	-3	-1	-4	2	1	-9	-16	5	-4	-17	-1	-11	-9	-9	-8		-19	-4
20	2	0	-13	-13	6	-15	-18	-18	-10	03		13	-10	-12	15	17	8	8	-13		39
21	3	-7	-12	-12	-1	-13	-16	-17	-9	2	5	15	-8	-10	8	9	9	18	-3	-23	

1...2 1 Genotipos de caña de azúcar empleados.
 f^M_{AB} Estimados ajustados de similitud genética X 100 (Triangular superior).⁹
 S_{AB} Indicador ajustado ($S_{AB} - d_{AB}$) X 100 (Triangular superior).

De éstas el 47 % no tiene relación genética conocida ($f_{AB} = 0$); el 44 % presenta nexos lejanos ($0 < f \leq 0,125$) y el 9 % restante tiene valores de f superiores ($0,125 < f \leq 0,25$).

Es bien conocido que el indicador S_{AB} brinda estimados sesgados del coeficiente teórico de coancestría (f_{AB}) y es especial, entre genotipos lejanamente relacionados y que el sesgo aumenta con el incremento de la proporción de variantes RFLP comunes entre líneas no-emparentadas (δ_{AB}).^{10-12,14,15}

En las condiciones de este trabajo, el coeficiente de consanguinidad promedio fue de $f_{AB} = 0,05$ y los de los indicadores SG, S_{AB} , f^M_{AB} y S'_{AB} de 0,7; 0,4; 0,1 y 0,06, respectivamente. Los valores notablemente menores de los dos últimos (Tabla III), corrigieron δ_{AB} (0,18-0,52) (Tabla IV), dando valores negativos en el 54 y 55 % de las combinaciones, respectivamente.

Bernardo considera que estimados negativos de f^M_{AB} pueden atribuirse a valores nulos o bajos de f_{AB} , o a la sobreestimación de δ_{AB} . Este indicador dió valores negativos en el 65 % de los casos con algún grado de parentesco. Por otra parte, los valores esperados de f_{AB} en la caña de azúcar son poco precisos debido a la complejidad de los esquemas de nobilización y de su naturaleza genética por lo cual se violan varias premisas resultando en la desigual contribución de los padres al genotipo de la progenie y en la presencia de bandas RFLP en diferentes dosis.¹⁶⁻¹⁸

La sobreestimación de S_{AB} se vió confirmada con el aumento de valores negativos de S'_{AB} (76 %) con relación a f^M_{AB} en combinaciones emparentadas, demostrando también un control parcial e inadecuado de este sesgo.

La comparación de los estimados de similitud genética de las combinaciones de pares de genotipos, consideradas en su totalidad, o por grado de parentesco arrojó estimados de correlación de rangos (r_s) no significativos entre éstos y la f_{AB} , lo cual evidencia la presencia de otras fuentes de sesgo no contempladas y ratifica que no pueden ser usados para calcular el grado de consanguinidad (Tabla V). Hubo correlación alta entre los dos indicadores ajustados y moderada, entre éstos y SG. A pesar de la ausencia de correlación de los valores S_{AB} con los de los indicadores ajustados mostró con éstos las mayores proporciones de coincidencia entre los agrupamientos multivariados.

En la figura 1 se presentan los genotipos agrupados según los estimados de distancia genética. Pueden observarse tres grupos:

Grupo I. Abarca genotipos con grados de parentesco, en su mayoría, lejanos ($f \leq 0,063$), mostrando la mayor similitud genética dentro de grupo (Tabla VI). Se caracteriza por tener relaciones de parentesco sólo con parte de la población.

Grupo II. Está formado por los genotipos con menores S_{AB} y la línea no-emparentada. Presenta las mayores distancias promedio.

Grupo III. Agrupa los restantes genotipos los cuales guardan nexo de parentesco con la mayoría de la población.

Se puede observar que los testigos no-emparentados 5, 12 y 13 se distribuyen sobre la base de las características comunes de sus variantes RFLP en los grupos anteriores para los indicadores moleculares. Los genotipos 1, 4, 7, 9, 15, 16, 17, y 18 mantuvieron grupos estables para todos los indi-

cadore moleculares, coincidentes con sus valores de f_{AB} , y de forma similar, el 2, con una excepción. Los genotipos 3, 6, 8, 13, 14, 20 y 21 mantuvieron un agrupamiento estable sólo para los indicadores moleculares. Los grupos de los indicadores ajustados coincidieron totalmente.

La estabilidad de los tres grupos puede ser explicada principalmente por la característica de mantener los individuos de cada grupo patrones de RFLP similares entre ellos para las combinaciones enzima de restricción-sonda evaluadas.

TABLA IV
Distancias genéticas basadas en RFLP de los genotipos estudiados con relación a los testigos no emparentados

Genotipos	1- δ_i	Genotipos	1- δ_i
1	0,55	12	0,59
2	0,52	13	0,68
3	0,61	14	0,82
4	0,61	15	0,45
5	0,64	16	0,48
6	0,58	17	0,48
7	0,70	18	0,48
8	0,69	19	0,79
9	0,67	200	,58
10	0,52	21	0,61
11	0,58		

δ_i Proporción promedio de variantes RFLP compartidas entre el i ésimo genotipo y los genotipos no relacionados.^{5, 12, 13}

TABLA V
Correspondencia entre los indicadores de similitud genética de las combinaciones de pares de genotipos estudiados

f^{AB}	f_{AB}	f^{M}_{AB}	S'_{AB}	S_{AB}	SG
f^{AB}_{mAB}	1,00	0,15	0,11	0,13	0,07
f_{AB}	0,48	1,00	0,94*	0,04	0,68*
S'_{AB}	0,48	1,00	1,00	0,02	0,67*
S	0,48	0,90	0,90	1,00	0,03
SG	0,43	0,86	0,86	0,76	1,00

Triángulo superior: Coeficiente de correlación de rangos de Sperran (rs).

Triángulo inferior: Frecuencia relativa de genotipos con agrupamiento coincidente.

* Significativo para $P \leq 0,05$.

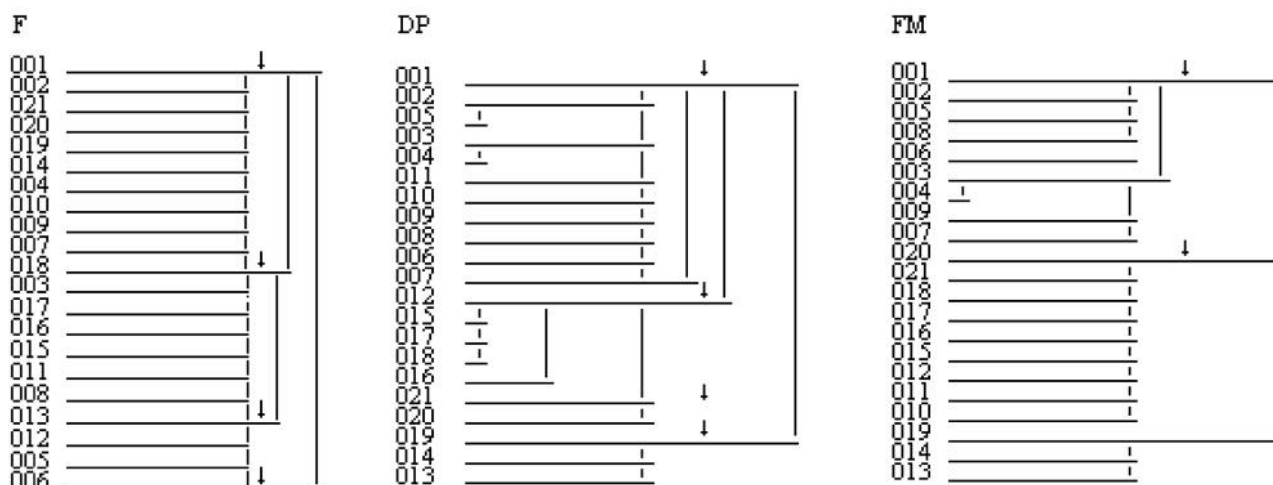


Fig. 1. Dendrogramas representativos de la distribución de los genotipos estudiados según los estimados de distancias genéticas F, DP y FM.

TABLA VI
Valores promedios de los indicadores de similitud genética de los grupos de genotipos

Población	SG	S _{AB}	S' _{AB}	f ^M _{AB}	f _{AB}
Grupo I	0,73	0,46	0,07	0,12	0,04
Grupo II	0,62	0,27	0,01	0,02	0,05
Grupo III	0,09	0,41	0,07	0,11	0,06
Dentro grupos					
Grupo I	0,82	0,64	0,15	0,30	0,04
Grupo II	0,64	0,23	0,01	0,02	0,08
Grupo III	0,76	0,53	0,18	0,25	0,06
Entre grupos					
Grupos I-II	0,68	0,33	0,05	0,03	0,04
Grupos I-III	0,69	0,37	0,01	0,02	0,05
Grupos II-III	0,55	0,30	0,002	0,003	0,06

En la figura 2 se presentan los ejes principales del análisis factorial de correspondencias realizado con las 15 bandas polimórficas de mayor contribución a su diferenciación según los análisis previos por subgrupos de CES. Los tres primeros ejes responden por el 57,2 % de la variabilidad. En el eje 1 tienen las contribuciones relativas mayores a la inercia explicada por éste y una mejor calidad de representación, las bandas polimórficas a5, d4 y c11 (Bam HI-p11; Bam HI-CSU 70 y Bam HI-GL26) con el 21, 17 y 14 %; y en el eje 2 las bandas a15, b13 y c8 (Eco RV-E10; Eco RV-GL26 y BamHI-GL26 con el 19, 17 y el 14 %, respectivamente. Puede observarse que el grupo III se caracteriza por presentar las bandas a5 y d14 (eje 1) en tanto que la mayoría de los genotipos de los otros dos grupos expresan las bandas c11 y e3. El mayor polimorfismo basado en la conformación de estas bandas se presenta en el grupo I. Los miembros del grupo II tienen un patrón similar para las tres bandas más importantes de la serie 3.

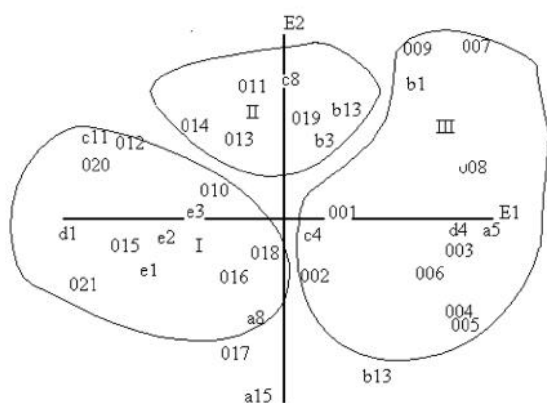


Fig. 2. Representación gráfica de los genotipos estudiados según el análisis factorial de correspondencias simples (FAC) de las quince bandas polimórficas de mayor contribución a su diferenciación.

CONCLUSIONES

La presencia de variantes RFLP comunes entre genotipos no-emparentados (δ_{AB}) conduce a la sobreestimación de los estimados de similitud genética por lo que debe tenerse en cuenta en estudios de caracterización genética de la caña de azúcar.

Los indicadores f^M_{AB} y S'_{AB} , no controlaron este sesgo adecuadamente, obteniéndose estimados negativos en combinaciones con y sin grado de parentesco.

La comparación de los rangos de las combinaciones de pares de genotipos entre los estimados de f^M_{AB} y eS'_{AB} y los valores esperados de f_{AB} mostró la ausencia de correlación entre éstos, confirmando la presencia de otros sesgos que impiden estimar con precisión f_{AB} y el uso de éstos como indicadores de la verdadera proporción de alelos (ibd) entre dos genotipos de caña.

Las distancias genéticas basadas en estimados de los indicadores estudiados permitieron diferenciar en tres grupos de genotipos mediante 9 bandas polimórficas de siete CES.

BIBLIOGRAFIA

- Lu, Y.H., D'Hont A, Walker D.I.T., Rao P.S., Feldmann P. and Glaszmann J.C. *Euphytica*, **78**, 7, 1994.
- Debener T., Salamini F. and Gebhard C. *Theor. Appl. Gen.*, **79**, 360-368, 1990.
- Miller J.C. and Tanksley S.D. *Theor. Appl. Gen.*, **80**, 437, 1990.
- Song K.M., Osborn T.C and Williams D.H. *Theor. Appl. Gen.*, **79**, 497, 1990.
- Wang Z.Y. Second G. and Tanksley S.D. *Theor. Appl. Gen.*, **83**, 565, 1992.
- Simmonds N.W. Sugarcane. In: Simmonds N.W. Ed., Evolution of Crop Plants. Longman, London, 104-108, 1977.
- Berding N. and B.T. Roach. Germplasm collection, maintenance and use. In: Heinz D.J. ed. Sugarcane Improvement Through Breeding. Elsevier Press, Amsterdam, 143-210, 1987.
- Nei M. & Li W. H. *Proc. Natl. Acad. Sci.*, USA, **76**, 5269, 1979.
- Bernardo R. *Theor. Appl. Gen.*, **85**, 1055, 1993.
- Falconer D. S. Introduction to Quantitative Genetics. Longman. 2nd Ed. London, 382, 1981.
- Sneath and Sokal R. R. Numerical Taxonomy. The principles and practice of numerical classification. W. H. Freeman and Co. S. Fco., 573, 1973.
- Cox T. S, Kiang Y. T., Gorman M. B. and Rodgers D. M. *Crop Sci.*, **25**, 529, 1985.
- Sigarroa A. y Cornide M.T. *Revista CENIC Ciencias Biológicas*, **27**, 1996.
- Deren C.W. *Crop Sci.*, **35**, 1195, 1995.
- Lynch M. *Mol. Biol. Evol.*, **5**, 584, 1988.
- Burnquist W. Development and application of restriction fragment length polymorphism technology in sugarcane (*Saccharum spp.*) breeding. Ph. D. Dissertation. Cornell Univ. Ithaca, USA. 1991.
- Grivet L., D'Hont A., Dufour P., Hamon P., Roques D. and Glaszmann J.C. *Heredity*, **73**, 500, 1994.
- Wu K. K., Burnquist W., Sorrells M. E., Tew T. L., Moore and Tanksley S. D. *Theor. Appl. Gen.*, **83**, 294, 1992.