



8:30 às 9:10h: Palestra 1 – TEMA “Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”

Cosme Damião Cruz
Universidade Federal de Viçosa
cdcruz@ufv.br
<http://www.ufv.br/dbg/biodata.htm>
<https://www.facebook.com/GenesNews>



Melhoramento Genético



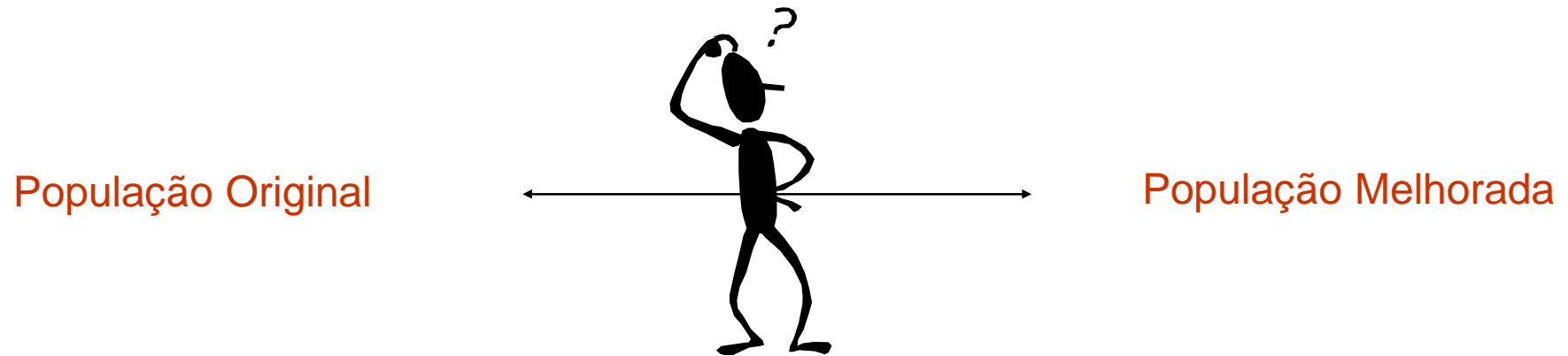
Passarei hoje por todo o teu rebanho, separando dele todos os salpicados e malhados, e todos os morenos entre os cordeiros, e os malhados e salpicados entre as cabras; e isto será o meu salário.

Gênesis 30:32



1866: Gregor Mendel, monge Austríaco, publicou as leis da hereditariedade baseado nos resultados de seus experimentos, iniciados em 1857, com ervilhas.

1. MELHORAMENTO GENÉTICO



Observador

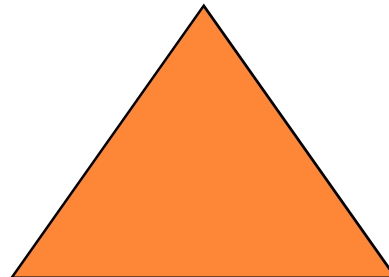


Observador + Conhecimento científico



Observador + Conhecimento científico + Informação Processada

Recursos



Mão de obra

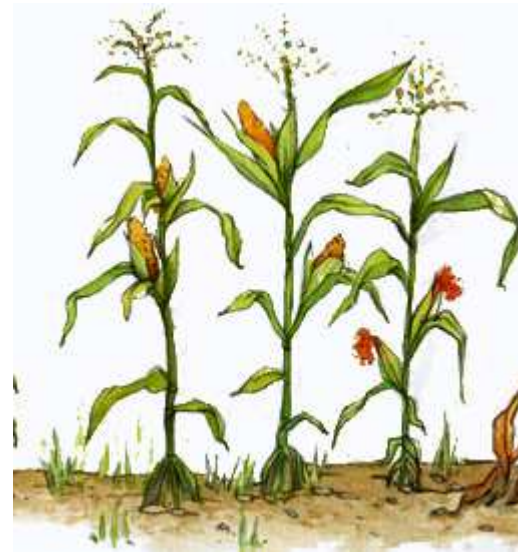
Tempo



2. IDENTIFICAÇÃO DE GENÓTIPOS SUPERIORES



Ampla variabilidade



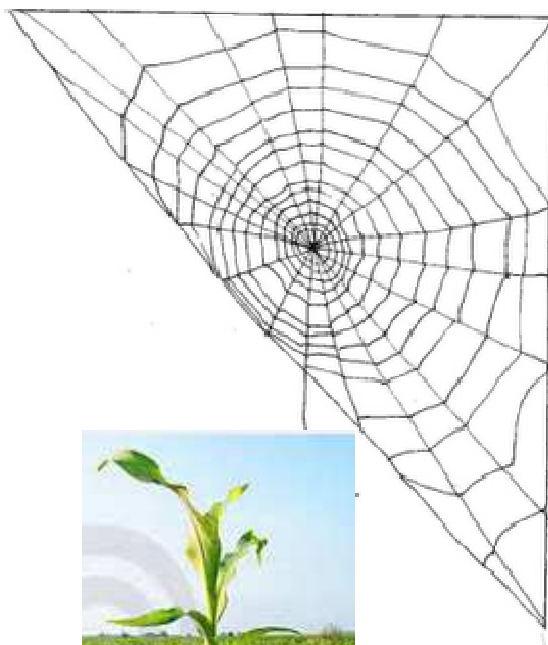
A	B	C
10	9.8	4.2



Pouca variabilidade



3. OTIMIZAÇÃO NA IDENTIFICAÇÃO DE GENÓTIPOS



Experimentação

Genética Quantitativa

Genômica : SAM e GWS

Melhoramento

Redes Neurais

Biometria





A	B	C
10	9.8	4.2

“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



EXPERIMENTAÇÃO



4. EXPERIMENTAÇÃO



Bloco 1



T 3

T 1

T 2

T 8

T 9

T 5

T 6

T 10

T 7

T 4

Bloco 2



T 2

T 10

T 1

T 7

T 5

T 3

T 4

T 8

T 9

T 6

Bloco 3



T 1

T 8

T 5

T 10

T 3

T 6

T 2

T 7

T 4

T 9

Princípio

Repetição

Casualização

Controle Local

Delineamentos

Processamento de dados

Qual a finalidade das análises estatísticas ?



Acesso à
Informação



Experimento => Dados => Processamento => Informação

ANÁLISE DE VARIÂNCIA

FV	GL	SQ	QM	F	Probabilidade
BLOCOS	2	11244.2580	5622.1290	12.3929	
GENÓTIPOS	9	22027.2483	2447.4720	5.3950	.0012
RESÍDUO	18	8165.8487	453.6583		
TOTAL	29	41437.3550			

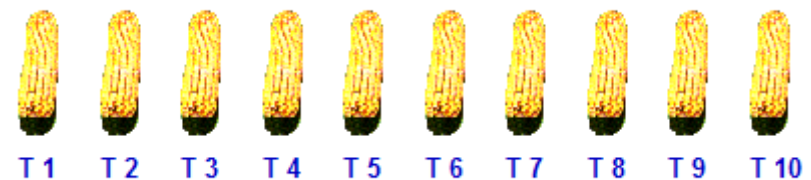
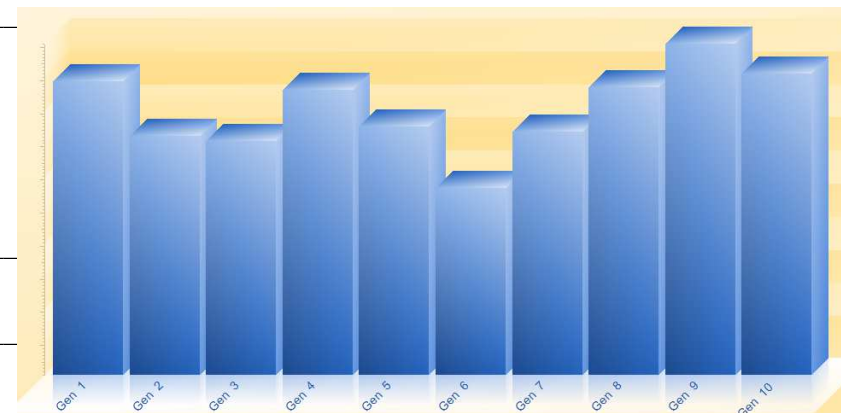
DESCRIÇÃO ESTIMATIVA

MÉDIA	199.9500
S ² GENÉTICA	664.6046
S ² RESIDUAL	453.6583
H² (MÉDIA)	81.4642
H ² (IND.)	59.4319
CV exp	10.6523
CV gen	12.8932

Genótipos Média Classificação

1	186.50	7º
2	225.60	3º *
3	208.50	5º
4	210.27	4º *
5	189.50	6º
6	161.80	10º
7	168.27	9º
8	243.53	1º *
9	172.43	8º
10	233.10	2º *

Média da população	199.95
Média dos selecionados	228.13
Diferencial de seleção	28.18
Herdabilidade	.81
Ganhos por seleção	22.95
Ganhos por seleção (%)	11.48
Média da população melhorada	222.90



Algumas alternativas de esquemas experimentais:

Objetivos:

- Gerar informações
- Reduzir custos

a) Blocos Aumentados de Federer

A, B, C +
1, 2,...20

A, B, C +
21, 22,...40

A, B, C +
41, 42,...60

The screenshot shows a software window titled "Blocos Aumentados de Federer". It contains a file path "c:\dados\plano38.dat" under "Arquivo de Dados". Under "Parâmetros", there are four dropdown menus: "Número de Variáveis" (set to 2), "Número de Genótipos (g) - não comuns" (set to 100), "Número de Testemunhas (t)- comuns" (set to 5), and "Número de Blocos" (set to 5). A hand cursor is pointing at the "Número de Genótipos" dropdown. On the right, under "Opções", there are two buttons: "Retornar" and "Ler Dados".

Informações Geradas

ANÁLISE DE VARIÂNCIA

FV	GL	SQ	QM	F	Probabilidade(%)
Blocos	4	2517170.768	629292.692		
Trat.(Ajust.)	104	4478123.36	43058.878462	7.4858	.002403 **
Resíduo	16	92033.76	5752.11		
Total	124	7087327.888			

Média geral	424.336
Média comuns - testemunhas	257.72
Média não comuns - genótipos	465.99
Média ponderada - μF (Federer)	456.072381
CV(%) geral	17.873258
CV(%) comuns	29.42832
CV(%) não comuns	16.275599
DP tratamentos comuns	47.967114
DP tratamentos não comuns de um mesmo bloco	107.257727
DP tratamentos não comuns de blocos diferentes	117.494953
DP tratamentos não comuns e comuns	91.011199

PARÂMETROS GENÉTICOS E AMBIENTAIS- Obtidos para genótipos


VARIÂNCIA FENOTÍPICA	43232.08596
VARIÂNCIA AMBIENTAL	5752.11
VARIÂNCIA GENOTÍPICA	37479.97596
HERDABILIDADE (%)	86.6948
COEFICIENTE DE VARIAÇÃO GENÉTICO (%)	41.5454
RAZÃO CVg/CVe	2.5526



b) Análise de Grupos de Experimentos

	Experimento 1	Experimento 2	Experimento 3
Rep 1	A, B,C + 1, 2,...20	A, B, C + 21, 22,...40	A, B, C + 41, 42,...60
Rep 2	A, B, C + 1, 2,...20	A, B, C + 21, 22,...40	A, B, C+ 41, 42,...60

Análise de Grupos de Experimentos




Arquivo de Dados

c:\dados\plano39.dat

Parâmetros

Número de Variáveis	3
Número de Experimentos (e)	3
Número Total de Genótipos não-comuns (g)	18
Número de Tratamentos Comuns (c)	2
Número de Repetições (r)	6
Esquema da Anova	1



Opções

Retornar

Esquemas

Ler Dados



Esquemas de Análise de Variância

Retornar

FV	GL	SQ
Blocos/Ensaio	$(t-1)e$	SQB
Ensaio (E)	$e-1$	SQE
Testemunhas (Te)	$t-1$	SQTe
E x Te	$(t-1)(e-1)$	SQTexE
Genótipos (G)	$\sum_{k=1}^e g_k - e$	SQG
(Te vs G)/E	e	SQGrupo
Resíduo	$(r-1)\left(\sum_{k=1}^e g_k + et + e\right)$	SQR
Total	$\left(ret + r\sum_{k=1}^e g_k\right) - 1$	

☒ Esquema 1

☐ Esquema 2

☐ Esquema 3

c) Látices

B1	1	2	3	4
B2	5	6	7	8
B3	9	10	11	12
B4	13	14	15	16

Repetição I

B6	1	5	9	13
B7	2	6	10	14
B8	3	7	11	15
B9	4	8	12	16

Repetição II





A	B	C
10	9.8	4.2

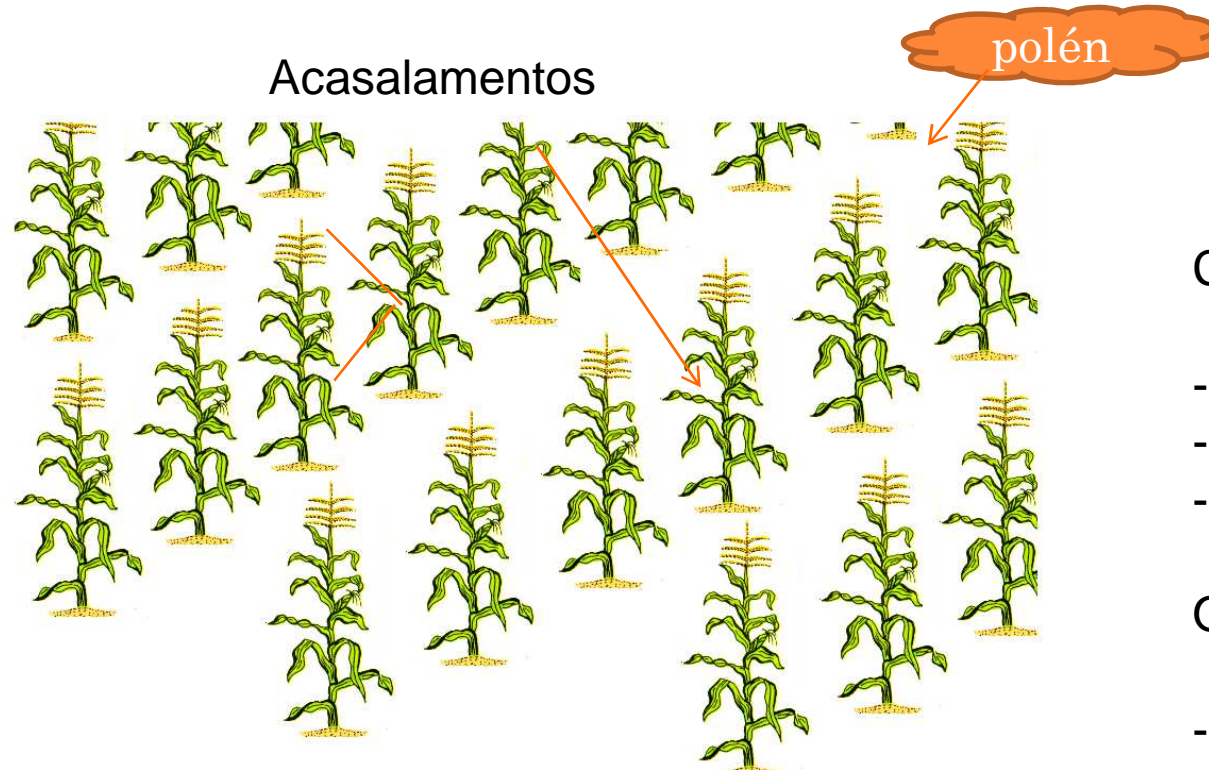
“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



ESTRATÉGIAS DE MELHORAMENTO
GENÉTICA QUANTITATIVA



5. ESTRATÉGIAS DE MELHORAMENTO GENÉTICO



Cruzamentos:

- Ao acaso
- Híbridas
- Autofecundações

Objetivo

-Manipular a variabilidade disponível



População Original



População Estruturada em Família

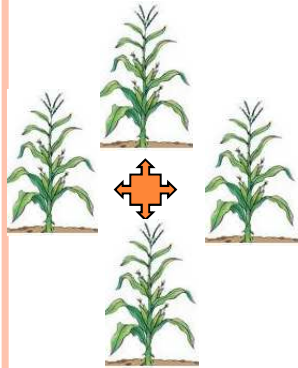
Progênie: FMI, FIC, S1
Combinações de progênie:
Dialelo
Delineamentos I e II



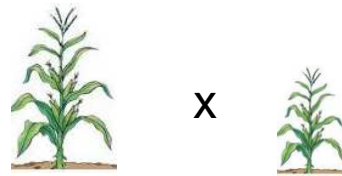
Melhoramento



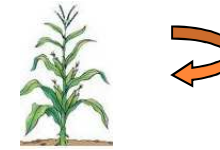
Genética Quantitativa



Acasalamento
ao acaso



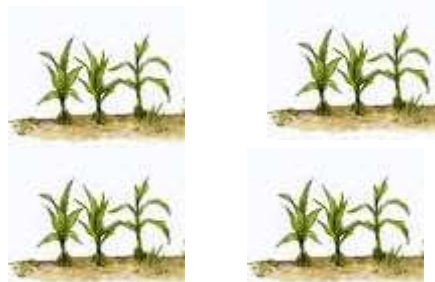
Hibridação



Autofecundação

Derivação de famílias

FMI
FIC
FS₁



Média

$$\mu_g = u + a(D - R) + dH$$

$$\mu_g = u + a(p - q) + 2pqd$$

Variância

$$\sigma_g^2 = a^2[(D + R) - (D - R)^2] + d^2H(! - H) - 2ad(D - R)H$$

$$\sigma_g^2 = 2pq\alpha^2 + (2pqd)^2$$

Covariância

$$\text{Cov}(X, Y) = 2r_{XY}\sigma_A^2 + U_{XY}\sigma_D^2$$

$$\text{Cov}(MI) = \frac{1}{4}\sigma_A^2$$

$$\text{Cov}(IC) = \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2$$





A	B	C
10	9.8	4.2

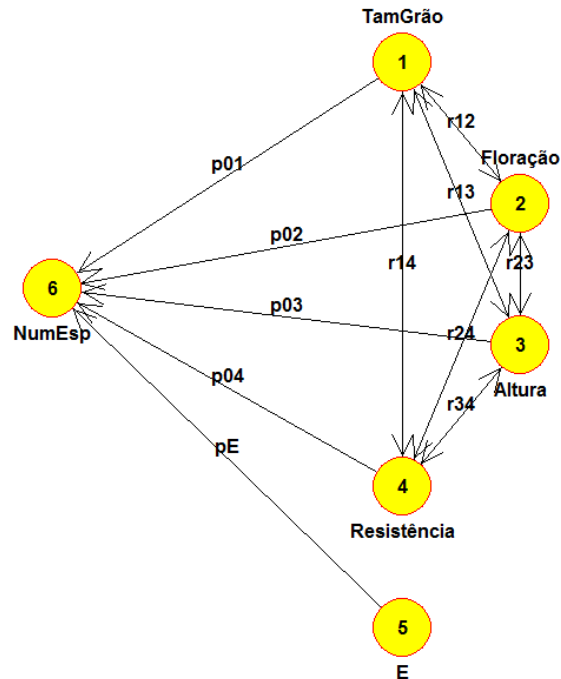
“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



BIOMETRIA



6. UTILIZAÇÃO DA BIOMETRIA



Objetivos (Informações):

Estudo da Hereditariedade
Herança dos caracteres

Ganhos por seleção
Diretos
Indireta
Simulânea

Genótipos
Interação GxA
Adaptabilidade e Estabilidade
Capacidade Combinatória
Heterose
Valor genético -BLUP

Ambiente
Estratificação

Características
Parâmetros genéticos
Correlações
Repetibilidade



a. Resposta correlacionada

Seleção Direta e Indireta

Número de indivíduos selecionados : 4

SELEÇÃO DIRETA SOBRE ==> Produção

VARIÁVEL	Vg(y)	rg	h ² (y) %	GS	GS %
Produção	12782.63368	1.0	87.19794	101.97541	19.44
Altura	2902.47403	-.6289	26.34404	-30.55978	-6.92
Floração	4004.00431	-.66527	82.6898	-37.96939	-9.92
Nespigas	4833.89055	.0902	92.03905	5.65638	1.72
Resist.	30436.79188	.48662	87.48611	76.57293	20.11

x : caráter sob seleção

GS direto em x : = i . Sgx . hx

GS indireto em y(x) = i . hx . rg . Sgy

Caráter principal hx : 0.933798366116691

Intensidade de seleção : 0.9659

SELEÇÃO DIRETA SOBRE: Produção

VARIÁVEL	Xo	Xs	h ² %	GS	GS %
Produção	524.46433	617.43667	87.19794	81.06996	15.46
Altura	441.58767	454.58917	26.34404	3.42512	.78
Floração	382.69533	375.25333	82.6898	-6.15378	-1.61
Nespigas	328.02	343.75333	92.03905	14.48081	4.41
Resist.	380.848	412.78833	87.48611	27.94336	7.34
GANHO TOTAL				120.76547	26.38

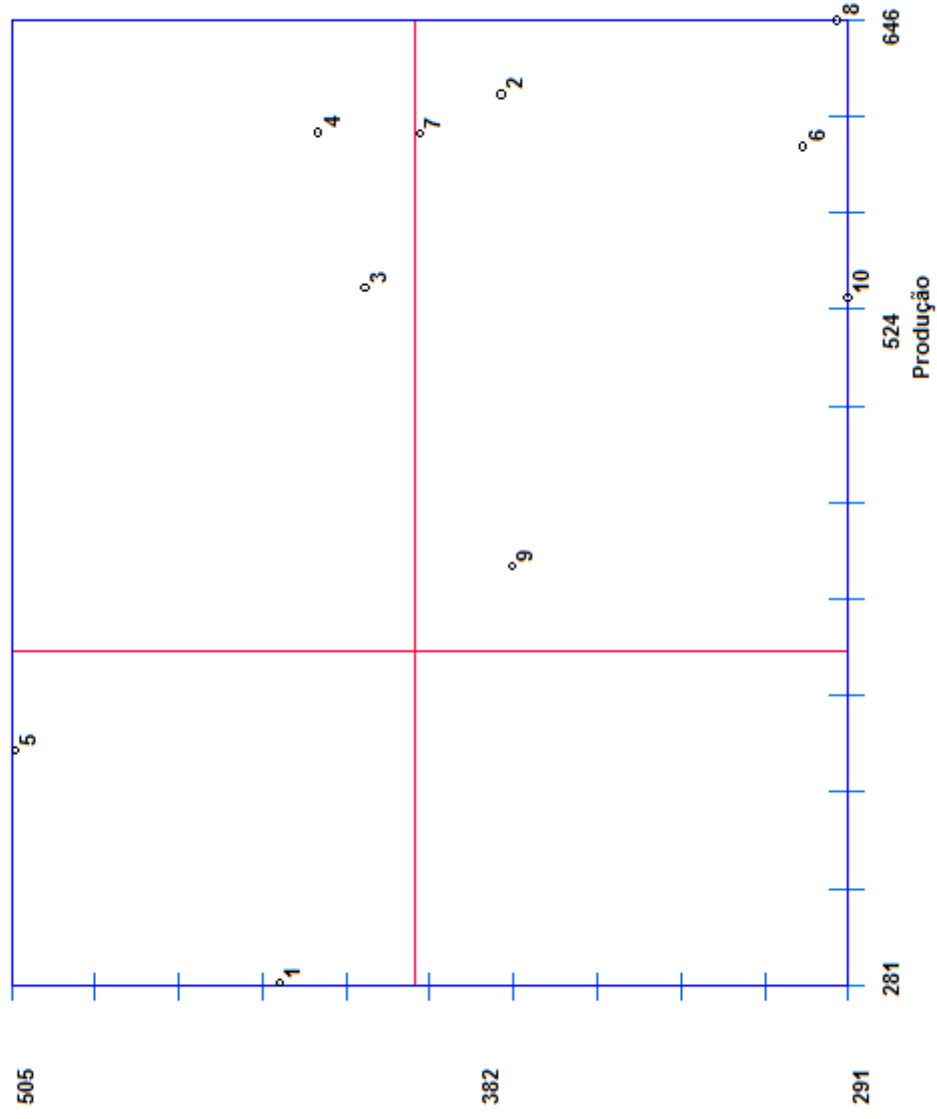
Progenies selecionadas :

8 2 4 7

	Produção	Altura	Floração	Nespigas	Resist.
8	645.77	438.56	293.91	271.58	486.63
2	617.69	357.27	379.88	348.69	521.26
4	603.34	373.31	426.63	495.47	212.18
7	602.94	649.21	400.59	259.28	431.09



Floração



b) Índice de seleção – Seleção simultânea de caracteres

Índice Clássico - Smith (1936) e Hazel (1943)

Número de indivíduos selecionados : 4

VARIÁVEL	PESO GENOTÍPICO	PESO FENOTÍPICO
Produção	3.0	2.29903
Altura	1.0	.11819
Floração	1.0	.01019
Nespigas	3.0	2.67221
Resist.	1.0	.71975

Determinante da matriz de covariância fenotípica : 34598887829608480241.76491

SELEÇÃO BASEADA NO ÍNDICE CLÁSSICO - Acréscimo

VARIÁVEL	Xo	Xs	h ² %	GS	GS %
Produção	524.46433	601.94417	87.19794	67.56082	12.88
Altura	441.58767	397.93333	26.34404	-11.50032	-2.6
Floração	382.69533	347.9075	82.6898	-28.76599	-7.52
Nespigas	328.02	364.18	92.03905	33.28132	10.15
Resist.	380.848	472.7675	87.48611	80.4168	21.12
GANHO TOTAL				140.99263	34.03

Variável assinalada com o símbolo # apresenta variância genética negativa.

Progênies selecionadas :					
4	2	10	8		
		Produção	Altura	Floração	Nespigas Resist.
4		603.34	373.31	426.63	495.47 212.18
2		617.69	357.27	379.88	348.69 521.26
10		540.97	422.59	291.21	340.99 671.
8		645.77	438.56	293.91	271.58 486.63



c) Seleção combinada

ANÁLISE DE VARIÂNCIA DA VARIÁVEL => Altura

FV	GL	SQ	QM	F	Probabilidade(%)
BLOCOS	5	174.10528	34.82106	3.15586	1.07294
GENÓTIPOS	7	437.29556	62.47079		
ENTRE PARCELAS	35	692.83028	19.79515		
DENTRO PARCELA	240	73.0933	.30456		
MÉDIA	4.26389				
CV(%)	42.59885				

PARÂMETROS GENÉTICOS E AMBIENTAIS

S ² blocos	.31304
S ² genotípica entre famílias	1.18543
S ² genotípica dentro famílias	3.5563
S ² FENOTÍPICA DENTRO DE FAMÍLIAS	.30456
S ² residual - (ambiental entre)	3.24843
S ² TOTAL	5.05146
HERDABILIDADE (US = MÉDIA FAMÍLIAS)	0.683129513220146
HERDABILIDADE (US = DENTRO FAMÍLIA)	11.6770327395651
HERDABILIDADE (US = INDIVÍDUO NO BLOCO)	1.00069972484068
HERDABILIDADE (US = INDIVÍDUO NO EXPERIMENTO)	0.938686249799856
CV EXPERIMENTAL - 100*RAIZ(QME/N))/μ (CV1)	42.5988527409803
CV EXPERIMENTAL - 100*RAIZ(S ² ENTRE)/μ (CV2)	42.2698832783177
CV GENÉTICO ENTRE (CV3)	25.5348262191831
CV GENÉTICO DENTRO (CV4)	44.227616374067
CV3/CV2	0.604090294052957
CV4/CV2	1.04631508165894

SELEÇÃO ENTRE FAMÍLIAS

PORCENTAGEM DE SELEÇÃO	25 (%)
NÚMERO DE FAMÍLIAS SELECIONADAS	2
FAMÍLIAS => 3 8	

MÉTODO 1 - $G_{Se} = DS h^2$	
MÉDIA DAS FAMÍLIAS SELECIONADAS	5.78056
DIFERENCIAL DE SELEÇÃO	1.51667
GANHO POR SELEÇÃO	1.03608
GANHO POR SELEÇÃO (%)	24.29894

MÉTODO 2 - $GS = i h DP_{ga}$	
INTENSIDADE DE SELEÇÃO	1.2711
GANHO POR SELEÇÃO	1.14385
GANHO POR SELEÇÃO (%)	26.82651

SELEÇÃO DENTRO - INDIVÍDUOS SELECIONADOS

MÉTODO 1 - $G_{Sd} = DS_{m h^2}$	
DIFERENCIAL DE SELEÇÃO MÉDIO	.56944
GANHO POR SELEÇÃO	.56944
GANHO POR SELEÇÃO (%)	13.35505

MÉTODO 2 - $G_{Sd} = i h DP_g$	
INTENSIDADE DE SELEÇÃO	1.2711
GANHO POR SELEÇÃO	2.39706
GANHO POR SELEÇÃO (%)	56.21772



SELEÇÃO COMBINADA

ÍNDICE : $G_{ijk} = b_1(Y_{ijk} - Y_{.j.}) + b_2(Y_{i..} - Y_{...})$

COEFICIENTE b1	1.27187
COEFICIENTE b2	.01
RELAÇÃO b1/b2	127.18651

RESPOSTA A SELEÇÃO COMBINADA

MÉTODO 1 : GS = DS	
GANHO POR SELEÇÃO	5.91142
GANHO POR SELEÇÃO (%)	138.63911
EFICIÊNCIA SELEÇÃO COMBINADA/SELEÇÃO ENTRE e DENTRO	3.68192
MÉTODO 2 : GS = i r(HI) DP(Var.Gen.)	
PORCENTAGEM DE SELECIONADOS (%)	4.16667
INTENSIDADE DE SELEÇÃO (tabelada)	2.13903
INTENSIDADE DE SELEÇÃO (estimada)	2.60298
GANHO POR SELEÇÃO	4.85779
GANHO POR SELEÇÃO (%)	113.92858

SELEÇÃO MASSAL

MÉDIA DOS SELECIONADOS	9.34167
DIFERENCIAL DE SELEÇÃO	5.07778
HERDABILIDADE	.93869
GANHO POR SELEÇÃO	4.76644
GANHO POR SELEÇÃO (%)	111.78622

SELEÇÃO MASSAL ESTRATIFICADA

MÉDIA DOS SELECIONADOS	8.1
DIFERENCIAL DE SELEÇÃO	3.83611
HERDABILIDADE	1.0
GANHO POR SELEÇÃO	3.83611
GANHO POR SELEÇÃO (%)	89.96743





A	B	C
10	9.8	4.2

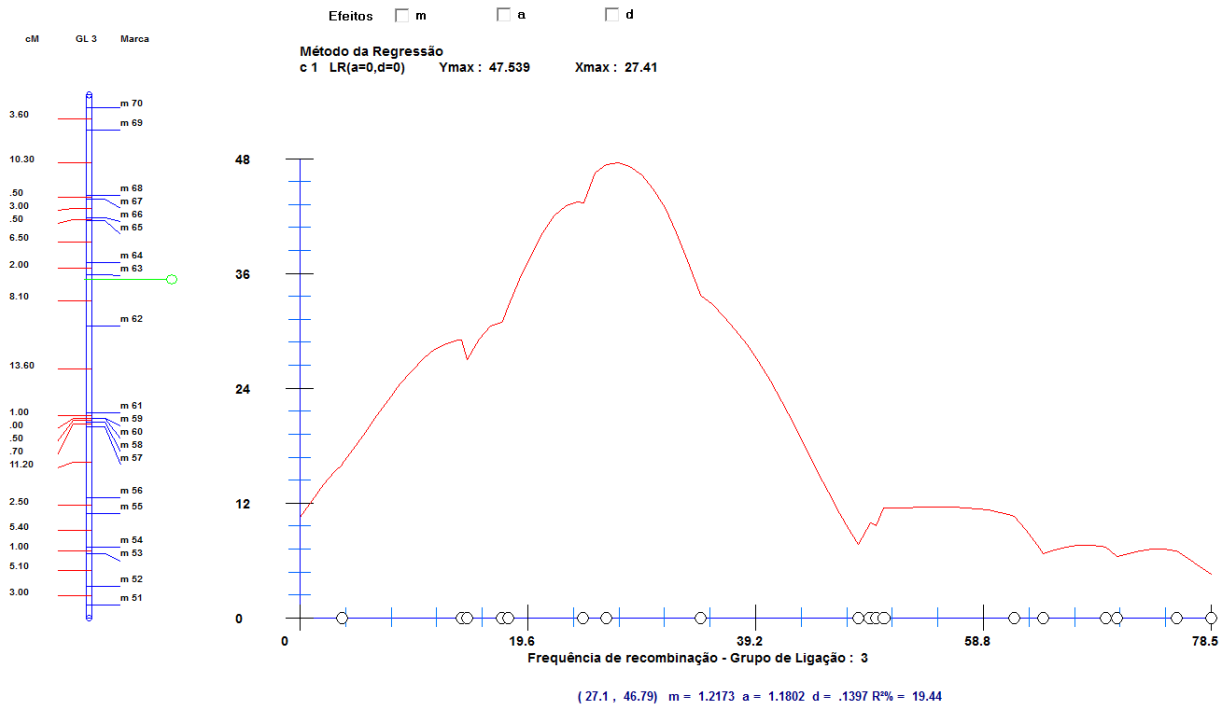
“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



GENÔMICA



7. UTILIZAÇÃO DA GENÔMICA



- Mapa genético
- Detecção de QTLs
- SAM
- GWS



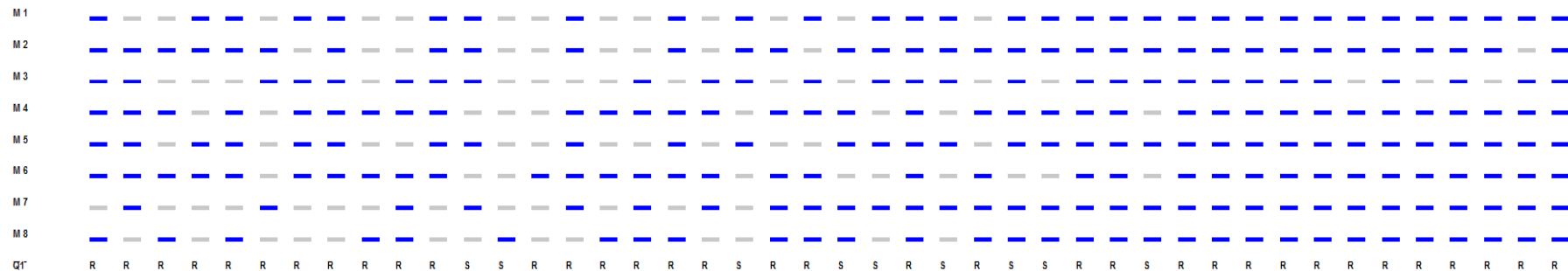
Exemplo:

População : F2

Número de marcadores : 8

Número de indivíduos = 44

Número de características fenotípicas : 1 (Resistência: R e S)



a) Teste de Segregação









Teste de segregação - Esperado 3:1

Marca	Obs1	Obs2	X ²	Probab. (%)
M 1	31	13	.48485	48.62344 ns
M 2	34	10	.12121	72.77235 ns
M 3	28	16	3.0303	8.17228 ns
M 4	35	9	.48485	48.62344 ns
M 5	32	12	.12121	72.77235 ns
M 6	34	10	.12121	72.77235 ns
M 7	31	13	.48485	48.62344 ns
M 8	31	13	.48485	48.62344 ns
Q1	35	9	.48485	48.62344 ns



b) Porcentagem de recombinação entre pares de locos

b1) Entre os marcadores M1 e M2

Teste de Ligação Gênica			
M 1	M 2	Observado	Esperado
		28	24.75
		3	8.25
		6	8.25
		7	2.75
Total		44	44

$\chi^2 = 10.94$ GL = 3 P = 1.2%
 Distância = 20.45

b2) Entre os pares de marcadores

Mapa de Ligação - Matriz de Distância Recombinação máxima: 30 LOD mínimo : 3

	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	Q1
M 1	100	22.59	29.93	47.77	6.72	50	41.39	48.19	57.35
M 2	1.96	100	48.19	50	14.71	65.83	34.93	50	63.36
M 3	1.17	0	100	53.97	33.36	55.73	29.93	60.24	53.97
M 4	0	0	0	100	45.39	12.83	39.69	20.22	16.03
M 5	6.84	3.58	0.75	0.03	100	57.35	39.34	53.97	67.98
M 6	0	0	0	3.74	0	100	59.39	34.93	2.41
M 7	0.17	0.49	1.17	0.2	0.25	0	100	35.12	57.35
M 8	0	0	0	2.31	0	0.49	0.54	100	39.69
Q1	0	0	0	2.82	0	8.03	0	0.2	100

Valores de distância (%) acima da diagonal e do LOD abaixo da diagonal.

b2) Entre os pares de marcadores

Teste de Ligação - Distância entre Marcas

Marcas		++	+-	-+	--	X²(9:3:3:1)	Probab.(%)	Dist.(MMV)	LOD	Dist.fhr
M 1	M 2	28	3	6	7	10.94949	1.20021	22.59	1.96	20.22
M 1	M 3	23	8	5	8	11.43434	.95948	29.93	1.17	14.71
M 1	M 4	25	6	10	3	1.0101	79.88078	NL	47.77	0
M 1	M 5	30	1	2	11	36.9697	.0	6.72	6.84	0
M 1	M 6	24	7	10	3	.60606	89.50436	NL	50	0
M 1	M 7	23	8	8	5	1.9798	57.66103	NL	41.39	.17
M 1	M 8	22	9	9	4	1.0101	79.88078	NL	48.19	0
M 1	Q1	24	7	11	2	1.33333	72.12334	NL	57.35	0
M 2	M 3	22	12	6	4	3.19192	36.2971	NL	48.19	0
M 2	M 4	27	7	8	2	.60606	89.50436	NL	50	0
M 2	M 5	30	4	2	8	18.06061	.04274		14.71	3.58
M 2	M 6	25	9	9	1	1.25253	74.04361	NL	65.83	0
M 2	M 7	26	8	5	5	3.19192	36.2971	NL	34.93	.49
M 2	M 8	24	10	7	3	.60606	89.50436	NL	50	0
M 2	Q1	26	8	9	1	1.25253	74.04361	NL	63.36	0
M 3	M 4	22	6	13	3	3.67677	29.85492	NL	53.97	0
M 3	M 5	23	5	9	7	8.0404	4.5184		33.36	.75
M 3	M 6	21	7	13	3	3.51515	31.88023	NL	55.73	0
M 3	M 7	23	5	8	8	11.43434	.95948		29.93	1.17
M 3	M 8	18	10	13	3	4.9697	17.40296	NL	60.24	0
M 3	Q1	22	6	13	3	3.67677	29.85492	NL	53.97	0
M 4	M 5	26	9	6	3	.76768	85.71817	NL	45.39	.03
M 4	M 6	32	3	2	7	16.76768	.07889		12.83	3.74
M 4	M 7	26	9	5	4	1.9798	57.66103	NL	39.69	.2
M 4	M 8	29	6	2	7	12.64646	.5467		20.22	2.32
M 4	Q1	32	3	3	6	12.64646	.5467		16.03	2.82
M 5	M 6	24	8	10	2	.60606	89.50436	NL	57.35	0
M 5	M 7	24	8	7	5	2.06061	55.99234	NL	39.34	.25
M 5	M 8	22	10	9	3	.76768	85.71817	NL	53.97	0
M 5	Q1	24	8	11	1	2.06061	55.99234	NL	67.98	0
M 6	M 7	23	11	8	2	1.25253	74.04361	NL	59.39	0
M 6	M 8	26	8	5	5	3.19192	36.2971	NL	34.93	.49
M 6	Q1	34	0	1	9	32.28283	.00005	2.41	8.04	9.54
M 7	M 8	24	7	7	6	4.24242	23.64486	NL	35.12	.54
M 7	Q1	24	7	11	2	1.33333	72.12334	NL	57.35	0
M 8	Q1	26	5	9	4	1.9798	57.66103	NL	39.69	.2
Esperado		24.75		8.25		8.25	2.75			

MMV : método da máxima verossimilhança.

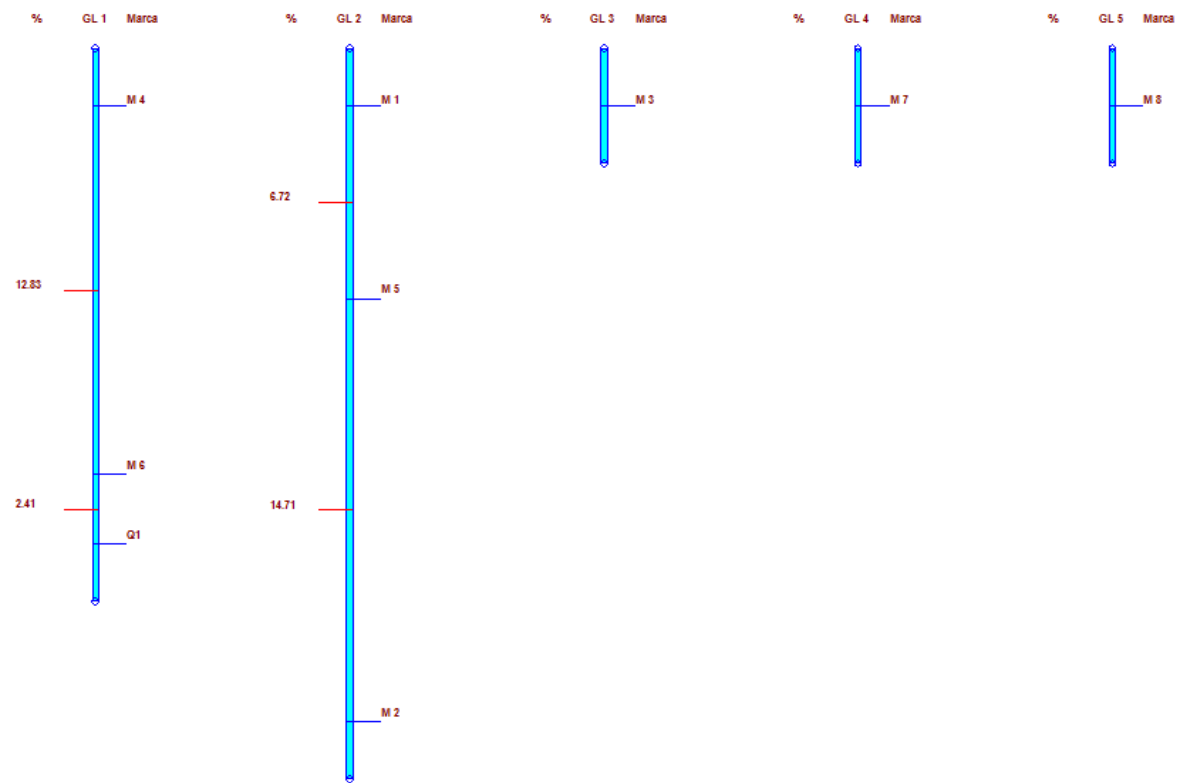
Dist.fhr : Distância calculada com base na frequência do homozigoto recessivo.

NL : Não ligado. Nível de significância estimado acima de 5% de probabilidade pelo teste X²






c) Mapa de ligação

Grupos de Ligação - Melhor Ordem
Distâncias expressa em : Frequência de recombinação (%)

Grupo	:	1	:	[M 4]	12.8	[M 6]	2.4	[Q1]
Grupo	:	2	:	[M 1]	6.7	[M 5]	14.7	[M 2]
Grupo	:	3	:	[M 3]				
Grupo	:	4	:	[M 7]				
Grupo	:	5	:	[M 8]				



d) Utilização da genômica para fins de seleção

Teste de Ligação Gênica			
M 6	Q1	Observado	Esperado
	R	34	24.75
	S	0	8.25
	R	1	8.25
	S	9	2.75
Total		44	44
 $\chi^2 = 32.28$ GL = 3 P = 0% Distância = 2.27			

Sem auxílio da genômica:

- 44 plantas seriam levadas a campo
- 11 plantas (25%) seriam descartadas

Com auxílio da genômica:

- 34 plantas seriam levadas a campo
- Nenhuma planta descartada





A	B	C
10	9.8	4.2

“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



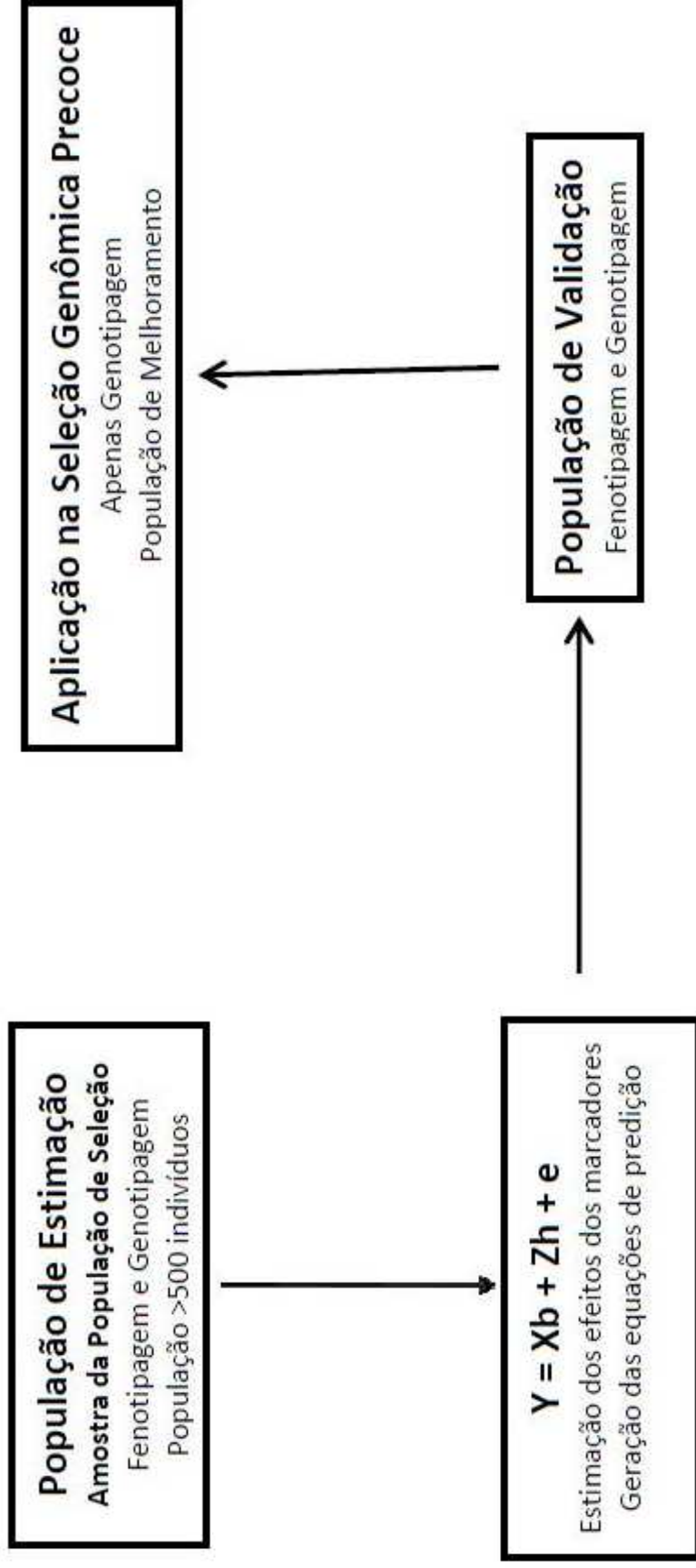
SELEÇÃO GENÔMICA AMPLA



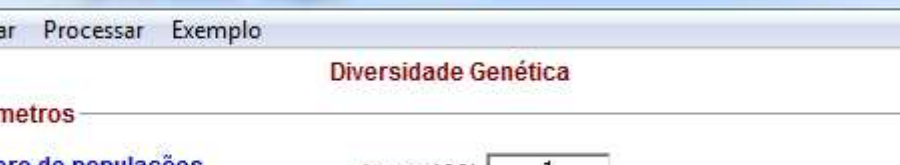
8. UTILIZAÇÃO DA GWS

- A Seleção Genômica Ampla (*Genome Wide Selection* - GWS), proposta inicialmente no ano de 2001 por Meuwissen et al. (2001),
- Consiste na análise de um grande número de marcadores amplamente distribuídos no genoma.
- Após a obtenção destes marcadores, seus efeitos são estimados baseados em dados fenotípicos de uma população conhecida usada para estimação.
- Uma vez que seus efeitos são modelados e estimados, estes são testados em uma população de validação e, então, seleciona-se os marcadores que explicam parte da variância genética do caráter em estudo para que sua informação seja efetivamente incorporada à etapa de seleção do Programa de Melhoramento





a. Genotipagem



Simulação de Populações e Marcadores Dialélicos

Finalizar Processar Exemplo

Diversidade Genética

Parâmetros

Número de populações (max=100)

Número de indivíduos/população Variação

Número locos (max = 2000)

Tipo de marcador

Arquivo de saída dos dados

genotipo - Bloco de notas																																
Arquivo	Editar	Formatar	Exibir	Ajuda																												
12	12	22	22	12	22	11	11	12	11	11	22	11	11	22	22	12	22	11	12	11	12	22	12	12	22	12	12	22	12	22	12	22
12	12	12	22	22	12	22	12	11	11	12	11	12	22	22	12	12	11	11	11	12	22	12	22	12	22	12	11	11	11	11	22	22
22	12	22	12	11	22	12	11	22	22	11	11	12	11	11	12	22	11	12	11	12	22	22	12	11	22	22	12	12	11	11	22	11
22	22	22	12	12	22	22	22	11	12	22	22	11	22	22	12	22	22	22	22	22	12	11	22	22	11	11	12	12	22	12	22	12
22	12	22	11	22	22	22	22	22	22	11	12	12	12	11	11	12	22	12	22	11	22	11	22	11	12	12	22	12	22	22	22	12
22	12	11	11	11	12	22	12	11	11	22	12	11	11	22	11	11	22	11	12	12	12	11	22	22	22	22	11	22	22	22	12	12
22	22	22	11	22	11	11	12	12	12	22	11	12	12	22	22	11	12	11	22	11	12	11	12	22	22	22	12	22	22	22	11	11
11	11	12	11	11	11	22	22	22	11	11	22	12	12	12	22	12	11	22	12	22	12	12	22	12	22	12	11	12	22	22	11	11
22	11	12	11	12	11	22	12	11	12	22	22	12	12	11	22	12	11	11	22	11	22	22	22	11	22	11	12	11	12	11	12	12
11	12	11	12	12	22	11	22	11	12	12	22	11	12	12	11	12	11	11	12	11	12	12	22	12	12	12	22	22	11	22	11	22
11	11	11	12	22	22	11	11	11	11	22	22	22	22	22	12	22	12	22	22	22	11	22	11	12	12	12	22	11	22	11	22	22
22	22	22	12	12	11	12	11	22	22	11	22	22	12	22	12	12	22	12	12	12	12	12	11	12	12	12	12	22	22	22	22	22
12	12	11	22	22	12	12	11	11	12	22	12	22	12	12	22	12	12	11	11	12	11	11	22	11	22	11	22	11	12	22	11	11
11	22	12	11	11	22	12	12	11	22	12	22	11	12	11	22	12	22	11	12	11	12	11	22	12	12	12	11	12	11	12	11	11
11	12	12	11	12	12	12	22	11	11	22	22	11	12	12	11	11	22	11	11	22	22	11	11	11	11	12	12	12	12	12	12	12
22	12	12	22	11	11	22	12	12	12	12	11	22	22	12	11	12	22	12	22	11	12	11	12	11	12	22	11	22	11	12	22	11
12	22	11	22																													

b. Fenotipagem

Simulação de Caracteres Quantitativos

Arquivo de dados
C:\Nova pasta\genotipo.txt

Parâmetros

Número de indivíduos (linhas) 1000

Número de marcadores (colunas) 500

Opções

Retornar

Ler Dados

Arquivo de saída dos dados

Fenotípico C:\Nova pasta\genotipo_feno.txt

Genotípico C:\Nova pasta\genotipo_geno.txt

Número de Locus Máximo 500

Loci selecionados

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68
69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Ação gênica

Efeito a e d ☐ Uniforme ☒ Binomial ☐ Poisson ☐ Exponencial

Efeito d ☒ Nulo

Herdabilidade da característica (0-1)

Constante (u)

Programa GENES	Simulação
Arquivo de dados	C:\Nova pasta\genotipo.txt
Número de indivíduos	1000
Número de marcadores	500
Arquivo de saída	C:\Nova pasta\genotipo_feno.txt
Arquivo de saída	C:\Nova pasta\genotipo_geno.txt
Data	05-01-2014



Efeito do loco

Loco	Uniforme	Binomial	Poisson	Exponencial
1	0.01	0.	0.000454	0.39346934
2	0.01	0.	0.00227	0.23865122
...				
44	0.01	0.03895256	0.	0.
45	0.01	0.0484743	0.	0.
46	0.01	0.0579584	0.	0.
47	0.01	0.0665905	0.	0.
48	0.01	0.07352701	0.	0.
49	0.01	0.07802866	0.	0.
50	0.01	0.07958924	0.	0.
51	0.01	0.07802866	0.	0.
52	0.01	0.07352701	0.	0.
53	0.01	0.0665905	0.	0.
54	0.01	0.0579584	0.	0.
55	0.01	0.0484743	0.	0.
56	0.01	0.03895256	0.	0.
...				
99	0.01	0.	0.	0.
100	0.01	0.	0.	0.
soma	1.	1.	0.9999546	1.

Herdabilidade : .4
Distribuição dos efeitos a :Binomial
Distribuição dos efeitos d :Nulo

Estimativa	Genética	Fenotípica
Média	.148137	.148137
S²	.018634	.047055
Correlação	.637773	
R²	.406754	
n	1000	

Seleção Genômica Ampla



Arquivo de dados

Fenotípico

C:\Nova pasta\genotipo_feno.txt

Genotípico

C:\Nova pasta\genotipo_geno.txt

Procedimentos


☐ RR BLUP

☐ G BLUP

☐ LASSO

Proporção de dados para validação

Resultado

 Script Dependência

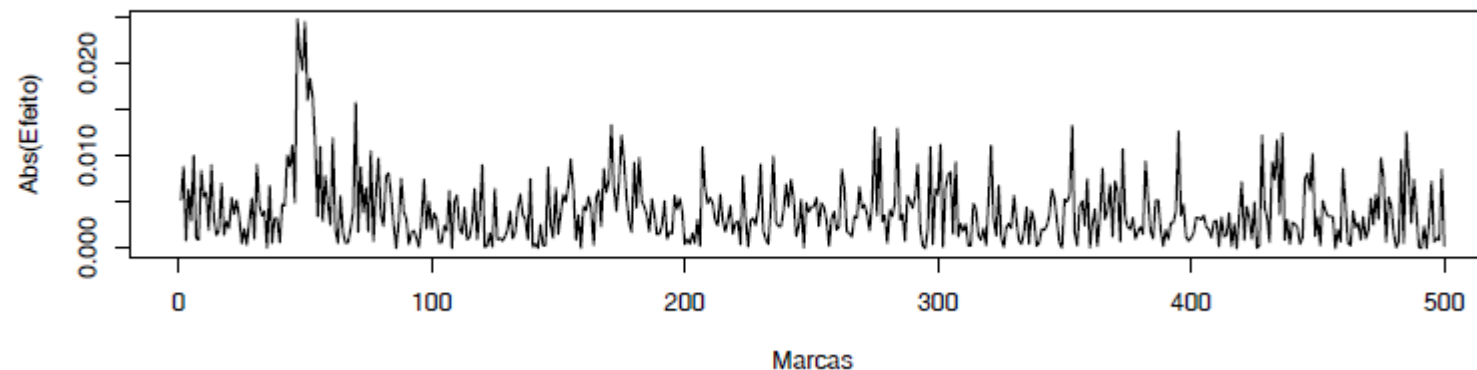
C:\dados\ScriptsR\GWS_RRblup.txt
rrBLUP BLR MASS



Variável = vfen
Média = 100.1505
Desvio-padrão 0.2168797
variância = 0.04703679
Num. obs = 800
DP média 0.007667854

Efeito dos marcadores

[1]	5.271359e-03	8.851225e-03	-8.640793e-04	6.366001e-03	2.994299e-03
[6]	1.003776e-02	-1.087664e-03	-9.526799e-04	8.428165e-03	5.577558e-03
[11]	6.047749e-03	1.965340e-03	-9.039372e-03	-3.278473e-03	-1.426062e-03
[16]	1.912324e-03	-7.038966e-03	1.462921e-03	2.922860e-03	-2.186472e-03
[21]	-5.470340e-03	-3.763219e-03	-5.189021e-03	-3.399641e-03	5.566117e-04
[26]	-2.098626e-03	3.799496e-04	-2.885364e-03	-5.402590e-03	-1.041173e-03
[31]	-9.089106e-03	-4.659429e-03	-3.555734e-03	3.997355e-03	4.845967e-05
[481]	-8.503819e-05	-7.284646e-04	9.622414e-03	5.687507e-04	1.260996e-02
[486]	7.939012e-03	-2.749067e-03	7.444189e-03	3.925074e-03	-1.080992e-04
[491]	-6.163062e-05	2.393033e-03	-4.272430e-05	2.801869e-03	7.250899e-03
[496]	7.172935e-04	-1.099128e-03	9.258061e-04	-8.529307e-03	-2.660772e-04



Valores Genômicos

[,1]

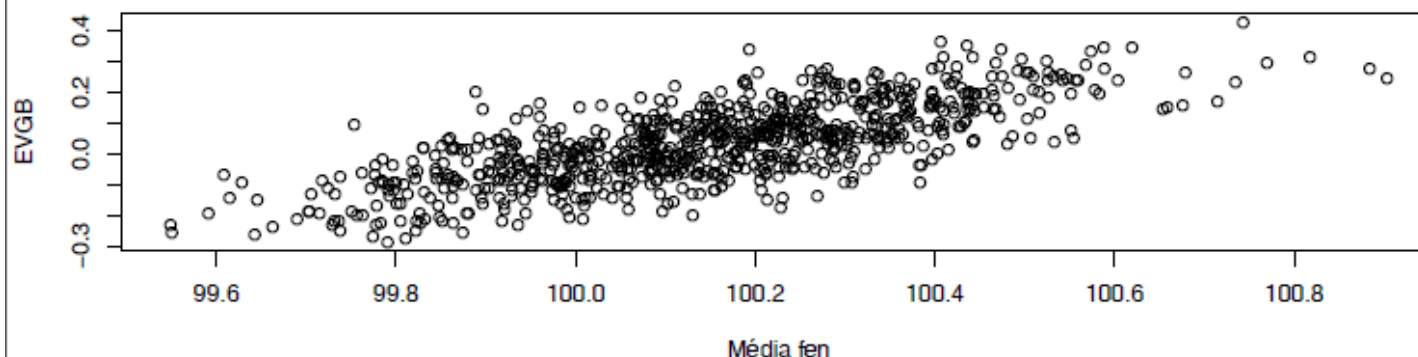
[1,] -5.669537e-02
[2,] 1.559289e-01
[3,] 3.394038e-02
[4,] 1.182204e-01
[5,] 1.797679e-01
[6,] -1.346971e-01

[797,] -2.215003e-01
[798,] 6.727659e-02
[799,] -1.182489e-02
[800,] -3.634309e-02

$R^2(\text{EGBV}, \text{Média}) = 0.5434341$

Valores Genômicos ordenados - 10%

210	441	712	370	774	690	530	552	
0.4294374	0.3646347	0.3508671	0.3459773	0.3450094	0.3411930	0.3363909	0.3344783	
328	281	170	635	137	17	614	450	
0.3139108	0.3133662	0.3108285	0.3081521	0.3039571	0.2935751	0.2932576	0.2865229	
471	631	151	527	306	241	681	601	
0.2812745	0.2798698	0.2744829	0.2744171	0.2741257	0.2728953	0.2706511	0.2669322	
770	290	167	367	38	168	410	311	
0.2654402	0.2636590	0.2623694	0.2609954	0.2601749	0.2601114	0.2600782	0.2584279	
795	454	652	548	389	761	650	294	
0.2554688	0.2518861	0.2517003	0.2500791	0.2488590	0.2481960	0.2449519	0.2445772	
531	497	401	730	721	232	785	146	
0.2434685	0.2432996	0.2419275	0.2413827	0.2403742	0.2365950	0.2363990	0.2362662	
125	671	490	510	266	511	561	49	
0.2362227	0.2361805	0.2358452	0.2352163	0.2348824	0.2336138	0.2315015	0.2307644	
130	285	457	709	224	670	669	186	
0.2300195	0.2280260	0.2272132	0.2266744	0.2259765	0.2228409	0.2225686	0.2207622	
113	765	207	325	608	327	411	432	
0.2200083	0.2174747	0.2158119	0.2149029	0.2148542	0.2134010	0.2125621	0.2111937	
189	177	208	610	346	121	97	777	
0.2106342	0.2096100	0.2091435	0.2077397	0.2053947	0.2045838	0.2035963	0.2034071	





A	B	C
10	9.8	4.2

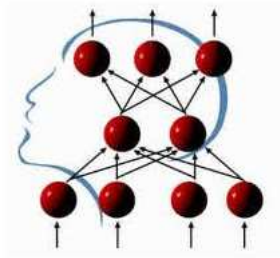
“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



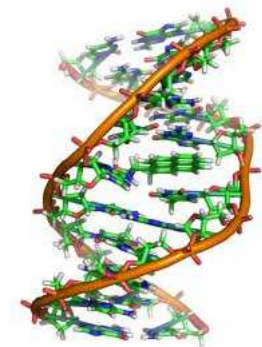
REDES NEURAIS



9. INTELIGÊNCIA ARTIFICIAL – REDES NEURAIS



- Análise Classificatória
- Predição de Valores Genéticos
- Ajuste de modelos
- Previsões temporais



Redes Neurais são modelos computacionais não lineares, inspirados na estrutura e operação do cérebro humano, que procuram reproduzir características humanas, tais como:

*aprendizado,
associação,
generalização e
abstração.*



Devido à sua estrutura, as Redes Neurais são bastante efetivas no aprendizado de padrões a partir de dados não-lineares, incompletos, com ruído e até compostos de exemplos contraditórios.



a. Inteligência computacional

Busca, através de técnicas inspiradas na Natureza, o desenvolvimento de sistemas inteligentes que imitem aspectos do comportamento humano, tais como: aprendizado, percepção, raciocínio, evolução e adaptação.

O que é Inteligência Computacional?

- *"O ramo da ciência da computação preocupada com a automação de comportamento inteligente." [LUGER & STUBBLEFIELD, 93]*
- *"É o estudo de como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor." [RICH & KNIGHT, 93]*

Objetivo científico:

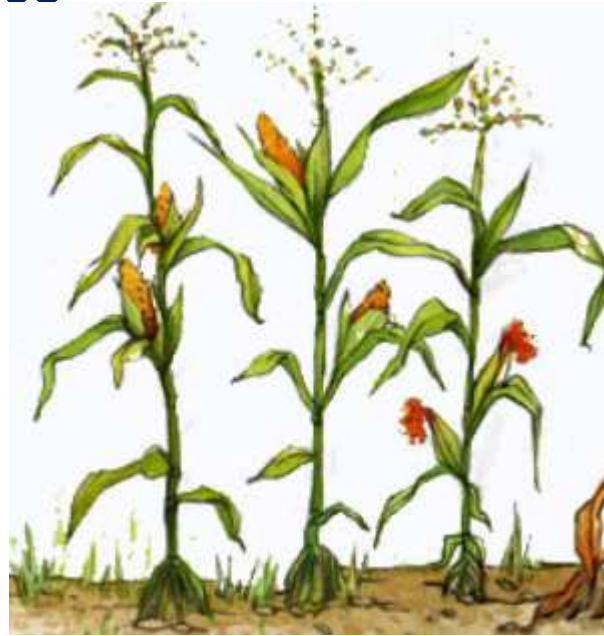
- Entender os princípios que tornam possíveis comportamentos inteligentes.

Objetivo de engenharia:

- Especificar métodos para projeto de artefatos inteligentes úteis.



b. Modelo Estatístico

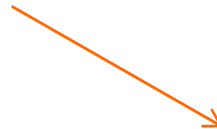
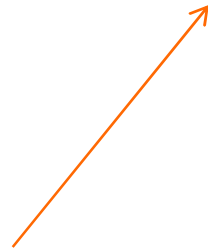
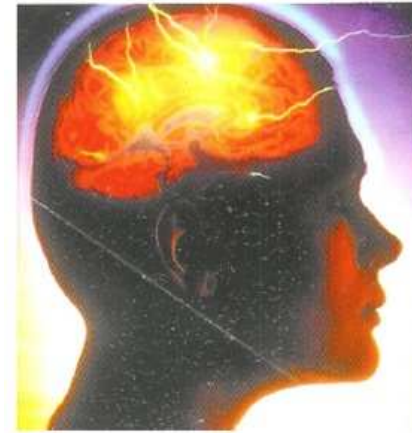


$$Y_{ij} = \mu + G_i + B_j + E_{ij}$$

- a. Modelo estatístico
- b. Pressuposições e distribuições
- c. Análises estatísticas



c. Modelo biológico





Bloco 1



T3 T1 T2 T8 T9 T5 T6 T10 T7 T4

Bloco 2



T2 T10 T1 T7 T5 T3 T4 T8 T9 T6

Bloco 3



T1 T8 T5 T10 T3 T6 T2 T7 T4 T9



Trat	Médias
T1	10,0
T2	9,8
...	
T10	7,5



Paradigma atual

Processamento

Resumo das informações

Tomada de decisão





Média
Máximo
Mínimo
DP
etc

Objetivo

Avaliar a eficiência da RNA em prever valores genéticos em comparação com o valor fenotípico médio, tendo por base um experimento em blocos ao acaso, com 6 repetições e avaliação de 100 genótipos em relação a uma característica de herdabilidade igual a 40%. A hipótese é a de que o pesquisador possa substituir a média por um outro critério de seleção que seria o valor da rede treinada a partir de informações estatísticas relevantes, tais como:

- A média do genótipo (critério convencional)
- Os valores máximos de cada genótipo nos b blocos
- Os valores mínimos de cada genótipo nos b blocos
- Os valores do DP de cada genótipo
- Os valores dos coeficientes de variação de cada genótipo
- Os valores da soma de rank obtida em cada bloco pelos g genótipos
- Os valores da estatística P_i que mede a distância do genótipo i ao genótipo ideal caracterizado por produzir o máximo em todos os blocos
- A média do genótipo transformada em variável qualitativa considerando a sua variação em torno da média nos limites, inferior e superior, a 1, 2 e 3 desvio-padrão

Arquivo de dados:

Arquivo original (teste) :

Informações da avaliação de 100 genótipos, em DBC, com 6 blocos.
Característica com H^2 de 40%, média igual a 100 e CV igual a 12.

Arquivos de validação:

Informações da avaliação de 100 genótipos simulados, em DBC, com 6 blocos. A simulação mantém a H^2 de 40%, média igual a 100 e CV igual a 12 e a mesma matriz de variâncias e covariância entre dados dos diferentes blocos. Também é predito o valor genético.

Arquivos de treinamento:

Informações da avaliação de 5000 genótipos simulados, em DBC, com 6 blocos. A simulação mantém a H^2 de 40%, média igual a 100 e CV igual a 12 e a mesma matriz de variâncias e covariância entre dados dos diferentes blocos. Também é predito o valor genético.

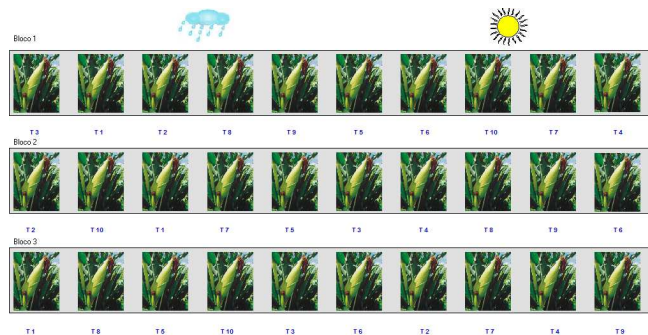




Ensaio de treinamento – Histórico ou *in silico*



Ensaio - Teste



...



Ensaio de validação

Resultados

I. **Análise dos dados de teste**

A análise do conjunto dos dados comprova a eficácia da simulação em prover informações de genótipos com os valores da média (igual a 100), coeficiente de variação (igual a 12) e herdabilidade (igual a 40%) tal como estabelecidos pelo pesquisador.

ANÁLISE DE VARIÂNCIA - ARQUIVO : C:\dados\IA\ProjetosRNA\redel.dat

FV	GL	SQ	QM	F	Probabilidade
BLOCOS	5	512.576	102.5152	.71192	
GENÓTIPOS	99	23760.48157	240.00486	1.66672	.00023
RESÍDUO	495	71279.2115	143.99841		
TOTAL	599	95552.2691			

DESCRIÇÃO	ESTIMATIVA
MÉDIA	99.99985
VARIÂNCIA GENÉTICA	16.00108
VARIÂNCIA RESIDUAL	143.99841
H ² (MÉDIA)	40.00188
H ² (INDIVÍDUOS)	10.0007
CV Experimental	11.99995
CV genético	4.00014



II. Qualidade das estatísticas auxiliares para a predição do valor genético

Algumas estatísticas chegam a ser critérios mais confiáveis em prever o valor genético do que a média do genótipo, conforme pode ser observado a seguir:

r^2 : Vg x Média	.40	<= herdabilidade do caráter
r^2 : Vg x Máximo	.2929	
r^2 : Vg x Mínimo	.2416	
r^2 : Vg x DP	.0015	
r^2 : Vg x CV	.0061	
r^2 : Vg x Pi	.3621	
r^2 : Vg x SRank	.3843	
r^2 : Vg x Classe	.4289	

Vg = valor genético verdadeiro (conhecido por ser gerados por simulação)



III - Replicação e Ampliação de Arquivo de Dados

Por simulação foram geradas 10 réplicas do arquivo original (redelvall.dat a redelvall10.dat) e uma ampliação (redelTr.dat) com a avaliação de 5000 genótipos. O resultado da análise deste arquivo ampliado é apresentado a seguir, podendo ser verificado que todos os parâmetros da simulação (média, cv, herdabilidade e matriz de variâncias e covariâncias entre informações dos b blocos) foram preservados.

ANÁLISE DE VARIÂNCIA - ARQUIVO :

C:\dados\IA\ProjetosRNA\redelTr.dat

FV	GL	SQ	QM	F	Probab
BLOCOS	5	25628.80008	5125.76002	35.59595	
GENÓTIPOS	4999	1199784.3157	240.00486	1.66672	.19346
RESÍDUO	24995	3599240.19244	143.99841		
TOTAL	29999	4824653.30822			
DESCRIÇÃO		ESTIMATIVA			
MÉDIA		99.99985			
VARIÂNCIA GENÉTICA		16.00108			
VARIÂNCIA RESIDUAL		143.99841			
H ² (MÉDIA)		40.00188			
H ² (INDIVÍDUOS)		10.0007			
CV Experimental		11.99995			
CV genético		4.00014			

A qualidade das informações das estatísticas auxiliares também é mantida conforme ilustrado a seguir:

Valores de r^2

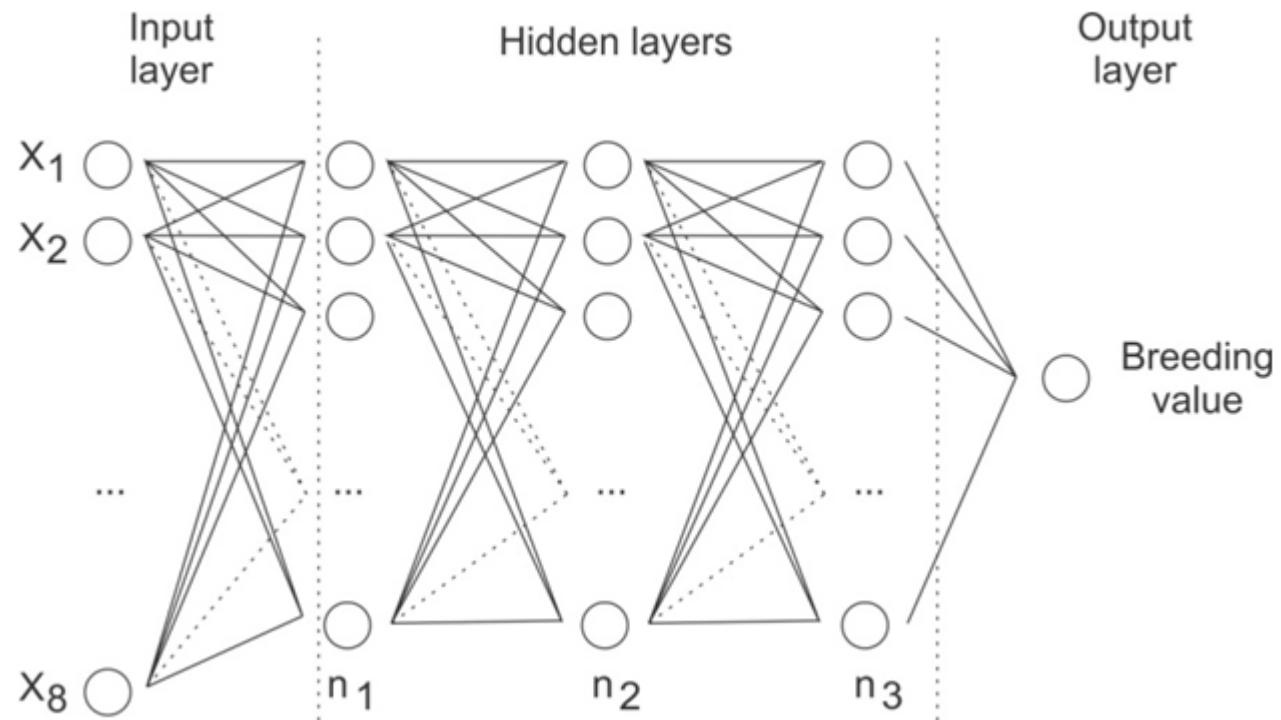
r^2 : Vg x Média	.40
r^2 : Vg x Máximo	.2901
r^2 : Vg x Mínimo	.275
r^2 : Vg x DP	.0047
r^2 : Vg x CV	.0123
r^2 : Vg x Pi	.3856
r^2 : Vg x SRank	.3697
r^2 : Vg x Classe	.3418

Para os arquivos de validação foram obtidos os seguintes resultados:

	Val1	Val2	Val3	Val4	Val5	Val6	Val6	Val8	Val9	Val10
Média	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Máximo	0.2812	0.2761	0.2277	0.2191	0.3625	0.1214	0.4266	0.3212	0.3244	0.2772
Mínimo	0.3334	0.3015	0.3072	0.1954	0.2995	0.369	0.1005	0.1661	0.2159	0.262
DP	0.0565	0.002	0.0491	0.0001	0.0015	0.0319	0.1361	0.0244	0.0002	0.0044
CV	0.1426	0.0334	0.1464	0.0294	0.0346	0.1183	0.0417	0	0.018	0.0446
Pi	0.4034	0.3864	0.4063	0.3708	0.3864	0.4195	0.3176	0.3547	0.359	0.3705
SRank	0.3321	0.3764	0.3767	0.3534	0.3531	0.3622	0.3711	0.3947	0.4199	0.3261
Classe	0.3663	0.4247	0.1693	0.3207	0.2838	0.4074	0.3934	0.3584	0.3691	0.4981



IV – Uso da Rede Neural



Arquitetura

Treinamento: C:\dados\IA\ProjetosRNA\rede1Tr.dat

EFICÁCIA DA REDE

Correlação VRede x VG: 0.66818816

R² VRede x VG: 0.44647542

EFICÁCIA DA MÉDIA

Correlação Média x VG: 0.63244036

R² Média x VG: 0.39998081

EFICÁCIA DO TREINAMENTO REDE

Correlação VRede máxima x VG: 0.69181707

R² VRede máxima x VG: 0.47861086

CONFIGURAÇÃO DA REDE

Neurônios camada 1: 5

Neurônios camada 2: 5

Neurônios camada 3: 1

Função de ativação - camada 1: 2

Função de ativação - camada 2: 1

Função de ativação - camada 3: 3

(*) Função de ativação 1=tansig 2=logsig 3 = purelin

RedexVG	MédiavG	RedexMédia
0.4709	0.4000	0.8457
0.3397	0.4000	0.8403
0.4403	0.4000	0.8788
0.4857	0.4000	0.8607
0.4107	0.4000	0.8738
0.3518	0.4000	0.9330
0.4229	0.4000	0.8604
0.4431	0.4000	0.8820
0.4807	0.4000	0.8359
0.6279	0.4000	0.7341

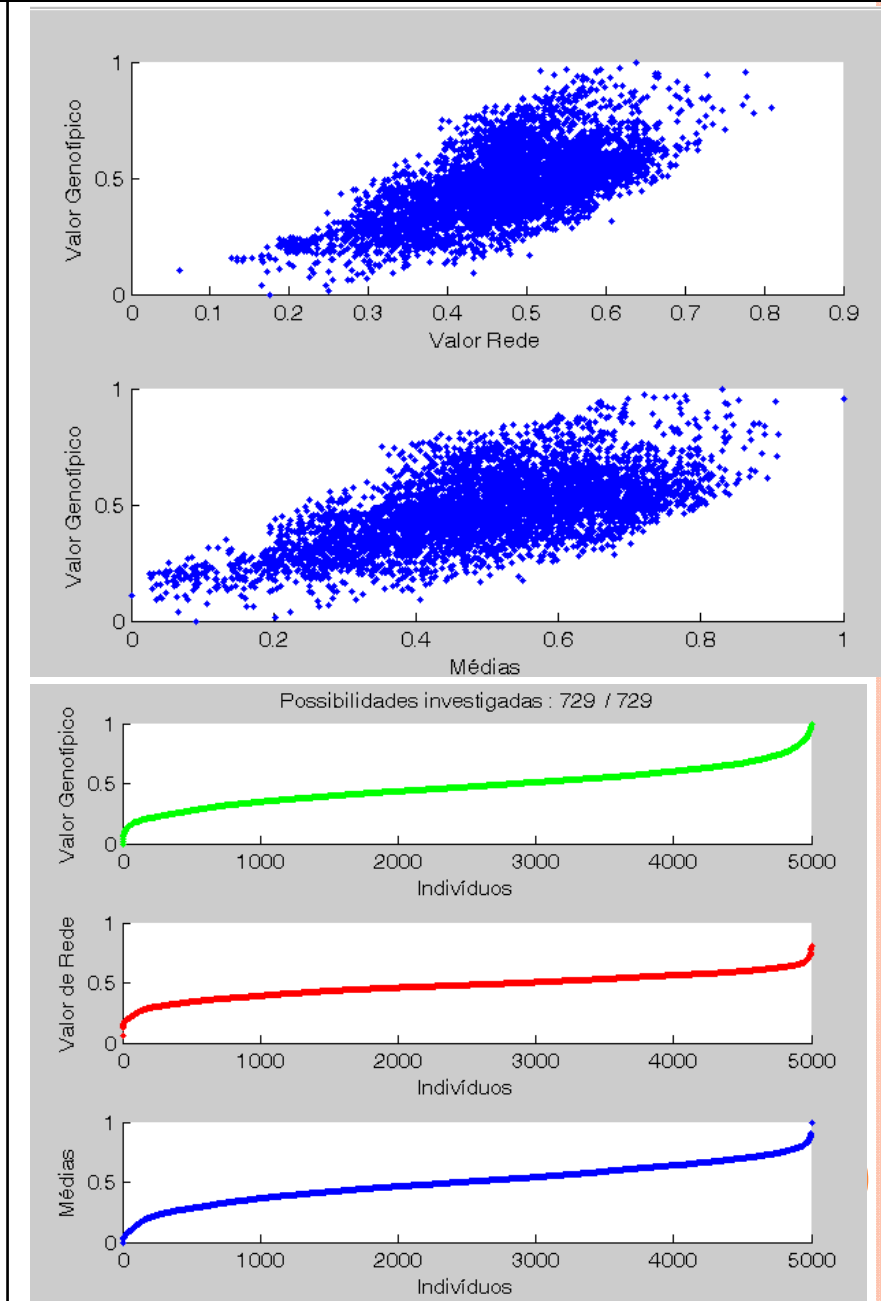
respostafinal =

Md 0.4474 0.4000 0.8545 8.0000

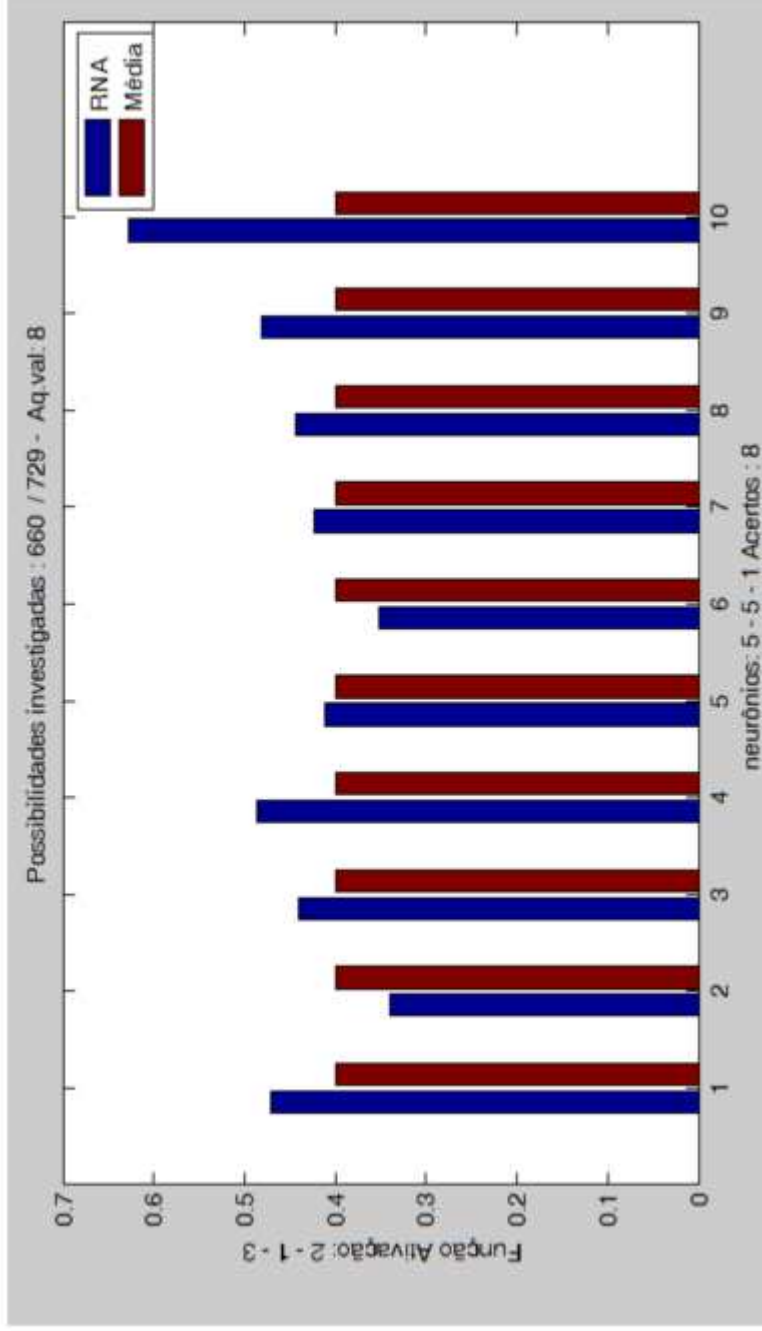
Max 0.6279 0.4000 0.9330 8.0000

Min 0.3397 0.4000 0.7341 8.0000

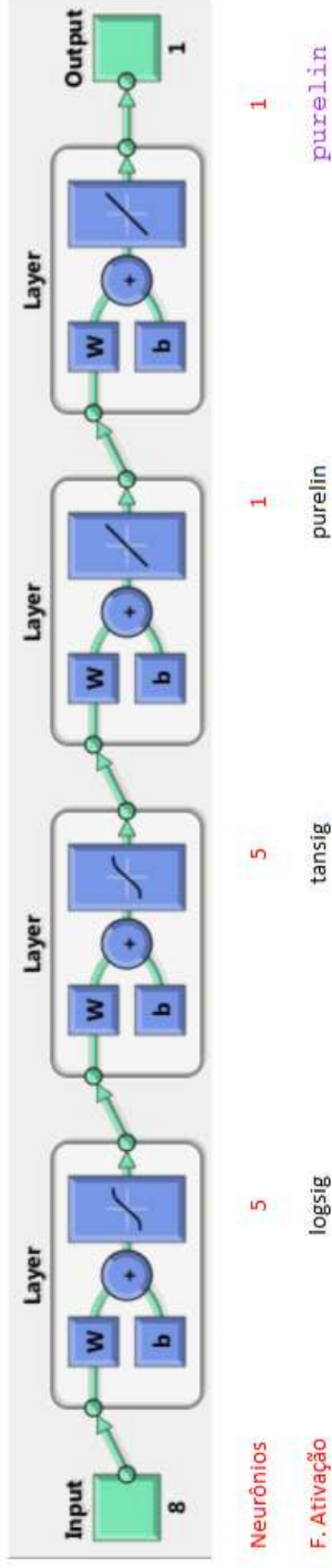
Resultado do Treinamento



Acurácia



Arquitetura da Rede

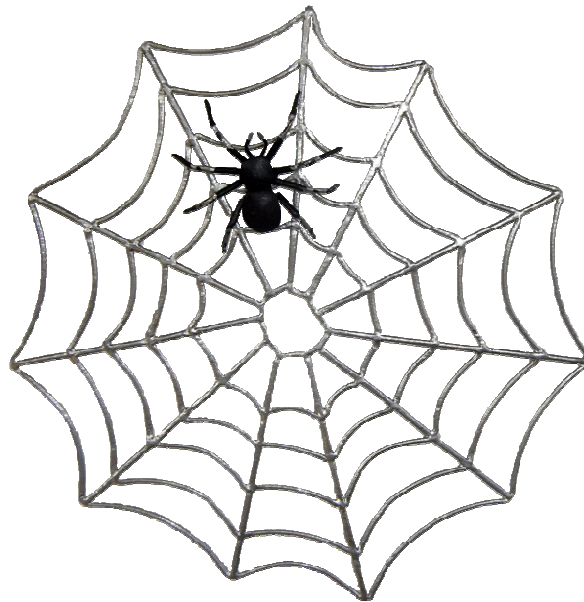


CONCLUSÃO

“Alternativas de análise de experimentos visando à otimização na identificação de genótipos superiores de milho”



Genética Quantitativa



Melhoramento

Experimentação

Redes Neurais

Biometria

Genômica : SAM e GWS



MUITO OBRIGADO

