

# CONSANGUINIDADE EM PLANTAS

## INBREEDING IN PLANTS

POR

PAULO DE OLIVEIRA\*

### I. INTRODUÇÃO

Consanguinidade, ou o termo que em sua substituição se prefira utilizar tanto para animais como para plantas [1], refere-se à união de gâmetas geneticamente aparentados. Segundo a biologia da reprodução de cada espécie haverá maior ou menor incidência de consanguinidade, e no caso das plantas cultivadas oferece-se um leque muito largo de perspectivas devido à multiplicidade de estratégias de reprodução sexuada que podem ser encontradas, desde as espécies anuais autogâmicas às florestais de ciclo de vida longo e com mecanismos de auto-incompatibilidade.

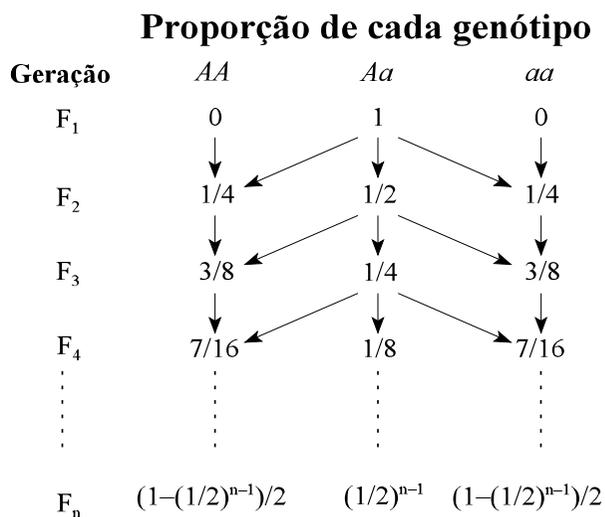
A consanguinidade pode ter uma grande importância para o melhoramento de plantas e para a conservação de recursos genéticos de plantas cultivadas. O presente sumário de conceitos e situações tem a sua origem em apontamentos preparados para o ensino de Genética a alunos de Engenharia Agrícola da Universidade de Évora. O propósito foi, e continua a ser, o de transmitir uma base teórica geral e facultar uma introdução à sua aplicação prática, organizando e dispondo os temas de maneira a colocar a ênfase nos fenómenos biológicos, procurando fazer com que a (inevitável) formulação matemática não obscureça a atenção a dar aos mesmos.

#### I.a. Fixação de alelos por *selfing*

Mendel [2], como corolário da 1ª lei da hereditariedade, deduziu que na continuação do

processo de *selfing* (auto-cruzamento, ou cruzamento entre indivíduos com genótipos idênticos) desde a  $F_1$  até à  $n$ -ésima geração ( $F_n$ ), a proporção de heterozigóticos para cada *locus* híbrido na  $F_1$  se reduzia, uniformemente, para metade em cada geração (esquema 1; *locus* com dois alelos,  $A$  e  $a$ ): os homozigóticos reproduziam apenas os mesmos homozigóticos, enquanto os heterozigóticos reproduziam-se em apenas metade de heterozigóticos, produzindo mais homozigóticos.

Tratando-se de um híbrido em  $k$  *loci* de segregação independente, este processo repete-se em  $k$  pares de cromossomas, pelo que o resultado de Mendel estende-se para a seguinte formulação: na  $F_n$ , apenas  $(1/2)^{n-1}$



esquema 1

\*

dos *loci* (ou, de uma maneira geral, dos grupos de ligação) que eram heterozigóticos na  $F_1$  ainda o são em cada indivíduo da  $F_n$ . Por exemplo para  $n = 8$ , este valor é de 0,78%, equivalendo a dizer que aproximadamente 99% dos *loci* heterozigóticos na  $F_1$  foram fixados (isto é, são homozigóticos), para um dos alelos, na  $F_8$ .

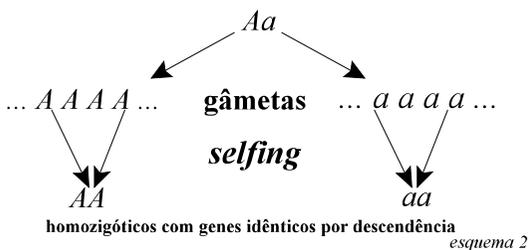
Em espécies autogâmicas (onde a polinização cruzada é muito reduzida), praticamente todos os *loci* estão fixados por este mecanismo, acontecendo até certas populações serem, para além disso, isogênicas (virtualmente constituídas por um único tipo de linha pura). Os *loci* heterozigóticos nos indivíduos destas populações são devidos praticamente só à ocorrência de mutação (como será evidenciado posteriormente). Portanto, a “linha pura” já existe na natureza; mas pode ser obtida experimentalmente por cruzamentos consanguíneos, de que o *selfing* é o exemplo extremo.

**II. Genealogias: identidade por descendência**

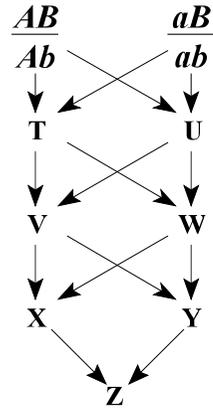
Quando a partir de um heterozigótico  $Aa$  se produz um descendente homozigótico, os dois genes deste último são idênticos por descendência: são duas réplicas de um só cromossoma presente no heterozigótico (esquema 2):

Neste esquema, a probabilidade de ter homozigóticos com genes idênticos por descendência através de *selfing* ( $AA$  ou  $aa$ ) é, pela 1ª lei de Mendel,  $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$  ao fim de uma geração; e como Mendel demonstrou (esquema 1), a continuação do *selfing* leva à acumulação de genótipos com genes idênticos por descendência, à custa da redução dos heterozigóticos.

Mas como será noutros modelos de cruzamento?



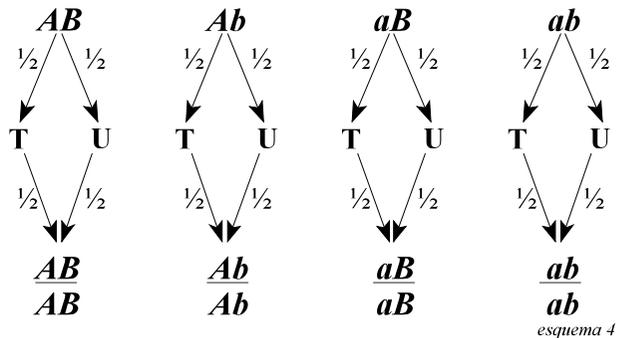
Suponhamos um *locus*  $A/a$ , e um *locus*  $B/b$  ligado no mesmo cromossoma (taxa de recombinação entre  $A$  e  $B$  virtualmente nula). O esquema 3 representa a sucessão de quatro gerações em que se faz sistematicamente o cruzamento entre irmãos ( $T \times U, V \times W, X \times Y$ ), e não há *selfing*, partindo de dois indivíduos heterozigóticos  $Bb$ .



esquema 3

Cada seta unindo 2 indivíduos indica a passagem de um gameta. Como é que poderia haver uma fixação do mesmo alelo em  $T$  e  $U$ ? Se ambos fossem  $BB$ , ou se ambos fossem  $bb$ . No entanto isto *não é* ainda identidade por descendência, porque nessa geração cada gene do *locus*  $B/b$  está ligado a um diferente gene do *locus*  $A/a$ , que serve neste exemplo para atestar a sua diferente origem.

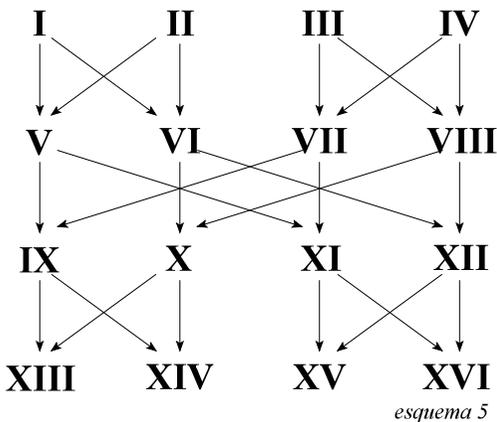
Já nos indivíduos  $V$  e  $W$  existe a possibilidade de exactamente um mesmo cromossoma (dos quatro presentes nos progenitores iniciais) aparecer duplamente, para o que se concebem quatro possibilidades (esquema 4):



A cada seta, segundo a 1ª lei de Mendel, corresponde uma probabilidade de  $\frac{1}{2}$  de o cromossoma em causa ser transmitido à geração seguinte. Daí se conclui que cada uma destas quatro alternativas de identidade por descendência tem uma probabilidade de  $(\frac{1}{2})^4 = 1/16$ ; mas como são independentes, a probabilidade de obter identidade por descendência (sem precisar *que* cromossoma aparece duplicado nesse indivíduo) é, *globalmente*,  $4 \times (\frac{1}{2})^4 = \frac{1}{4}$ .

Considere-se agora que eram utilizados 4 indivíduos em cada geração (cruzamentos entre primos-direitos, esquema 5): a possibilidade de identidade por descendência fica adiada para o terceiro cruzamento. Os progenitores directos do indivíduo XIII (ou de qualquer um dos da quarta geração) contêm em proporção probabilisticamente igual a herança dos quatro indivíduos iniciais I, II, III e IV. Mas o parentesco entre os indivíduos da terceira geração (IX - XII) já implica uma certa proporção de identidade por descendência na quarta, nesta sendo *possível* (por exemplo para o XIII) a duplicação de cromossomas de I via V e IX por um lado, e via VI e X pelo outro, como de IV via VII e IX e via VIII e X, etc.. Se os quatro indivíduos iniciais não tivessem parentesco entre si, então existiam à partida 8 cromossomas diferentes por grupo de ligação e a *probabilidade* de obter-se identidade por descendência na 4ª geração ficava  $8 \times (\frac{1}{2})^6 = 1/8$ .

Estes e outros modelos sistemáticos de cruzamento

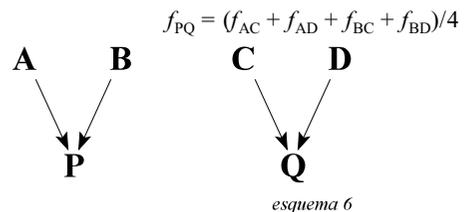


podem ser utilizados na prática em diversos contextos do melhoramento de plantas, especialmente para a obtenção mais ou menos acelerada de linhas puras: o aumento da percentagem de homocigóticos com genes idênticos por descendência será tanto mais rápido quanto menor o número de indivíduos envolvidos em cada geração (4, 2 ou 1), isto é, quanto menor a diversidade genética presente de início. E o usarem-se 2 ou mais indivíduos depende não só de ser prático ou não fazer-se o *selfing* sistematicamente: pode interessar não acelerar demasiado a fixação de alelos, pelas implicações que esta tem nalguns dos mecanismos postulados para a depressão de consanguinidade (que serão abordados mais adiante neste trabalho).

### II.a. Cálculo do coeficiente de parentesco

Em animais de criação, ou no estudo de árvores genealógicas na espécie humana, é mais provável que os esquemas de cruzamento sejam muito irregulares, pelo que se requer um método geral de cálculo da identidade por descendência. Como se viu nos esquemas 3 e 4, o parentesco entre os dois progenitores de um indivíduo é a condição para haver identidade por descendência entre os gâmetas que lhe deram origem; o cálculo do coeficiente de parentesco entre os progenitores [3] é uma das vias (mas não a única) para a determinação da identidade por descendência que recai sobre cada indivíduo. Definem-se as seguintes regras e corolários para esse cálculo:

*Regra principal:* o coeficiente de parentesco, entre dois indivíduos P e Q ( $f_{PQ}$ ), é a média dos coeficientes de parentesco entre cada um dos progenitores de P e cada um dos progenitores de Q. Assim, na genealogia (esquema 6)



*Regra auxiliar:* o coeficiente de parentesco entre dois indivíduos é igual à média dos coeficientes de parentesco entre um desses indivíduos e cada um dos progenitores do outro. Por exemplo,  $f_{PQ} = (f_{QA} + f_{QB})/2$ , e ainda  $f_{QA} = (f_{AC} + f_{AD})/2$  e  $f_{QB} = (f_{BC} + f_{BD})/2$ ; destas igualdades resulta  $f_{PQ} = (f_{AC} + f_{AD} + f_{BC} + f_{BD})/4$  (q.e.d....).

*Corolários:* os símbolos A, B, C e D no esquema 6 não têm de representar indivíduos diferentes. Por exemplo se P e Q são meios-irmãos, pondo por exemplo D = A, a fórmula geral fica  $f_{PQ} = (f_{AC} + f_{AA} + f_{BC} + f_{AB})/4$ ; já se P e Q são irmãos por *selfing*, D = C = B = A, dando  $f_{PQ} = 4f_{AA}/4 = f_{AA}$ .

Note-se que o coeficiente de parentesco do indivíduo consigo mesmo ( $f_{AA}$ ), como vimos acima na definição de identidade por descendência (esquema 2), é em princípio  $1/2$ . Mais em rigor, porém, define-se  $f_{AA} = 1/2(1 + F_A)$ . O  $F_A$  é o coeficiente de consanguinidade (genealógico) do indivíduo A, e representa a probabilidade de já se encontrar identidade por

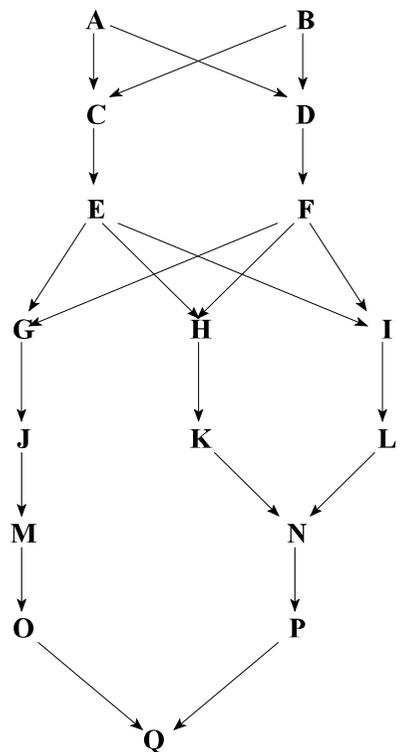
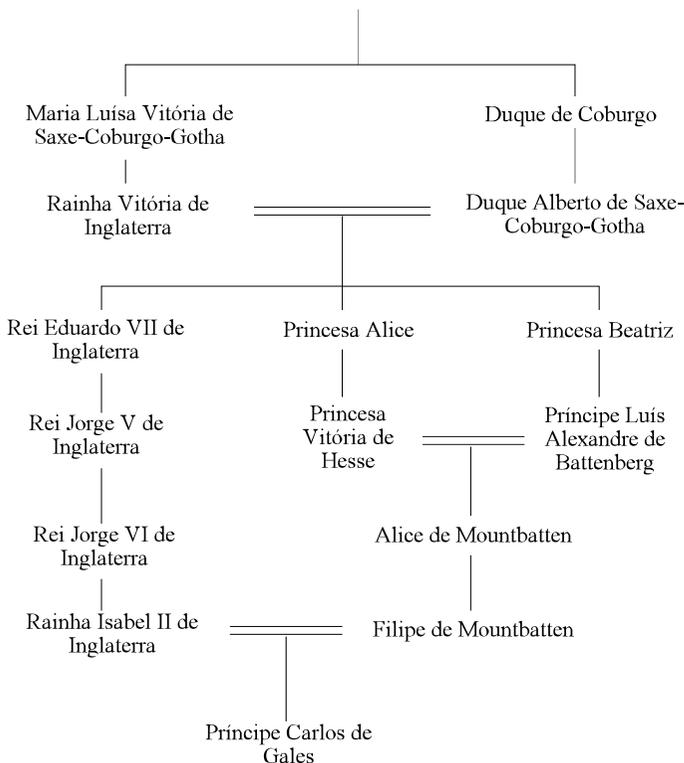
descendência nesse indivíduo (por isso  $f_{AA}$  varia entre  $1/2$  e 1). Por outras palavras,  $F_A$  é numericamente igual ao coeficiente de parentesco entre os progenitores de A. Aplicando este princípio ao parentesco entre progenitor e descendência, ou seja pondo B = Q, pode formular-se  $f_{PQ} = (f_{QA} + f_{QQ})/2$  e se  $f_{QQ} = (f_{CC} + 2f_{CD} + f_{DD})/4 = 1/2(1 + f_{CD})$ , assumindo  $F_C = F_D = 0$ , resulta  $f_{PQ} = (f_{AC} + f_{AD} + 1/2(1 + f_{CD}))/4$ .

Tome-se como exemplo a famosa árvore genealógica do esquema 7 [4]:

Considerando que  $f_{AB}, F_A, F_B, F_E$  e  $F_F$  são nulos, e que os progenitores não representados não têm parentesco com os seus parceiros, obtém-se

$$F_Q = f_{OP} = f_{MN}/4, f_{MN} = (f_{JK} + f_{JL})/4; f_{JK} = f_{GH}/4, f_{JL} = f_{GI}/4, f_{GH} = f_{GI} = (f_{EE} + 2f_{EF} + f_{FF})/4;$$

$$f_{EF} = f_{CD}/4, f_{CD} = (f_{AA} + 2f_{AB} + f_{BB})/4 = 1/4; \text{ donde } F_Q = f_{OP} = 9/1024 = 0,88\%.$$



esquema 7

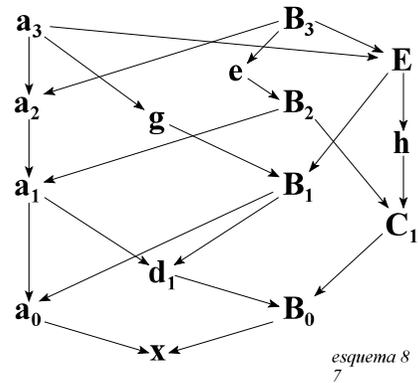
Examinando o *pedigree* do esquema 8 [5], que é muito mais complexo:

Pela regra principal enunciada, o coeficiente de parentesco entre  $a_0$  e  $B_0$  é dado pela resolução da fórmula  $f_{a_0B_0} = (f_{a_1d_1} + f_{a_1C_1} + f_{B_1d_1} + f_{B_1C_1})/4$ .

Através do cálculo dos coeficientes de parentesco necessários, dispostos no quadro 1, e assumindo que o parentesco com progenitores não representados (os segundos progenitores dos indivíduos e,  $B_2$ , g e h) é 0, fica  $f_{a_1d_1} = 0,3203$ ,  $f_{a_1C_1} = 0,1797$ ,  $f_{B_1d_1} = 0,3359$ , e  $f_{B_1C_1} = 0,0938$ , donde  $f_{a_0B_0} = 0,2324$ .

As regras enunciadas também se aplicam a genealogias regulares como a do esquema 3. Assim,  $f_{TU} = 1/4$ , ou seja a identidade por descendência já calculada para qualquer um dos seus descendentes (V e W). Quanto ao valor da identidade por descendência de Z:  $f_{VW} = (f_{TT} + 2f_{TU} + f_{UU})/4$ , e dado que  $f_{TT} = f_{UU} = 1/2$ , obtém-se  $f_{VW} = 3/8$  (o valor do coeficiente de consanguinidade de X e também o de Y); continuando,  $f_{XY} = (f_{VV} + 2f_{VW} + f_{WW})/4 = (1/2(1 + F_V) + 6/8 + 1/2(1 + F_W))/4 = (0,625 + 0,75 + 0,625)/4 = 1/2 = F_Z$ . De igual modo, para a genealogia dos primos-direitos XIII até XVI (esquema 5), verifica-se que na segunda geração são irmãos dois a dois (V com VI e VII com VIII), dando um  $f_{IX,X} = f_{XI,XII} = 1/8 = F_{XIII} = F_{XIV} = F_{XV} = F_{XVI}$ .

Comparando esses dois exemplos com o modelo de *selfing*, constata-se que ao fim de 3 gerações se



atingem valores de identidade por descendência de 0,875 no *selfing* ( $2 \times 7/16$ , esquema 1), 0,375 entre irmãos ( $f_{VW}$ , esquema 3), e 0,125 entre primos-direitos ( $F_{XIII}$ ). Trata-se de linhagens isoladas reprodutivamente, isto é, sem intervenção de indivíduos de outras proveniências, e que em cada geração têm um número fixo de indivíduos utilizados na reprodução, respectivamente 1, 2 e 4. O material de partida, sejam 2, 4 ou 8 cromossomas por grupo de ligação, é pois o único que pode continuar a considerar-se nas sucessivas gerações. Daqui se confirma que quanto menor é o número de indivíduos utilizados por geração, mais rapidamente aumenta a proporção de *loci* com identidade por descendência.

Porquê fazerem-se acasalamentos que aumentam a identidade por descendência? Recordando o esquema 2, nota-se que a identidade por descendência se dá com

Quadro 1 — Coeficientes de parentesco utilizados no cálculo do  $F_x$  referente ao esquema 8.

<i>f</i>	$a_1$	$a_2$	$a_3$	$B_1$	$B_2$	$B_3$	e	E	g	h
$a_1$	17/32	9/32		7/64						
$a_2$				3/16	1/16			1/4	1/8	1/8
$a_3$			1/2		0	0	0	1/4	1/4	
$B_1$				9/16						
$B_2$					1/2			1/16	0	1/32
$B_3$						1/2	1/4	1/4		
e								1/8		
E								1/2	1/8	1/4
g										1/16

Quadro 3 — Diferentes fórmulas para as frequências genótípicas do quadro 2

Genótipo	Frequências genótípicas		
	Referidas ao H-W	Média ponderada	Desvio da fixação
<i>AA</i>	$p^2 + p(1-p)F$	$p^2P + pF$	$p-p(1-p)P$
<i>Aa</i>	$2p(1-p)-2p(1-p)F$	$2p(1-p)P$	$2p(1-p)P$

igual probabilidade para qualquer dos alelos presentes no heterozigótico, e por isso da mesma maneira que se podem fixar genótipos desvantajosos, como é o caso do fenómeno de depressão de consanguinidade, o mesmo poderá dar-se com genótipos vantajosos para o melhoramento. Os acasalamentos consanguíneos dos *pedigrees* animais são acompanhados de uma cuidadosa selecção dos parceiros de modo a tentar fixar um máximo de genes favoráveis; aqui, portanto, a consanguinidade é uma característica desejável.

### III. Populações: coeficiente de fixação

Em rigor, o coeficiente de parentesco de cada indivíduo, medido através da sua genealogia, apenas representa a acumulação de identidade por descendência em relação a uma população inicial de referência, cujos valores de coeficiente de parentesco se convencionam serem 0. Mas de facto qual é a consanguinidade dos indivíduos dessa população de referência?

O modelo definido para o equilíbrio de Hardy-Weinberg estabelece que numa população panmíctica, isolada e infinitamente grande, a distribuição dos genótipos para cada *locus* depende apenas das frequências dos genes respectivos. Assim, para um alelo *A* de frequência *p*, a frequência dos homozigóticos *AA* é  $p^2$  e a dos heterozigóticos com *A* é  $2p(1-p)$ . Embora se trate de condições só concebíveis em abstracto, elas são informativas sobre a consanguinidade nesse modelo: com um número de indivíduos infinito, há um número infinito de cromossomas diferentes por grupo de ligação, pelo que em panmixia a probabilidade de resultar identidade por

descendência é nula em todos os indivíduos. Por outras palavras, *as distribuições genótípicas definidas pelo equilíbrio de Hardy-Weinberg correspondem exactamente a um  $F = 0$  para toda a população*. Mais ainda, se se puder partir de uma população razoavelmente grande, independentemente do parentesco entre os indivíduos à partida, ao fim de uma geração após cruzamentos ao acaso obtém-se uma população com  $F = 0$  em todos os *loci*, portanto uma população de referência com parentesco verdadeiramente nulo entre os seus membros [6:178].

No caso geral [6:174], as distribuições genótípicas que se verificam para cada *locus* numa população são (quadro 2):

Quadro 2 — Distribuições genótípicas para um gene *A* em relação ao modelo de Hardy-Weinberg

Genótipo <sup>1</sup>	Frequência	Desvio do H-W <sup>2</sup>
<i>AA</i>	$p^2 + p(1-p)F$	$p(1-p)F$
<i>Aa</i>	$2p(1-p)-2p(1-p)F$	$-2p(1-p)F$

<sup>1</sup> O símbolo *a* representa todos os alelos diferentes de *A* presentes na população

<sup>2</sup> H-W: modelo de Hardy-Weinberg

Ou seja, pelo facto de uma população ter um número finito de indivíduos, mesmo que os cruzamentos sejam ao acaso, a proporção de homozigóticos para um determinado gene é superior à esperada, em detrimento da de heterozigóticos com esse gene. Dito doutra maneira, *F mede a proporção de heterozigóticos esperados que deixa de haver por consanguinidade*. Esta definição do *F* como coeficiente de fixação de cada *locus* adianta a noção que o número finito de indivíduos acarreta também deriva genética. Por isso a distribuição dos genótipos pode também ser

escrita (quadro 3) como média ponderada entre uma componente “panmítica”, afectada do chamado índice panmítico  $P = 1 - F$ , e uma componente de fixação representada por este  $F$ , ou ainda como um desvio (dependente de  $P$ ) da fixação total do *locus* [6:174]:

Dentro deste contexto, o coeficiente de fixação calculado numa população é por sua vez o valor médio dos coeficientes de parentesco dessa população na geração precedente. No entanto, isso só é válido quando a única excepção ao modelo de Hardy-Weinberg é o número finito de indivíduos; nas populações reais é de presumir a acção de outros factores, entre os quais a selecção, a mutação e a migração. Então o valor do  $F$  que se calcula em referência ao esperado pelo modelo de Hardy-Weinberg terá uma interpretação muito mais complexa, como será exemplificado mais adiante — mas em compensação, e complementado por outras informações, pode fornecer indicações interessantes sobre outros aspectos da biologia das populações onde foram obtidos [7]. Para já, no entanto, importa analisar a situação mais simples, isto é, sem estas interferências adicionais, e portanto continuar a considerar  $F$  e  $P$  como estimadores do desvio, em relação à população ideal, resultante do número finito de indivíduos presentes nas populações.

**III.a. Relação entre o efectivo populacional e a consanguinidade**

Considerando a situação da maior parte das plantas cultivadas, isto é, cada indivíduo produzir elevado número de gâmetas dos dois sexos e com a possibilidade de auto-fertilização (na proporção  $h$  de todos os zigotos formados), e considerando ainda a separação de gerações como é o caso das plantas anuais, o coeficiente de fixação na  $n$ -ésima geração é dado por

$$F_n = F_{n-1} \left(1 + \frac{h}{2} - \frac{1}{N}\right) + F_{n-2} \left(\frac{1}{2N} - \frac{h}{2}\right) + \frac{1}{2N}$$

e o respectivo índice panmítico por

$$P_n = P_{n-1} \left(1 + \frac{h}{2} - \frac{1}{N}\right) + P_{n-2} \left(\frac{1}{2N} - \frac{h}{2}\right)$$

com  $N$  a designar o efectivo populacional participante na reprodução sexuada nas gerações  $n-2$  e  $n-1$  (assumido como constante por simplicidade) [6:194].

Estas duas fórmulas simplificam-se muito quando  $h = 1/N$ , isto é, quando a probabilidade de auto-fertilização depender apenas da frequência de gâmetas do mesmo indivíduo. Embora este valor de  $h$  seja uma situação-padrão em muitos dos desenvolvimentos teóricos nesta área, é considerado altamente improvável nas populações reais [6:194], pois ou há preferência pelos próprios gâmetas como nas autogâmicas ou existem mecanismos de interferência com a auto-fertilização (por exemplo auto-incompatibilidade), dando respectivamente valores de  $h$  mais altos ou mais baixos que  $1/N$ .

Embora dependendo ligeiramente dos valores de  $h$ , atinge-se assintoticamente, para a fórmula geral, uma relação

$$P_n = P_{n-1} \times \left[1 - \frac{1}{N} + \sqrt{1 + 1/N^2}\right] / 2$$

que é aproximada por  $P_n = P_{n-1} [2N / (2N + 1)]$ . Da adaptação desta última, para  $2N = (1 - F_{n-1}) / (F_n - F_{n-1})$ , é fácil calcular estimativas do efectivo populacional de que descendeu uma determinada geração: por exemplo, se na geração  $n-1$  se observassem 250 indivíduos  $AA$ , 100  $Aa$  e 50  $aa$ ,  $p(A) = 0,75$ ,  $q(a) = 0,25$ , e  $F_{n-1} = 0,3(3)$ ; supondo que na geração  $n$  se observavam 255  $AA$ , 90  $Aa$  e 55  $aa$ , o resultado seria um  $F_n = 0,4$ , donde se tira  $2N = 0,6/0,06(6)$  e  $N = 4,5$ .

III.a.i. Varição do  $N$

Haverá populações com um  $N$  constante durante

longos períodos de tempo? Se bem que as condições ambientais possam variar marcadamente entre ciclos de reprodução sucessivos, com incidências, por exemplo, no número de indivíduos que atingem a maturidade se desenvolvem em cada geração (em espécies anuais), ou na intensidade de floração feminina em cada ano (em espécies perenes), pode assumir-se que essas variações se compensam entre si, resultando presumivelmente um  $N$  médio representativo para a população num período de tempo alargado.

Casos há, por exemplo na colonização por uma espécie pioneira de um terreno desocupado, em que o valor de  $N$  pode aumentar rapidamente de geração em geração [6:209]. No modelo de expansão geométrica de uma população, em que na geração  $t$  se tem  $N = N_0(1+r)^t$ , sendo  $N_0$  o efectivo da geração 0 (em princípio representativo do grau de variabilidade genética com que foi iniciada) e  $r$  a taxa de acréscimo populacional por geração, obtém-se um índice panmíctico

$$P_t = P_0 e^{-\frac{1-e^{-rt}}{2N_0r}}$$

no de expansão aritmética,  $N = N_0(1+rt)$ , obtém-se

$$P_t = P_0(1+rt)^{-\frac{1}{2N_0r}}$$

A dinâmica de  $N$  vai determinar a rapidez com que se atingem efectivos suficientemente elevados para que o declínio de  $P$  entre gerações sucessivas fique quase negligenciável. Por exemplo, com  $N_0 = 5$  mas  $r = 1$  (crescimento para o dobro em cada geração),  $P_t$  estabiliza em poucas gerações em valores acima dos 90% de  $P_0$ ; por outro lado, mesmo com  $N_0 = 20$ , uma  $r = 0,1$  não chega para evitar um declínio de  $P_t$  que continua até para lá das 20 gerações, nesse caso para baixo dos 80% do  $P_0$  inicial. Em qualquer caso, a um baixo efectivo populacional nas primeiras gerações corresponde sempre uma quebra de  $P$ . Essa quebra

(portanto, o aumento de  $F$ ) mede a proporção de *loci* fixados por deriva, e neste contexto é importante realçar que, mesmo que ao fim de algumas gerações a população esteja a níveis muito próximos do modelo de Hardy-Weinberg e portanto os valores observados indiquem um  $F$  aproximadamente 0, no seu historial houve uma “crise” de consanguinidade que se traduziu na perda de alelos, a qual significa um empobrecimento genético da população. E para medir este último recorre-se a parâmetros como o  $D_{ST}$ , que é independente do  $F$  (ver adiante).

### III.b. Produtividade variável e tamanho efectivo

O parâmetro  $N$  utilizado nas fórmulas anteriores não é aquele que é mais exacto na estimativa da variabilidade genética que participa num dado ciclo reprodutivo, na medida em que se está a assumir uma igual participação de todos os  $N$  indivíduos no património genético da geração seguinte. Por outras palavras, quando a produtividade dos progenitores é variável, passa a considerar-se uma distribuição de número de gâmetas por progenitor, com valor médio  $k$  e variância  $\sigma_k^2$ . Essa variação de produtividade pode até ser meramente por erros de amostragem (distribuição Poisson com  $\sigma_k^2 = k$ ). Na situação-padrão  $h = 1/N$ , fica-se [6:216] com um tamanho efectivo

$$N_e = \frac{N(k-1)}{k-1 + \frac{\sigma_k^2}{k}}$$

isto é, uma representação da variabilidade genética que passa para a geração seguinte e não do número de indivíduos que a transmitiram. À excepção da distribuição de Poisson, espera-se um desvio mais ou menos marcado entre o  $N_e$  e  $N$ . Assim, com  $k = 2$ , que é a situação de uma população com um efectivo constante de geração em geração, se  $\sigma_k^2 = 2k = 4$  o  $N_e$  reduz-se para 66% de  $N$  ( $N = 100$ ; com  $N = 10$  a redução é para 63% — de facto, a diferença relativa

entre  $N_e$  e  $N$  depende pouco do valor de  $N$ ). Já considerando (sempre com  $N = 100$ ) uma ligeira redução da produtividade média, por exemplo  $k = 1,5$  mas com  $\sigma_k^2 = 3$ , o  $N_e$  situa-se a 60% do respectivo valor de  $N$ . Em contraste, com  $k = 2,5$  e uma  $\sigma_k^2 = 1,5$ , o  $N_e$  vai para 119% do  $N$ . Uma maior produtividade média, por isso, traduz-se até num aumento do  $N_e$ . Assim, com um  $N$  constante de geração em geração, uma variação de produtividade entre os progenitores de cada geração traduz-se em tendências para endocruzamento diferentes das que os valores de  $N$  dariam a entender.

A substituição de  $N$  por  $N_e$  tem grande significado, por estabelecer a distinção entre o número de indivíduos presentes e a variabilidade que de facto transmitem, pelo que é considerado o parâmetro de real interesse. Sendo uma aproximação da variabilidade genética que é herdada por uma determinada geração, o  $N_e$  é a melhor medida do que se pode esperar em termos de fixação de alelos e de todas as consequências que daí podem advir [10]. Existem muitas fórmulas para o  $N_e$ , algumas muito complexas, segundo os modelos populacionais para que foram deduzidas.

#### IV. Populações: correlação entre gâmetas

O conceito de correlação entre gâmetas numa população foi originalmente introduzido em função do desvio da frequência de heterozigóticos ao valor previsto pela distribuição de Hardy-Weinberg, portanto, identificou-se essa correlação com o coeficiente de fixação [6:174]. Tomando o exemplo de dois alelos,  $A$  e  $a$ , com frequências  $p$  e  $q$  respectivamente, obtém-se

$$r = 1 - \frac{\text{freq}(Aa)}{2pq} = F$$

(dado que a fracção é entre heterozigóticos observados e esperados, é comum usar-se a

nomenclatura  $F = 1 - H_o/H_e$ ). A correlação  $r$  indica a *semelhança estatística* para cada *locus* entre os gâmetas participantes em cada ciclo reprodutivo, e com esta interpretação para o valor de  $F$  abre-se uma nova perspectiva, pois *as correlações podem ser negativas* e de facto há situações em que a frequência de heterozigóticos é superior à da distribuição do modelo de Hardy-Weinberg.

Sendo o  $F$  uma correlação, que significa o valor  $F = 0$  nas populações reais, visto que têm  $N_e$  finitos? Dado que os mecanismos combinatorios tendem *sempre* para um aumento do  $F$ , o facto do valor de  $F$  ser 0 implica por isso a existência de mecanismos compensatórios desse aumento, e então *o valor de referência  $F = 0$  nas populações reais, finitas, não é de facto o do modelo de Hardy-Weinberg*, mas sim um equilíbrio entre factores que aumentam e diminuem o valor de  $F$  em cada geração. Por outras palavras, um  $F = 0$  é uma correlação entre gâmetas numericamente igual à do modelo de Hardy-Weinberg, num contínuo entre correlações positivas e negativas.

A complexidade das situações reais é bem ilustrada no seguinte exemplo, tirado dum estudo de aloenzimas feito com duas populações ( $A$  e  $B$ ) de azevém (*Lolium perene*), na Califórnia (quadro 4):

As plântulas da geração  $n$  foram criadas em viveiro (*ex situ*), constituindo amostragens na ordem dos milhares por população, e os seus genótipos permitiram deduzir os genótipos maternos, ou seja das plantas adultas da geração  $n-1$  que cresceram e foram polinizadas no campo (*in situ*), na ordem de centenas por população. Note-se o contraste entre os valores negativos de  $F$  tanto para o *locus PER* nas duas populações como para o *locus ACP* na população  $A$ , nas plantas adultas, e os respectivos valores (positivos) para as plântulas. No entanto, para o *locus PGI*, nas duas populações, os valores de  $F$  permanecem relativamente baixos e quase inalterados de uma geração para a outra. Portanto, segundo os *loci*, diferentes factores influenciam as distribuições genotípicas — isto sem que as diferenças da frequência

Quadro 4 — Resumo dos resultados do estudo publicado por Mitton [8]. Geração n-1: plantas adultas no campo (*in situ*); geração n: plântulas criadas em viveiro (*ex situ*).

População	Geração	<i>Locus</i>		<i>PGI</i>		<i>ACP</i>		<i>PER</i>	
		n-1	n	n-1	n	n-1	n	n-1	n
A	Freq. alelo 1	0,3426	0,3485	0,9350	0,9465	0,2037	0,2047		
	Freq. heterozigót.	0,4396	0,4401	0,1300	0,874	0,3333	0,2968		
	<b>F</b>	<b>0,240</b>	<b>0,309</b>	<b>-0,695</b>	<b>0,1367</b>	<b>-0,275</b>	<b>0,885</b>		
B	Freq. alelo 1	0,2486	0,2725	0,9514	0,9547	0,3541	0,3556		
	Freq. heterozigót.	0,3568	0,3866	0,865	0,739	0,4703	0,3767		
	<b>F</b>	<b>0,452</b>	<b>0,251</b>	<b>0,657</b>	<b>0,1468</b>	<b>-0,281</b>	<b>0,1781</b>		

de cada gene entre as duas populações, como é o caso nos *loci* *PGI* e *PER*, tenham, aparentemente, qualquer influência nos coeficientes de fixação obtidos.

A geração n criada em viveiro foi avaliada bastante cedo após a germinação e considera-se relativamente liberta de factores ecológicos que poderiam eliminar preferencialmente certos genótipos; poderá assim admitir-se que as distribuições genotípicas na geração n deram valores de F semelhantes aos que antes se verificaram na fase de semente da geração n-1. Sendo assim, os valores de F negativos sugerem que durante o desenvolvimento da geração n-1, até à maturidade, terá havido selecção (pós-zigótica) favorecendo os heterozigóticos em “certos” *loci*, possivelmente acumulada de vários anos em virtude de ser uma espécie perene. Tais *loci* estariam em desequilíbrio de ligação com *loci* como *PER* ou *ACP* — quando não fossem estes últimos os próprios alvos directos da selecção (o paralelismo entre as populações A e B favorece a segunda perspectiva). Utilizando a fórmula aproximada de  $2N$  em função dos valores de  $F$ , resultam estimativas de  $N_e$  da ordem de 2 a 5 para os *loci* *ACP* e *PER*, que seria talvez indicativa de uma elevada taxa de auto-polinização nesta espécie [8].

Porém, para que esta explicação fosse generalizável, o valor de F no *locus* *PGI* teria também

de ter aumentado em função deste  $N_e$ , e como tal não aconteceu (nem neste nem noutro *locus*, *GOT*, não incluído nesta tabela [8]), pode parecer que há uma contradição. Admitindo que o  $N_e$  é na realidade um valor baixo, poderia postular-se, por exemplo, um mecanismo compensatório que seleccionasse, não após as sementes germinarem mas antes (por exemplo, selecção pré-zigótica) a favor dos heterozigóticos em *loci* como o *PGI*; então, apesar do  $N_e$  ser baixo obter-se-ia uma distribuição genotípica nas plântulas mais próxima da do modelo de Hardy-Weinberg como é o caso. Em conclusão, se aquilo que se passa nas populações reais parecer diferente em função dos *loci*, isso patenteia a acção de diferentes mecanismos condicionantes das distribuições genotípicas, e pode acrescentar-se que *a priori* será essa a regra e não a excepção [9].

#### IV.a. *Inbreeding* e *outbreeding*

A mutação contribui, se bem que subtilmente, para manter uma certa variabilidade em cada *locus*. Segundo a Teoria Neutral, F tem um valor de equilíbrio  $F = 1/(1+u)$  — em que  $u = 4N_e u$  (é o chamado parâmetro neutral;  $u$  é a taxa de mutação no *locus*, representativa para qualquer dos alelos). Este valor indica que, especialmente em populações de  $N_e$

grande, é improvável a fixação completa de um alelo, ou seja, o limite teórico de  $F$  é ligeiramente abaixo de 1. Por exemplo para uma população com  $N_e = 1000$ , o valor de equilíbrio de  $F$  é aproximadamente 0,95 caso a taxa de mutação ronde os  $10^{-5}$ .

Dado que se pode presumir que a mutação é um mecanismo relativamente uniforme de interferência com a fixação, é possível fazer algumas generalizações. Em cada *locus* homocigótico com identidade por descendência, logo que ocorra uma mutação num desses alelos, esse indivíduo não só fica heterocigótico numa das suas linhagens somáticas como pode transmitir dois gametas diferentes para esse *locus* se essa mutação ocorrer em células precursoras dos gametas (recorde-se que nas plantas superiores não há uma separação precoce entre soma e germen). Considerando uma população muito grande, com  $N_e = 10^4$  e  $h = 1/N$ , há um acréscimo de  $5 \times 10^{-5}$  no valor de  $F$  por geração, na ordem de grandeza de muitas taxas de mutação, ou seja: a mutação cancela o aumento do  $F$  de tal modo que a correlação entre gametas deixa de tender a aumentar nessa população, daí resultar em distribuições genotípicas aparentemente de acordo com o modelo de Hardy-Weinberg.

Este valor de  $N_e = 10^4$  é assim utilizável como referência para as populações naturais, delimitando duas situações opostas [10]:

- a) *inbreeding* (tendência para uma correlação entre gametas positiva) correspondendo a valores de  $N_e < 10^4$ , sendo que abaixo de 100 já se falaria de *inbreeding* extremo;
- b) *outbreeding* (tendência para uma correlação entre gametas negativa), situação com  $N_e > 10^4$  que é exemplificada não tanto por populações excepcionalmente grandes mas antes pelo cruzamento *entre* indivíduos de populações diferentes, divergentes geneticamente, por exemplo entre subespécies ou mesmo entre espécies.

#### IV.b. Subdivisão da população

A correlação negativa entre os gametas, em

*outbreeding*, simula um “excedente” de mutação, e de facto a possibilidade de mutações independentes entre duas populações de origem comum, ocorrendo enquanto isoladas entre si, estaria na base da divergência genética entre elas. Talvez mais marcadamente, a possibilidade de terem estado sujeitas a pressões selectivas diferentes, ou aos efeitos da deriva genética, explicaria essa correlação negativa. Este fenómeno de divergência tem grande importância no melhoramento genético assim como na conservação de recursos genéticos.

Voltando à situação mais próxima do modelo de Hardy-Weinberg: se uma população tiver uma distribuição geográfica tão alargada que os cruzamentos se dêem sobretudo localmente, com a conseqüente redução do  $N_e$  e acção da deriva genética, então define-se essa população como um conjunto de demes, isto é, subdivisões idealmente panmícticas [6:291]. Diferentes demes podem ter diferentes frequências para o mesmo gene, ou seja, se considerarmos um gene  $A$  de frequência  $p_T$  em dada geração (o índice  $T$  refere-se ao total da população), a frequência global de heterocigóticos de  $A$  nessa geração é  $2p_T(1-p_T) - 2\sigma_{p(D)}^2$ , em que o segundo termo representa a variância das frequências  $p'$  (da geração anterior) nos diversos demes ( $D$ ). Por seu lado,  $\sigma_{p(D)}^2$  relaciona-se com a variância das frequências  $p$  (da própria geração e não da anterior) através da proporção de gametas de origem local,  $1-m$ , tal que  $\sigma_{p(D)}^2 = (1-m)^2\sigma_{p(D)}^2$  [6:293].

Pela definição de desvio do modelo de Hardy-Weinberg, obtém-se a equivalência  $F = \sigma_{p(D)}^2/[p_T(1-p_T)]$ . Por outras palavras, *maior  $\sigma_{p(D)}^2$  significa maior efeito da deriva acumulado até à geração anterior e com isso um maior valor de  $F$  na geração que se lhe segue*, para  $p_T$  constante [6:293].

Mas este modelo encerra uma contradição: se há deriva em cada subdivisão, então esta não se pode considerar em rigor um deme. Em lugar disso, a correlação entre os gametas de uma população é hierarquizada por subdivisões, sem implicar que são

demes [6:294]:

$F_{IS}$  — média das correlações entre os gâmetas produzidos por cada indivíduo (I) de uma subdivisão (S) e o conjunto dos gâmetas da respectiva subdivisão;

$F_{IT}$  — correlação entre os gâmetas produzidos por cada indivíduo e os gâmetas da população em geral (T);

$F_{ST}$  — correlação entre os gâmetas dentro das subdivisões, relativamente aos gâmetas da população em geral. Ou seja, assumindo que  $F_{IS}$  é independente das frequências dos genes entre subdivisões,  $F_{ST} = (F_{IT} - F_{IS}) / (1 - F_{IS})$ , ou usando índices panmícticos,  $P_{IT} = P_{IS}P_{ST}$ .

Esta hierarquização funciona da mesma maneira intercalando mais níveis entre I e T, por exemplo “raças” dentro das subdivisões [6:295], tais que  $P_{IS} = P_{IR}P_{RS}$ , etc..

Considerando que  $F_{ST} > 0$  [6:295], uma distribuição aparentemente de Hardy-Weinberg no conjunto da população, isto é com  $F_{IT} = 0$ , implica um  $F_{IS}$  negativo, ou seja o favorecimento dos heterozigóticos dentro das subdivisões. Uma das maneiras de se ter  $F_{IS}$  negativo é haver diferenciação genética dentro das subdivisões em “raças”, com  $F_{IR} > 0$  mas  $F_{RS} < 0$ , o que implica de alguma forma uma preferência, ao nível da subdivisão, pela conjugação entre gâmetas correlacionados negativamente, ou seja entre raças e não dentro de raças. Esta situação é bem exemplificada pelos mecanismos de auto-incompatibilidade em plantas, chamando “raça” a uma vizinhança de indivíduos aparentados (por exemplo, plantas descendentes da mesma mãe e localizadas perto umas das outras) e ao conjunto destas famílias, entre as quais as polinizações cruzadas são extensas, uma subdivisão [9].

#### IV.b.i. Identidade e diversidade genética

Nei [11] introduziu o coeficiente de diversidade genética  $G_{ST} = D_{ST}/H_T$ , definido como a proporção da diversidade genética total  $H_T = H_S + D_{ST}$  que diferencia as subdivisões de uma população.  $H_S$  define,

complementarmente, a proporção de  $H_T$  que constitui a diversidade genética média dentro das subdivisões (note-se que  $H_S/H_T = 1 - G_{ST}$ ). As fórmulas respectivas [12] baseiam-se no cálculo de identidades genéticas  $J = 1 - H$ , ao nível da população (índice T), das subdivisões (índice S), de “colónias” que compõem as subdivisões (índice C), etc. (entre outras propriedades, Nei demonstrou que  $H_T = H_C + D_{CT} = H_C + D_{CS} + D_{ST}$ , e  $H_C/H_T = (1 - G_{CS})(1 - G_{ST})$ , sendo  $G_{CS} = D_{CS}/H_S$ ).

Como as identidades genéticas se baseiam exclusivamente nas frequências dos genes e não dos genótipos,  $G_{ST}$  não é comparável com  $F_{ST}$ , embora o sentido (diferenciação entre subpopulações) seja análogo. Relembrando o estudo em *Lolium perene* (quadro 4), os coeficientes de fixação calculados como medidas de desvio em relação à distribuição de Hardy-Weinberg parecem ser a resultante não só da deriva genética (tendência para aumento da correlação entre gâmetas nas populações finitas) mas de factores selectivos actuando aos mais diversos níveis e podendo influenciar esse desvio em qualquer dos sentidos. Mas, em rigor, os coeficientes de fixação só pretendem medir a componente de deriva, isto é, com  $F_{IS}$  positivo. O coeficiente  $G_{ST}$  parte do princípio que uma população começou por ser homogénea e passou a ter uma diferenciação entre subdivisões em virtude de factores locais que podem não só ser resultantes das distâncias entre os indivíduos (que condicionam os valores de  $N_e$ ) mas também de factores selectivos, como características do solo, exposição a pragas, etc. que, variando entre subdivisões, podem determinar diferentes frequências dos genes. Por outras palavras,  $G_{ST}$  procura medir o mesmo que  $F_{ST}$  mas é mais genérico, daí ser considerado o melhor parâmetro de diversidade entre subdivisões de uma população.

Considerem-se 3 populações, A B e C, com 4 subdivisões cada uma (quadro 5).

Quadro 5 — Exemplo (fictício) de distribuições de frequências alélicas num *locus* em três populações (A, B e C) e nas suas subdivisões (numeradas de 1 a 4 para cada população)

Alelos	Subpopulações				Subpopulações				Subpopulações			
	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
1	0,4	0,35	0,4	0,45	0,7	0,4	0,6	0,3	0,6	0,35	0,5	0,2
2	0,3	0,35	0,35	0,25	0,25	0,45	0,4	0,6	0,15	0,4	0,3	0,5
3	0,1	0,05	0	0,05	0,05	0,05	0	0,1	0,1	0,05	0	0,05
4	0,05	0	15	0,05	0	0	0	0	0,05	0	0,1	0,05
5	0	0,1	0	0,1	0	0	0	0	0	0,1	0	0,1
6	0	0	0,05	0,1	0	0	0	0	0	0	0,05	0,1
7	0,05	0,1	0	0	0	0	0	0	0,05	0,05	0	0
8	0,1	0,05	0,05	0	0	0	0	0	0,05	0,05	0,05	0

Para cada alelo apresentam-se as respectivas frequências dentro de cada subpopulação (cada coluna totaliza 100%). Os parâmetros calculados a partir destes “dados” encontram-se no quadro 6:

Quadro 6 — Parâmetros de identidade e diversidade genética calculados [12] a partir das frequências do quadro 5.

	População A	População B	População C
$J_S$	<b>0,286</b>	<b>0,475</b>	<b>0,342</b>
$D_{ST}$	<b>0,014</b>	<b>0,042</b>	<b>0,048</b>
$H_S$	0,714	0,525	0,658
$H_T$	0,728	0,567	0,705
$G_{ST}$	<b>0,020</b>	<b>0,074</b>	<b>0,067</b>

Comparando a população A, onde estão presentes os 8 alelos, com a população B, apenas com 3, o valor mais elevado de  $J_S$  (identidade genética dentro das subdivisões) nesta última reflecte a relativa monotonia de cada uma das suas subdivisões; em contrapartida, as frequências dos alelos predominantes, 1 e 2, bastante semelhantes entre subdivisões na população A, variam bastante mais na população B, o que se reflecte num maior valor de  $D_{ST}$  (diferenciação entre subdivisões) nesta última. Daqui resulta uma elevada diversidade dentro das subdivisões ( $H_S = 1 - J_S$ ) na população A, e também um valor mais pequeno de  $G_{ST}$ . Note-se que na população B, apesar da menor diversidade genética total ( $H_T$ ), o valor  $G_{ST}$  é maior que na A porque este parâmetro mede especificamente a diferenciação genética entre subdivisões.

A população C é uma espécie de compromisso entre A e B: tem todos os alelos de A, mas os alelos 1

e 2 têm, como em B, frequências bastante diferentes entre subdivisões. Isto dá o valor  $D_{ST}$  mais elevado das três, mas o de  $G_{ST}$  é ligeiramente mais baixo que na população B porque o  $H_S$  também é mais elevado. Portanto note-se que um valor de  $G_{ST}$  elevado pode, como aqui é exemplificado na população B em relação às outras duas, estar associado a uma relativa falta de diversidade genética.

#### IV.c. Métodos moleculares

Os fenótipos morfológicos são pouco satisfatórios para a análise genética de populações por duas razões: são poucos os que numa dada espécie têm uma genética simples e conhecida (a maior parte destes fenótipos são poligénicos e muito susceptíveis a modificadores ambientais), e desses ainda menos se podem analisar em diferentes espécies da mesma maneira. Em contrapartida, fenótipos moleculares como o comportamento electroforético de diferentes isoenzimas [13] e as próprias sequências do DNA têm uma correspondência genotípica simples (isto é, fácil atribuição de genótipos a cada tipo enzimático, perfil de restrição enzimática, tamanho de fragmento amplificado por PCR, etc.), e a metodologia que serve para uma espécie pode servir para outra, geralmente com leves adaptações (por exemplo, a revelação de uma actividade enzimática, um *primer* PCR para o rDNA, etc.).

Muitos dos enzimas que revelam polimorfismos em electroforese ou focagem isoeléctrica são não só abundantes e correspondem a passos reguladores de vias metabólicas (glutamato desidrogenase, álcool desidrogenase, isocitrato desidrogenase, etc.) ou estão associados a respostas a situações de stress (peroxidases, etc.). Estas correspondências parecem não ser acidentais e desde há muito se julga que estes polimorfismos resultam de um favorecimento dos heterozigóticos [8, 14, 15], isto é, estão-lhes associados mecanismos selectivos que implicam uma compensação da consanguinidade.

Muitos dos fenótipos DNA são considerados selectivamente neutros e são por isso indicadores mais específicos dos fenómenos relacionados com a deriva genética; os fenótipos identificáveis com o número de repetições em cadeia de sequências nucleotídicas simples (“micro-satélites”), onde as taxas de mutação ( $10^{-3}$  ou mesmo  $10^{-2}$  por geração) são muito mais elevadas do que é comum às mutações pontuais em sequências codificantes, são especialmente úteis. A sua análise, correntemente, vai aos poucos sendo alargada a um grande número de espécies, designadamente as cultivadas, e deverá permitir conhecer de uma maneira mais directa, se de facto na ausência de factores selectivos, os mecanismos genéticos que prevalecem nas populações e suas implicações com o sistema de reprodução sexuada de cada espécie.

## V. Depressão de consanguinidade

A principal importância prática da identidade por descendência reside na expressão fenotípica em que costuma traduzir-se, a chamada depressão de consanguinidade. Ela é definível, nos indivíduos consanguíneos, por reduzido *fitness*, isto é, por redução da adaptabilidade às condições naturais ou, no contexto do melhoramento, por redução do vigor (seja a nível somático, seja em fertilidade), ou ainda por aumento da susceptibilidade às flutuações ambientais. Opostamente, o “vigor híbrido” ou heterose define-se

nos mais heterozigóticos pelo aumento de vigor e/ou pela maior uniformidade fenotípica face às flutuações ambientais.

A existência de estratégias compensatórias da identidade por descendência, assentes em mecanismos em si mesmos independentes da capacidade de adaptação ao meio ambiente, como é o caso da maior parte dos mecanismos de auto-incompatibilidade em plantas (e o tabu do incesto na espécie humana), ou ainda a alopoliploidia em espécies autogâmicas, parece evidenciar bem a importância que o problema da depressão de consanguinidade tem para a adaptação das populações na natureza. E, por extensão, para o melhoramento. Depressão de consanguinidade e heterose são manifestações do mesmo fenómeno genético — a maior estabilidade do desenvolvimento dos caracteres fenotípicos com impacto no *fitness* — e reconhecem-se pelo menos três classes de interacção genética, que provavelmente coexistem nos diversos casos, para explicá-lo:

- i) hipótese da sobredominância: há *loci* que contribuem maximamente para o *fitness* apenas em genótipos heterozigóticos [3, 8].
- ii) hipótese da dominância: em diversos *loci* existem genes recessivos desfavoráveis que se traduzem em deficiências fisiológicas nos indivíduos que neles são homozigóticos, enquanto nos que são heterozigóticos essa redução não se manifesta devido à presença dos alelos favoráveis. Ao contrário da hipótese anterior, nesta é possível em teoria obter linhas puras de *fitness* máximo [3, 8].
- iii) interacções entre *loci*: certos alelos de diferentes *loci*, por epistasia ou complementaridade, ou por constituírem haplótipos com diferenças substanciais na expressão fenotípica, definem entre si conjuntos genotípicos que se dizem coadaptados; esta classe de interacções parece ser a mais plausível em relação a certas experiências, com *Drosophila* e outros organismos, onde se observou um retardamento da redução do *fitness* à medida que a identidade por descendência se ia

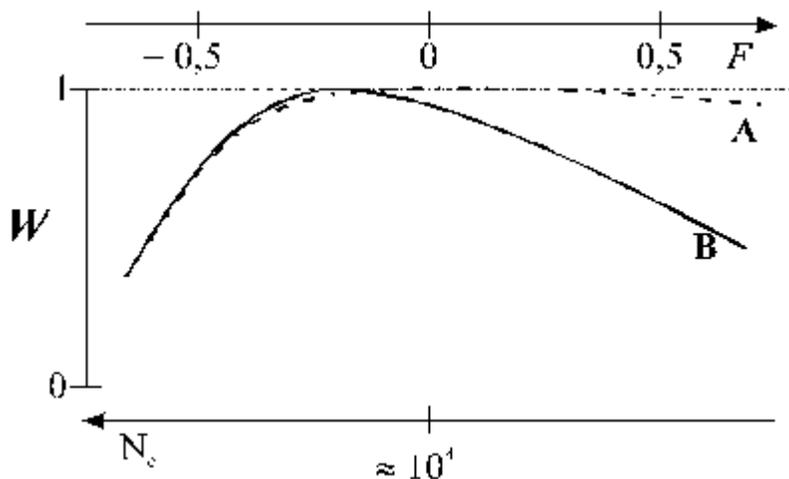


Figura 1 — Dependência do *fitness* ( $W$ ) em função da correlação entre os gâmetas em duas populações de espécies diferentes (A e B), e com indicação da correspondente variação no  $N_c$  na delimitação entre *inbreeding* e *outbreeding*.

acumulando [7].

A figura 1 mostra dois exemplos hipotéticos de curvas de *fitness* (símbolo  $W$ ) em função da correlação entre gâmetas [adaptado de 8 e 10]. Na curva A os valores de  $W$  não são particularmente penalizados pelo aumento da correlação, e a população só não resvalará para  $F = 1$  se tiver um  $N_c$  grande (por efeito da mutação). Já na curva B o *fitness* é máximo em valores de  $F$  ligeiramente negativos, mas desce marcadamente à medida que  $F$  sobe: a depressão de consanguinidade é capaz de manter o  $F$  próximo de 0 numa população destas, mesmo que a  $N_c$  não seja muito elevada.

Nestes dois exemplos, o *fitness* desce também quando  $F$  tende para  $-1$  [9, 10]: o “excesso” de heterozigóticos é prejudicial, por exemplo porque certos *loci* dão *fitness* máximo quando homozigóticos para certos alelos, ou porque as combinações alélicas obtidas entre gâmetas demasiado contrastantes levam a conflitos de expressão e consequente baixa do *fitness*. Por outras palavras, há *loci* sujeitos a uma forte selecção disruptiva, actua a favor de um ou mais homozigóticos.

As curvas A e B da figura 1 apenas indicam dois exemplos: cada espécie tem uma estratégia reprodutiva

evolutivamente estável [9], que se traduz em *nuances* adaptativas próprias. A sua caracterização ao detalhe, para cada espécie, pode determinar o sucesso ou o insucesso de certas manipulações genéticas. De notar que, com o exaustivo conhecimento prático que se foi acumulando sobre as espécies mais manipuladas no melhoramento, essa caracterização está bastante avançada para elas. Tome-se o exemplo de duas espécies com características reprodutivas muito diferentes entre si: o trigo é autogâmico e cada planta é praticamente homozigótica em todo o genoma e, embora se note por vezes o maior vigor das  $F_1$  de cruzamentos experimentais, é pouco significativo e não chega a ser explorável no melhoramento; aliás, todas as variedades de trigo comerciais são homozigóticas (usa-se como critério de linha pura seleccionar até à  $F_8$ ) e é evidente que não são afectadas por qualquer depressão de consanguinidade; o milho, planta alogâmica, é notório pelos efeitos da heterose (de tal maneira que foi nesta espécie que primeiro se caracterizou este fenómeno e se explorou para o melhoramento [3, 6]), e ao mesmo tempo nele é conhecida uma forte depressão de consanguinidade.

Os métodos moleculares anteriormente referidos podem vir a revelar-se úteis no levantamento de

indicadores genotípicos para o *fitness*, e mesmo estabelecer estimativas de *F* relevantes para a previsão da adaptabilidade a determinadas condições ambientais.

### V.a. Carga genética (*genetic load*)

O ambiente natural é limitativo do crescimento das populações e a competição intraspecífica inerente levaria a supor que uma população estabelecida há bastantes gerações, num local de condições estáveis, se encontra muito perto do óptimo adaptativo para essas condições. No entanto, a presença de genótipos com *fitness* sub-óptimo é inevitável, isto é, existe uma carga genética, como que um “lastro” de genes (ou haplótipos) desfavoráveis que persiste na população. Os tipos de carga genética que se conhecem são [7, 10]:

Mutacional: se o *fitness* máximo num determinado locus é representado por um dos homozigotos, a mutação (especialmente para um gene que não seja recessivo) introduz heterozigotia de *fitness* sub-óptimo. Este tipo de carga é especialmente importante em espécies clonais, embora também se faça sentir em populações de  $N_e$  grande. Segundo Wright [6:363], com uma taxa de mutação para um alelo letal recessivo da ordem dos  $10^{-5}$ , se  $N_e = 10^4$  a expectativa média de frequência desse gene é de 0,2% (com probabilidade de 15% de uma população nessas condições nem o ter). Nos casos de *fitness* máximo no heterozigótico, Wright sugere que tanto o  $N$  como os coeficientes de selecção dos homozigóticos terão maior ou menor peso sobre a distribuição das frequências dos genes em causa em função dos valores das taxas de mutação: por exemplo  $4N = 1/u$  pode dar praticamente qualquer valor dessas frequências, pela acção combinada dos coeficientes de selecção dos genótipos homozigóticos e da deriva genética.

Segregacional: produção de homozigóticos para genes desfavoráveis, nomeadamente letais ou semi-letais, ou inférteis, a partir de heterozigóticos portadores. Em espécies de ciclo de vida longo,

especialmente se com  $N_e$  pequeno, pode assumir grande importância.

Recombinatória: haplótipos que determinam maior *fitness* tendem a manter-se mais frequentes e a produzir por isso um desequilíbrio de ligação: esta é uma das facetas da co-adaptação entre diferentes *loci*, e pode ser substancialmente afectada pela recombinação se se produzirem haplótipos menos favoráveis — esta situação é exemplificada pela selecção disruptiva.

Dispersiva: resultante da impossibilidade dos descendentes desenvolverem-se na vizinhança dos progenitores; quando melhores genótipos em cada geração não se podem desenvolver por este constrangimento, não chegam a expressar-se genótipos mais favoráveis e assim persistem genes desfavoráveis nas populações.

### V.b. Eficiência fisiológica e versatilidade ecológica

Mitton [8] procurou fundamentar a heterose em termos de maior eficiência fisiológica dos indivíduos, postulando um menor “custo metabólico” basal (isto é, maior consumo de  $O_2$  em repouso) nos indivíduos menos consanguíneos, diferença especialmente notável em condições ambientais de stress. Estudos em animais (que provavelmente terão a sua correspondência em plantas) mostraram uma menor perda de peso, nos indivíduos menos consanguíneos, quando sujeitos a períodos de restrição alimentar, assim como uma diferença de consumo de  $O_2$  entre repouso e exercício muito mais marcada (talvez correlacionada com um crescimento mais rápido desses indivíduos). O facto de os menos consanguíneos terem proteínas com um período de semivida mais longo pode estar estreitamente relacionado com essa redução do metabolismo basal, mas ignora-se o porquê dessa diferença.

A menor eficiência fisiológica nos mais consanguíneos resulta num menor vigor somático, donde menor fecundidade e fertilidade... e menor

*fitness*. A nível populacional, uma das consequências da menor fertilidade desses indivíduos será uma maior variância de produtividade média ( $\sigma_k^2$ ) que tenderá a uma substancial redução do  $N_c$  a não ser que o “espaço” deixado aos indivíduos mais férteis permita um aumento compensatório do próprio  $k$  — naquilo que é designado por “homeostase populacional” — e, com isso, uma acelerada eliminação dos génotipos menos férteis.

Talvez em interligação com estas diferenças esteja um outro postulado para o maior *fitness* dos heterozigóticos em certos *loci*: os dois alelos de um heterozigótico complementam-se entre si na interacção com certos factores ambientais, o produto proteico de um conferindo as mesmas vantagens que o do outro mas para condições diversas — por exemplo de temperatura, potencial hídrico, intensidade luminosa, etc. — a que o indivíduo chega a estar sujeito. É como uma codominância ao nível metabólico a traduzir-se numa sobredominância em termos de *fitness*.

## VI. Aplicações

Da exposição feita até aqui resulta que as consequências que a nível fisiológico pode acarretar o aumento do coeficiente de consanguinidade requer que esse aumento seja controlado ou, muitas vezes, corrigido. As opções a tomar são forçosamente ditadas especificamente pelas necessidades de cada situação, por isso a discussão que se segue pretende somente apontar alguns princípios básicos a ter em mente.

### VI.a. Recursos genéticos

Os melhoradores de plantas compreendem há muito a necessidade de localizarem, caracterizarem e manterem as populações mais antigas das espécies de seu interesse, porque é nestas que os efeitos (relativamente lentos) da selecção natural podem ter resultado numa riqueza máxima de soluções genóticas relevantes para a sobrevivência em condições naturais e potencialmente importantes do

ponto de vista agronómico. A acessibilidade ao germoplasma dessas populações continua a ser o principal meio de obter novas variedades com acrescido interesse agronómico, nomeadamente pela introgressão de novos alelos nas variedades existentes. Noutra perspectiva, a heterose resultante do cruzamento entre variedades melhoradas e as formas “primitivas” presentes na natureza, ou cultivares delas obtidas independentemente, pode beneficiar a produção.

#### VI.a.i. Preservação do germoplasma: *ex situ* e *in situ*

Os bancos de germoplasma são repositórios da diversidade genética que vai sendo recolhida da natureza ou dos programas de melhoramento. Devido à enorme quantidade e diversidade de material que se acumula, a preocupação principal é a de manter o germoplasma (geralmente sementes) vivo, em câmaras refrigeradas.

Porém, mesmo nas melhores condições de preservação, e com maior ou menor rapidez consoante as espécies, o germoplasma vai perdendo a viabilidade, acabando por exigir o “rejuvenescimento” de cada lote com pelo menos um ciclo de sementeira. A nova geração de sementes, que se colhem para nova preservação, tenderá a ter maior consanguinidade, e em certas circunstâncias não é de excluir a possibilidade de genes importantes para o melhoramento — em futuro próximo ou longínquo — estarem perdidos (por fixação dos seus alelos) com a repetição desses ciclos.

Mantendo o germoplasma nas estações de melhoramento, ou seja longe dos locais onde as espécies devem ser semeadas para produção (daí o termo *ex situ*), pode ainda vir a favorecer génotipos que não interessam aos agricultores. É natural que, pela elevada qualidade dos terrenos dessas estações e/ou pelo uso de fertilizantes e outros químicos, etc., a manifestação de maior vigor por parte de certos génotipos resulte no desalojar progressivo de outros, esses sim úteis para a produção. Acresce que, na

ausência de factores ecológicos condicionantes da sobrevivência dos diferentes genótipos (por exemplo competição, parasitismo, predação, etc.), a preservação *ex situ* pode conduzir à acumulação de genótipos que seriam eliminados nas condições naturais a que a espécie está adaptada. Além disso, implica que o germoplasma conservado não tem a oportunidade de coevolução com alguns desses factores. Estes factores em conjunto podem resultar na erosão progressiva da adaptabilidade das colecções de germoplasma à produção agrícola “real”, e no limite tornar as colecções existentes inúteis.

A preservação de germoplasma *ex situ* pode ainda recorrer à criação das chamadas “metapopulações”, obtidas pelo cruzamento (ao acaso ou segundo procedimentos controlados) entre lotes de origens distintas. Com esta estratégia visa-se contrabalançar os efeitos da deriva genética para, de modo geral, evitar a perda de genes; mas a grande proporção de genótipos *outbred* nestas metapopulações implica uma elevada carga recombinatória, acaba por ser necessário libertá-las para que se recuperem, *nos locais de produção*, genótipos bem adaptados. Mesmo assim, há que prever um reduzido sucesso adaptativo nas primeiras gerações, ou seja: apesar do elevado  $N_e$  destas metapopulações, no campo ele baixa de novo, com a possibilidade de perda da diversidade genética.

Em conjunto com a preservação da diversidade genética duma espécie, é muito importante a preservação dos habitats onde essa diversidade se encontra estabilizada. A pressão da actividade humana tende, senão a fazê-los desaparecer, pelo menos a reduzir a sua extensão a tal ponto que, para algumas espécies, resultam drásticas reduções do  $N_e$ . Outra componente a preservar, com as espécies cultivadas, são as práticas tradicionais dos agricultores nos centros de origem e noutros locais onde cada espécie seja particularmente rica em diversidade genética. Não que isso exclua a possibilidade de melhorar essas práticas com base nos conhecimentos científicos, mas a sua substituição por práticas de cultura não-tradicionais, às

quais nem todos os genótipos dão a mesma resposta em termos de produção, pode implicar a rápida erosão genética dessa diversidade, senão mesmo extinção de uma componente importante desse repertório genético. A resposta a estas necessidades e aos problemas inerentes à preservação *ex situ* podem passar por uma estratégia totalmente diferente, a preservação *in situ*.

Contraopondo-se à estratégia “estática” da preservação *ex situ*, a preservação *in situ* é uma estratégia “dinâmica”, na qual se opta por manter o germoplasma no campo, isto é, semeando-o continuamente e tentando controlar o aumento da consanguinidade. A preservação *in situ* implica que o agricultor, seja a título individual seja através das comunidades onde pertence, participa na preservação do germoplasma de uma espécie, cultivando em parcelas do seu terreno genótipos que lhe são entregues pelo melhorador. Actuando na coordenação de redes de agricultores, apesar de porventura ser uma tarefa complicada, os melhoradores têm o potencial de otimizar a preservação dinâmica de germoplasmas, pela manutenção, *conjuntamente*, dos genótipos adaptados naturalmente e das condições ambientais que se relacionam com essas adaptações. Por isso, a sua implementação para dada espécie é especialmente apropriada em regiões onde a mesma exiba maior diversidade genética, nos centros de origem das espécies e não só — onde o grau de diversidade existente implique uma responsabilidade de conservação envolvendo melhoradores e produtores. No caso de Portugal, espécies como o trigo rijo e o feijão constituem exemplos disso mesmo [17].

#### VI.b. Programas de melhoramento por selecção

O impressionante progresso conseguido nas características das novas variedades saídas de programas de melhoramento não dá sinais de vir a parar tão cedo. Deve-se tal facto à não menos impressionante diversidade genética presente nos germoplasmas disponíveis, assim como ao cuidado que

os melhoradores se vêem obrigados a ter de modo a preservar essa diversidade.

Um desses cuidados tem a ver com a aparente interdependência dos caracteres alvo de melhoramento com certos fenótipos vitais para a sobrevivência e reprodução dos indivíduos. O progresso no melhoramento de um fenótipo traz consigo, muito frequentemente, efeitos adversos de perda de fertilidade ou aumento da letalidade; as causas possíveis são muito diversas, e podem ser o resultado da perda fortuita de genes (às vezes envolvidos pleiotropicamente na expressão dessas características), seja por deriva genética, seja por ligação cromossômica; a vantagem dos heterozigóticos em certos *loci* pode também implicar um carga genética segregacional que se manifesta através da selecção artificial. No seu conjunto, reflectem uma selecção natural contra os indivíduos seleccionados artificialmente, e requerem um relaxamento da selecção durante uma ou duas gerações, ou então medidas de “abertura” das populações onde se faz a selecção.

Trabalhar com populações “fechadas”, isto é, começar com uma população-base e produzir uma única linha de melhoramento sem introduzir diversidade por cruzamentos com outras populações, é ignorar que os  $N_e$  são forçosamente finitos e a fixação de alelos acaba por ter os reflexos mais imprevisíveis (perda de flexibilidade adaptativa, perda de fertilidade, etc.) sobre as características das populações que se supõe estarem melhoradas. Duas medidas de “abertura” contrabalançam esses efeitos, e que cabe aos melhoradores saber introduzir na altura certa:

Primeiro, e a exemplo do que é prática corrente em melhoramento animal, o cruzamento com *stocks* genéticos diferentes, como que a introduzir “sangue novo” na população; para isso importa, em plantas, ter acesso a um germoplasma capaz de corresponder a essa necessidade; bem entendido, a escolha de *stocks* introduzidos é regida pelos objectivos de cada programa de melhoramento, pelas características das

populações disponíveis, etc..

Segundo, a manutenção em simultâneo de diversas linhas sujeitas ao mesmo procedimento de melhoramento, na expectativa que os efeitos da deriva genética não se tenham repetido exactamente de umas para as outras — ou seja, diversos *loci* podem ficar fixados em cada linha, mas com alelos diferentes de umas para as outras. Por outras palavras, os gâmetas tendem a ficar com uma correlação negativa entre diferentes linhas. Cruzando-as entre si, durante o programa de selecção restabelece-se temporariamente um  $N_e$  relativamente elevado que permite recuperar muita da diversidade para a nova geração. Equivale isto a dizer que, comparado com uma só população de  $N_e$  custosamente alto, é preferível fazer o melhoramento por selecção dentro de  $n$  populações pequenas ( $N_e/n$  para cada uma, por exemplo) em paralelo, onde a deriva assume um papel mais preponderante, mas com muito menor risco da perda irreversível de alelos de interesse, que a mutação muito provavelmente não iria repor. Como para tudo, há limites para esta estratégia, dado que a excessiva fragmentação do *stock* a ser melhorado torna as respostas demasiado imprevisíveis, e aí o progresso do melhoramento dentro de cada linha seria inviabilizado.

#### VI.c. Heterose

A produção de híbridos (*crossbreeding*) é o único meio universal de obter populações melhoradas geneticamente, homogéneas e heterozigóticas [3, 16]. Explorando duas componentes da heterose — o acréscimo de vigor e da uniformidade fenotípica — para otimizar a produção, a qualidade, o tempo de maturação, etc., tem dado grandes resultados no caso do milho e doutras espécies, não só em termos de quantitativos de produção como na maior homogeneidade das culturas, facilitando a mecanização das colheitas.

A heterose pode ter interesse económico de várias maneiras. A mais evidente reside em vir a dispor-se de

populações homogêneas que produzem mais e/ou resistem melhor do que qualquer das populações anteriormente existentes, por via de fortes efeitos de (sobre)dominância e interações entre *loci*. Porém os híbridos a comercializar, apesar de muito férteis, não se utilizam para reprodução: a segregação dos genes na meiose acarreta a produção de homozigóticos e um declínio progressivo da heterose, e a selecção nas populações recombinantes não se revela a curto prazo útil. Deste modo, em termos de germoplasma, a manutenção resume-se a linhas puras progenitoras dos híbridos. A outra face da moeda, é que essas linhas puras estão exclusivamente na posse de produtores de semente híbrida, pelo que os agricultores dependem sempre do fornecimento de novas sementes todos os anos.

Mesmo quando o fenótipo de interesse não exhibe os efeitos da heterose por si mesmo, a melhoria do vigor e fertilidade pela heterose amplifica os ganhos económicos de utilizar os híbridos como progenitores (geralmente femininos). Por exemplo, a descendência do cruzamento triplo em gado produtor de carne,  $A \times (B \times C)$ , apesar da diminuição de 6% do peso médio dos descendentes na altura do desmame em relação à raça A, traduziu-se num acréscimo de 18% na produção total graças à fertilidade 25% mais alta das mães  $B \times C$  [3].

Após uma primeira fase de híbridos de cruzamentos quádruplos do tipo  $(A \times B) \times (C \times D)$ , os melhoradores de milho decidiram aproveitar algumas das linhas puras, consideradas mais interessantes, polinizando-as entre si para resultarem metapopulações donde se foram seleccionando novas linhas puras ainda melhores, não só em termos de multiplicação das sementes como dos muito importantes efeitos maternos (a teoria genética sobre os valores aditivos e genotípicos envolvidos neste processo de selecção pode ser consultada no livro de Falconer e MacKay [3], por exemplo). Graças a isso as  $F_1$  comerciais passaram a ser obtidas numa só geração, e passou a ser essa a regra também noutras espécies.

Constata-se hoje que uma grande parte da semente melhorada produzida anualmente, de quase todas as espécies que não são rigorosamente autogâmicas, é obtida por este procedimento, daí que se fale não só do milho híbrido, mas também da soja híbrida, etc.. Quase todo o material transgénico (isto é, hemizigótico para genes introduzidos por métodos moleculares) que é ou venha a ser comercializado também é híbrido.

## VII. Resumos

### VII.a. Português

Nesta revisão sobre consanguinidade em plantas parte-se do conceito (genealógico) de identidade por descendência e do cálculo de coeficientes de parentesco, para em seguida rever conceitos a nível populacional, através das duas perspectivas do parâmetro F de medida de consanguinidade: a de um coeficiente de fixação, que traduz o afastamento em relação ao modelo de Hardy-Weinberg das distribuições genotípicas nas populações reais, pelo facto de não serem infinitamente grandes; e a de correlação entre gâmetas, que entra em linha de conta com factores (como mutação, migração e selecção) que podem compensar a fixação de genes e assim fazer populações finitas parecerem estar de acordo com esse modelo. Exemplificam-se ainda os cálculos de identidade e diversidade genética nas populações. Analisam-se as bases genéticas e fisiológicas da depressão de consanguinidade (a redução de *fitness* frequentemente observada em indivíduos consanguíneos), e referem-se as implicações destes conceitos em aplicações como a gestão de recursos genéticos e melhoramento de plantas.

### VII.b. Inglês

This review of inbreeding in plants analyses the (genealogical) concept of identity by descent and the computation of relationship coefficients, followed by the two perspectives for the F parameter as measure of

inbreeding at the population level: as a coefficient of fixation measuring the distortion, in relation to the Hardy-Weinberg model, of genotype distributions in real populations, due to the fact that they are finite; and as correlation between gametes, where the compensatory effects of other factors (such as mutation, migration and selection) on gene fixation can also be taken into account, sometimes resulting in genotype distributions, in finite populations, that mimic those predicted by that model. An example of the computation of genetic identity and diversity in populations is included. The genetic and physiological bases for inbreeding depression (the decrease in fitness often observed in inbred individuals) are analysed, and the implications of this knowledge on the management of genetic resources and plant breeding are addressed.

### VIII. Agradecimentos

Aos colegas do departamento que participam na cadeira de Genética para Engenharia Agrícola da Universidade de Évora, pelo estímulo, perspectivas e bibliografia facultada.

### XI. Notas e referências

1. Segundo o Méd. Vet. Prof. Sieuve Monteiro, a consanguinidade é uma «Expressão originada pela concepção clássica (pré-mendeliana) de hereditariedade em que se admitia serem os fenómenos hereditários determinados nos descendentes pela fusão do 'sangue' dos progenitores...» (*Enciclopédia Luso-Brasileira de Cultura*, 5º volume, página 1422, Verbo, Lisboa, 1967), preconizando-se o uso do termo endocruzamento, ou por vezes endogamia, para tradução literal do termo inglês *inbreeding*. O termo consanguinidade, porém, parece ainda ser o único consagrado pelo uso em Português.

2. Mendel, G., 1865 — Experiências sobre híbridos de plantas. *Naturalia* (Supl.), edição comemorativa do centenário, traduzida para Português.

3. Falconer, D. S., Mackay, T. F. C., 1996 — *Introduction to quantitative genetics*, 4ª ed.. Longman, Harlow.

4. Strickberger, M. W., 1976 — *Genetics*, 2ª ed..

MacMillan.

5. Plum, M., 1954 — Computation of inbreeding and relationship coefficients in populations with a relatively small number of different male ancestors. *J. Hered.* 45, 92–94.

6. Wright, S., 1969 — *Evolution and the Genetics of Populations*, Vol. 2: *The Theory of Gene Frequencies*. The University of Chicago Press, Chicago & London (com indicação do número de página em cada referência).

7. Templeton, A. R., Read, B., 1994 — Inbreeding: one word, several meanings, much confusion. In: *Conservation Genetics*, pp. 91–105, ed. V. Loeschcke, J. Tomiuk, S. K. Jain. Birkhäuser, Basel.

8. Mitton, J. B., 1993 — Theory and data pertinent to the relationship between heterozygosity and fitness. In: *The Natural History of Inbreeding and Outbreeding. Theoretical and Empirical Perspectives*, Cap. 2, ed. N. W. Thornhill. Univ. Chicago, Chicago & London.

9. Waser, N. M., 1993 — Population structure, optimal outbreeding, and assortative mating in angiosperms. In: *The Natural History of Inbreeding and Outbreeding. Theoretical and Empirical Perspectives*, Cap. 9, ed. N. W. Thornhill. Univ. Chicago, Chicago & London.

10. Shields, W. M., 1993 — The natural and unnatural history of inbreeding and outbreeding. In: *The Natural History of Inbreeding and Outbreeding. Theoretical and Empirical Perspectives*, Cap. 8, ed. N. W. Thornhill. Univ. Chicago, Chicago & London.

11. Nei, M., 1973 — Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci.* 12, 3321–3.

12. Considerando um número  $n$  de alelos em dado *locus*, e a divisão da população em  $s$  subpopulações de peso igual, define-se  $J_s$  como sendo a média das identidades genéticas ( $J$ ) das subpopulações, ou seja  $J_s = (\sum J_i)/s$  ( $i = 1, \dots, s$ ), e  $D_{ST}$  como a diversidade média entre subpopulações tomadas duas a duas, ou seja  $D_{ST} = (\sum \sum D_{ij})/s^2$ , em todas as combinações de  $i$  e  $j$  ( $i, j = 1, \dots, s$ ). Calcula-se cada  $J_i$  pela fórmula  $J_i = \sum p_{ik}^2$  ( $k = 1, \dots, n$ ) sendo  $p_{ik}$  a frequência do alelo  $k$  na subpopulação  $i$ . O cálculo dos valores  $D_{ij}$  faz-se pela fórmula  $D_{ij} = \frac{1}{2}(J_i + J_j) - J_{ij}$ , em que  $J_{ij} = \sum p_{ik}p_{jk}$  ( $k = 1, \dots, n$ ) é a identidade genética entre as subpopulações  $i$  e  $j$ , expressa pelos produtos das frequências para cada alelo  $k$  nessas subpopulações. Note-se que quando  $i = j$ ,  $D_{ij} = 0$ . Nas outras combinações, Nei demonstrou que  $D_{ij} = \frac{1}{2}\sum (p_{ik} - p_{jk})^2 \geq 0$  [11].

13. Hamrick, J. L., Linhart, Y. B., Mitton, J. B., 1979 —

Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Ann. Rev. Ecol. Syst.* 10, 173-200.

*Rev. Ecol. Syst.* 10, 173-200.

14. Johnson, G. B., 1974 — Enzyme polymorphism and metabolism. *Science* 184, 28-37.

15. Lewontin, R. C., 1973 — Population genetics. *Ann. Rev. Genet.* 7, 1-17.

16. Eberhart, S. A., 1977 — Quantitative genetics and practical corn breeding. In: *Proceedings of the International Conference on Quantitative Genetics* (E. Pollack, O. Kempthorne, T. B. Bailey, Jr., eds.), Iowa State University, pp. 491-502.

17. Comunicação pessoal, Eng<sup>o</sup> Maçãs (ENMP, Elvas) e Eng<sup>o</sup> Téc. Bettencourt (EAN, Oeiras).