

Terrence P. McGarty
Telmarc Gardens

BACKCROSSING: A MATHEMATICAL MODEL IN DETAIL

Problem Statement

- ⦿ We have a plant, say “Bess Ross” which is red but not well branched.
- ⦿ We have another plant, say H multiflora which is branched but not red.
- ⦿ We desire a red-branched plant which has all the characteristics of Bess Ross but it must have branching
- ⦿ We want to “drive” branching into Bess Ross

Restatement in Genes

Recurrent
Assume a Red
“Bess Ross” with no
branching.

$$x_1 x_2 \dots x_N$$

$$x_1 x_2 \dots x_N$$

Non Recurrent:
Assume a species H
multiflora with
branching. Assume
y2 is the branching
gene.

$$y_1 y_2 \dots y_N$$

$$y_1 y_2 \dots y_N$$

Desired: A Red
Bess Ross like
plant but with
branching.

$$x_1 y_2 \dots x_N$$

$$x_1 y_2 \dots x_N$$

To get F1 we cross the two original plants to obtain:

$$x_1 x_2 \dots x_N$$

$$y_1 y_2 \dots y_N$$

No we assume that x_2 is recessive and y_2 dominant. This means that we get the branching we were seeking. However we get everything else in the mix. Namely we have all the other y genes. We want to have only y_2 and no other y genes.

F2: We cross the F1 with the Recurrent parent.

$$\begin{array}{ccc} x_1 x_2 \dots x_N & & x_1 x_2 \dots x_N \\ & \times & \\ y_1 y_2 \dots y_N & & x_1 x_2 \dots x_N \end{array}$$

Then select the branched plants only. This yields genes which are a mix. We assume we have no linkages for simplicity. Furthermore we see we have a 25% chance of there being a y gene at each of the loci except the y2 position because we deliberately always choose branching.

F2 gene sequences may be the following if we let N=2

 $x_1 x_2$
 $y_1 y_2$
 \times
 $x_1 x_2$
 $x_1 x_2$

	x1	y1
x1	x1x1	x1y1
x1	x1x1	x1y1

	x2	y2
x2	x2x2	x2y2
x2	x2x2	x2y2

Total of 4 possible in the 1 cross and 4 in the 2 cross for a total of 16 possible crosses. However we seek only the y2 crosses since they give branching. Thus there are 8 possible outcomes that retain branching.

F2 gene sequences may be the following if we let $N=2$

x_1x_2

y_1y_2

X

x_1x_2

x_1x_2

Retain Branching

x_1x_2

x_1y_2

4 of
These

x_1x_2

y_1y_2

4 of
These

F3: We cross F2 with the Recurrent Parent.

$x_1 x_2 \dots x_N$

$y_1 y_2 \dots y_N$

X

$x_1 x_2 \dots x_N$

$x_1 x_2 \dots x_N$

$x_1 x_2 \dots x_N$

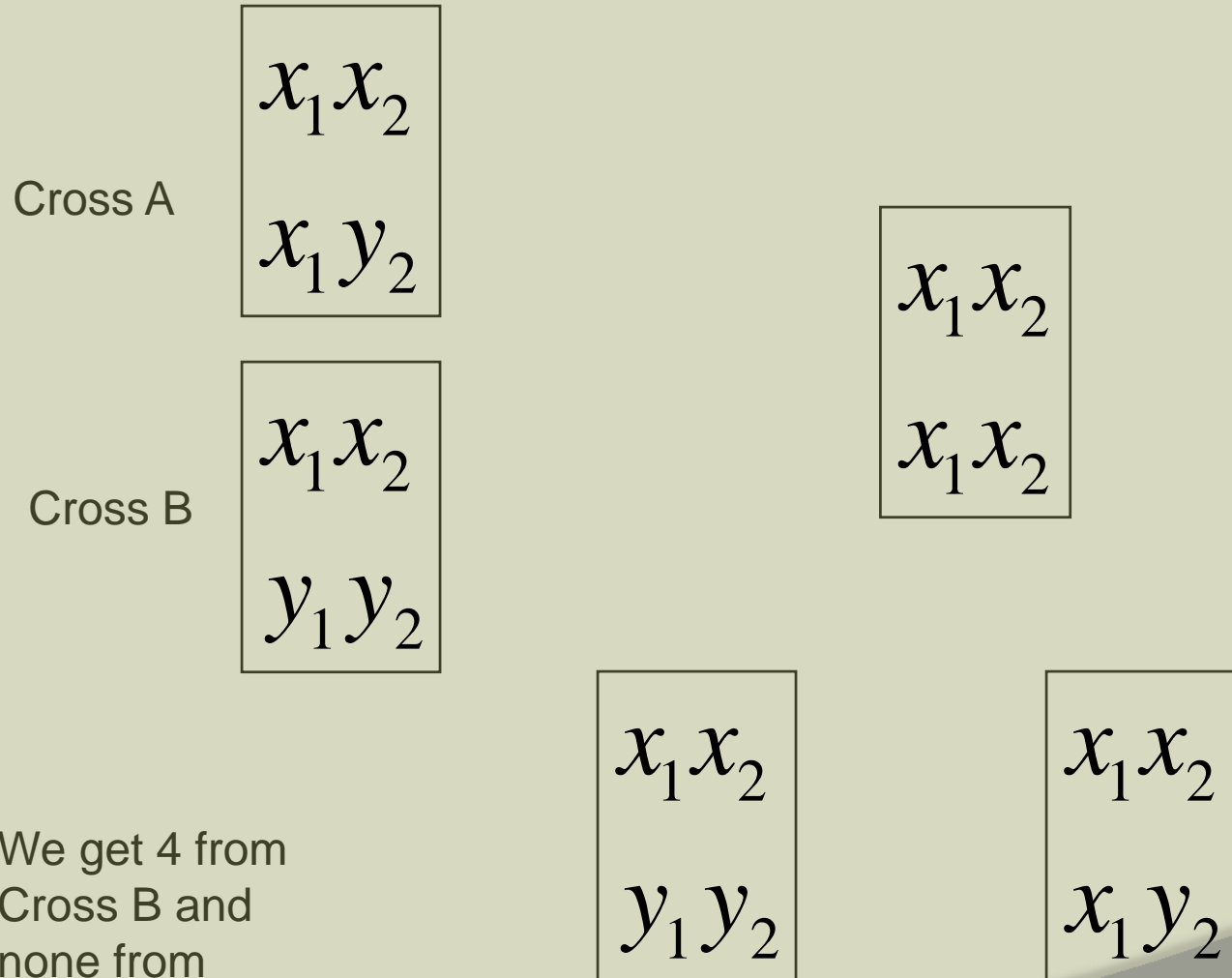
$x_1 y_2 \dots y_N$

$x_1 x_2 \dots x_N$

$x_1 y_2 \dots x_N$

By back crossing but by selecting the branched one each time to backcross we retain y_2 but we slowly drive out the other y genes.

For F3: We consider the 2 gene case again.

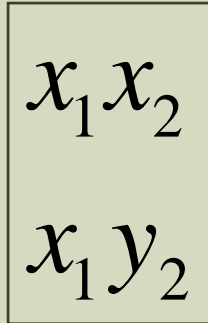


We get 4 from
Cross B and
none from
Cross A

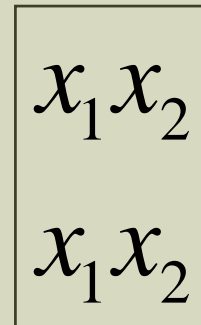
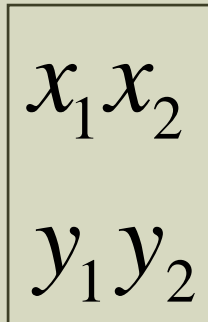
We get 4 from
Cross B plus 8
from Cross A
for a total of
12 or 75% are
homozygous
on gene 1.

For F4: We consider the 2 gene case again and we then iterate.

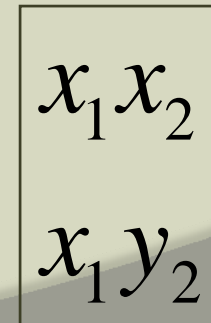
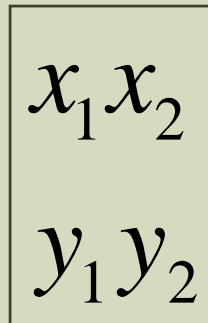
Cross A
25% of
population



Cross B
50% of
population



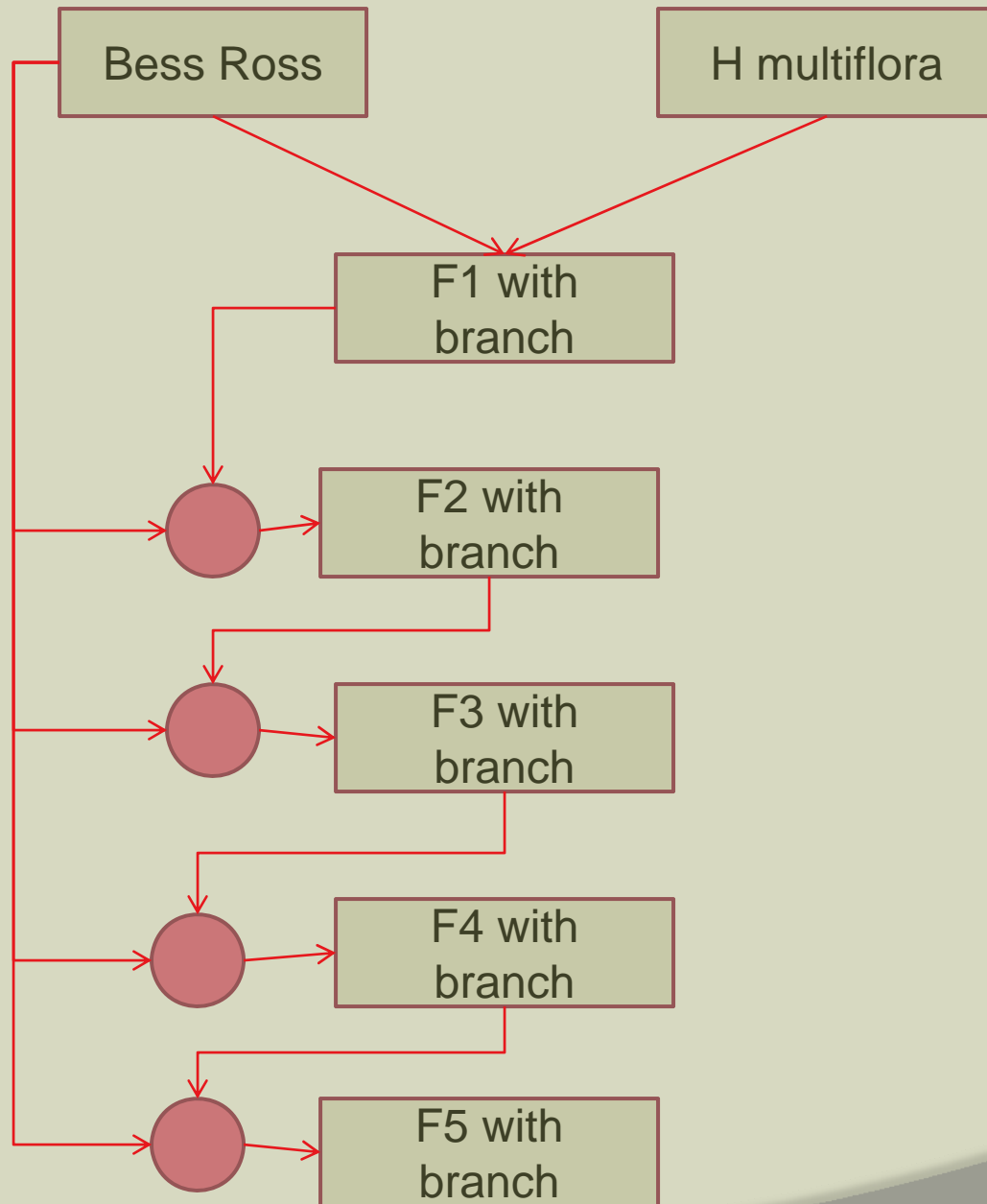
We get 1/2 of
1/4 or 12.5%
from Cross B
and none from
Cross A



We get 87.5%
which are
homozygous
on gene 1.

- We get more and more homozygous on gene 1 and we keep y2 on gene 2.
- The same percent would be the case across all genes!
- Consider the model using 4 Genes

Algorithm for Backcrossing



Details

Detailed Example

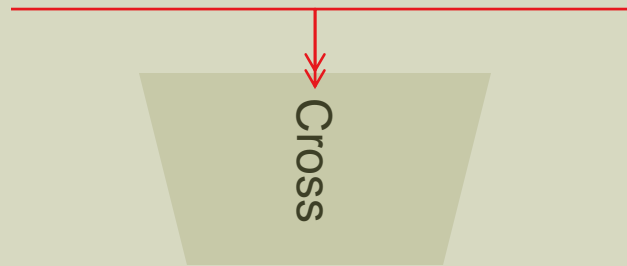
- Consider a 4 Gene Case. Assume we want to insert y_1 into the genome of the x sequence. Assume further that y_1 is dominant.
- For example, we want branching from a *H. multiflora* to be placed into a red “Bess Ross”
- Example can be generalized to N genes and even M characteristics to be “driven: in from Non Recurrent into the Recurrent.

The F1 cross is as follows. All F1 are identical. We assume that both initial parents are homozygous. Namely they have identical genes on both chromosomes. We further assume that there is no linkage.

$x_1 x_2 x_3 x_4$

$x_1 x_2 x_3 x_4$

Bess Ross



$y_1 y_2 y_3 y_4$

$y_1 y_2 y_3 y_4$

H multiflora

$x_1 x_2 x_3 x_4$

$y_1 y_2 y_3 y_4$

$$\text{Let } X = \begin{Bmatrix} x_1 x_2 x_3 x_4 \\ x_1 x_2 x_3 x_4 \end{Bmatrix}, Y = \begin{Bmatrix} y_1 y_2 y_3 y_4 \\ y_1 y_2 y_3 y_4 \end{Bmatrix}, XY = \begin{Bmatrix} x_1 x_2 x_3 x_4 \\ x_1 x_2 x_3 x_4 \end{Bmatrix}$$

define

$$\tilde{X}_0 = \begin{Bmatrix} x_2 x_3 x_4 \\ x_2 x_3 x_4 \end{Bmatrix}$$

$$\tilde{X}_1 = \begin{Bmatrix} x_2 x_3 x_4 \\ y_2 x_3 x_4 \end{Bmatrix} \text{ or } \begin{Bmatrix} x_2 x_3 x_4 \\ x_2 y_3 x_4 \end{Bmatrix} \text{ or } \begin{Bmatrix} x_2 x_3 x_4 \\ x_2 x_3 y_4 \end{Bmatrix}$$

$$\tilde{X}_2 = \begin{Bmatrix} x_2 x_3 x_4 \\ y_2 y_3 x_4 \end{Bmatrix} \text{ or } \begin{Bmatrix} x_2 x_3 x_4 \\ y_2 x_3 y_4 \end{Bmatrix} \text{ or } \begin{Bmatrix} x_2 x_3 x_4 \\ x_2 y_3 y_4 \end{Bmatrix}$$

$$\tilde{X}_3 = \begin{Bmatrix} x_2 x_3 x_4 \\ y_2 y_3 y_4 \end{Bmatrix}$$

or any variant of a y_k in each of the entries as shown above.

Then using the previous notation we have for F1:

$$X \oplus Y = XY$$

And then for F2 if we choose the plant with the y_1 expression and any other genes we can write for F2:

$$X \oplus XY = XY_1 \tilde{X}_k$$

and

$$XY_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Then for F2, if we select the y1 dominant ones only we have as possible results:

$$X \oplus XY = \begin{cases} XY_1\tilde{X}_0; & \text{there is just one of these, with a probability of } 1/8 \\ XY_1\tilde{X}_1; & \text{there are three of these each with a probability } 1/8 \\ XY_1\tilde{X}_2; & \text{there are three of these each with probability } 1/8 \\ XY_1\tilde{X}_3; & \text{there is one of these with probability } 1/8 \end{cases}$$

These are the resulting possible gene collections at F2 when we select only one with y1 dominant and assume no linkage and equal probability of linking. Thus at F2 the probability of a y1 plus homozygous for all x elsewhere is only 12.5%

Some Mathematical Detail

- Note that genes x_n and y_n are equally likely and have probability $\frac{1}{2}$
- Note that if we look at the gene tails, if they are M in length then we have $[1/2]^M$ for any one of them
- Note that for the combinations of 0, 1, 2, 3, etc we have the binomial distribution to provide the probability.

We now move to F3. At this point we take only the F1 with y1 and then again cross with pure X. We have the following possible crosses:

$$X_0 \oplus X_0 = \{ X_0$$

$$X_0 \oplus X_1 = \begin{cases} X_0; \text{with probability } 1/2 \\ X_1; \text{with probability } 1/2 \end{cases}$$

$$X_0 \oplus X_2 = \begin{cases} X_0; \text{with probability } 1/4 \\ X_1; \text{with probability } 1/2 \\ X_2; \text{with probability } 1/4 \end{cases}$$

$$X_0 \oplus X_3 = \begin{cases} X_0; \text{with probability } 1/8 \\ X_1; \text{with probability } 3/8 \\ X_2; \text{with probability } 3/8 \\ X_3; \text{with probability } 1/8 \end{cases}$$

Observations

- ◉ Note that we have probabilities of a specific form in a current generation being contingent upon what the form was on the prior generation.
- ◉ The prior discussion can be generalized for any set of complex generations.
- ◉ However these transition probabilities remain fixed from generation to generation. That is they are Markov.

Thus we can calculate for F3 the number or percent or probability of a homozygous set of X:

<i>Cross</i>	<i>Prob of This Cross</i>	<i>Prob of X0 in this Cross</i>	<i>Prob X0 at F3</i>
X0	1/8	1	1/8
X1	3/8	1/2	3/16
X2	3/8	1/4	3/32
X3	1/8	1/8	1/64
Total Prob X0 in F3			27/64

Note that with this back cross we have Y1 but we now have a Prob of a homozygous of X at 0.42 from 0.125 at the prior cross.

Now we can do this do the other three possibilities as we did before and use this in F4. Thus:

<i>Cross</i>	<i>Prob of This Cross</i>	<i>Prob of X1 in this Cross</i>	<i>Prob X1 at F3</i>
X0	1/8	0	0
X1	3/8	1/2	3/16
X2	3/8	1/2	3/16
X3	1/8	3/8	3/64
Total Prob X1 in F3			27/64

<i>Cross</i>	<i>Prob of This Cross</i>	<i>Prob of X2 in this Cross</i>	<i>Prob X2 at F3</i>
X0	1/8	0	0
X1	3/8	0	0
X2	3/8	1/4	3/32
X3	1/8	3/8	3/64
Total Prob X2 in F3			9/64

<i>Cross</i>	<i>Prob of This Cross</i>	<i>Prob of X3 in this Cross</i>	<i>Prob X3 at F3</i>
X0	1/8	0	0
X1	3/8	0	0
X2	3/8	0	0
X3	1/8	1/8	1/64
Total Prob X3 in F3			1/64

We now do F4, and again we select the plants expressing Y1 and we again back cross with the homozygous X. This follows the same logic we did for F3.

<i>Cross</i>	<i>Prob of This Cross in F3</i>	<i>Prob of X0 in this Cross</i>	<i>Prob X0 at F4</i>
X0	27/64	1	27/64
X1	27/64	1/2	27/128
X2	9/64	1/4	9/256
X3	1/64	1/8	1/512
Total Prob X0 in F4			343/512= 0.67

This then yields a 67% Homozygous for F4 with three genes other than the one we want impressed. The Table above can then be iterated again and again. We simply use 342/512 in the second column.

<i>Cross</i>	<i>Prob of This Cross in F3</i>	<i>Prob of X1 in this Cross</i>	<i>Prob X1 at F4</i>
X0	27/64	0	0
X1	27/64	1/2	27/128
X2	9/64	1/2	9/128
X3	1/64	3/8	3/512
Total Prob X1 in F4			147/512= 0.287

<i>Cross</i>	<i>Prob of This Cross in F3</i>	<i>Prob of X2 in this Cross</i>	<i>Prob X2 at F4</i>
X0	27/64	0	0
X1	27/64	0	0
X2	9/64	1/4	18/512
X3	1/64	3/8	3/512
Total Prob X2 in F4			21/512= 0.041

<i>Cross</i>	<i>Prob of This Cross in F3</i>	<i>Prob of X3 in this Cross</i>	<i>Prob X3 at F4</i>
X0	27/64	0	0
X1	23/64	0	0
X2	9/64	0	0
X3	1/64	1/8	1/512
Total Prob X3 in F4			1/512

Algorithm

- ⊙ We now have a simple algorithm
 - The column for the last cross must be iteratively calculated for every prior step as shown
 - The column for the probability at the current cross can be calculated once, they will be binomial in form
 - The probabilities for the current and then next cross can be calculated by summing the products
 - Note that the larger the genome in the Recurrent the more complex and the longer the convergence