AN ALGEBRAICALLY EXACT EXAMINATION OF JUNCTION FORMATION AND TRANSMISSION IN PARENT-OFFSPRING INBREEDING

Sir RONALD A. FISHER
Department of Genetics, University of Cambridge

1. INTRODUCTION

THE Theory of Junctions in inbreeding was first sketched in the third chapter of the author's *Theory of Inbreeding* (1949). This was by no means a thorough treatment, being confined to the case of sib-mating. It was intended to illustrate the method by which the extent to which the germ-plasm is subjected to recombination in the course of a complete inbreeding programme, and thence the frequency with which at each stage the entire line becomes homogenic, can be calculated. In the author's opinion the course of events cannot be halted, or even greatly retarded, by moderate differences in viability; but, in the case of such bisexual organisms as the house-mouse must often be completed in forty or fifty generations.

In 1953 J. H. Bennett published in *Genetica* a paper on "Junctions in Inbreeding" giving comparative results for three other cases, namely (*a*) self-fertilisation in disomics, (*b*) self-fertilisation in tetrasomics, and (*c*) alternate parent-offspring mating in bisexual forms. In the last case it was remarkable that complete homogeneity appeared to set in some three generations earlier than in the case of sib-matings, which in many other respects it closely resembles.

The author has been struck by some minor discrepancies in the last series of results and, since the case is in some respects of especial simplicity, has been led to explore so far as to see if exact expectations at all stages could not be calculated instead of the asymptotic formulae he had previously used. Some inaccuracies in the original discussion have in the meanwhile been corrected in "A fuller theory of junctions in inbreeding" in *Heredity* (Fisher, 1954).

As will be seen in the following account, Bennett's conclusion that homogeneity is attained, at each level of probability, rather earlier by parent-offspring than by sib-matings is confirmed, but the difference appears to be less than was thought, ranging in the relevant region from about 1·8 to about 1·6 generations.

The symbols *a*, *b*, *c*, *d* are used to specify the tracts, in the neighbourhood of any particular point, introduced into the line by the foundation mating. The symbol *j* is used for a junction always taken to be between tracts *a* and *b*, so as to resemble *a* on one side and *b* on the other. With these conventions the various possible types of mating may be specified concisely. It seems probable that the small discrepancies mentioned above are due in part to inconsistencies in the numbering of the generations.

Without junctions, parent-offspring inbreeding starts with an unrelated mating, generation zero, leading to the first mating of offspring with parent. This is typically trigenic, but will lead to segregating digenic types of three kinds in accordance with the standard matrix.

$$
\begin{array}{lll}
\text{Generation} & 0 & ab = cd \\
\text{,,} & 1 & ac = ab
\end{array}
$$

	Matrix			
Offspring	ab	ab	ab	aa
Parent	ac	ab	aa	ab
Frequency	x	v	u'	u
$ab = ac$	2	–	–	–
$ab = ab$	1	1	1	–
$ab = aa$	–	–	–	1
$aa = ab$	1	1	1	–
	4	2	2	1

From this it appears, since in generation (1) the mating must be of trigenic type, that

$$x_n = \left(\tfrac{1}{2}\right)^{n-1}$$

$$v_n = u_n = -\tfrac{1}{2}\left(\tfrac{1}{2}\right)^{n-1} + \tfrac{1}{10}(1 + 6\epsilon)\epsilon^{n-1} + \text{conjugate}$$

$$u'_n = -\tfrac{1}{2}\left(\tfrac{1}{2}\right)^{n-1} + \tfrac{1}{10}(2 + 2\epsilon)\epsilon^{n-1} + \quad \text{,,}$$

Hence for $s > 0$, the fraction of the strands still heterogenic is

$$-\left(\tfrac{1}{2}\right)^s + \tfrac{1}{10}(6 + 16\epsilon)\epsilon^s + \tfrac{1}{10}(14 - 16\epsilon)\left(\tfrac{1}{2} - \epsilon\right)^s$$

TABLE 1

s	Numerator	Percentage
0	1	100
1	2	100
2	4	100
3	7	87·5
4	12	75·0
5	20	62·5
6	33	51·5625
7	54	42·1875
8	88	34·375
9	143	27·92969
10	232	22·65625
11	376	18·35938
12	609	14·86816
13	986	12·03613
14	1596	9·72900
15	2583	7·88269

Table 1. Percentage of map length still heterogenic, illustrating the continuous method of calculation of the numerator

$$N_s = N_{s-1} + N_{s-2} + 1$$

to be divided by 2^s.

2. INTERNAL JUNCTIONS OF TYPE (a, b, j)

The trigenic and digenic types produce the internal junction type (δ), in particular,

$$aj = ab$$

with frequency $\quad u'_s + 2v_s + \frac{1}{2}x_s$

or, appearing in generation $(s+1)$

$$-2(\tfrac{1}{2})^s + \tfrac{1}{10}(6 + 16\epsilon)\epsilon^s + \text{conjugate,}$$

and of these the fraction $(\tfrac{1}{2})^{n-s-1}$ survive as internal junctions (a, b, j) to generation n. But

$$_1 S^{n-1}(\tfrac{1}{2})^{n-s-1}\{-2(\tfrac{1}{2})^s + \tfrac{1}{10}(6 + 16\epsilon)\epsilon^s + \tfrac{1}{10}(14 - 16\epsilon)(\tfrac{1}{2} - \epsilon)^s\}$$

is

$$\tfrac{1}{10}(14 + 44\epsilon)\epsilon^{n-1} + \tfrac{1}{10}(36 - 44\epsilon)(\tfrac{1}{2} - \epsilon)^{n-1} - (2n + 3)(\tfrac{1}{2})^{n-1}.$$

Table 2 shows the calculation of these expected frequencies.

TABLE 2

Generation	N	(δ)	(a)
1	0	0	1·00000
2	2	·50000	·50000
3	8	1·00000	·25000
4	20	1·25000	·12500
5	42	1·31250	·06250
6	80	1·25000	·03125
7	144	1·12500	·01562
8	250	·97656	·00781
9	424	·82812	·00391
10	708	·69141	·00195
11	1170	·57129	·00098
12	1920	·46875	·00049
13	3136	·38281	·00024
14	5106	·31165	·00012
15	8296	·25317	·00006

Table 2. Expected numbers for 100 cM of map length of internal junction (a, b, j), with two original elements a, b, and a third derived from them by recombination. If either a or b is lost the junction is transmitted as an external junction.

Recurrence relation :—

$$\mathcal{N}_s = \mathcal{N}_{s-1} + \mathcal{N}_{s-2} + (4s - 6)$$

3. OTHER TYPES OF INTERNAL JUNCTIONS

The trigenic junctions (a, b, j) derived from

$$aj = ab$$

are the only kinds derivable from digenic matings, yet the foundation mating,

$$ab = cd,$$

and trigenic types,

$$ab = ac,$$

derived from it, yield three other kinds of internal junction.

(α) The type

$$cj = ab$$

involving both original strands a and b, the crossover strand j derived from them, and a strand c having nothing in common with the others, is completely transient, leading never to itself, but always to

$$ac = cj$$

or

$$aj = cj.$$

Being transient, the frequency is simply that of production by crossing-over in the previous generation. But, in the foundation mating, crossing-over occurring in the mate to be used a second time, leads to this type, while crossing-over in the other mate is ineffectual, since both the original strands are simultaneously lost.

Moreover, from the trigenic matings

$$ab = ac$$

it arises with frequency $\frac{1}{2}$ for each unit of map length, and therefore from generation 2 onwards, it must appear with frequency

$$2(\tfrac{1}{2})^n,$$

as also in the first generation.

These also are shown in table 2.

(β) The types (a, c, j) constitute the total output of (α), and therefore appear new in generation $n(n > 1)$ with frequency

$$4(\tfrac{1}{2})^n.$$

They are also half the output by recombination from the trigenic types and this supplies as much again by a different path.

These are $aj = ac$ and $cj = ac.$

The new production (for $s > 1$) is then

$$8(\tfrac{1}{2})^s,$$

and the type is maintained in half the descendants in each generation; consequently by generation n there has accrued

$$\underset{s=2}{S^n} \, 8(\tfrac{1}{2})^s(\tfrac{1}{2})^{n-s}$$

or

$$8(n-1)\,(\tfrac{1}{2})^n.$$

(γ) These types β, or (a, c, j) are of six kinds, with the generation matrix shown below

Offspring	Offspring Parent	r ac cj	s aj cj	t aj ac	u cj ac	v ac aj	w cj aj
ac	cj	.	.	.	1	.	1
aj	cj	.	.	.	1	.	1
aj	ac	1	.	.	.	1	.
cj	ac	1	.	.	.	1	.
ac	aj	.	1	1	.	.	.
cj	aj	.	1	1	.	.	.
cj	cj	.	.	.	1	.	1
cc	cj	.	.	.	1	.	.
jj	cj	.	.	.	.	.	1
		4	4	4	4	4	4

Latent Root	Principal Component
0	$r - s$
0	$t - u$
0	$v - w$
$-\frac{1}{4}$	$2r - s - t - u + 2v - w$
$-\frac{1}{4}$	$r - u + v - w$
$\frac{1}{2}$	$r + s + t + u + v + w$

The first four are supplied anew at rate $2(\frac{1}{2})^s$ each, when $s > 1$; after $n - s$ generations

$$u_n + w_n = \tfrac{1}{3}\{4(\tfrac{1}{2})^{n-s} - (-\tfrac{1}{4})^{n-s}\}(\tfrac{1}{2})^{s-1},$$

summed from $s = 2$ to n, this is

$$\tfrac{4}{3}(n - \tfrac{7}{6})(\tfrac{1}{2})^{n-1} + \tfrac{2}{9}(-\tfrac{1}{4})^{n-1},$$

so that a new production of type (γ) appears in the nth generation with frequency

$$\tfrac{4}{3}(n - \tfrac{13}{6})(\tfrac{1}{2})^{n-1} - \tfrac{4}{9}(-\tfrac{1}{4})^{n-1}$$

from $n = 3$.

For the last enumeration then we may note that the probability of these remaining after $s - n$ further generations, as a junction (c, j) of type (γ), is

$$\tfrac{1}{10}(3 + 8\epsilon)\epsilon^{s-n} + \tfrac{1}{10}(7 - 8\epsilon)(\tfrac{1}{2} - \epsilon)^{s-n}.$$

The process of summation from $n = 3$ to $n = s$ is somewhat intricate, but the result is similar to those of the other cases. Algebraically, it is

$$-(4n + 2)2^{-n} + \tfrac{1}{5}(6 + 16\epsilon)\epsilon^n + \tfrac{1}{5}(14 - 16\epsilon)(\tfrac{1}{2} - \epsilon)^n,$$

which can be evaluated continuously by the recurrence formula

$$\mathcal{N}_s = \mathcal{N}_{s-1} + \mathcal{N}_{s-2} + (4s - 10)$$

dividing at each stage by 2^s.

Table 4 summarises the expected numbers per 100 cM of strand, of the four types of internal junctions.

TABLE 4

	α $cj = ab$	β (a, c, j)	γ (c, j)	δ (a, b, j)	Total internal junctions
1	1·00000	—	—	—	1·00000
2	·50000	2·00000	—	·50000	3·00000
3	·25000	2·00000	·25000	1·00000	3·50000
4	·12500	1·50000	·50000	1·25000	3·37500
5	·06250	1·00000	·62500	1·31250	3·00000
6	·03125	·62500	·65625	1·25000	2·56250
7	·01562	·37500	·62500	1·12500	2·14062
8	·00781	·21875	·56250	·97656	1·76562
9	·00391	·12500	·48828	·82812	1·44531
10	·00195	·07031	·41406	·69141	1·17773
11	·00098	·03906	·34570	·57129	·95703
12	·00049	·02148	·28564	·46875	·77637
13	·00024	·01172	·23438	·38281	·62915
14	·00012	·00635	·19141	·31165	·50952
15	·00006	·00342	·15582	·25317	·41248
Asymptotic	$2 \cdot 2^{-n}$	$8(n-1)2^{-n}$	$-(4n+2)2^{-n}$ $\frac{1}{5}(6+16\epsilon)\epsilon^n$	$-(4n+6)2^{-n}$ $\frac{1}{5}(8+28\epsilon)\epsilon^n$	$-14\cdot2^{-n}$ $\frac{1}{5}(14+44\epsilon)\epsilon^n$

TABLE 5

	Total internal	Total junctions	External
1	1·00000	1·00000	—
2	3·00000	3·00000	—
3	3·50000	4·50000	1·00000
4	3·37500	5·25000	1·87500
5	3·00000	5·56250	2·56250
6	2·56250	5·56250	3·00000
7	2·14062	5·35938	3·11875
8	1·76562	5·03125	3·26562
9	1·44531	4·63281	3·18750
10	1·17773	4·20312	3·02539
11	·95703	3·76855	2·81152
12	·77637	3·34668	2·57031
13	·62915	2·94849	2·31934
14	·50952	2·58032	2·07080
15	·41248	2·24524	1·83276
	$-14\cdot2^{-n}$ $+\frac{1}{5}(14+44\epsilon)\epsilon^n$	$\frac{1}{5}(-2+8\epsilon)\epsilon^n$ $+\frac{n-1}{5}(6+16\epsilon)\epsilon^n$	$14\cdot2^{-n}$ $+\frac{1}{5}(6+16\epsilon).$ $(n-5+2\epsilon)\epsilon^n$

The total of junctions current in generation n is

$$\tfrac{1}{5}(-2+8\epsilon)\epsilon^n + \tfrac{1}{5}(2-8\epsilon)\left(\tfrac{1}{2}-\epsilon\right)^n$$

$$+\frac{n-1}{5}\left\{(6+16\epsilon)\epsilon^n+(14-16\epsilon)\left(\tfrac{1}{2}-\epsilon\right)^n\right\}.$$

If from this is subtracted the total of internal junctions, the difference representing at each stage the expected number of external junctions, is given by the general formula

$$14\cdot2^{-n}$$
$$+\tfrac{1}{5}(n-1)(6+16\epsilon)\epsilon^n+\text{conjugate}$$
$$+\tfrac{1}{5}(16+36\epsilon)\epsilon^n \qquad + \qquad ,,$$

but
$$16+36\epsilon = (6+16\epsilon)(4-2\epsilon)$$

so the expected number of external junctions may be written

$$14 \cdot 2^{-n}$$
$$\tfrac{1}{5}(n-5+2\epsilon)\,(6+16\epsilon)\epsilon^n + \text{conjugate term}.$$

TABLE 6

m		Probability e^{-m}
25	5·30832	·00495
26	4·48609	·01126
27	3·78430	·02272
28	3·18695	·04130
29	2·67973	·06858
30	2·25001	·10540
31	1·88669	·15157
32	1·58007	·20596
33	1·32176	·26667
34	1.10448	·33139
35	·92198	·39773
36	·76891	·46352
37	·64068	·52693
38	·53338	·58662
39	·44369	·64167
40	·36881	·69155
41	·30635	·73613
42	·25429	·77547
43	·21095	·80982
44	·17488	·83955
45	·14490	·86511
46	·11999	·88693
47	·09931	·90546
48	·08215	·92114
49	·06792	·93434
50	·05614	·94541
51	·04637	·95468
52	·03829	·96244
53	·03160	·96889
54	·02608	·97375
55	·02151	·97872
56	·01773	·98242
57	·01461	·98549
58	·01204	·98803
59	·00992	·99013
60	·00816	·99187

Note.—The distribution of the probability of elimination of the last heterogenic tract by generations would be very similar to that shown in Fig. 1 of Fisher (1954) for sib-mating, but about $1\tfrac{3}{4}$ generations earlier.

4. THE NUMBERS OF HETEROGENIC TRACTS

For a number of generations exceeding 20, only the leading terms need be considered. If L is the total length of strand, the number of external junctions expected will be

$$\tfrac{1}{5}(n-5+2\epsilon)\,(6+16\epsilon)\epsilon^n L$$

and if ν is the number of chromosomes, the number of chromosome ends still heterogenic will be

$$\tfrac{1}{5}(6+16\epsilon)\epsilon^n\,\nu.$$

The expected number of heterogenic tracts will be half the sum of these, or

$$m = \tfrac{1}{5}(3+8\epsilon)\,\epsilon^n\{(n-5+2\epsilon)L+\nu\}.$$

Putting, as is approximately appropriate for mice,

$$L = 25, \; \nu = 20,$$

it appears that at 25 generations m has fallen to about 5·3, and the probability that the whole material is homogenic, e^{-m}, is just under 1 per cent. The value up to $n = 60$ of m and of the derived probability, are shown in table 6.

The values shown in table 6 correspond closely to the corresponding probabilities for sib-matings published in 1954. The first quartile, 25 per cent., is passed between the 32nd and the 33rd generations, whereas with sib-mating it falls between the 34th and the 35th. Fifty per cent. is passed between generations 36 and 37 instead of between 38 and 39, while the third quartile, 75 per cent., is between generations 41 and 42, instead of 43 and 44. These comparisons suggest a difference of two generations, while a more accurate interpolation of the probabilities gives a difference of about 1·8 at 25 generations falling very slightly to about 1·6 at 60 generations, when in both cases complete homozygosis is almost certainly attained.

5. SUMMARY

For the comparatively simple system of inbreeding by alternate parent-offspring matings exact algebraic expressions can be found for the proportion of the germinal strands expected at each stage to be heterogenic, and for the expected number of junctions, external and internal, at each stage. The internal junctions are of four kinds, three of which are derived in a kind of cascade, requiring fairly careful analysis. It seemed therefore useful to explore the possibility of such an exact examination.

6. REFERENCES

BENNETT, J. H.   1953.   Junctions in inbreeding.   *Genetica XXVI*, 392-406.
FISHER, R. A.   1949.   *The Theory of Inbreeding.*   Oliver and Boyd Ltd., Edinburgh.
FISHER, R. A.   1954.   A fuller theory of "junctions" in inbreeding.   *Heredity, 8*, 187-197.