

A simple method for estimating recombination percentages and linkage intensities from F_2 data: examples from *Triticum monococcum* and other self-fertilizing diploid plant species

J. Kuspira and R. N. Bhambhani

Department of Genetics, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Accepted September 29, 1983 Communicated by K. Tsunewaki

Summary. We offer an alternative approach to the extensively used maximum likelihood and product methods for calculating recombination values and linkage intensities from F_2 data. This new method which we designate as the square root approach is simpler than the ones in current use in that it obviates the need for formulae and tables. It can be applied to autosomal F_2 data from F₁'s heterozygous in both the coupling and repulsion phases. It has greater applicability than the product method in that it can be used in all cases involving 2-, 3-, 4-, 6- and 9-class segregations regardless whether gene interaction occurs or not, provided the double recessive and other specific phenotypes are each determined by one particular genotype. The proposed method is based on the same well established genetic facts as the other two approaches. Percent recombinant gametes and therefore percent recombination are calculated by deriving the square root of the proportion of the F_2 population that expresses the double recessive or equivalent phenotype. The recombination values obtained by our method are compared with those derived by product method for 17 crosses in 7 different species and were found to be insignificantly different from the latter. The advantages and disadvantages of the square root method compared with the two most used ones are discussed.

Key words: F_2 data – Recombination values – Linkage – Square root method – Coupling – Repulsion – Self-fertilizer– *Triticum monococcum*

Introduction

The first method for the measurement of linkage intensities was presented by Bateson and Punnett in 1911. Since then various approaches have been proposed and utilized in the measurement of genetic distances between genes. In haploid eukaryotes, depending upon whether the products of each meiocyte in a given species are arranged in a linear or non-linear fashion, ordered and unordered tetrads respectively are analyzed to determine recombination values and linkage intensities (Ebersole 1956; Lindegren 1933, 1936; Papazian 1952; Pascher 1916; Perkins 1953). In cross-fertilizing species of diploid and polyploid eukaryotes, and where possible in self-fertilizing ones, testcross data are used for linkage analysis (Bateson et al. 1905; Morgan 1911a, 1911b; Punnett 1917). These two approaches provide direct information on the genetic kinds and proportions of F₁ gametes produced by dihybrids and therefore on recombination values. In self-fertilizing species however, it is often impossible and/or impractical to obtain testcross data. In such situations F_2 (and even F_3) results must be relied on to determine recombination values and linkage distances.

Several methods for calculating recombination and linkage values from F_2 data have been proposed and used (Bateson et al. 1905; Bateson and Punnett 1911; Collins 1912; Emerson 1916; Fisher 1928; Fisher and Balmukand 1928; Haldane 1919; Oweven 1928; Wellensiek 1927). However, only the maximum likelihood (Haldane 1919) and product (Fisher and Balmukand 1928) methods have been found to be efficient and are widely used. Allard (1956), Immer (1930), Immer and Henderson (1943) and Stephens (1939) have provided formulae and tables to facilitate the calculation of recombination values, using these latter two approaches.

We offer yet another approach to calculating recombination values and linkage distances from F_2 results. The method proposed is simpler than the ones in current use in that it circumvents the need for formulae and tables. In addition it is useful in many cases where gene interaction is involved as well as in all cases where the different pairs of alleles determine the expression of different pairs of traits.

The square root method for estimating linkage intensities from F_2 data: theory and method

1 Theory

The proposed method is based on well established genetic facts:

(a) Gametes unite at random.

(b) The likelihood of occurrence of a given F_2 genotype is the product of the probabilities of occurrence of the combining F_1 male and female gametes.

(c) If two pairs of alleles determine the expression of different pairs of traits, regardless of the specific allelic relationship at each locus there is only one F_2 phenotype, the double recessive one, that is always specified by one particular F_2 genotype, the one homozygous for recessive alleles at both loci.

(d) Unless there is evidence to the contrary, the frequency of crossing over is the same in both male and female meiocytes of the species under investigation.

(e) A single reciprocal exchange (crossing over) event in dihybrids between two pairs of loci in a meiocyte results in two complementary recombinant type gametes.

Cross in Coupling

2 Method

If all assumptions are correct, to determine percent recombination simply derive the square root of the proportion of the F_2 population that expresses the double recessive phenotype.

(a) If the cross is in coupling – this derived value (which is designated x), e.g., 0.40 represents the proportion of F_1 male and female gametes with the parental genotype carrying recessive alleles at both loci. The proportion of the F₁ male and female gametes with the corresponding parental genotype carrying dominant alleles at both loci will have the same frequency (0.40). The recombinant gametes would constitute the remaining fraction of F_1 gametes, e.g., 1-2x=0.20, and since both complementary recombinant genotypes are products of the same reciprocal exchange event they will occur in equal proportions, e.g., 0.10 and 0.10. This derived fraction of gametes indicates percent recombination and represents the genetic distance between linked genes. For example, if the F_1 dihybrids *BP/bp* produced parental BP and bp and recombinant Bp and bPgametes in a ratio of 0.4:0.4:0.1:0.1, percent recombination and map distance is 0.20 (20%).





Fig. 1. Calculation of percent recombination from F_2 progenies from dihybrid F_1 's in coupling and repulsion for the partially linked allele pairs *Bb* and *Pp* in *Triticum monococcum*

(b) If the cross is in repulsion the value x obtained by calculating the square root of the proportion of the F_2 population that expresses the double recessive phenotype indicates the proportion of recombinant F_1 gametes that are double recessive in both F1 male and female meiocytes, e.g., 0.1. Of the remaining fraction (0.9) of the F₁ gametes, the corresponding double dominant F₁ gamete type would be present with the same frequency, e.g., 0.1, as the recombinant double recessive gamete, for reasons stated above. The remaining fraction (0.8) of F_1 gametes would consist of the two parental types dominant-recessive and recessive-dominant in equal frequencies. For example, if the dihybrid $F_1 Bp/$ bP produced parental Bp and bP and recombinant BPand bp gametes in a ratio of 0.4:0.4:0.1:0.1, the percent recombination and map distance between the genes would be 2x = 0.2 (20%).

3 Application of method

The calculation of percent recombination from F_2 data is illustrated in Fig. 1 which shows unpublished personal F_2 results from dihybrid F_1 's in both coupling and repulsion phases for the two partially linked pairs of alleles *Bb* (black vs. white glume) and *Pp* (pubescent vs. glabrous glumes) in *Triticum monococcum* which are approximately 20 map units apart.

4 Reliability of proposed method

The reliability of our method for calculating recombination values is compared with the product approach for actual data from several authors and seventeen crosses and given in Table 1.

The product method (Fisher 1928; Fisher and Balmukand 1928) which is as efficient as the maximum likelihood one for calculating linkage intensities is relatively simple and straightforward when appropriate tables are available. The T. monococcum example discussed previously (Fig. 1) will be used to indicate how recombination values are calculated using this approach. Since one allele is completely dominant to the other at each of the two loci, e.g., B is dominant to b and P is dominant to p and F_2 populations from self-fertilization or intercrossing F_1 's heterozygous in the same phase (coupling or repulsion) will consist of four phenotypic classes: double dominant determined by B-P-, dominant-recessive specified by B-pp, recessive-dominant given by bbP- and double recessive determined by *bbpp* which are designated a, b, c and d respectively. The observed frequencies for the four classes are then substituted in appropriate formulae to determine the p value (Allard 1956; Emerson 1916; Fisher and Balmukand 1928; Immer 1930; Owen 1928).

In repulsion p is the percent recombination and in coupling 1-p is the recombination percentage, ex-

pressed in decimal fractions. Tables have been calculated by Immer (1930); also see Allard (1956) giving the values of the ratio of the products ad/bc (for repulsion) and bc/ ad (for coupling) for different values of p, thus significantly reducing the labor involved in determining linkage values. The determination of linkage intensities by the product method simply resolves itself into calculating the ratio of products from F₂ data and finding the recombination percentage by interpolation in the appropriate table, a small sample of which is given in Table 1. The probable errors are obtained by dividing the probable error factor corresponding to the calculated recombination value by the square root of the number of F₂ individuals designated by N.

As an example of the use of the product method, consider the F_2 data from the *T. monococcum* F_1 's heterozygous in the coupling phase (Fig. 1). The ratio of products is $bc/ad = 15 \times 20/134 \times 31 = 0.07222$. By interpolation in Table 2 we find the recombination value to be 0.195 (19.5%). From the same table, by interpolation we find that the proper factor for the probable error is 0.3039 which, divided by $\sqrt{200}$ gives a probable error of the above linkage intensity of 0.021 or 2.1% (19.5±2.1).

Discussion

Unlike tetrad and testcross data, F_2 results provide indirect information on the kinds and proportions of parental and recombinant type gametes produced by dihybrid F_1 's. Because analysis of F_2 results is more difficult, tedious and time consuming, whenever possible, testcross data is analyzed to determine recombination and linkage values. However, in self-fertilizing species (rarely in cross-fertilizing ones) it is often impossible and/or impractical to obtain testcross data to calculate percent recombination and F_2 (and even F_3) results must be relied on for this purpose. The most useful methods for estimating recombination values using F_2 data are the product and maximum likelihood methods. The product method is relatively simple to apply with tables such as that in Table 2. The method is limited to F_2 's where four phenotypic classes occur. The method of maximum likelihood is not as easy to apply as the product one, but is quite general and may be applied to all testcross, F_2 and F_3 genetic data.

The square root method, proposed in this paper, is, in our opinion, also a satisfactory approach to calculating recombination values and linkage intensities. Moreover, unlike the product and maximum likelihood methods, it does not necessitate the use of formulae and tables. The conditions for its use are the same as those for the two currently used approaches. Like the other two methods, the proposed one can be applied to

-dib	(
lizing)
lf-ferti	
ent se	
ı diffeı	
I sever	
leles ir	
s of al	
17 pair	•
d for	
metho	
coduct	
and pu	•
nethod	
root m	
quare	•
osed s	
e prop	•
1 by th	•
derive	
alues	
ation v	
ombin	
of rec	
oarison	s
A com	specie
ole 1. /	l plant
Tab	loid

64

									Percent	recombin	ation	
Species	Phase of	Cross	F.	F_2					Product	method	Square	Source of
	6010			ង	q	ა	d D	rotal(N)	Value	Probable error	method	1 2 uala
1. Pisum sativum	Coupling	Round Yellow X Wrinkled Green RR YY rr yy	Round Yellow Rr Yy	315	101	108	32	556	51.10	<u>+</u> 2.2	52.02	Mendel (1866)
2. Hordeum vulgare	Repulsion 7	Vormal, male-sterile × Liguleless, male fertile Li Li, ms3 ms3 li li, Ms3 Ms3	<i>Normal, male-fertile</i> Li li, Ms ₃ ms ₃	126	43	39	12	220	48.50	<u>+</u> 3.5	46.71	Kasha and Walker (1960)
3. Brassica oleracea	Coupling	Pubescent, dark green × Glossy, pale green Hr-2 Hr-2, Pg-1 Pg-1 hr-2 hr-2, pg-1 pg-1	Pubescent, dark green Hr-2 hr-2, Pg-1 pg-1	279	98	16	32	500	49.98	<u>+</u> 2.3	49.40	Sampson (1978)
4. Lycopersicon esculentum	Repulsion	Narrow leaf, wiry × Broad leaf, normal EE wi wi ee Wi Wi	Narrow leaf, normal Ee Wi wi	161	263	346	57	1457	40.15	<u>+</u> 1.5	39.55	Butler (1963)
5. Brassica oleracea	Coupling	Pubescent ^a X Pubescent ^a Hr-2 Hr-2, Hr-1 Hr-1 hr-2, hr-1 hr-1	<i>Pubescent</i> Hr-2 hr-2, Hr-1 hr-1	549	131	178	77	935	41.79	<u>+</u> 1.5	42.61	Sampson (1978)
6. Hordeum vulgare	Repulsion	Normal, absent ^b X "Uzu" present ^b Uz Uz, als als uz uz, Als Als	Normal, present Uz uz, Als als	150	99	67	×	291	32.69	<u>+</u> 3.5	33.16	Kasha and Walker (1960)
7. Hordeum vulgare	Coupling	Normal, six-row X Male-sterile, two-row Ms Ms VV ms ms vv	<i>Normal, six-row</i> Ms ms Vv	954	169	176	178	1477	27.85	± 0.95	30.58	Kasha and Walker (1960)
8. Hordeum vulgare	Repulsion	Covered, long ^e × Naked, short ^e NN lb ₃ lb ₃ nn Lb ₃ Lb ₃	Covered, short Nn Lb3 lb3	643	363	319	16	1341	20.03	+1.8	21.84	Kasha and Walker (1960)
9. Lycopersicon esculentum	Coupling	<i>Normal, narrow leaf X Divergens, broad leaf</i> Di Di EE di di ee	<i>Normal, narrow</i> 1 Di di Ee	(851	279	260	415	2805	21.79	+ 0.6	23.07	Butler (1963)
10. Triticum monococcum	Coupling	Black, pubescent glumes × White, glabrous glumes BB PP bb pp	Black, pubescent Bb Pp	134	15	20	31	200	19.50	±2.1	21.26	Personal unpublished data
11. Triticum monococcum	Repulsion	Black, glabrous × White, pubescent BB pp bb PP	Black, pubescent Bb Pp	101	45	53	7	201	19.60	+4.5	19.96	Personal unpublished data
12. Triticum monococcum	Repulsion	Compact, awned × Lax, awnless C2 C2, ga ga c2 c2, Ga Ga	<i>Compact, awnless</i> C ₂ c ₂ , Ga ga	491	198	255	4	948	13.62	<u>+</u> 2.1	13.0	Moseman and Smith (1954)

J. Kuspira and R. N. Bhambhani: Recombination estimation from F_2 data

									Percent 1	ecombina	ation	
Species	Phase of	Cross	F	F_2					Product	method	Square	Source of
	CI (155			а	p	c	d T	otal(N)	Value 1	Probable error	method	I'2 uata
13. Lathyrus odoratus	Coupling	Purple, long × Red, short PP LL pp ll	Purple, long Pp Ll	4831	390	393	1338	6952	12.00	±0.3	12.26	Batcson <i>et</i> <i>al</i> (1905) Punnett (1917)
14. Hordeum vulgare	. Coupling	Solid, normal X Variegated, ribbon grass Va Va, Rb Rb va va, rb rb	<i>Solid, normal</i> Va va, Rb rb	379	6	14	123	525	4.33	+0.6	3.20	Walker et (1963)
15. Triticum monococcum	Coupling	Normal, green × Creamex, yellow Cx Cx YY cx cx yy	Normal, green Cx cx Yy	1432	31	48	479	1990	3.90	±0.3	1.88	Moseman and Smith (1954)
16. Hordeum vulgare	Repulsion	Green, long × Glossy, short Gl ₂ Gl ₂ , lb ₂ lb ₂ b ₂ Lb ₂ Lb ₂	<i>Green, short</i> Gl ₂ gl ₂ , Lb ₂ lb ₂	76	37	34	0	147	— (1.5) ^d		!	Kasha and Walker (1960)
17. Lycopersicon esculentum	Coupling	Short, beaked sepal × Long. non-beaked sepal Sx Sx, En En sx sx, en en	Intermediate, beakea Sx sx, En en	f shor beaked be 59 1	int long aked beah 02 23	shori ced non- beaked b 2	int loi non-nc aked bea 22 4:	250 ked	20.17°	± 2.36	18.02	Bouwkamp and Honm (1971)

J. Kuspira and R. N. Bhambhani: Recombination estimation from F₂data

Table 2. Product method formulae (ratio of products), their associated recombination fractions, and factors to be divided by N to obtain probable errors

Recombi-	Ratio of pr	roducts	Factor to be \sqrt{N} to abte	e divided by
fraction	ad/bc	bc/ad	error	in probable
	repuision	couping	F ₂ repulsion	F₂ coupling
0.00	0.000000	0.000000	0.000000	0.000000
0.05	0.005031	0.003629	0.6724	0.1515
0.10	0.02051	0.01586	0.6662	0.2153
0.15	0.04763	0.03915	0.6560	0.2651
0.20	0.08854	0.07671	0.6422	0.3079
0.25	0.1467	0.1328	0.6253	0.3464
0.30	0.2271	0.2132	0.6055	0.3820
0.35	0.3377	0.3259	0.5833	0.4153
0.40	0.4898	0.4821	0.5592	0.4469
0.45	0.7013	0.6985	0.5333	0.4771
0.50	1.0000	1.0000	0.5059	0.5059
0.55	1.4317	1.4260	0.4771	0.5333

autosomal F_2 data from F_1 's heterozygous in both the coupling and repulsion phases. Unlike the product method which is limited to F_2 data where 4-class segregations occur, our method can be applied to all cases involving 2-, 3-, 4-, 6- and 9-class segregations regardless whether gene interaction occurs or not provided the double recessive and other specific phenotypes are each determined by one particular genotype. For example, if one allele is completely dominant to the other at both loci or if this relationship occurs at one locus and incomplete dominance of one allele over the other or codominance occurs at the second locus, the double recessive or equivalent phenotype can be used to estimate percent recombination. If incomplete dominance and/ or codominance occurs at both loci then each of the 4 different phenotypes each specified by one specific homozygous genotype can be used for this purpose. This is also true for those forms of gene interaction such as dominant epistasis, duplicate dominant epistasis (duplicate genes) and cumulative complementary dominant alleles of different genes which give rise to an F₂ recessive phenotype which is always determined by one specific (double recessive) genotype.

Like the product and maximum likelihood methods, the proposed approach cannot be used to estimate recombination values when the frequency of the double recessive or equivalent phenotype is 0. In such cases one must resort to analysis of F_3 results. Obviously, the value of 0 itself is indicative of close linkage. Thus, although the square root method does not have as general an applicability as the maximum likelihood method it can be used in a greater variety of cases than the product method.

Neither the square root nor the other two methods can be used to calculate percent recombination and linkage values from F_2 data for sex (X, Z)-linked genes and autosomal genes if there is no crossing over in meiocytes of either or both sexes. For example, there is no crossing over in Drosophila melanogaster males (Morgan 1912, 1914). In such organisms F_2 constitutes a testcross with respect to sex-linked genes. Half of the F₂ progeny, the heterogametic individuals, can be used for direct estimation of percent recombination if the dihybrid F₁ homogametic parent is mated with a heterogametic individual that does not carry recessive alleles at both loci. The entire F_2 (both homogametic and heterogametic individuals) constitutes a testcross if the heterogametic F_1 parent carries recessive alleles at both loci. If there is no crossing over in at least one of the sexes, F₂ results from dihybrid F₁'s heterozygous in both coupling and repulsion phases can be used to determine whether genes are syntenic or not and those from dihybrid F_1 's in coupling can also be used to determine whether linkage exists and if so the degree thereof. This is possible because half of the F_2 constitutes a testcross progeny with respect to the genes being studied since all these individuals receive a gamete with recessive alleles at both loci from the parent in which crossing over does not occur.

For genes that are closely linked the proposed approach is more accurate in establishing percent recombination in coupling than in repulsion phase since the chance of obtaining the expected number of double recessives is greater than in repulsion. Also if the progeny size is small, the square root method may not provide as accurate a measure of amount of recombination as the other two approaches since a small deviation from the expected number of double recessives is proportionally large and can lead to a relatively large deviation from the actual percent recombination.

In conclusion, the proposed method, since it utilizes only one F_2 phenotypic class, may appear to be a less reliable method than the conventional ones which use all four F₂ phenotypic classes in estimating recombination values. Nonetheless, its applicability, dependability and accuracy, particularly in large F₂ populations, are demonstrated by the similarities of the recombination values to those using the product method as shown in Table 1. Like other methods, the one being proposed here also suffers from limitations inherent in small F₂ populations. However, for a given F₂ population, depending on the phenotypic class(es) whose value(s) deviate from expected, and the direction and extent of deviation, either the product and maximum likelihood methods or the proposed square root approach can provide a more accurate estimate of percent recombination.

J. Kuspira and R. N. Bhambhani: Recombination estimation from F₂ data

References

- Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278
- Bateson W, Punnett RC (1911) On gametic series involving reduplication of certain terms. J Genet 1:293-302
- Bateson W, Saunders ER, Punnett RC (1905) Experimental studies in the physiology of heredity. Rep Evol Commun Soc 2:1-55 and 80-99
- Bouwkamp JC, Honma S (1971) Estimation of linkage between several genes in the tomato. J Hered 62:37–40
- Butler L (1963) Five genes located on chromosome 4 of the tomato. Can J Bot 41:1159–1164
- Collins GN (1912) Gametic coupling as a cause of correlation. Am Nat 46:569–590
- Ebersole RA (1956) Biochemical mutants of *Chlamydomonas* reinhardi. Am J Bot 43:404-407
- Emerson RA (1916) The calculation of linkage intensities. Am Nat 50:411-420
- Fisher RA (1928) Statistical methods for research workers, 2nd edn. Oliver and Boyd, Edinburgh, pp 291-324
- Fisher RA, Balmukand B (1928) The estimation of linkage from the offspring of selfed heterozygotes. J Genet 20:79-92
- Haldane JBS (1919) The probable errors of calculated linkage values, and the most accurate method of determining gametic from certain zygotic series. J Genet 8:291–297
- Immer FR (1930) Formulae and tables for calculating linkage intensities. Genetics 15:81–98
- Immer FR, Henderson MT (1943) Linkage studies in barley. Genetics 28:419-440
- Kasha KJ, Walker GWR (1960) Several recent barley mutants and their linkage. Can J Genet Cytol 2:397-415
- Lindergren CC (1933) The genetics of Neurospora. 3. Pure bred stocks and crossing over in Neurospora crassa. Bull Torrey Bot Club 60:133-154
- Lindegren CC (1936) A six-point map of the sex chromosome of *Neurospora crassa*. J Genet 32:243–356

- Mendel G (1866) Versuche über Pflanzenhybriden. Verh Naturforsch, Verh Brünn Abhandlungen, IV 3047; translated and reprinted in Peters JA (ed). Classic papers in genetics. Prentice-Hall, Englewood Cliffs NJ, pp 1–20
- Morgan TH (1911a) Random segregation versus coupling in Mendelian inheritance. Science 34:384
- Morgan TH (1911b) An attempt to analyse the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila*. J Exp Zool 11:365–414
- Morgan TH (1912) Complete linkage in the second chromosome of the male of *Drosophila*. Science 36:719–720
- Morgan TH (1914) No crossing over in the male of Drosophila of genes in the second and third pairs of chromosomes. Biol Bull 26: 195–204
- Moseman JG, Smith L (1954) Gene location by three-point test and telocentric half-chromosome fragment in *Triticum* monococcum. Agron J 46:120-124
- Owen FV (1928) Calculation of linkage intensities by product moment correlation. Genetics 13:80-110
- Papazian HP (1952) The analysis of tetrad data. Genetics 37:175-188
- Pascher A (1918) Über die Beziehung der Reduktionsteilung zur Mendelschen Spaltung. Ber Dtsch Bot Ges 36:163
- Perkins DD (1953) The detection of linkage in tetrad analysis. Genetics 38:187-197
- Punnett RC (1917) Reduplication series in sweet peas. 2. J Genet 6:185–193
- Sampson DR (1978) A second gene for hairs in Brassica oleracea and its tentative location in linkage group 4. Can J Genet Cytol 20:101–109
- Stephens WL (1939) Tables of the recombination fraction estimated from the product ratio. J Genet 39:171-180
- Walker GWR, Dietrich J, Miller R, Kasha K (1963) Recent barley mutants and their linkage. 2. Genetic data for further mutants. Can J Genet Cytol 5:200–219
- Wellensiek SJ (1927) Methods for calculating the actual gametic F_2 series from a given zygotic series. Genetica 9:329-340