A Mathematical Model for Solving Four Point Test Cross in Genetics

Jugal Gogoi
Department of Mathematics
Dibrugarh University
Dibrugarh, Assam, India

Tazid Ali
Department of Mathematics
Dibrugarh University
Dibrugarh, Assam, India

ABSTRACT
During meiosis cell division, chromosomes are replicated and form a tetrad. While the chromosomes are lined up close to one another, they have the opportunity to interact with one another, get tangled, and recombine to form new combinations of alleles. Four point cross refers to using 4 points (genes) to determine the order and distance between the genes. This test cross is also the method of choice in determining linkage in organisms with many genetic markers. In this paper an attempt is made to solve the four point test cross problem.

Keywords
Gene, Meiosis, crossing-over, linkage, recombination, three point test cross, four point test cross.

1. INTRODUCTION
The principle of independent assortment [1] in genetics assumes that different genes followed in patterns of inheritance analysis are on different chromosomes. Mendel studied seven traits [2] in peas: each trait is now known to be located on separate chromosomes. (Mendel's work was rediscovered by Bateson [3] when his research gave progeny ratios that Bateson could not interpret. While Mendel's work confirmed independent assortment ratios, it was the work of Morgan that finally provided an explanation for deviation from dihybrid ratios). By solving a four point cross we can determine two important things:

1. Order of the genes on a chromosome.

2. The distance (in map units) between each pair of genes.

A successful finding of map distance (i.e., distance between two genes) can also help for findings the probability of recombination [4].

2. DEFINITIONS AND ASSUMPTIONS
2.1. Gene
A gene is a hereditary factor [5] that had two or more alleles which determined the difference between two or more alternative phenotypes. Different genes controlled different aspects of phenotype. Different genes could be separated by recombination. So gene is the unit of recombination also. These units usually agreed with each other, until genetic analysis was extended to bacteria and viruses in which rare genotypes can be selected and detected. Mutations [6] within a gene, even in adjacent base pairs, can be separated by recombination. Two different genes identified by recombination or mutation may control the same phenotype. Genes on the same chromosome are physically connected or linked. Gene pairs that are close together on the same chromosome are genetically linked because the alleles on the same homolog are transmitted together (parental types) into gametes more often than not during meiosis [7]. Genes may assort independently either because they are on different chromosomes or because they are far apart on the same chromosome.

2.2. Recombination Frequency
Recombination frequency is a measure of genetic linkage [7], [8] and is used in the creation of a genetic linkage map. Recombination frequency (denoted by θ) is the frequency with which a single chromosomal crossover will take place between two genes during meiosis. A centimorgan (cM) is a unit that describes a recombination frequency of 1%. In this way we can measure the genetic distance between two loci, based upon their recombination frequency. Double crossovers (or even numbers of crossovers) would turn into no recombination. In this case we cannot tell if crossovers took place. If the loci we're analyzing are very close (less than 7 cM) a double crossover is very rare. When distances become higher, the likelihood of a double crossover increases. During meiosis, chromosomes assort randomly into gametes, such that the segregation of alleles of one gene is independent of alleles of another gene. This is stated in Mendel's Second Law [8] and is known as the law of independent assortment. The law of independent assortment always holds true for genes that are located on different chromosomes, but for genes that are on the same chromosome, it does not always hold true.

2.3. Recombination Mapping
Recombination frequencies may be used to map the position of genes (loci) on linear linkage groups. The order of genes and the relative distances between them in a linkage group corresponds to their order and relative distances on a chromosome. Recombination can be used to map genes. Morgan hypothesized that the further apart two genes are, the more likely they are to have a crossover between them. Sturtevant [10] worked out method to map genes on chromosomes. Genes far enough apart on same chromosome can show recombination frequencies of 50%. Without addition all information, one cannot distinguish these cases from loci on independently segregating chromosomes. Map function is a mathematical formula that relates the observed number of crossovers to the real number, which is a function of the physical distance [11]. It assumes a simple model in which crossovers are distributed randomly on the chromosome. The maximum recombination frequency is 50%: as distance increases, crossovers between any two markers increases, but some of these are double and other even-numbered crossovers. The recombination frequencies of pairs of genes indicate how often 2 genes are transmitted together. For linked genes, the frequency is less than 50%. The greater the distance between linked genes, the higher the recombination frequency ($r.f$). Recombination frequencies

become more inaccurate as the distance between genes increases.

The relationship between relative recombination frequency and distance is used to create genetic maps. The greater the density of genes on the map, and the smaller the distances between the genes results more accurate the map.

2.4. Existing Method for solving Three-Point Test Crosses

Step 1. In a three point cross first rewrite the classes of progeny (data) assigning genotypes to each trait.

Step 2. The offspring are generated by a heterozygous individual. Therefore, all classes (recombinant, parental, etc.) will occur as reciprocal pairs of progeny. These reciprocal pairs will be both genetic reciprocals and numerically equivalent.

Step 3. Designate the different gametes or offspring as noncrossover (parental), single crossover or double crossover. The noncrossover classes are those classes of progeny who have one of the intact, non-recombinant homologs from the parent. The noncrossover classes will be represented by the greatest numbers of offspring. The double crossover classes will be represented by the smallest numbers of offspring. Sometimes one or both double crossover classes are missing because they are rare.

Step 4. By examining the pattern of data seen in a problem, one can often start solving the problem with a basic understanding of the linkage relationships of the genes. Some of the more common patterns of data are:

- 3 linked genes give 8 classes of data that occur as 4 reciprocal pairs genetically and numerically;

- 3 unlinked genes gives 8 classes of data that occur as 4 genetically reciprocal pairs, but all classes are seen in a 1:1:1:1:1:1:1:1 ratio;

Step 5. Begin the process of mapping the genes by ordering the genes. To figure out which gene is in the middle of a group of three genes, choose one of the double crossover classes. Compare it to the most similar parental class of progeny where two of the three genes will have the same combination of alleles. The gene that differs is the gene in the middle.

Test problem 1

To draw a map of these 3 genes ($vg, b,$ and pr) showing the distances between all pairs of genes of the following:

Drosophila autosomal genes:

vg = vestigial wings; $vg+$ = normal

b = black body; $b+$ = normal body

pr = purple eyes; $pr+$ = normal eyes.

GAMETE OBSERVED	FREQUENCY
vg b pr	1779
vg b+ pr	118
vg+ b pr	252

vg+ b+ pr	9
vg+ b pr+	131
vg+ b+ pr+	1654
vg b pr+	13
vg b+ pr+	241

Where,

- All three alleles that are on a single chromosome are written together.

- SCO- single crossover

- DCO- double crossover

- First, the offspring need to be arranged into reciprocal pairs of no crossovers, SCO, and DCO as follows:

vg b pr	1779
vg+ b+ pr+	1654
vg+ b pr	252
vg b+ pr+	241
vg+ b pr+	131
vg b+ pr	118
vg b pr+	13
vg+ b+ pr	9

- Because there is a higher probability of having no crossovers, the two genotypes with the highest number of offspring are NCO (no crossover). There are two sets of SCO, and the two with the lowest offspring will be DCO.

Next we need to determine which order of alleles on the chromosome will result in the DCO.

- Here the no crossovers are (i.e., highest number of offspring)

$$vg \quad b \quad pr$$
$$vg+ \quad b+ \quad pr+$$

- And the DCO are (i.e., lowest number of offspring)

$$vg \quad b \quad pr+$$
$$vg+ \quad b+ \quad pr$$

- In order for this to be a double crossover, the pr allele must be in the center, i.e., they are as follows:

$$vg \quad pr \quad b$$

3. FOUR-POINT CROSSES

A four point test cross can reveal map order [Figure 1]. A successful approach is to analyze results by calculating recombination frequencies per gene pair. If the parental cross is not known, then the parental types are those progeny classes that are greater than 50% of the total. The occurrence

of one cross-over event in one region of a chromosome decreases the likelihood of another cross-over in an adjacent region. This is known as interference [12], and can be calculated based on observed and expected recombination frequencies.
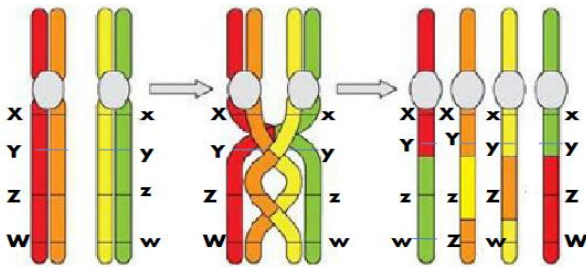

Figure 1: Arrangement of four genes (or points) in the chromosomes

It is also known that mapping distances are not linear with recombination frequencies over large chromosomal distances. A mathematical expression, called a mapping function, is used to correct for this. Genetic linkage in humans is determined by the frequency of recombination between two traits that are passed from parent to child. The traits used as markers have been limited to observable phenotypic characteristics until recently. Traits used today include chromosome bands, RFLPs [13], and other physical landmarks generated through recombinant DNA techniques [14].

3.1 Algorithm for solving Four-Point Test Cross

Step1. We have to first calculate the map distances (recombination frequency) between each of the two gametes by analyzing as two-factor crosses.

Step2. Compare the distances and arrange them according to their distances.

Test Problem 2

Four point testcross data from the cross of **Gl Lg A Df//gl lg a N** are (downloaded from www.agron.missouri.edu/mnl/47/41 Rhoades.pdf, dated 01/12/2014).

GAMETE OBSERVED	FREQUENCY
Gl Lg A Df	63
gl lg a N	249
Gl lg a N	134
gl Lg A Df	41
Gl Lg a N	174
gl lg A Df	58
Gl Lg A N	59
gl lg a Df	37
Gl lg A Df	19
gl Lg a N	75

GAMETE	FREQUENCY
Gl lg a Df	22
gl Lg A N	41
Gl Lg a Df	12
gl lg A N	8
Gl lg A N	8
gl Lg a Df	4
	Total = 1004

Now analyze data as two-factor crosses:

Gl-Lg	Gl-A	Gl-Df	Lg-A	Lg-Df	A-Df
134	134	134	174	174	59
41	41	41	58	58	37
19	174	174	19	59	22
75	58	58	75	37	41
22	22	59	12	19	12
41	41	37	8	75	8
8	12	8	8	22	8
4	8	4	4	41	4
Total =344	total =490	Total =515	Total = 358	Total =485	Total =191

The distance between $Gl - Lg$ is $(344/1004) = .343$
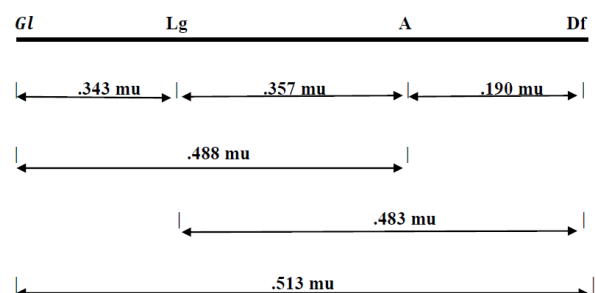
The distance between $Gl - A$ is $(490/1004) = .488$

The distance between $Gl - Df$ is $(515/1004) = .513$

The distance between $Lg - A$ is $(358/1004) = .357$

The distance between $Lg - Df$ is $(485/1004) = .483$

The distance between $A - Df$ is $(191/1004) = .190$

Longest distance is $Gl - Df$, so these are the outside markers on the map. According to their distances the map should be as below:

4. CONCLUSION

The techniques for finding the mapping of three point testcross cannot be applied in case of four point testcross, as there is more than one gene in the middle, so a new approach for finding the same for four point test cross is needed. By this approach the reader can not only map the four point test cross but also map the n-point test cross data and hence generalize further. It is also verified for three point test cross (the method has been verified for a large number of three point test cross problems including the above test problem 1). The main advantage of this approach is that, this method does not require any arrangement of the progeny data as DCO, SCO or NCO types. In the prior method, when there are equal numbers of progeny, there arises a mystification to judge which are DCO, which are SCO or which are NCO. So this method is free from this mystification. But it should also note that as even number of crossovers (2, 4….etc) returns no detectable recombination, so the map distances are not additive for the large number of distances. This is one disadvantage of this method. But this can   be overcome as practically it happens very rare.

5. REFERENCES

[1] Balzer, W. and Dawe, C.M. 1990. Models for Genetics. Peter Lang publishing co, New York. First edition. Chapter 1.

[2] Hartl, D.L et al. 2001. Genetics: Analysis of Genes and Genomes, Jones & Bartlett Publishers. 8th edition. Chapter 1.

[3] Bateson, W. and Punnett, R.C. 1905. Experimental Methods in the Physiology of Heredity, Report to the Evolution Committee Royal Society II, London: Harrison and Sons.

[4] Etheridge, A. (2012): *Some Mathematical Models from Population Genetics.* Lecture notes in Mathematics. Springer publication, Chapter 2.

[5] Klug, W.S., Sarah, M. W. 2009. Concepts of Genetics. Pearson Benjamin Cummings, San Francisco, pp.113-117.

[6] Hastings, P. et al.2009. Mechanisms of change in gene copy number. Nature Reviews Genetics. London, Nature Publishing Group. Vol. 10, pp.551-564.

[7] Terwiliger, Joseph, D. 1994. Handbook of human genetic linkage. The John Hopkins University Press. 4th edition.

[8] Griffiths, AJF. et al. 1993. An Introduction   to Genetic Analysis. 5th edition. New York: W.H. Freeman and Company.

[9] Van der Vaart, A.W. 2006. Statistics in Genetics. Fall publication. 1st edition, chapters 1,2,11.

[10] Sturtevant, A.H. and Morgan, T.H., 1923. Reverse Mutation of the Bar Gene Correlated with Crossing Over, Science 57, pp.746-747.

[11] Gogoi, J., Ali, T., Sarma, R. 2013. Stochastic Model for Genetic Recombination because of Crossing over and Chromatid Exchange. International Journal of Computer Applications, Vol.80.pp.1-3

[12] Kenneth, L. 2003. Mathematical and statistical methods for genetic analysis, Springer Publication, 2nd edition.

[13] Rao, C. R. 1973. Linear Statistical Inference and its Applications, 2nd edition. Wiley, New York.

[14] Gusfield, D. 2014. Recombinatories. The MIT press, Cambridge London. 1st edition. Chapter 3.