IJIRSET

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Estimation of Genetic Recombination Frequency with the Help of Logarithm Of Odds (LOD) Method

Jugal Gogoi [1], Tazid Ali [2]

Research Scholar, Department of Mathematics, Dibrugarh University, Assam, India[1]

Professor, Department of Mathematics, Dibrugarh University, Assam, India[2]

ABSTRACT: In living beings, all characters are controlled by genes, each of which is a segment of DNA (Deoxyribonucleic acid) in a chromosome. Huge numbers of genes are present per chromosome. Thus, many of the genes are linked and they tend to segregate together into the same cell during cell division. To identify association between different characters, it is important in many cases to estimate how far genes are apart in the same chromosome. This is the problem of estimating the linkage proportion or recombination fraction for linked loci. In this paper, we will illustrate how to estimate the linkage between two loci in the same chromosome with the help of LOD (Logarithm of Odds) score for 6 phenotypes

KEYWORDS: Genes, chromosome, linkage, recombination fraction, phenotypes, LOD score.

I. INTRODUCTION

Genetic linkage is the tendency of genes that are located proximal to each other on a chromosome to be inherited together during meiosis. Genes whose loci are nearer to each other are less likely to be separated onto different chromatid during chromosomal crossover, and are therefore said to be genetically linked [1].The definition of linkage is that the proportion of recombinant classes is less than 50% indicates linkage.

During the past years numerous methods have been developed to test for the occurrence of recombination and linkage, to identify the parental and recombinant individuals, and to determine the location of the recombinational break-points. These techniques differ in approach and applicability, but may be classified into four nonexclusive general categories:

1. Distance-based[2,3]
2. Substitution distribution-based[4],
3. Compatibility-based[5,6] and
4. Phylogenetic-based[7],

Much of the biological researches on recombination and linkage are based on experimental data analyzed using certain statistical measures. However, presently mathematical models are widely been used for understanding of biological processes and in linkage analysis.

II. DEFINITION AND ASSUMPTIONS FOR THE MODEL

a. Recombination frequency

Recombination frequency is a measure of genetic linkage [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

Copyright to IJIRSET

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

this way we can measure the genetic distance between two loci, based upon their recombination frequency. Double crossovers would turn into no recombination. In this case we cannot tell if crossovers took place. If the loci we are 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 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.

b. LOD score method for estimating recombination frequency

The LOD score (logarithm (base 10) of odds), developed by Newton E. Morton [9], is a statistical test can be used for linkage analysis in human, animal, and plant populations. The LOD score compares the likelihood of obtaining the test data if the two loci are really linked, to the likelihood of observing the same data purely by chance. Positive LOD scores favor the presence of linkage, whereas negative LOD scores indicate that linkage is less likely. Computerized LOD score analysis is a simple way to analyze complex family pedigrees in order to determine the linkage between Mendelian traits (or between a trait and a marker, or two markers).
The method is described in detail as follows:

1. Start with a model of inheritance for the gene of interest, and work out an equation that gives the expected frequency of various types of offspring given an arbitrary value of θ.
2. Then, using a form of the binomial expansion, we can determine the likelihood of our data (family) at a number of different values of θ.
3. Then, determine the odds (likelihood ratio), i.e., likelihood at each value of θ divided by the likelihood at $\theta = 0.5$ (unlinked).
4. Then, take the base 10 logarithm of the odds ratio. This is the log of the odds, the LOD score for each value of θ.
5. Add LOD scores for all θ values between families. Thus, data from many small families can be added to achieve a statistically significant value for θ.

c. Statistical significance

A LOD score $\geq$ 3.0 is considered evidence for linkage. On the other hand, a LOD score less than -2.0 is considered evidence to exclude linkage. Generally more than one value of θ will go over the 3.0 level. The θ with the highest LOD score is the point estimate of the true map distance. All other adjacent θ values with a LOD score of at least 1 less than the maximum value are considered the “support interval”, the region in which the true linkage value is found. The LOD score in $[-2, 3)$ is consider as not sufficient data for testing linkage.

III. THE MODEL

The point of recombination mapping is to determine the frequency of different kinds of gametes. This situation is mostly done in a test cross, where meiosis in only one parent needs to be considered. We should first calculate the expected offspring frequencies

We are now going to consider what happens when meiosis in both parents is relevant. We proceed as following three steps:

1. Calculate gamete frequencies as a function of θ
2. Calculate offspring genotype frequencies using a Punnett Square
3. Use a spreadsheet varies the value of θ and sees what the resulting expected frequencies for the phenotypes are.

Now, consider a cross with 2 linked genes. Let the disease gene has alleles R (dominant, normal) and r (recessive disease allele) and the marker gene is co-dominant, with alleles M1 and M2.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

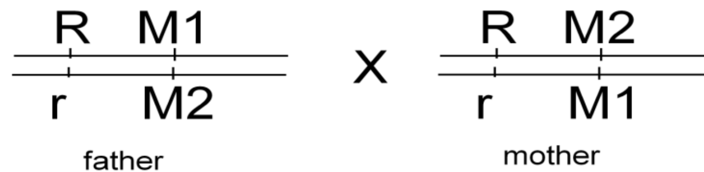
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Assume that, we know the linkage phase in both parents:

i. the father is R M1 / r M2
ii. the mother is R M2 / r M1



So the gametes frequency, for the father is parental gametes: RM1 and rM2 and recombinant gametes: R M2 and r M1; and for the mother, parental gametes: R M2 and rM1 whereas recombinant gametes: RM1 and rM2. The proportion of recombinant gametes is θ. Since there are two recombinant gametes, each has a proportion of $\left(\frac{1}{2}\right)\theta$. The proportion of parental gametes is (1- θ). Each of the two parental gametes has a proportion of $\frac{1}{2}(1-\theta)$. So for the father, R M1 and r M2 gametes have a proportion of $\frac{1}{2}(1-\theta)$ and R M2 and r M1 have $\frac{1}{2}\theta$. For the mother, R M2 and r M1 gametes have a proportion of $\frac{1}{2}(1-\theta)$ and R M1 and r M2 have $\frac{1}{2}\theta$. Now combine the gametes in each row and column, also multiply the gamete frequencies at each intersection of row and column. There are 3 possibilities:

i. Two parental gametes combining have a frequency of (¼)$(1-\theta)^2$
ii. Two recombinant gametes combining have a frequency of (¼)θ^2 .
iii. A parental gamete and a recombinant gamete combining have a frequency of $\frac{1}{4}\theta(1-\theta)$. Put them on a Punnett Square

Mother \ Father	RM1 (1/2)(1-θ)	RM2 (1/2)θ	rM1 (1/2)θ	rM2 (1/2)(1-θ)
RM1 (1/2)θ	RRM1M1 (1/4)θ(1-θ)	RRM1M2 (1/4)θ^2	RrM1M1 (1/4)θ^2	RrM1M2 (1/4)θ(1-θ)
RM2 (1/2)(1-θ)	RRM1M2 (1/4)$(1-\theta)^2$	RRM2M2 (1/4)θ(1-θ)	RrM1M2 (1/4)θ(1-θ)	RrM2M2 (1/4)$(1-\theta)^2$
rM1 (1/2)(1-θ)	RrM1M1 (1/4)$(1-\theta)^2$	RrM1M2 (1/4)θ(1-θ)	rrM1M1 (1/4)θ(1-θ)	rrM1M2 (1/4)$(1-\theta)^2$
rM2 (1/2)θ	RrM1M2 (1/4)θ(1-θ)	RrM2M2 (1/4)θ^2	rrM1M2 (1/4)θ^2	rrM2M2 (1/4)θ(1-θ)

Since r is the recessive disease allele and R is dominant normal: thus RR and Rr give the same normal phenotype (Designated as R_). rr gives the mutant disease phenotype. M1 and M2 are co-dominant, so M1M1, M1M2, and M2M2 are all distinct phenotypes. Therefore a total of 6 phenotypes divided among the 16 cells of the Punnett Square.

Now combining the equations for all cells that give the same phenotype and then summing we get:

R_M1M1: $\frac{1}{4}(1-\theta)^2+\frac{1}{4}\theta^2+\frac{1}{4}\theta(1-\theta)=\frac{1}{4}(\theta^2-\theta+1)$

R_M1M2: $4\times\frac{1}{4}\theta(1-\theta)+\frac{1}{4}(1-\theta)^2+\frac{1}{4}\theta^2=\frac{1}{4}(2\theta-2\theta^2+1)$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

R_M2M2: $\frac{1}{4}(1-\theta)^2+\frac{1}{4}\theta^2+\frac{1}{4}\theta(1-\theta)=\frac{1}{4}(\theta^2-\theta+1)$

rrM1M1: $\frac{1}{4}\theta(1-\theta)$

rrM1M2: $\frac{1}{4}(1-\theta)^2+\frac{1}{4}\theta^2=\frac{1}{4}(2\theta^2-2\theta+1)$

rrM2M2: $\frac{1}{4}\theta(1-\theta)$

Use the equations in a spreadsheet to calculate offspring frequencies for all recombination fractions from 0 (completely linked) to 0.5 (unlinked). We now have the expected frequencies of all possible phenotypes at different values of θ.

θ	0	0.1	0.2	0.3	0.4	0.5
R_ M1M1	0.25	0.2275	0.21	0.1975	0.19	0.1875
R_ M1M2	0.25	0.295	0.33	0.355	0.37	0.375
R_ M2M2	0.25	0.2275	0.21	0.1975	0.19	0.1875
rr M1M1	0	0.0225	0.04	0.0525	0.06	0.0625
rr M1M2	0.25	0.205	0.17	0.145	0.13	0.125
rr M2M2	0	0.0225	0.04	0.0525	0.06	0.0625

a. **Likelihood of a Family**

Likelihood functions determine the probability of the observed data in terms of the parameter being estimated. For LOD scores, a version of the binomial expansion is used. The binomial describes the probability of families with two different phenotypes.

b. **Likelihood Ratio**

We first calculate the expected frequency of each type of offspring at different values of θ. Then we use the data from actual families to calculate the likelihood of each family at each value of θ. After that we take the likelihood ratio and take the logarithm (base 10) of each likelihood.

IV. TEST PROBLEM

Consider a family of 7 children, where

R_ M1M1: 2 children
R_ M1M2: 1 child
R_ M2M2: 1 child
rr M1M1 : 0 child
rr M1M2 : 2 children
rr M2M2 : 1 child

The expressions we will use to determine likelihood is $p^2q^1r^1s^0t^2u^1$ where p, q, r, s, t, and u are the probabilities of the 6 types of offspring.

V. EXPERIMENTAL RESULTS

θ	0	0.1	0.2	0.3	0.4	0.5
Likelihood	0	3.28×10^{-6}	3.53×10^{-6}	3.02×10^{-6}	2.57×10^{-6}	2.41×10^{-6}
Ratio	0	1.36	1.46	1.25	1.07	1
LOD	infinity	.134	.164	.097	.029	0

IJIRSET

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

The maximum LOD for this example is at 0.2. So, more families are needed to reach the significance level of 3.0.

VI. CONCLUSION

A large data sample is needed to determine a **LOD score**. In this type of statistical analysis, when the sample is big, the inferences will be better. This method can be applied on organisms like fruit, flies, human etc. For the purpose of studying the **LOD score** for specific traits, while with populations such as humans, we must give the importance to data collected through observation. Huge samples tend to yield more statistically meaningful information because they reduce the risk of flukes and clusters which twist the data. This method has been successfully applied for 4 phenotypes till now. In this paper an algorithm is made for apply this method more than 4 phenotypes (Here 6 phenotypes).

REFERENCES

1. Yuan, T.W., "Stochastic Models with Applications to Genetics, Cancers, AIDS and Other Biomedical Systems", World Scientific Publishing Co. Pte. Ltd , Vol.4, 2000.
2. Weiller, G.F., "Phylogenetic profiles: a graphical method for detecting genetic recombination in homologous sequences", Journal of Molecular Biology and Evolution, Vol.15, pp.326-335, 1998.
3. Etherington, G.J., Dicks, J. and Roberts, I.N., "Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination", Journal of Bioinformatics, Vol. 21, pp.278-281, 2005.
4. Sawyer., " Statistical tests for detecting gene conversion", Journal of Molecular Biology and Evolution, Vol.6, pp.526-538, 1989.
5. Jacobsen, I.B. and Easteal, S. "A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences", Journal of Computer Applications in the Biosciences, Vol.12, pp.291-295, 1996
6. Bruen, T.C., Philippe, H. and Bryant, D. "A simple and robust statistical test for detecting the presence of recombination", Journal of Genetics, Vol. 172, pp.2665-2681, 2006.
7. Hein.,"Reconstructing evolution of sequences subject to recombination using Parsimony", Journal of Mathematical Biosciences, Vol.98, pp.185-200, 1990.
8. Griffiths, AJF., Miller, JH., Suzuki, DT., Lewontin, RC., Gelbart ,WM. "An Introduction to Genetic Analysis (5th ed.)", New York: W.H.
9. Freeman and Company, chap.5, 1993.
10. Morton, N.E., " Sequential tests for the detection of linkage," American Journal of Human Genetics , Vol. **7** , pp. 277–318, 1995.