Gene Locater: Genetic linkage analysis software using three-point testcross

Muhammad Zohaib Anwar[1], Anoosha Sehar[1], Inayat-ur Rehman[2], Mohammad Hassan Khalid[1]*

[1]Comsats Institute of Information Technology, Islamabad, Pakistan; [2]Department of Computer Science, Comsats Institute of Information Technology, Islamabad, Pakistan; Mohammad Hassan Khalid – Email: hassankhalid@comsats.edu.pk; *Corresponding author

Abstract:
Locating genes on a chromosome is important for understanding the gene function and its linkage and recombination. Knowledge of gene positions on chromosomes is necessary for annotation. The study is essential for disease genetics and genomics, among other aspects. Currently available software's for calculating recombination frequency is mostly limited to the range and flexibility of this type of analysis. GENE LOCATER is a fully customizable program for calculating recombination frequency, written in JAVA. Through an easy-to-use interface, GENE LOCATOR allows users a high degree of flexibility in calculating genetic linkage and displaying linkage group. Among other features, this software enables user to identify linkage groups with output visualized graphically. The program calculates interference and coefficient of coincidence with elevated accuracy in sample datasets.

Availability: http://www.moperandib.com.

Background:
Three point Test Cross has been in extensive use to determine the linkage between two or more genes. Viruses, bacteria, fungi, tomato, barley, wheat, rice, sorghum, morning glory, garden pea, maize, Drosophila, etc. are the examples of the organisms with constructed linkage or genetic maps using this approach [1]. Linkage is an essential technique to locate genes on a chromosome. This helps us to understand the mode of transmissions for many genetic disorders as well as many other aspects of a gene, especially for those which are not extensively studied yet. Moreover, the evolutionary behavior exhibited by the genes in a population is also affected by the linkage [2]. Most existing software (Map Distance Calculator at Chang biosciences & Knud Christensen's Calculator) find out the map distance by the number of parental progeny and recombinants progeny taken as input. Recombination is used to determine map distance between two loci [3]. They calculate map distance on the proportional basis as represented in equation (I). For small distance, the recombination frequency (RF) is proportional to the map distance **(see equation 1 in supplementary material).**

Methodology:
The Gene Locator uses formulae (II), (III), and (IV) to find out the Recombination frequency and map the genes on the linkage group to display the sequence of the genes graphically **(see supplementary material).**

Input:
Input the traits, their dominant and recessive alleles in the form of a label or name of the traits and their forms. This input is then processed into possible number of crossovers. Then the user is required to input the number of physical progeny against each of the trait thus formed. These values are then sorted by progeny or score. Those with the maximum progeny are classified as "Parental", those with minimum are classified as "Double Crossovers", and the rest as "Single Crossovers". The processing of the input entered in the previous steps gives the output in a three panned frame.

Output:
The top left contains the recombination frequency between the genes sequenced in their exact physical position. The second pane is for the Coefficient of coincidence, and interference. The map is printed in the last pane.
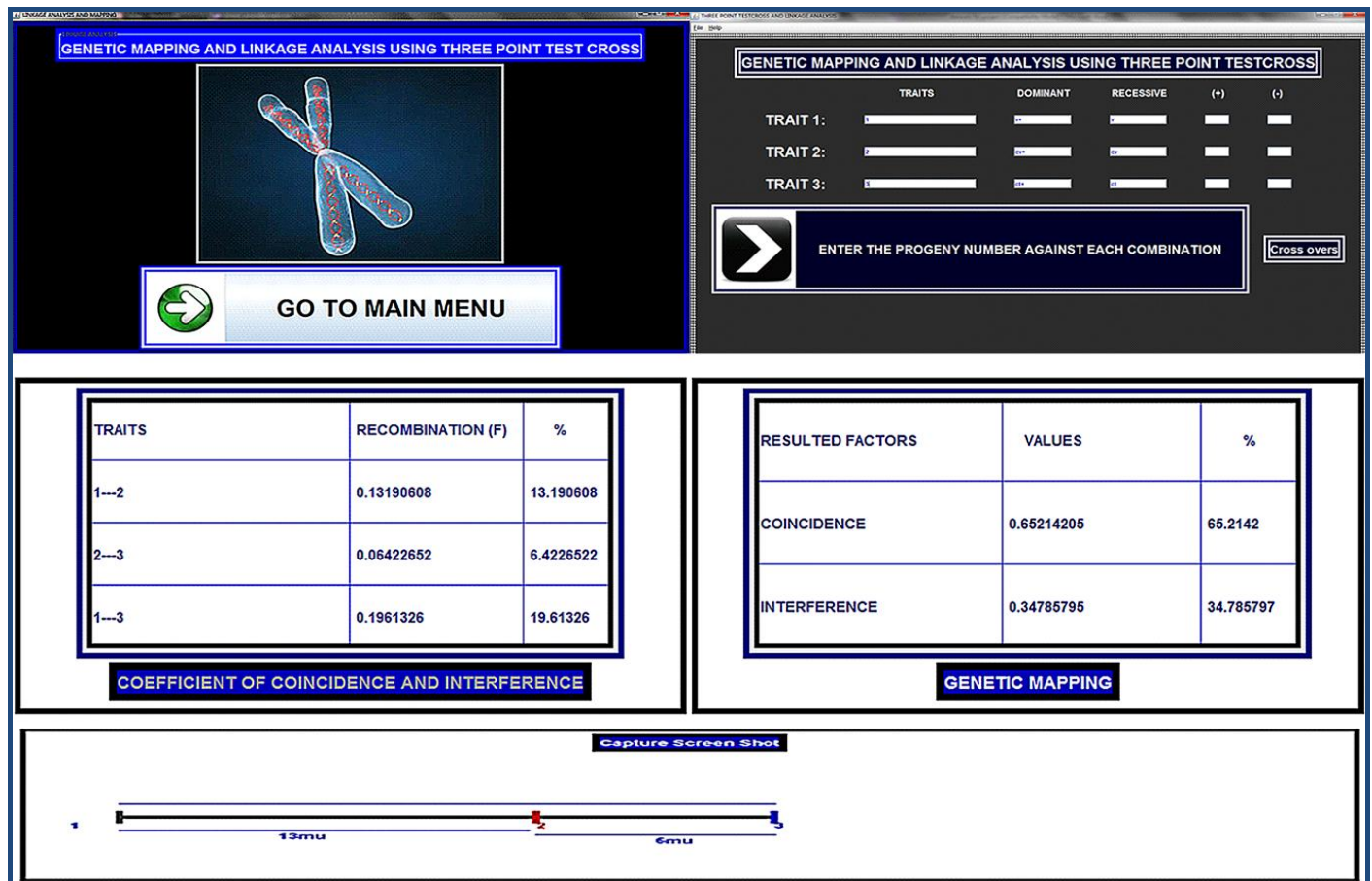

Figure 1: [Left panel] GUI interface of GENE LOCATOR showing the main page. [Top-right panel] GUI interface for taking the input in the text fields, such as the name of the traits and the recessive and dominant alleles for the traits to represent them. [Bottom-right panel] GUI displaying the results of the data set from the NCBI BOOK SHELF which is displaying the recombination frequency, coefficient of coincidence and interference.

Implementation:
Gene Locator has been developed integrally in JAVA as a standalone application for Microsoft Windows operating systems. The program operates on two main inputs as mentioned previously: Traits and their Dominant and recessive alleles. Progeny can be entered in integer form and the incorrect data or input is trashed by the software with a message of wrong input. Other user-customized options such as "save" the data, use a "default" data set, "copy", "paste" make it more flexible to use to calculate the results **(Figure 1).** The Bottom-right Panel is an example of such options as it shows the output which was saved through the options.

Performance:
The Software describes the prediction of Genetic Linkage using three-point testcross. The tool was tested using 45 different datasets (available with the authors). **Table 1 (see**

supplementary material) provides with the comparison of the Gene locator with two similarly designed working programs to elaborate the functional capabilities of this software.

Utility:
This software can be used for identifying and evaluating the results of Plant breeding labs accurately and faster than manual calculation. Moreover, it can help in educational purposes as testcross calculator and recombination frequency calculator.

References:
[1] Lobo I. & Shaw K. (2008) *Nature Education* 1: 1
[2] http://staff.jccc.net/pdecell/transgenetics/linkage.html#definition
[3] http://www.changbioscience.com/stat/rf.html

BIOINFORMATION

Supplementary material:

The following formulae were used to calculate the recombination frequency:

$$RF = (1 - e^{-2*\text{ map distance}})/2 \qquad \longrightarrow \qquad (I)$$

$$RF(i,j) = \frac{\text{Single crossovers between i \& j}}{\text{Total Numbers of progeny}} \times 100 \qquad \longrightarrow \qquad (II)$$

$$\text{Coefficent of coincidence} = \frac{\text{Observed Double Cross overers}}{\text{Expressed Double Crossovers}} \qquad \longrightarrow \qquad (III)$$

$$\text{Interference} = 1 - \text{Coefficent of coincidence} \qquad \longrightarrow \qquad (IV)$$

Table 1: Comparison of Gene locator with existing software

Parameters	Map Distance Calculator At Chang biosciences	Knud Christensen's Calculator	Gene Locator
Implicit Estimation Of Type Of Cross (Repulsion/Coupling)			
User Define Progeny	✓	✓	✓
Implicit Estimation Of Gene Sequence			✓
Self-Identification Of Type Of Crossover (Single, Double, Parental)			✓
Determination Of Cross Over Points			✓
Estimation Type Of Gamete (Heterozygous, Homozygous)			✓
Display Cross Over Combinations			✓
Display Gene Sequence			✓
Calculate and Display Recombination Frequency	✓	✓	✓
Calculate and Display Coefficient Of Coincidence			✓
Calculate and Display Interference		✓	✓
Builds and Display Graphical Genetic Map			✓

Additional material:
Sample data sets taken from the following links:

[1] http://staff.jccc.net/pdecell/transgenetics/linkage.html#definition
[2] http://www.bios.niu.edu/johns/genetics/map_ex
[3] http://www.ag.ndsu.nodak.edu/plantsci/adv_genetics/genetics/tpt/tpt01.htm
[4] http://www.ndsu.edu/pubweb/~mcclean/plsc431/linkage/linkage3.htm
[5] http://www.protocolonline.org/?f
[6] http://www.ncbi.nlm.nih.gov/books/NBK21408/
[7] http://staff.jccc.net/pdecell/transgenetics/linkage.html
[8] http://www.ncbi.nlm.nih.gov/books/NBK21408/
[9] http://www.education.com/study-help/article/linkage-chromosome-mapping-solved-problems/?nocache
[10] http://www.education.com/study-help/article/genetic-maps/?nocache