



Development of Web based Tool for Analysis of Triple Test Crosses in Plant Breeding

Vinay Kumar¹, O.P. Sheoran², Komal Malik³

^{1,2}Department of Mathematics and Statistics

College of Basic Sciences and Humanities

CCS Haryana Agricultural University, Hisar-125001

³Department of Economics, Govt. College, Nalwa

Email: vinay.stat@gmail.com

ABSTRACT

An algorithm for the analysis of triple test cross data has been developed using Active Server Pages (ASP) and a server-client architecture. It consists of six main steps, including reading data on the number of crosses, parents, and environments; computing correction factors and total sum of squares; performing an analysis of variance for parents and crosses; testing for epistasis using an F test; estimating additive and dominance components of genetic variation; and preparing an analysis of variance table for detection of epistasis and its interaction with the environment. The algorithm is demonstrated using data from an experiment on bread wheat varieties conducted at the Department of Genetics and Plant Breeding at Chaudhary Charan Singh Haryana Agricultural University.

Keywords: Algorithm, Triple test cross, Active Server Pages (ASP), Server-client architecture, Epistasis, Genetic variation

Received 04.05.2023

Revised 16.06.2023

Accepted 29.07.2023

INTRODUCTION

In the realm of plant and animal breeding, the utilization of specific mating designs has become a vital tool for evaluating the genetic architecture of a trait of interest by leveraging quantitative genetic variation. One such design, the triple test cross (TTC), proposed by Kearsey and Jinks in 1968, offers a powerful means of statistically testing the significance and estimating the additive and dominant components of variation. However, to fully comprehend the epistatic component of genetic variation, more complex designs are required.

The TTC, as first introduced by Kearsey and Jinks in 1968, is particularly useful for estimating additive and dominant components, as well as testing the significance of the epistatic variance component in the absence of epistasis. Over the years, various modifications and extensions of the TTC have been developed to improve its applicability in both animal and plant breeding, as exemplified in the works [2-8].

The Kearsey and Jinks [1] method offers a clear and precise way to detect and separate epistasis and its relationship with the environment, as well as providing accurate estimates of the additive (D) and dominance (H) components of variation and their relationship with the environment in the absence of epistasis. Furthermore, this approach requires fewer assumptions, and is not reliant on allelic frequency, gene correlation, or mating system, making it highly applicable to both segregating and non-segregating plant populations. Additionally, this method does not significantly increase the number of crosses even when the number of parents increases, making it easy to manage. This method also enables testing for the suitability of testers and provides unbiased estimates of D and H components when the inadequacy of testers is the primary cause for the failure of the additive dominance model. Additionally, in the presence of epistasis, this method provides more accurate estimates of D and H components than other mating designs.

The use of scientifically valid methods, such as the TTC method recommended by Singh and Panwar [7], for identifying and assessing various components of genetic variation and their susceptibility to environmental change in both segregating and non-segregating plant materials can greatly assist breeders in planning and implementing their breeding programs. This approach can facilitate the identification of key genetic factors influencing the trait of interest, and the optimization of breeding strategies to improve the genetic potential of plants and animals.

In the field of plant and animal breeding, the utilization of advanced technologies such as web-based tools for the analysis of data is becoming increasingly important. The complexity of the triple test cross (TTC) method, along with the various modifications and extensions that have been made to it over the years, make a web-based tool for the analysis of the TTC highly beneficial for plant and animal breeders. Such a tool would provide a clear and precise way to detect and separate epistasis and its relationship with the environment, as well as provide accurate estimates of the additive (D) and dominance (H) components of variation and their relationship with the environment in the absence of epistasis. Additionally, a web-based tool would enable breeders to easily manage the large number of crosses required for the TTC method, and would provide unbiased estimates of D and H components when the inadequacy of testers is the primary cause for the failure of the additive dominance model. Furthermore, in the presence of epistasis, a web-based tool would provide more accurate estimates of D and H components than other mating designs, greatly assisting breeders in planning and implementing their breeding programs. The utilization of cutting-edge technologies such as Active Server Pages (ASP) and server-client architecture in the development of such a tool, allows for streamlined data management, computation of correction factors and total sum of squares, and the performance of sophisticated analyses, including the detection of epistasis and its interaction with the environment. The implementation of such a tool serves to augment the rigor and precision of genetic and breeding studies, ultimately leading to a deeper understanding of the underlying biology and genetics of the system under investigation. In present paper a web-based tool for the analysis of triple test crosses have been developed and demonstrated on a dataset, which is an indispensable asset for contemporary research in genetics and plant breeding.

MATERIAL AND METHODS

The TTC design involves crossing m randomly selected male parents from the population being studied to two homozygous parents, P_1 and P_2 , and the resulting F_1 offspring. These offspring are referred to as L_1, L_2 , and L_3 testers. In addition to the two types of crosses present in the North Carolina design III, each parent is also crossed with the F_1 to produce L_{3i} families. This results in the production of $3m$ progeny families, consisting of mL_{1i}, mL_{2i} , and mL_{3i} .

In the TTC design, a random sample of plants from an F_2 generation produced by crossing homozygous parents are backcrossed to the original parents and the F_1 to produce $3m$ crosses. Here, m represents the number of plants randomly selected from the F_2 as male parents. The full TTC analysis can be broken down into two parts for ease of understanding: (1) identification of epistasis and (2) identification and quantification of additive (D) and dominance (H) components.

To detect epistasis using the TTC design, a cross between two distinct loci A-a and B-b is made and the F_2 generation is produced. The nine genotypes produced will be backcrossed to the two parents (AABB and aabb) and the F_1 (AaBb). By calculating the quantity $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ for each genotype, the additive gene effects and dominance gene effects will be cancelled out, leaving only the interaction terms. If there is no epistasis present in the material being studied, this quantity $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ should not be significantly different from zero for any number of segregating loci.

In Kearsey and Jinks' method for detecting epistasis, m values of $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ ($\bar{L}_{11} + \bar{L}_{21} - 2\bar{L}_{31}, \bar{L}_{12} + \bar{L}_{22} - 2\bar{L}_{32}, \dots, \bar{L}_{1m} + \bar{L}_{2m} - 2\bar{L}_{3m}$) are calculated for each block. The sum of squares due to epistasis is calculated using m degrees of freedom. Two types of errors are typically considered: block error and error within crosses (families) within blocks, with $m(b-1)$ and $3mb(k-1)$ degrees of freedom, respectively. The significance of block error is first tested against the within crosses error using a chi-squared test with $m(b-1)$ degrees of freedom. If block error is significant, the significance of epistasis is tested against it using an F-test. If block error is not significant, the significance of epistasis is also tested using a chi-squared test with m degrees of freedom against the within crosses error.

The TTC method, as proposed by Singh and Panwar [7], is a useful tool for plant breeders in detecting and estimating the different components of genetic variation, as well as determining the relative sensitivity of these components to environmental change. The method involves randomly selecting m male parents from an F_2 population or homozygous lines, and crossing them with the two homozygous parent lines (P_1 and P_2) as well as the F_1 produced by them (L_1, L_2 , and L_3 testers). This results in $3m$ progeny families (mL_{1i}, mL_{2i} , and mL_{3i}). The full TTC analysis can be divided into two parts: detection of epistasis and detection and estimation of additive (D) and dominance (H) components. The Kearsey and Jinks method is used to detect epistasis by calculating m values of $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$, and testing the significance of block error against within crosses error. Jinks and Perkins (1970) proposed a modification of this method based on the orthogonality of components among L_{1i}, L_{2i} , and L_{3i} family means, allowing for partitioning of epistasis into fixable and unfixable types, as well as block interactions.

Triple Test Cross Analysis Algorithm

Step 1: This algorithm starts by reading the data pertaining to number of crosses, Number of parents, number of replications and number of environments. This data is stored in various arrays which are then used for computing correction factor and total sum of squares of Parents, crosses and their interactions. Finally, the algorithm prepares the analysis of variance for parents and crosses and computes the variances of the parents for estimating additive variance (D).

Step 2: Compute $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ from crosses data for all the replications and environments and sum of squares due to epistasis, flexible epistasis, unfixable epistasis, block errors, block \times fixable epistasis and blocks \times unfixable epistasis is obtained. The testing of the epistasis is carried using F test.

Step 3: The test $\bar{L}_{1i} + \bar{L}_{2i} - \bar{P}_i$ when applied together with $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ test, provides test of adequacy of testers if epistasis is absent. The values of test $\bar{L}_{1i} + \bar{L}_{2i} - \bar{P}_i$ is computed from crosses data for all the replications and environments and sum of squares due to test $\bar{L}_{1i} + \bar{L}_{2i} - \bar{P}_i$, blocks and their interaction is computed and their significance is tested.

Step 4: In this step the analysis of sums ($\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ or $\bar{L}_{1i} + \bar{L}_{2i}$) is carried out as it provides unbiased estimate of additive genetic component if epistasis is absent and testers are adequate. Similar to step 2 and 3 the values of $\bar{L}_{1i} + \bar{L}_{2i}$ and $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ have been computed from crosses data and sum of squares due to sums, blocks and their interactions have been computed and tested for significance using F test. Also, this step provides the unbiased estimate of additive genetic component D.

Step 5: If testers are adequate and epistasis is absent, the analysis of differences ($\bar{L}_{1i} - \bar{L}_{2i}$) provide unbiased estimate of dominance component (H) of genetic variation. The values of ($\bar{L}_{1i} - \bar{L}_{2i}$) have been obtained for crosses data and sum of squares due to difference, blocks and blocks \times differences have been computed and their significance is also tested using F test.

Step 6: The algorithm finally prepares analysis of variance table for detection of epistasis and its interaction with environment. This step computes various sum of squares due to epistasis, unfixable epistasis, environments, unfixable epistasis \times environments, blocks within environment \times fixable epistasis, blocks within environment \times unfixable epistasis, sums \times environments, differences \times environments.

Web-based tool Development

The development of a web-based tool for the analysis of triple test cross data necessitates a comprehensive and robust methodology that incorporates a variety of advanced techniques. The primary method employed in this tool is the utilization of Active Server Pages (ASP) and a server-client architecture. This architecture allows for seamless data transfer between the client and server, facilitating efficient data processing and analysis.

The algorithm used in this tool comprises of six main steps, including:

1. Reading data on the number of crosses, parents, and environments.
2. Computing correction factors and total sum of squares.
3. Performing an analysis of variance for parents and crosses.
4. Testing for epistasis using an F test.
5. Estimating additive and dominance components of genetic variation.
6. Preparing an analysis of variance table for detection of epistasis and its interaction with the environment.

In terms of materials, the tool requires a dataset of triple test cross data as input. The dataset should include information on the number of crosses, parents, and environments, as well as any other relevant biological or genetic information. This data can be easily obtained through experimental studies or by utilizing existing databases.

Algorithm Implementation and Data Analysis

The methodology for utilizing this web-based tool for the analysis of Triple Test Cross data involves a meticulous process of data organization and inputting into the interface. This process ensures that the data for all characters being analyzed is nested within appropriate parameters, such as replications, genotypes, and environments. The data is inputted into the system via a designated text area, where it must be separated by at least one space and organized in a specific manner. Alternatively, the data can also be entered via an Excel file, where each row's columns contain data for a specific character (Fig 1). This data can then be imported into the system through the process of copying and pasting (Fig 2).

| | A | B | C | D | E | F | G | H | I | J | K |
|----|------------|----|-----|----|-----|--------|------|-------|---|---|---|
| 1 | Treat comb | DH | PH | TN | GPE | GW1000 | BYPP | Yield | | | |
| 2 | E1G1R1P1 | 96 | 90 | 20 | 77 | 45.5 | 111 | 53.7 | | | |
| 3 | E1G1R1P2 | 96 | 90 | 30 | 82 | 41.6 | 130 | 71.5 | | | |
| 4 | E1G1R1P3 | 96 | 96 | 14 | 86 | 40.3 | 181 | 60.6 | | | |
| 5 | E1G1R1P4 | 96 | 103 | 31 | 82 | 43 | 205 | 34.9 | | | |
| 6 | E1G1R1P5 | 96 | 105 | 32 | 84 | 42.5 | 184 | 43.5 | | | |
| 7 | E1G1R1P6 | 93 | 87 | 24 | 73 | 37.2 | 65 | 45 | | | |
| 8 | E1G1R2P1 | 94 | 94 | 26 | 71 | 39.4 | 145 | 45 | | | |
| 9 | E1G1R2P2 | 93 | 92 | 17 | 62 | 42.4 | 121 | 43.5 | | | |
| 10 | E1G1R2P3 | 94 | 91 | 28 | 72 | 40.1 | 122 | 33 | | | |
| 11 | E1G1R2P4 | 94 | 90 | 32 | 67 | 43.6 | 75 | 47.6 | | | |
| 12 | E1G1R2P5 | 95 | 82 | 23 | 76 | 37.4 | 136 | 31 | | | |
| 13 | E1G1R2P6 | 95 | 84 | 34 | 78 | 37.4 | 138 | 49.8 | | | |
| 14 | E1G1R3P1 | 95 | 91 | 25 | 99 | 42.45 | 159 | 49 | | | |
| 15 | E1G1R3P2 | 96 | 94 | 24 | 82 | 44 | 161 | 43.8 | | | |
| 16 | E1G1R3P3 | 96 | 88 | 29 | 76 | 43.45 | 166 | 40.6 | | | |
| 17 | E1G1R3P4 | 94 | 100 | 26 | 42 | 47.6 | 120 | 40.7 | | | |
| 18 | E1G1R3P5 | 94 | 105 | 20 | 42 | 44 | 110 | 35.8 | | | |
| 19 | E1G1R3P6 | 94 | 99 | 22 | 43 | 43.6 | 89 | 31.5 | | | |
| 20 | E1G2R1P1 | 95 | 106 | 21 | 44 | 43 | 119 | 43.6 | | | |
| 21 | E1G2R1P2 | 95 | 106 | 20 | 49 | 43.2 | 124 | 32.6 | | | |
| 22 | E1G2R1P3 | 94 | 90 | 21 | 74 | 41.4 | 60 | 40.5 | | | |

Data for each character must be arranged in separate column

Here P, R, G and E represents plant no, Replication, Genotype (Crosses, Parents) and Environment respectively

Copy only data and Paste it in the text area provided for this purpose. The character names must be entered in separate text area

Fig 1: Data arrangements of TTC in Excel Sheet

Once the data is entered or imported, the user must provide additional information, such as the number of genotypes, replications, environments, and plants selected per replication, in the designated text boxes as shown in fig 3. Upon providing the necessary data, the user can initiate the analysis process by clicking on the "Analyze" button.

Data Entry Form

127.0.0.1/ttc/tripletest.html

Data Entry interface for Triple Test Cross Analysis

Please Enter or paste your data in the text area given below

Enter data for each plant sampled in a replication in first line and each observation must be separated by at least single space. Similarly data for another replication must be entered in separate line

Enter Character Names

| | | | | | | |
|----|-----|----|----|-------|-----|------|
| 96 | 90 | 20 | 77 | 45.5 | 111 | 53.7 |
| 96 | 90 | 30 | 82 | 41.6 | 130 | 71.5 |
| 96 | 96 | 14 | 86 | 40.3 | 181 | 60.6 |
| 96 | 103 | 31 | 82 | 43 | 205 | 34.9 |
| 96 | 105 | 32 | 84 | 42.5 | 184 | 43.5 |
| 93 | 87 | 24 | 73 | 37.2 | 65 | 45 |
| 94 | 94 | 26 | 71 | 39.4 | 145 | 45 |
| 93 | 92 | 17 | 62 | 42.4 | 121 | 43.5 |
| 94 | 91 | 28 | 72 | 40.1 | 122 | 33 |
| 94 | 90 | 32 | 67 | 43.6 | 75 | 47.6 |
| 95 | 82 | 23 | 76 | 37.4 | 136 | 31 |
| 95 | 84 | 34 | 78 | 37.4 | 138 | 49.8 |
| 95 | 91 | 25 | 99 | 42.45 | 159 | 49 |
| 96 | 94 | 24 | 82 | 44 | 161 | 43.8 |
| 96 | 88 | 29 | 76 | 43.45 | 166 | 40.6 |
| 94 | 100 | 26 | 42 | 47.6 | 120 | 40.7 |
| 94 | 105 | 20 | 42 | 44 | 110 | 35.8 |
| 94 | 99 | 22 | 43 | 43.6 | 89 | 31.5 |
| 95 | 106 | 21 | 44 | 43 | 119 | 43.6 |
| 95 | 106 | 20 | 49 | 43.2 | 124 | 32.6 |
| 94 | 90 | 21 | 74 | 41.4 | 60 | 40.5 |
| 94 | 100 | 15 | 64 | 42.1 | 181 | 37.8 |
| 93 | 98 | 26 | 46 | 39.7 | 76 | 31.7 |
| 94 | 96 | 28 | 47 | 39.9 | 79 | 49.9 |
| 94 | 96 | 29 | 58 | 47 | 80 | 42.9 |
| 94 | 97 | 19 | 91 | 42.4 | 200 | 37.8 |

Character Names: DH, PH, TN, GPE, GW1000, BYPP, Yield

Submit Reset

Fig 2: Interface for Data Entry of Triple test Cross Analysis.

Triple Test Cross

Note: Please Enter 1 in No of plants selected in Replication text box if only single observation is taken

| | | |
|--|---------------------------------|---|
| Enter number Parents(Lines) | <input type="text" value="16"/> | If Data for Parents (Testers) is not include enter 0 here |
| Enter number Parents(Testers) | <input type="text" value="0"/> | |
| Enter number of Replications | <input type="text" value="3"/> | |
| No. of Plants selected in Replications | <input type="text" value="5"/> | |
| Enter number of Environments | <input type="text" value="2"/> | |
| Enter number of Characters | <input type="text" value="7"/> | |

Enter only numeric values in these text boxes without putting any space before the figure

Fig 3: Interface for entering parameters of triple test cross

RESULTS AND DISCUSSION

The results of the analysis are then displayed on a separate web page, which includes an analysis of variance for the parents, families, a test of epistasis, an evaluation of the adequacy of testers, sums, and differences for each environment, as well as a comprehensive table for the genotype and macro-environment interactions. This robust tool is designed to provide a comprehensive statistical analysis of the data, while also being user-friendly and easily accessible via a web-based platform. A user can access this tool using url <http://14.139.232.166/opstat/default.asp> and clicking on the tab "Triple Test Cross Analysis".

Example Analysis Demonstration

The analysis will be demonstrated using data from an experiment conducted in the Department of Genetics and Plant Breeding at Chaudhary Charan Singh Haryana Agricultural University. The dataset includes information on 18 varieties of bread wheat, including DPW621-50, HD2967, Tobari, WH1080, DBW17, PBW550, Aus15854, WH147, WH711, WH1021, WH542, Veery's, WH730, Raj3765, Raj MR-1, WH525, WH1105 and WH283. These varieties were used to create 48 triple test cross families, using two agronomically superior varieties, WH 1105 and WH 283, as male testers in a triple test cross fashion. The 48 progeny families and 16 parents were evaluated in a randomized block design with three replications in two different environments, one with a timely sowing and the other with a late sowing. Data on 7 characters such as Days to heading, Plant height, Number of tillers, Grains per ear head, 1000 grains weight, Biological yield per plant and Yield was collected from 5 randomly selected plants from each replication. The analysis of this data using the developed web-based tool for triple test cross analysis will be demonstrated using the data for the "Days to heading" character. The analysis will be displayed on separate web page the part of analysis is shown in Fig 4

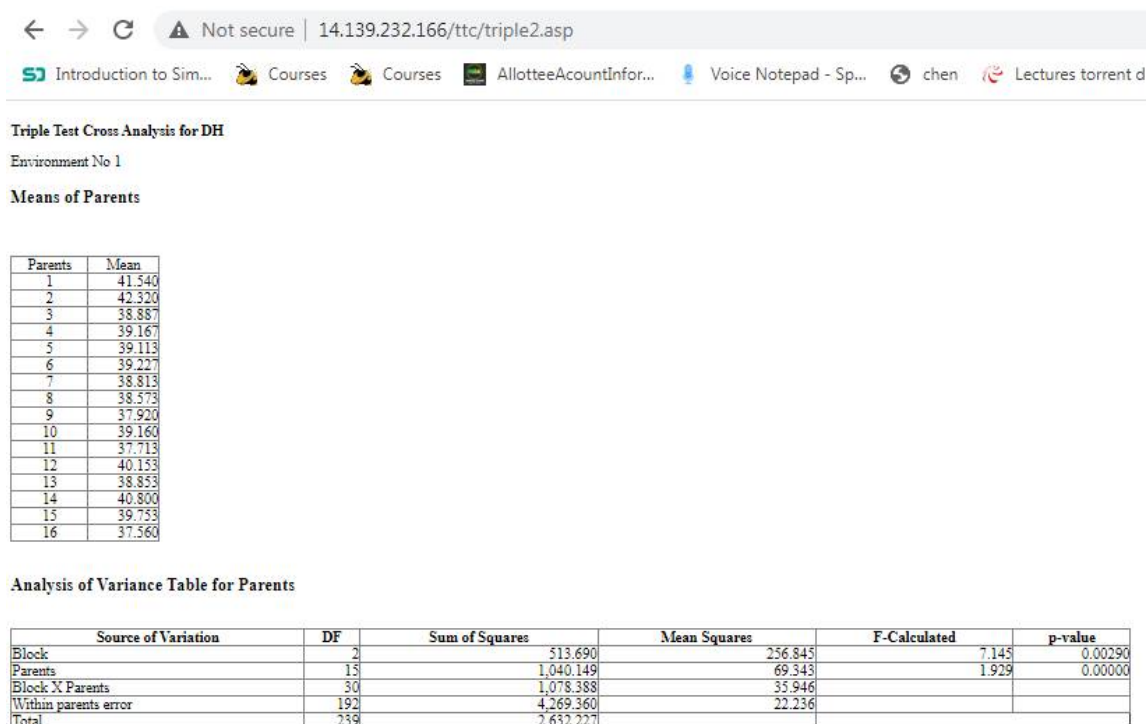


Fig 4: The output generated from the analysis of triple test cross (only few lines)

Outcomes of Triple Test Cross Analysis:

The Triple Test Cross Analysis for all seven characters was conducted in two distinct micro-environments and the results were meticulously analyzed to provide a comprehensive understanding of the genotype x micro-environment interaction. The script provides a detailed account of the means of the parents and crosses for each environment, along with an Analysis of Variance Table. The significance of blocks and parents were tested against Block x Parents mean squares and that of blocks and families were tested against Block x Families mean squares. Furthermore, the script also calculates the Additive component D value for each environment, providing a deeper insight into the genetic architecture of the characters.

The Analysis of Variance (ANOVA) table for Genotype x micro-environment interaction is a crucial aspect of the script, as it provides information about the sources of variation in the data and their significance. The table lists different sources of variation and their respective degrees of freedom (DF), sum of squares, mean squares, F-calculated value, and significance level. The sources of variation include Epistasis, Fixable, Unfixable, Fixable epistasis x environment, Unfixable epistasis x environment, Block within Env x fixable epistasis, Block within Env x unfixable epistasis, Within family error, Sums, Sums x Environments, Block within Env x Sums, Within Family error, Differences, Diff x Environments, Block within Env x Differences, and Within Family Error.

The F-Calculated value, which is the ratio of the mean square of a source of variation to the mean square of error, is a measure of the strength of the association between the independent variable and the dependent variable. The significance column, on the other hand, shows the p-value, a measure of the probability that the variation observed is due to chance. A low p-value indicates that the variation is unlikely to be due to chance and is therefore statistically significant.

CONCLUSIONS

The Triple Test Cross Analysis tool presented in this script is a powerful and sophisticated tool for analyzing genetic data. It provides a comprehensive analysis of several characters in any number of different environments, and presents the results in a clear and easy-to-interpret format. The tool utilizes a variety of statistical techniques, including Analysis of Variance, to provide detailed information about the sources of variation in the data, and their significance. The tool also provides an Additive component D value for each environment, providing a quantitative measure of the genetic variation present in the data. Overall, the Triple Test Cross Analysis tool is a valuable tool for any researcher studying the genetics of complex traits, and provides a wealth of information that can be used to further understand the underlying genetic mechanisms that govern these traits.

REFERENCES

1. Kearsy, M. J., & Jinks, J. L. (1968). Triple test cross: a new method for measuring dominance and epistasis. *Heredity*, **23(2)**: 237-253.
2. Jinks, J. L., Kearsy, M. J., & Quirke, J. (1969). Triple test cross: estimation of dominance and epistasis. *Annals of human genetics*, **32(3)**: 305-317.
3. Jinks, J. L., & Perkins, J. (1970). Triple test cross: estimation of dominance and epistasis. *Annals of human genetics*, **33(3)**: 305-317.
4. Poonie, K. S., Jinks, J. L., & Patel, C. B. (1980). Triple test cross: estimation of dominance and epistasis. *Heredity*, **45(3)**: 331-339.
5. Goldringer, I., Jinks, J. L., & Kearsy, M. J. (1997). Triple test cross: estimation of dominance and epistasis. *Heredity*, **78(3)**: 356-365.
6. Singh, R. K., & Panwar, H. S. (1998). Triple test cross: estimation of dominance and epistasis. *Euphytica*, **97(1)**: 3-11.
7. Singh, S. and Pawar, I.S.,(2005). Theory and Application of Biometrical Genetics. CBS Publishers & Distributors, New Delhi. pp 467
8. Phougat, D. (2016). Triple Test Cross Analysis in Bread Wheat (*Triticumaestivum* L.) Unpublished Ph.D. Dissertation, Department of Genetics and Plant Breeding, Chaudhary Charan Singh Haryana Agricultural University, Hisar.
9. Phougat, D., Panwar, I.S. and Singh V. (2016). Detection of genotype× environment interaction in triple test cross families in bread wheat. *Bangladesh Journal of Botany*,**45(5)**:1225-1228.

CITATION OF THIS ARTICLE

Vinay Kumar, O.P. Sheoran, Komal Malik- Development of Web based Tool for Analysis of Triple Test Crosses in Plant Breeding . Bull. Env.Pharmacol. Life Sci., Vol 12 [8] July 2023: 94-100