Experimental designs for precision in phenotyping

Murari Singh and Khaled El- Shama'a

International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan

Address for correspondence:

Dr. M. Singh

Biometrics and Statistics Section, DDG-R

Bldg. no. 15, Khalid Abu Dalbouh Str., Abdoun

P.O. Box 950764 - Code No. 11195

Amman – Jordan

Tel . +962 6 5903120, Mobile: (+962)79 933 8547

Email: m.singh@cgiar.org

Prepared 10 November 2013

Abstract

Precision phenotyping is evaluation of the genotype's expression in a given environment with minimum influence of the experimental error. This chapter presents the basic principles of experimental designs and lists commonly used experimental designs for phenotyping crop genotypes. Experimental designs include unreplicated designs, incomplete block designs and variable replication block designs along with some selected software that can be used to generate the designs. Some illustrations experimental designs and key directives of the software have also been included.

Key words: phenotyping, experimental designs, statistical analysis

Introduction

Phenotyping stands for observing or evaluating a genotype(s) in an environment, with least effect due to experimental error, while genotyping stands for observing and describing primarily the genetic make-up of the genotype which is done in terms of using various molecular markers such as amplified fragment length polymorphism (AFLP), simple sequence repeats (SSR) and single nucleotide polymorphism (SNP). A phenotype is an expression of the molecular construct of a genotype in a given environment and depends on the various sources that govern the expression. Thus, if a genotype is to be phenotypically evaluated in a specified-factor controlled/designated environment, for example a drought-stressed environment, effort should be made to eliminate the

effects of all the other factors which influence the phenotypic expression. We will discuss designs commonly used for phenotyping in grain legumes or for crop variety evaluation in general. The experimental designs may depend on the nature of genetic material and its availability.

The selection of the traits for phenotyping is important from various perspectives. Tuberossa (2011) has discussed key concepts, issues and approaches for phenotyping for drought-stressed crops. The role of phenotyping of drought-adaptive traits and use of germplasm resources and genomics methods has been emphasized to improve drought-resistance, and important elements of field plot techniques for obtaining phenotypic data under water-limited conditions have been listed. Recent interests have been found in identifying traits that can be used to identify genotype for climate change using climatic ad agro-ecological information (Bari et al. 2012). The field-based precision phenotyping may be used to generate high-quality and large-scale datasets under managed stressed environments providing valuable guidance for drought screening (Campos et al. 2011). Depending on the trait, the mechanism of phenotyping could vary. The phenotyping can take place in petri dishes in a temperature controlled room, pots in a green house or plots in a field at a location with known biotic, abiotic and edaphic conditions/factors. The phenotypic expression of the traits of interest of what is being phenotyped, for example the genetic material, requires the identification of the population of the responding units, for example the field plots under an environment with known stress levels.

The objective of this chapter is to briefly discuss basic principles of experimental designs and provide examples of various experiment designs used in phenotyping the crop genotypes at various stages of plant growth. We give main features of statistical analysis of data generated using such designs. We also overview some main statistical software which are used to generate these designs.

Design of Experiments

Experimental design for phenotyping will depend on the experimental material and sources of variation there in which are likely to distort the genotypic value of the genotypes. Experimental design is a mechanism to generate scientific evidences for collecting statistically valid and reliable pieces of evidence on the phenotype of the underlying genotype and is guided by the level of variability within the experimental material, size and shape of the experimental unit (e.g. a pot in a greenhouse and a plot in a field), operational convenience and cost. The experimental material may be seeds of a genotypes kept in petri-dishes for dormancy and germination is the trait for phenotyping; seedlings grown in tubes for tolerant to salinity levels; plants in pots kept in the green-house response to controlled application of stress—moisture stress, insect, disease infection, or field plots for yield and yield components evaluation.

An experimental unit is smallest division of the experimental material to which a genotype is assigned recognizing the fact that any neighbouring experimental units may be assigned to a

different genotypes. Set of all the experimental units form the experimental material. In simplest terms, an experimental design is an assignment of treatments to the experimental units and is implemented using the principles of randomization, replications and local control of experimental error or reduction of errors with a view to obtain valid and precise evaluation of the treatments under investigation. These three basic principles of experimental designs are also known as 3Rs of Sir R.A. Fisher (1990). Randomization is a random assignment of genotypes (treatments) to the experimental unit. It is a key element for assigning validity to the information on phenotype and forms the basis for describing the phenotype using a statistical model. Replication, the number of experimental units assigned to a given genotype, is essential for estimating the experimental error or experimental error variance which is variation arising from the responses of the same genotypes on homogeneous experimental units. In reality experimental material is not homogenous, the effort is made to eliminate the effect of any systematic factor using proper field plot management techniques and or by accounting for these systematic factors, which helps in reducing the experimental error variance. The experimental error variance also depends on the size and shape of the experimental units determined by the nature of the experimental material required for phenotyping and the treatments applied.

The precision of the treatment performance or effect depends on the variability in the experimental material and number of replications, and can be increased by reducing the error variability and or by increasing the replication. When the error variability cannot be reduced further, the number replications (r) can be chosen to achieve estimates with a given precision of

the treatment estimates using the following commonly used expression

$r = \frac{4\theta^2}{\varepsilon^2}$, where θ is the coefficient of variation of the trait of interest for the population or the experimental material after eliminating the effects of every systematic factors, and ε is the relative absolute difference in the observed treatment mean from the r replications and the true treatment mean. The basics of the principles of experimental designs are described in standard texts Federer (1955), Cochran and Cox (1957), Kempthorne (1983), Cox and Reid (2000), Mead et al (2002), Hinkelmann and Kempthorne (2005, 2008) among others. A number of specific situations related experimental designs are given in Hinkelmann and Kempthorne (2012). We also refer to a checklist of questions experimenters are advised to answer, were provided by Jeffers (1978).

There are primarily two types of effects assumed for the treatments (genotypes) which form the basis for developing the criterion for which the designs are constructed. Under genotypes effects assumed as fixed, designs are developed by minimizing the average variance of estimated difference between effects of pairwise treatments and the resulting designs are called A-optimal (Kiefer, 1959). Under this set-up one evaluates the phenotypes in form of best linear unbiased estimates (BLUEs). The crop variety trial process comprises of selecting a number of desired genotypes from a much larger number under evaluation and, therefore, the genotypes keep varying with time and the prediction of future performance of a genotype is needed. In this situation genotypes are seen to have been randomly drawn from a population or a process resulting from a breeding strategy, and the genotype effects may more appropriately could be

assumed as random. Maximization of genetic gain or heritability are the parameters of interest. These lead to developing experimental designs which could optimize for average variance of predicted difference between the best linear unbiased predictors (BLUPs) (Cullis et al. 2006).

At various stages during plant development, observations are recorded on the expressions or responses in the field-books or in an electronic form using a hand-held palmtop or other electronic devices. The data are then subjected to transformation, e.g. yields recorded at plot basis are transformed to yield per hectare, before using them in statistical analysis.

Software for generating experimental designs

There are several statistical packages such as GenStat (Payne 2011), SAS (SAS Institute Inc. 1989), CycDesigN (Whitaker et al 2002), Agrobase (Agronomix Software, Inc. 1999) etc. that can be used to generate randomized plans. Design for partial replications can be generated using codes of DiGGer, an R-package (Coombes 2009).

Data Analysis Procedures

Statistical analysis is a procedure to draw inference on the genotypes by searching pattern in the phenotypic evidences and assessing the strength of the pattern relative to the noise arising from experimental errors. Power of the evidence on the genotype effects can be enhanced by incorporating any features inherited in the experimental material at the design and analysis stages. The data or response values are generally modelled using the following representation:

Data or function of (Data) = Pattern (experimental factors, environmental patters, any other systematic feature in the experimental material) + random error,

The total variability in the data is then partitioned into that due to various components of the pattern and errors. The error variance, measured by error mean-squares, is used to assess the significance or contribution of the components of interest in the pattern. Often we use analysis of variance (ANOVA) and estimate means with standard errors, perform multiple comparisons and residual plot analysis is used to examine the validity of assumptions underlying the ANOVA. We will now discuss a number of commonly used experimental designs for phenotyping in a wide range of disciplines, such as plan breeding and genetics, physiology, pathology and entomology.

Experimental Designs for Phenotyping of Crop Genotypes

For phenotyping of improved genetic material, resistant to biotic and abiotic stress factors, through collection-selection missions, crossing, and evaluation in field conditions, the experimental designs are needed for preliminary screening, advanced yield trials, multi-locational trials, international nurseries, as given in the following. The necessary codes and steps for using GenStat menus and R-package DiGGer are given in the Appendix.

Preliminary screening/unreplicated trials

At the preliminary stage of genetic material development or the early generation variety trials, the number of genotypes is reasonably large with seeds limited to one two replications. Further, seeds

of a number of genotypes, called checks with similar maturity level, are available in sufficient number for required number replications for evaluation of experimental errors. A number of experimental designs that are available include reinforced block designs (Das 1958), augmented designs of Federer (1961) in one-way blocks, and those due to Federer and Raghavarao (1975), and Lin and Poshinsky (1983) in two-way blocks. A randomized plan for 45 unreplicated test genotypes in total in 9 incomplete blocks of size 8 and comprising 3 checks and 5 test entries each is given in Table 1.

[Insert Table 1 here]

The statistical analysis model accounts for the effects of incomplete blocks, or row and column effects, and genotype effects. Interest lies in estimates of adjusted means for the genotypes and their standard errors, along with estimate of error variance, coefficient of variation CV%, standard errors of comparisons of two test entries, test and check entries, two check entries. The software that could be used includes GenStat (REML command), SAS (PROC MIXED) and ICARDA programs using GenStat software codes.

Advanced yield trials/replicated trials

Majority of research has gone into developing experimental designs for situations in crop variety evaluations where seeds are available to implement replicated trials. Designs with high efficiency factor are available for almost any number of genotypes evaluated in practices. Our experience indicates that the following types of designs have been found to be used frequently. However, these

are not our recommendations.

Small number of genotypes (V<8)

Often the experimental units within small sized blocks can be expected to be homogeneous. Phenotyping a relatively small number of genotypes in tubes or pots in green house or in plots in the field one may use Randomized Complete Block (RCB) designs with larger number of replications resulting into error degrees of freedom around 30. For controlling experimental error variation in two directions, for example in the field, Latin Square (LS) designs and Youden Square designs are found suitable. LS designs, the number of replications is equal to the number of genotypes while in RCB designs they can be chosen at will. An example is given in Table 2.

[Insert Table 2 here]

Moderate number of genotypes (V≤15)

While scope lies in having a better control of variability, with moderate number of genotypes frequent use of randomized complete block (RCB) designs can be found with three or more replications. An example is given in Table 3.

[Insert Table 3 here]

Large number of genotypes (V>15)

In field trials, the plot-to-plot variability within block increases with size of the block. If a large number of genotypes are experimented using complete blocks then plot-to-plot variability within the large sized blocks could be perceived to be considerably high and thus RCB design may not

give precise estimates unless replications are increased at added cost. Experimentation in relatively smaller sized blocks i.e. use of an incomplete block design seems to be a favorable alternative. Further, it is possible to find designs in incomplete blocks such that if we rearrange/position the incomplete blocks in such a way that the group of incomplete blocks placed physically together on the layout also form full replicates. Such designs are called resolvable block designs. An advantage of resolvable block design is that effectiveness of incomplete blocks can be assessed in relation to complete blocks. Literature contains several classes of resolvable incomplete block designs: balanced incomplete block designs, square lattice designs, rectangular lattices, α-designs, also called α- lattices, in one-dimension; these designs are based on the structure of number of genotypes, e.g. it may be a square number or a rectangular number. The α- designs (Patterson and Williams 1976) are available for almost every practical number of genotypes, with a small difference in block sizes, and suit most of the field configuration. The number of replications can also be chosen at will.

α -Designs: A class of Resolvable Incomplete Block Designs

Patterson and Williams (1976) introduced a class of resolvable incomplete block designs for any number of varieties v and block size k such that v is a multiple of k, i.e., v=ks where s is the number of incomplete blocks of the same size k. Thus the square lattices, rectangular lattices, and resolvable cyclic designs are the special cases of α - designs. Construction of these designs required knowledge of generation array, a combinatorics concept and the methods are given in Patterson and Williams (1976), Patterson et al

230 (1978), and John and Williams (1995). However, these can be obtained by using
231 CycDesigN software (Whitaker et al 2002) and GenStat (Payne 2011) for number of
232 genotypes less than 100. These computer generated methods have shown to provide high
233 efficiency factors within their comparable class of designs for a wide range of parameter
234 values. There may also be situation where the number of genotypes is not a multiple of
235 block size, i.e v ≠ ks. Suppose the number of treatments v is represented by $v = k_1 s_1 + k_2 s_2$;
236 k_1 , k_2 , s_1 , s_2 being positive integers. Every replication has s_1 blocks of size k_1 each
237 and s_2 blocks of size k_2 each. In such situations, it is possible to develop designs with
238 two block sizes k_1 and k_2 where k_1 and k_2 are very close, say have a difference $|k_1 - k_2|$
239 equal to 1 or 2. The small difference in the block size may still support the homogeneity
240 of experimental error variances within such blocks. For example for evaluating v = 23
241 genotypes, one may use $v = 23 = 4 \times 5 + 3 \times 1 = k_1 \times s_1 + k_2 \times s_2$, thus using 5 blocks of size 4
242 and 1 block of size 3 in each replicate. Such designs are derived by omitting one or more
243 varieties of the α - designs with $v = k s$. Two examples of α designs are given in Table 4
244 and Table 5. In case of the designs in Table 5, the empty cell need not be retained or if
245 required for keeping the planting machinery or any other logistics then it could be filled
246 by a filler check genotype.
247 [insert Table 4 here]
248

[insert Table 5 here]

Designs Eliminating Heterogeneity in Two Directions

When the direction of soil fertility is unknown or if variability exists in two perpendicular directions in the field it is often helpful to use two-way blocks in the field to reduce the experimental error. There are several designs controlling variability in two directions. Some of the frequently discussed designs are row-columns (Pearce, 1975), Youden-squares (Youden 1940), lattice squares, (Yates, 1940; Cochran and Cox 1957), lattice rectangles (Federer and Raktoe, 1965), row-column α-designs (John and Eccleston, 1986), incomplete block designs with nested rows and columns (Singh and Dey, 1979).

In recent years, a more realistic approach has been suggested for searching experimental designs using a criterion which maximizes genetic gain due to selection (Kempton 1984). Another related criterion, minimize average pairwise prediction error variance is presented in Cullis et al (2006). These designs were obtained for an early generation variety trials (EGVTs), called p-rep designs are alternative to augment designs in blocks (referred as grid-plots). Simulation studies, based on 1000 runs and 12 different combinations of genetic variance ratio and spatial autocorrelation parameters along rows & columns, have shown that p-rep designs resulted in higher genetic gain. In variety evaluation, a more practical situation shows that different sets of genotypes could have

seeds available for varying replications. Further, in the field layout the spatial variability exists and the plot errors are generally correlated (Singh et al 2003). To generate experimental designs incorporating the need of variable replications and correlated errors, Coombes (2009) has developed an R-program package called, DiGGeR. An example of p-rep design is in Table 6 for 20 test genotype with no replications and 10 test genotypes with 2 replications and 3 check genotypes. To generate randomized plans for p-rep designs, DiGGeR package in R- language programs has been developed by Coombe (2009).

[Insert Table 6 here]

Multi-environment Trials

Multi-environment trials (MET), normally designed in replicated designs, e.g., RCB or α-design, are conducted over multi-locations and multi-years to obtain information on the variety response to the environments and study the nature of the genotype x environment (GxE) interaction. Main objectives of METs are selection of varieties for high and stable yield, and their adaptability to specific zones (clusters) of the environments. The number of and variability due to the locations, years and experimental error may be used to determine the number of replications per trial. However, for moderately large number of locations and years, two replications per trial have been found to be optimal (Kempthorne 1983). A large list of methods of analyses can been found in literatures and in several review papers (Lin et al 1986, Westcott 1986, Smith et al 2005). The methods for analysis of GxE interaction studies have used based on extracted patterns in form of multiplicative models for GxE interaction (Gauch 1988), multiplicative model for G+GxE

interactions (Weikai and Hunt, 2001), factor analytic representations of GxE interactions using fixed genotype effects and random environmental effects (Piepho 1997) and fixed environment effects and random variety effects (Smith et al 2001biometrics). See Smith et al (2005) for a review of mixed models used in multi-environment variety trials. Singh et al. (1996) using information on genotype means and standard errors in multi-location trials assessed thee varieties using indices measuring inter-site transferability of varieties. The combined analysis at plot levels used to be under similar designs and under the assumption of homogeneous error variances, primarily due to limitations of computational software, but in the recent years, a much more complex models can be fitted at the plot level data with complex structures of variance-covariance matrices using GenStat, ICARDA modules in GenStat, SAS, AGROBASE and ASReml (Gilmour et al., 2009).

A large number of phenotyped data are obtained through the International Nurseries, with specific purposes, which facilitate screening and evaluation of genetic material across a wide range of environments. Experimental designs such as RCB, α- designs, augmented designs are used. Trials should have independent randomizations. CGIAR (Consultative Group for International Agricultural Research) centers use an ICIS (International Crop Information System) for generating randomizations and storage and retrieval of crop information in terms of genotype pedigree and response data.

Inheritance Studies form a part of the genetics of the traits used in phenotyping through the use of specific mating designs such as such as complete/partial diallel crosses and line x tester to generate

information on the gene actions controlling the traits in terms of genetic ratios, genetic variance and its components (such as additive, dominance, and allelic interactions of various orders). Embedding of mating and environmental designs derived from incomplete crosses and blocks are discussed and reviewed in Singh et al. (2012).

REFERENCES

Agronomix Software, Inc. (1999). Agrobase 99: User's Guide & Reference Manual. Winnipeg, Manitoba. Canada.

Bari, A., Street, K., Mackey, M., Endresen, D.T.F., De Pauw, E. & Amri, A. Focused Identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. Genet Resour Crop Evol 59, 1465–1481 (2012).

Campos, H., Heard, J.E., Ibañez, M., Luethy, M. H., Peters, T. J. and Warner, D. C. (2011). I.2 Effective and efficient platforms for crop phenotype characterisation under drought. Chapter 2 of "Drought phenotyping in crops: from theory to practice" (Philippe Monneveux and Jean-Marcel Ribaut, Eds.) http://generationcp.org/onlinepubls/chapteri1/index.html.

Cochran, W.G. and Cox, G.M. (1957). Experimental designs. John Wiley & Sons.

Cox, D. R. and Reid, N. (2000). The Theory of the Design of Experiments. Chapman & Hall/CRC .

Coombes, N. (2009). DiGGer design search tool in R. (http://www.austatgen.org/software/)

Cullis, B.R., A.B. Smith, and N.E. Coombes (2006). On the design of early generation variety

trials with correlated data. Journal of Agricultural, Biological and Environmental Statistics 11:381-393.

Das, M. N. (1958). On reinforced incomplete block designs. Journal of Indian Society of Agricultural Statistics, 10:73 -77

Federer, W. T. (1955). Experimental design: theory and application. Macmillan.

Federer, W.T. (1961). Augmented designs with one-way elimination of heterogeneity. Biometrics 17: 447-473.

Federer, W. T. and Raktoe, B. L. (1965). General Theory of Prime-Power Lattice Designs. Lattice Rectangles for v = sm Treatments in sr Rows and sc Columns for $r + c = m$, $r \neq c$, and $v < 1000$. Journal of the American Statistical Association, 60: 891-904.

Federer, W.T. and Raghavarao, D. (1975). On augmented designs. Biometrics 31:29-35.

Fisher, R.A. (1990). Statistical methods, experimental design, and scientific inference: A re-issue of statistical methods for research workers, the design of experiments, and statistical methods and scientific inference. Oxford University Press, Oxford.

Gauch, H.G. (1988). Model selection and validation for yield trials with interaction. Biometrics 44:705–715.

Gilmour, A.R., Cullis, B.R. and Thompson, R. (2009). ASReml User Guide Release 3.0. VSN International Ltd.: Hemel Hempstead, HP1 1ES, UK

Hinkelmann, K. and Kempthorne, O. (2005). Design and analysis of experiments, Vol. 2: Advanced experimental design. Wiley-Interscience, New York.

Hinkelmann, K. and Kempthorne, O. (2008). Design and Analysis of Experiments, Vol.1:

Introduction to Experimental Design; 2nd edition, Wiley-Interscience, New York.

Hinkelmann, K. (ed.) (2012). Design and Analysis of Experiments, Vol.3: Special Designs and Applications, Wiley, Hoboken, NJ.

Jeffers, J. N. R. (1978) Design of experiments. (Statistical. Checklist 1). Institute of Terrestrial Ecology, Cambridge, U.K.

Jonh, J. A. and Eccleston, J. A. (1986). Row-column α designs. Biometrika, 3: 301-306.

John, J. A. and Williams, E. R. (1995). Cyclic and Computer Generated Designs (2nd ed). Chapman & Hall, London, U.K.

Kempthorne, O. (1983). The design and analysis of experiments. R.E. Krieger Publ., Malabar, Florida.

Kempton, R. A. (1984). The Design and Analysis of Unreplicated Field Trials. *Vortrage fur Pflanzenzuchtung*,7,219-242.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. B 21, 272-319.

Lin, C.S. , Binns, M.R. and Leftkovitch, L.P. (1986). Stability analysis: where do we stand?. Crop Science 26: 894-900.

Lin, C. S. and Poshinsky, G. (1983). A modified augmented designs for an early stage of plant selection involving a large number of test lines without replications. Biometrics 39:533 - 561.

Mead, R. Curnow, R. N., Hasted, A. M., Curnow, R. M. (2002). Statistical Methods in Agriculture and Experimental Biology, Third Edition. Chapman & Hall/CRC.

Payne, R.W. (editor). 2011. The guide to GenStat® Release 14. Part 2: Statistics. VSN

International, Hemel Hempstead, United Kingdom.

Patterson, H. D. and Williams, E.M. (1976). A new class of resolvable incomplete block design. Biometrika 63: 83-92.

Patterson, H. D., Williams, E. R. and Hunter, E. A. (1978). Block designs for variety trials. Journal of Agricultural Science, Cambridge 90: 395-400.

Piepho, H.-P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics 53, 761–767.

Pearce, S. C. (1975). Row-and-column designs. Applied Statistics 24: 60-74.

Singh, M. and Dey, A. (1979). Block designs with rested rows and column. Biometrika 66: 321-326.

Singh, M., Yau, S.K., Hamblin, J. and Porceddu, E. (1996). Inter-site transferability of crop varieties: another approach for analyzing multi-locational variety trials. Euphytica 89: 305-313.

Singh, M., R. S. Malhotra, S. Ceccarelli, A. Sarker, S. Grando, and W. Erskine (2003). Spatial variability models to improve dryland field trials. Journal of Experimental Agriculture, 39:151-160.

Singh, M., S. Gupta and R. Prasad (2012). Genetic Crosses Experiments. Pages 1-71 in Design and Analysis of Experiments, Volume 3: Special Designs and Applications, John Wiley & Sons (Editor: Dr K. Hinkelmann)

Smith, A. B., Cullis, B.R. and Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57,

1138–1147.

Smith A. B., Cullis, B. R. and Thomson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143: 449–462.

SAS Institute Inc. (1989) SAS/STAT User's Guide, Version 6, Fourth Edition, Volumes 1 & 2, Cary, NC: SAS Institute Inc.

Tuberossa, R. (2011). I. 1. Phenotyping drought-stressed crops: key concepts, issues and approaches. Chapter 1 of "Drought phenotyping in crops: from theory to practice" (Philippe Monneveux and Jean-Marcel Ribaut, Eds.) http://generationcp.org/onlinepubls/chapteri1/index.html.

Westcostt, B. (1986). Some methods of analyzing genotype-environment interaction. Heredity 56: 243-252.

Whitaker, D., E.R. Williams and J.A. John (2002). CycDesigN: A Package for the Computer Generation of Experimental Designs, Version 2. CSIRO Forestry and Forest Products, Canberra, Australia.

Yan, W., and L.A. Hunt. 2001. Interpretation of genotype by environment interaction for winter wheat yield in Ontario. Crop Sci. 41:19–25.

Yates, F. (1940). Lattice squares. Journal of Agricultural Science, 30: 672-687.

Youden, W. J. (1940). Experimental designs to increase accuracy of greenhouse studies. Contr. Boyce Thompson Inst. 11, 219-228.

Table 1. Layout of an augmented design in blocks containing test entries numbered from 1 to 45 and check entries numbered from 46 to 48.

Blocks Plots	1	2	3	4	5	6	7	8	9
1	47	23	41	43	14	36	25	21	22
2	46	35	26	48	47	13	48	48	1
3	9	47	46	47	16	4	47	3	5
4	28	46	33	15	31	47	12	47	47
5	48	48	47	29	48	48	32	24	48
6	38	19	6	46	34	46	46	46	44
7	7	45	48	17	46	20	11	18	46
8	39	10	30	37	2	40	27	8	42

Table 2. A randomized plan of a Latin Square design in six genotypes numbered 1 to 6

Columns	1	2	3	4	5	6
Rows						
1	1	2	4	5	6	3
2	4	5	1	2	3	6
3	6	4	3	1	2	5
4	2	3	5	6	4	1
5	5	6	2	3	1	4
6	3	1	6	4	5	2

Table 3 A randomized plan for a randomized complete block design in four replications and ten genotypes numbered 1 to 10.

Rep	1	2	3	4
Plots				
1	3	11	7	3
2	2	2	12	8
3	11	12	4	2
4	4	3	2	7
5	9	7	9	10
6	1	6	10	12
7	7	4	6	1
8	12	1	3	6
9	10	9	11	9
10	5	5	5	11
11	6	8	8	5
12	8	10	1	4

Table 4. A randomized plan for an alpha design in 40 genotypes, incomplete blocks of size 5 and 3 replications

	Plots	1	2	3	4	5
Replicates	Blocks					
1	1	7	40	12	30	36
	2	39	34	33	29	1
	3	26	32	4	31	14
	4	9	2	21	20	10
	5	25	17	6	23	19
	6	24	8	11	3	22
	7	15	37	13	5	28
	8	16	35	38	18	27
2	1	36	38	11	5	32
	2	3	13	16	39	30
	3	40	31	6	20	1
	4	27	8	25	28	33
	5	17	9	26	34	37
	6	2	29	24	23	15
	7	21	35	7	19	14
	8	12	22	10	4	18
3	1	33	2	17	16	32
	2	9	40	27	15	14
	3	20	8	26	23	30
	4	35	12	24	31	34
	5	11	4	19	39	28
	6	25	29	13	36	10
	7	37	22	6	7	38
	8	21	3	1	18	5

Table 5. A randomized plan for an alpha design in 29 genotypes, incomplete blocks of sizes 4 and 5, and 3 replications

Replicates	Plots / Blocks	1	2	3	4	5
1	1	17	15	18	5	29
	2	3	8	1	9	16
	3	28	12	#	22	7
	4	2	26	23	19	24
	5	11	21	27	4	25
	6	10	14	6	20	13
2	1	29	10	25	7	8
	2	#	21	15	1	26
	3	3	4	23	28	14
	4	5	22	19	20	9
	5	2	16	13	27	17
	6	12	6	11	18	24
3	1	16	6	15	25	22
	2	23	8	#	11	17
	3	18	26	20	7	4
	4	14	27	19	12	1
	5	28	13	9	29	24
	6	3	21	2	5	10

#: the empty plot need not be retained or, if required, could be filled by a suitable filler check.

Table 6. A randomized plan, on an 8 × 8 layout, of a p-rep design in 33 genotypes numbered 1 - 20 have no replications, 21 - 30 have two replications and 31 - 33 are checks with 8 replications

Rows\columns	1	2	3	4	5	6	7	8
1	4	33	20	29	24	5	31	32
2	31	9	25	23	1	32	19	27
3	14	30	21	10	31	18	33	28
4	32	13	33	17	21	27	24	31
5	6	28	23	32	12	26	30	33
6	31	16	22	3	33	11	32	15
7	7	32	26	33	22	31	8	32
8	33	29	31	25	32	2	33	31

Appendix

Some key codes used in generating the experimental designs under various tables.

A1. GenStat Code for Table 2 (`Geno` stands for genotypes)

```
AGLATIN [PRINT=design; ANALYSE=Yes] NROWS=6; NSQUARES=1; \
   TREATMENTFACTORS=!p(Geno); ROWS=Rows; COLUMNS=Columns; SEED=27257
```

A2. GenStat code for Table 3 (`Rep, Plots` and `Geno` stand for replications or complete blocks, plots within block and genotypes respectively)

```
 AGHIERARCHICAL [PRINT=design; ANALYSE=Yes;SEED=2534] \
  BLOCKFACTORS=Rep,Plots; TREATMENTFACTORS=*,!p(Geno); LEVELS=4,12
```

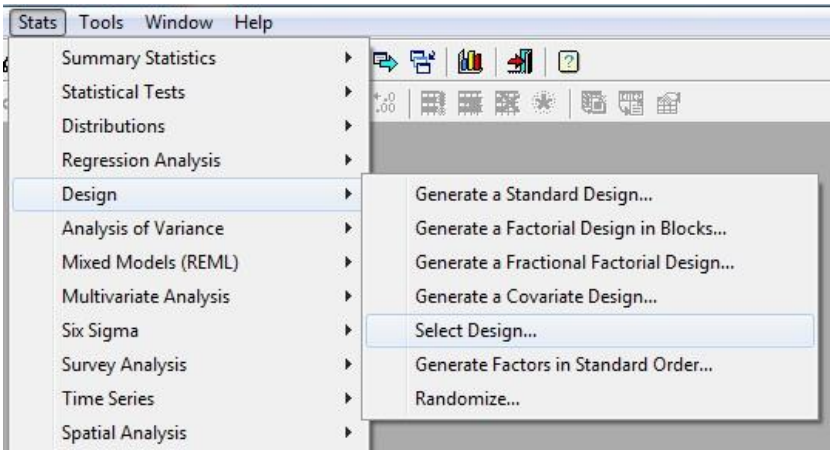
A3. R Language code for Table 6

```
library(DiGGer)
trep <- rep(c(1, 2, 8), c(20, 10, 3))
design <- DiGGer(33, 8, 8, TreatmentRep = trep)
design <- run(design)
getDesign(design)
layout <- getDesign(design)
des.plot(layout, seq(1, 20), col = 5, new = TRUE)
des.plot(layout, seq(21, 30), col = 6, new = FALSE)
des.plot(layout, seq(31, 33), col = 7, new = FALSE)
```

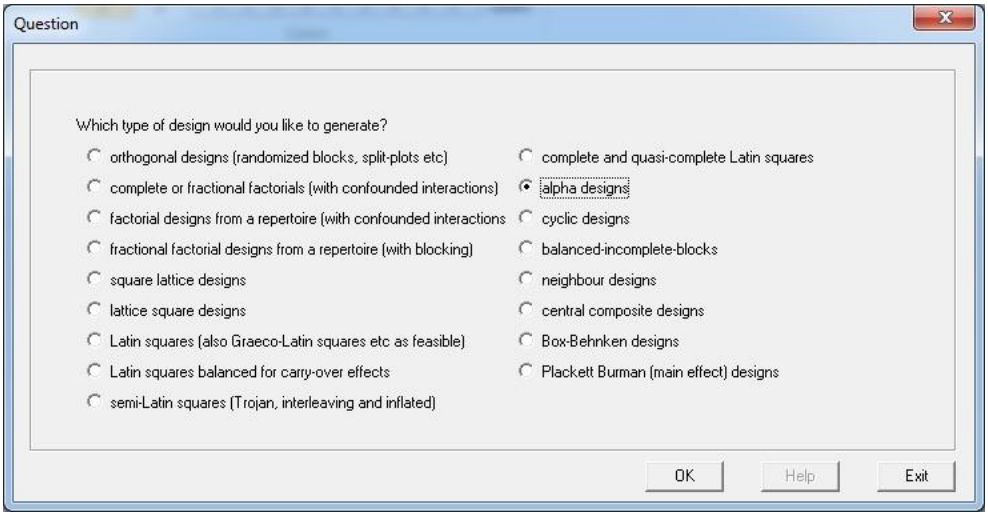
A4. Further details on GenStat menu and R-program

A4.1 ***Generate an α- design using GenStat:***

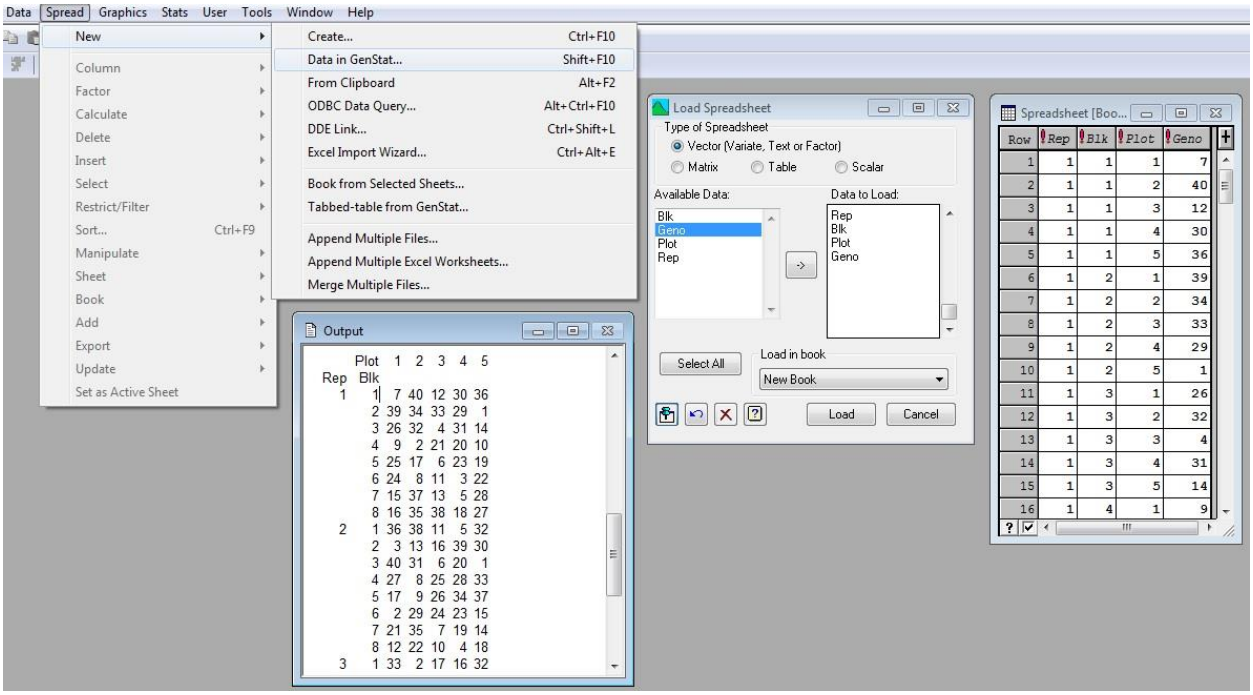
To generate randomizations using GenStat statistical package go to its “Stats” menu, “Design” sub menu, then “Select Design ...” item, (see the screenshot below):



This will pop-up the dialog box listing several special analyses (see the screenshot below):



Select “alpha designs” option then click “OK” button and answer the series of questions on number of treatments (within the range 20 - 100), number of blocks per replication, number of replications, and the labels that should be assigned to the factors. Using the “Spread” menu and further “Data in GenStat”, item from “New” sub menu, one can obtain the randomized plan in the GenStat spreadsheet as shown in the following screenshot. For more than 100 genotypes, one may use CycDesigN software (Whitaker et al. 2002).



The plan in Table 4, in 40 genotypes in blocks of size 5 and 3 replications, can be obtained by running the following code.

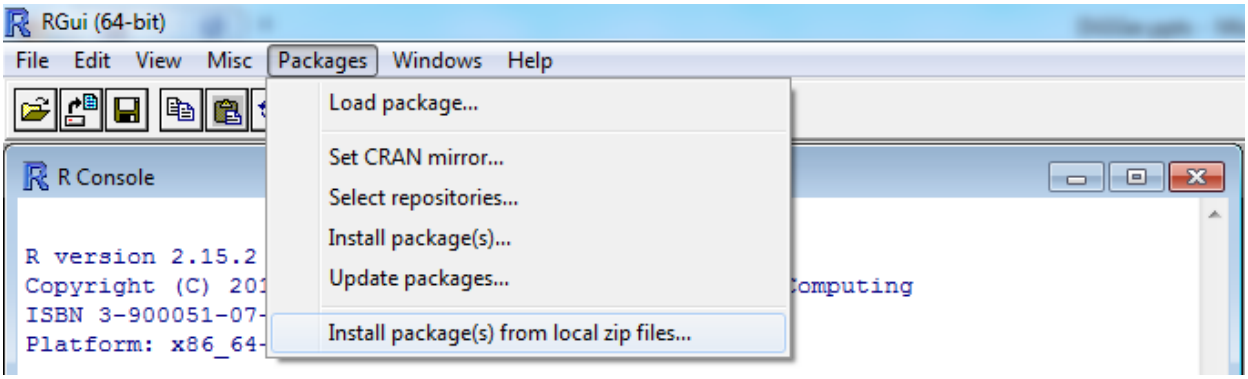
```
AGALPHA [PRINT=design] LEVELS=40; NREPLICATES=3; NBLOCKS=8;\
  TREATMENTS=Geno;\
  REPLICATES=Rep;\
  BLOCKS=Blk;\
  UNITS=Plot;\
  SEED=1592654
```

A4. 2. R- package DiGGer codes for Table 6.

Generate Design for Partial Replications using DiGGer and R Language:

To use DiGGer tool, one needs to carry out required installation for the R package and download

the following zip-files "R.methodsS3_*.zip"[1], "R.oo_*.zip"[2], and "DiGGer_*.zip"[3] where " * " in the filenames denotes the current version available. Then one may start the R program , go to the "Packages" menu and select "Install package(s) from local zip files…". Find the downloaded files and let R install them.



Once DiGGer packages are installed, the following codes are used to generate the experimental design in the Table 6.

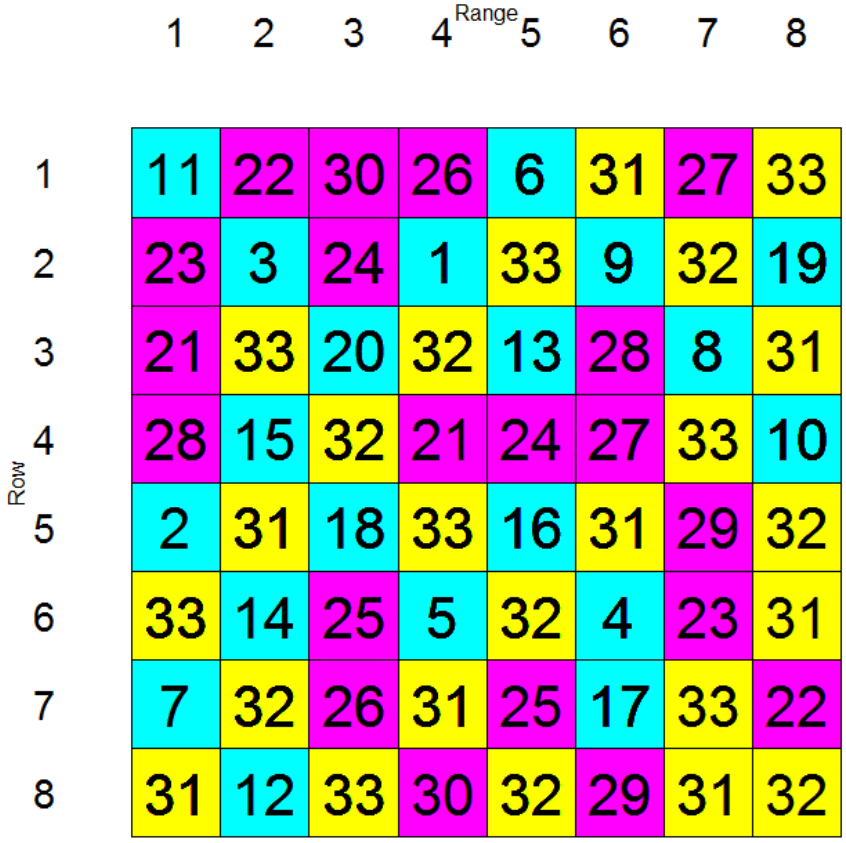
```
# load required package
library(DiGGer)

# 20 genotypes with no replications [1  - 20]
# 10 genotypes with 3  replications [21 - 30]
# 3  genotypes with 8  replications [31 - 33]
trep <- rep(c(1, 2, 8), c(20, 10, 3))

# in total we have 33 genotypes (i.e. 20 + 10 + 3)
# in total we have 64 plots (i.e. 20*1 + 10*2 + 3*8)
# field layout set as 8 rows x 8 columns
design <- DiGGer(33, 8, 8, TreatmentRep = trep)
```

[1] http://www.austatgen.org/files/software/downloads
[2] http://cran.rstudio.com/web/packages/R.methodsS3/index.html

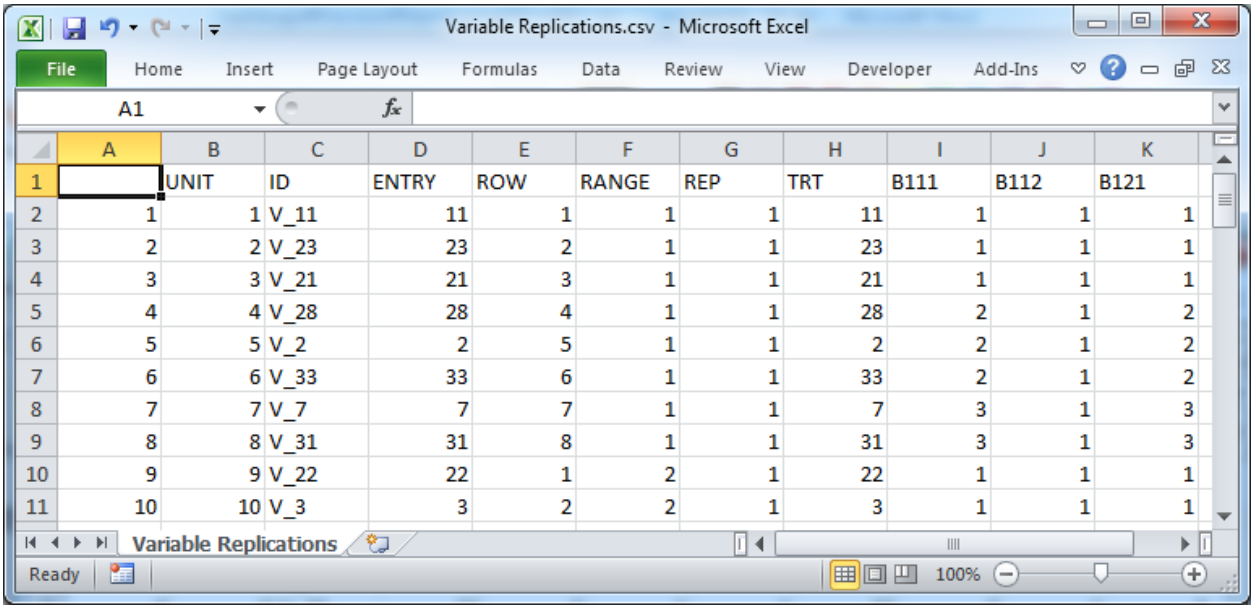
```
517  # once the design search object has been created
518  # we can produce the design
519  design <- run(design)
520
521  # extracting matrix of design numbers
522  layout <- getDesign(design)
523
524  # draw colored field layout
525  # or you may simply use plain plot(design) function in this case
526  des.plot(layout, seq(1, 20), col = 5, new = TRUE)
527  des.plot(layout, seq(21, 30), col = 6, new = FALSE)
528  des.plot(layout, seq(31, 33), col = 7, new = FALSE)
```




```
529
530  # export into CSV file
```

[3] http://cran.rstudio.com/web/packages/R.oo/index.html

```
write.csv(design$dlist, "Variable Replications.csv")
```



	A	B	C	D	E	F	G	H	I	J	K
1		UNIT	ID	ENTRY	ROW	RANGE	REP	TRT	B111	B112	B121
2	1	1	V_11	11	1	1	1	11	1	1	1
3	2	2	V_23	23	2	1	1	23	1	1	1
4	3	3	V_21	21	3	1	1	21	1	1	1
5	4	4	V_28	28	4	1	1	28	2	1	2
6	5	5	V_2	2	5	1	1	2	2	1	2
7	6	6	V_33	33	6	1	1	33	2	1	2
8	7	7	V_7	7	7	1	1	7	3	1	3
9	8	8	V_31	31	8	1	1	31	3	1	3
10	9	9	V_22	22	1	2	1	22	1	1	1
11	10	10	V_3	3	2	2	1	3	1	1	1