# IBFIELDBOOK, AN INTEGRATED BREEDING FIELD BOOK FOR PLANT BREEDING

## IBFIELDBOOK, UN SISTEMA INTEGRADO DE LIBROS DE CAMPO PARA MEJORAMIENTO DE PLANTAS

**Oziel Lugo-Espinosa[1], Tito M. Sánchez-Gutiérrez[1], J. Gamaliel Camarena-Sagredo[1], Mateo Vargas[1], Gregorio Alvarado[1], Diego Jarquin[1], Juan Burgueño[1], Jose Crossa[1] and Héctor Sánchez-Villeda[1, 2]***

[1]Centro Internacional de Mejoramiento de Maíz y Trigo. Km 45 carr. México-Veracruz. 56130, El Batán, Texcoco. Edo. de México. [2]G2Apps IT Engineering. Prolongación Corregidora Norte 909 – C. Villas del Parque. 76140, Querétaro ,Qro.

*Autor para correspondencia* (hector.sanchez@g2Apps.net)

## SUMMARY

The development of an integrated breeding field book (IBFieldbook) for different crops involves the generation, handling and analysis of large amounts of data. Managing the integration of environmental, pedigree, and phenotypic information for plant breeding data analyses requires appropriate and successful software that facilitates breeders, technicians, and researchers management of the vast collected field information in an easy, efficient and interactive way. Users may also need methods to exchange information with different devices used to record information in the field. Additionally, collected information needs to be analyzed inside or outside the application, and then generate reports for germplasm improvement.

**Index words:** data analysis, information collected, large amounts of data, methods to exchange information, successful software, freely available.

## RESUMEN

El desarrollo de un sistema integrado de libros de campo para mejoramiento (IBFieldbook, por sus siglas en inglés) para diferentes cultivos incluye la generación, manejo y análisis de grandes volúmenes de datos. El manejo de integración de la información fenotípica, ambiente y pedigrí para análisis de datos requiere de un programa computacional ("software") apropiado y exitoso, que facilite a los mejoradores, técnicos e investigadores la administración de la información recolectada en el campo de manera fácil, eficiente e interactiva. Los usuarios del programa también necesitan métodos para intercambiar información con diferentes dispositivos usados para almacenar información en el campo. La información recolectada necesita además ser analizada adentro y afuera de la aplicación, y generar reportes para el mejoramiento del germoplasma.

**Palabras clave:** análisis de datos, información recolectada, grandes volúmenes de datos, métodos para intercambiar información, software exitoso, libremente disponible.

# IBFIELDBOOK, AN INTEGRATED BREEDING FIELD BOOK FOR PLANT BREEDING

## IBFIELDBOOK, UN SISTEMA INTEGRADO DE LIBROS DE CAMPO PARA MEJORAMIENTO DE PLANTAS

**Oziel Lugo-Espinosa[1], Tito M. Sánchez-Gutiérrez[1], J. Gamaliel Camarena-Sagredo[1], Mateo Vargas[1], Gregorio Alvarado[1], Diego Jarquin[1], Juan Burgueño[1], Jose Crossa[1] and Héctor Sánchez-Villeda[1, 2]\***

[1]Centro Internacional de Mejoramiento de Maíz y Trigo. Km 45 carr. México-Veracruz. 56130, El Batán, Texcoco. Edo. de México. [2]G2Apps IT Engineering. Prolongación Corregidora Norte 909 – C. Villas del Parque. 76140, Querétaro ,Qro.

*\*Autor para correspondencia* (hector.sanchez@g2Apps.net)

## SUMMARY

The development of an integrated breeding field book (IBFieldbook) for different crops involves the generation, handling and analysis of large amounts of data. Managing the integration of environmental, pedigree, and phenotypic information for plant breeding data analyses requires appropriate and successful software that facilitates breeders, technicians, and researchers management of the vast collected field information in an easy, efficient and interactive way. Users may also need methods to exchange information with different devices used to record information in the field. Additionally, collected information needs to be analyzed inside or outside the application, and then generate reports for germplasm improvement.

**Index words:** data analysis, information collected, large amounts of data, methods to exchange information, successful software, freely available.

## RESUMEN

El desarrollo de un sistema integrado de libros de campo para mejoramiento (IBFieldbook, por sus siglas en inglés) para diferentes cultivos incluye la generación, manejo y análisis de grandes volúmenes de datos. El manejo de integración de la información fenotípica, ambiente y pedigrí para análisis de datos requiere de un programa computacional ("software") apropiado y exitoso, que facilite a los mejoradores, técnicos e investigadores la administración de la información recolectada en el campo de manera fácil, eficiente e interactiva. Los usuarios del programa también necesitan métodos para intercambiar información con diferentes dispositivos usados para almacenar información en el campo. La información recolectada necesita además ser analizada adentro y afuera de la aplicación, y generar reportes para el mejoramiento del germoplasma.

**Palabras clave:** análisis de datos, información recolectada, grandes volúmenes de datos, métodos para intercambiar información, software exitoso, libremente disponible.

## INTRODUCTION

Nowadays maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), and rice (*Oryza sativa* L.) are the three crops that feed 95 % of the population worldwide (FAO, 2013). To help researchers to genetically improve these crops, it is necessary to provide them with the appropriate tools to manage, use and exploit their field information. The main objective of all breeders is to improve the germplasm to reduce the hunger in the world. To achieve this goal, it is important to have software tools that allow automatic information recollection in the field, either typed, bar code scanned or acquired through Bluetooth devices. Additionally, information must be shared and made it accessible for easy management and analysis. To achieve these goals, the integrated breeding field book (IBFieldbook) was developed to facilitate breeders creation of their field books with export and import capabilities to hand held devices and Excel© spreadsheets and standard analyses with the collected information.

It is well known that breeders at different agricultural experiment stations have requirements that vary considerably depending on the objectives of their projects. It is possible to collect accurate phenotypic data with the information technology available today, which would make possible genomic selection and prediction of genetic values of germplasm. Precise phenotypic information is obtained from data collected in target environments by means of adequate experimental designs (Banziger *et al.*, 2000; Ortiz *et al.*, 2007).

Currently, many breeders and technicians at different experimental stations in the world use spreadsheets to manage their data, but in developing countries there is lack of information systems and analysis tools, along with a poor phenotyping infrastructure (McLaren *et al.*, 2005; Ribaut *et al.*, 2010). If breeders want to store and manage their data in an easy and efficient manner, they must either develop an information system, which can be cumbersome, time consuming and expensive, or they need to purchase a system. Therefore an open information system for managing breeding related data is needed.

This paper describes and discusses the integrated breeding field book (IBFieldbook) as one component of the

molecular breeding platform (Delannay *et al.,* 2011) that was developed to aid breeders during the plant breeding process: (a) To generate nurseries for line advancement with and without randomization; (b) To generate trials for germplasm evaluation in different locations using different randomizations; (c) To allow data management with record keeping, reporting, and retrieving; (d) To ensure data quality and accessibility to the scientific community, and (e) To generate standard analysis for one and multiple locations.

## MATERIALS, METHODS AND IMPLEMENTATION OF THE IBFieldbook

The integrated breeding fieldbook system described here is an example of an application of informatics to practical agronomic and plant breeding challenges. An overview of the IBFieldbook components and their functions is shown in Table 1.

Several open source software were used to develop the IBFieldbook system. The Java© programming language used provides users with a modular and intuitive user in-

terface (UI). The UI focuses on creating an easy to use system that facilitates information handling and analysis of information to help in breeding activities and improve the germplasm for different crops. The UI is a high-performing, cross-platform Java© client application, which can be executed in Microsoft Windows©, Mac OS X© or any type of IX Operating System. The IBFIeldbook system uses MySQL© for the backend, with a dual database (central and local) containing the same structure to keep track of public and private information. The client-server architecture system permits many concurrent users to access the central database and also makes their private local information available. Java© database connectivity (JDBC) provides interoperability with the databases and shells connectivity with R package. R software uses scripts developed to generate the standard analysis (See section of managing phenotypic data, below).

The programming language used to code the integrated breeding field book was Java 1. 6. 0© using NetBeans 7.1.2© as a developer platform. To collect the data, either Microsoft Access© or MySQL versions 5.0© were used as the backend; however, other relational databases can be

**Table1. IBFieldbook components and their functions.**

| IBFieldBook component | Description |
| --- | --- |
| Traits management | Manages traits, scales, and methods to collect information in the field or in the wet lab by specifying validation ranges. |
| Germplasm lists generation | Generates germplasm lists that are used for managing specific lines for evaluation or advancing germplasm. |
| Trials generation | Generates trials with specified designs, export them either to hand-held devices or excel spreadsheets to evaluate different germplasm for yield, diseases, or other important traits of interest in plant breeding or in the wet lab. This tool also helps to generate statistical analysis after collecting the information for further germplasm development. |
| Nurseries generation | Generates nurseries for advancing lines during the germplasm improvement process, as well as export and import functions to handle information in different hand-held devices or to spreadsheets. |
| Inventory management | Keeps track of the seed inventory using easy quick deposit withdrawal functions after harvesting or before planting. |

used with little tuning. The application interfaces with R Statistical Package© for field trial design and data analysis. IBFieldbook provides data management for the information collected in the field and generates either single or multi-location trials for data analysis. It also manages catalogs using UI components for easy information management located in the presentation layer.

This system provides a middleware, which is composed of the service layer and data access layer using the Spring© framework to send and receive objects from the service layer to the presentation layer, and it also uses Hibernate© to communicate between the data access layer and service layer through entities. Spring provides a lightweight container which allows injecting required objects into others objects; this results in a design where the Java Class are not hard coded. A design based on independent classes/components increases re-usability and software testability. The data access layer gets and receives information through records from the local and central databases using Microsoft Access® or MySQL© with Java© database connectivity (JDBC) as shown in Figure 1.

### Automation using Java© and R software for data analysis

There are several steps used to generate single and multilocation analyses that start after the information is collected in the field and stored into the database. First the information has to be exported to a coma separated values (csv) format file used by R software. The csv-file is automatically uploaded, the standard R scripts and the needed parameters are copied, and a batch file generated for Windows© or Mac OS X© operating systems. This file is passed on to R and processed using a shell command in the R environment without user interaction. Finally the result files are returned after the analyses are done (Figure 2).

### Managing phenotypic data by the IBFieldbook

The IBFieldbook stores a set of interrelated tables in a relational database (ICISDB) that supports pedigree, phenotypic, and inventory information for the germplasm. This provides easy editing, querying, importing and exporting functions from trials and nurseries, shown in the entity relationship diagram [Available at: http://www.revistafitotecniamexicana. org/documentos/36-3/ICISdbschema.pdf]. All of this makes the system a user-friendly field book generation tool, using Java© interface components to facilitate data access (Figure 3).

IBFieldbook accepts comma separated values format used in hand held devices, as well as a pre-established Mi-

crosoft Excel® format. Once the information is imported into the system, it can be edited in an intuitive user-friendly interface and also can be saved to the database to be exported to a specific R format for statistical analysis.

The steps to generate a trial or nursery for collecting information are as follows:

- Generate the germplasm list.
- Generate a template or use one that come as example with the traits to measure
- Generate a study; a set of trial and nurseries
- Generate a trial or a nursery using the wizard or quick creation interfaces
- Collect the phenotypic information in the system or export to hand held or Excel© for collection

The best possible way to collect phenotypic information and avoid errors is by means of automating the data collection as much as possible.

### Statistical models for single and multienvironment plant breeding trials

For the description of the mean response of genotypes over environments and for studying and interpreting genotype ´ environment interaction (GE) in multi environment plant breeding trials (MET), the baseline conventional two-way model used to explain the empirical mean response, $\bar{y}_{ij}$, of the $i_{th}$ genotype (I = 1, 2, …, I) in the $j_{th}$ environment (j = 1, 2,…, J) with n replications in each of the I x J cells is expressed as

$$\bar{y}_{ij} = \mu + \tau_i + \delta_j + (\tau\,\delta)_{ij} + \varepsilon_{ij} \qquad \text{(Eqn. 1)}$$

Where $\mu$ is the grand mean over all genotypes and environments, $\tau_i$ is the additive effect of the $i_{th}$ genotype, $\delta_j$ is the additive effect of the $j_{th}$ environment, $(\tau\delta)_{ij}$ is the non-additivity interaction GE of the $i_{th}$ genotype in the $j_{th}$ environment, and $\bar{\varepsilon}_{ij}$ is the average error assumed to be NID (0, $\sigma^2$ /n)(where $\sigma^2$ is the within-environment error variance, assumed to be constant). The model for a single environment is represented by $\bar{y}_i = \mu + \tau_i + \bar{\varepsilon}_i$. The single and multi-environment model can be used to analyze as many traits as it is required.

The simplest linear model for assessing GE was proposed by Yates and Cochran (1938), where the GE is expressed as the linear regression coefficient of the $i_{th}$ genotype value on the environmental mean. This approach was later used by Finlay and Wilkinson (1963) and modified by Eberhart and Russell (1966). Williams (1952) was the first author to link the fixed effects two-way model (Eqn. 1) with principal
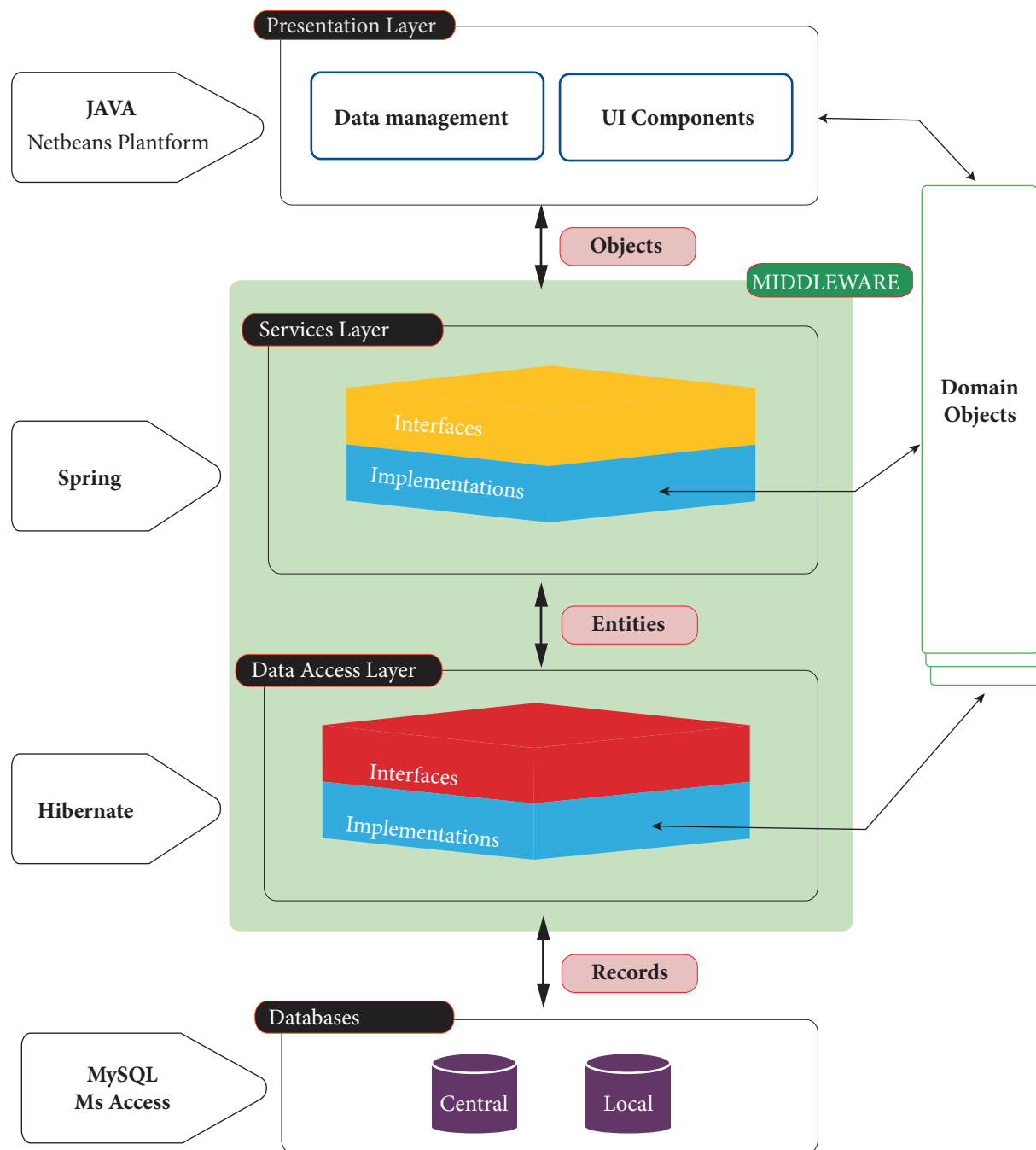
**Figure 1. Implementation of the IBFieldbook system. The presentation layer shows the user interface components and data management; the service layer shows the interfaces and implementations of objects; the data access layer shows the interfaces and implementations of entities; and the databases represent the storage of information. All the layers are connected through domain objects, and transfer objects, entities, and records.**
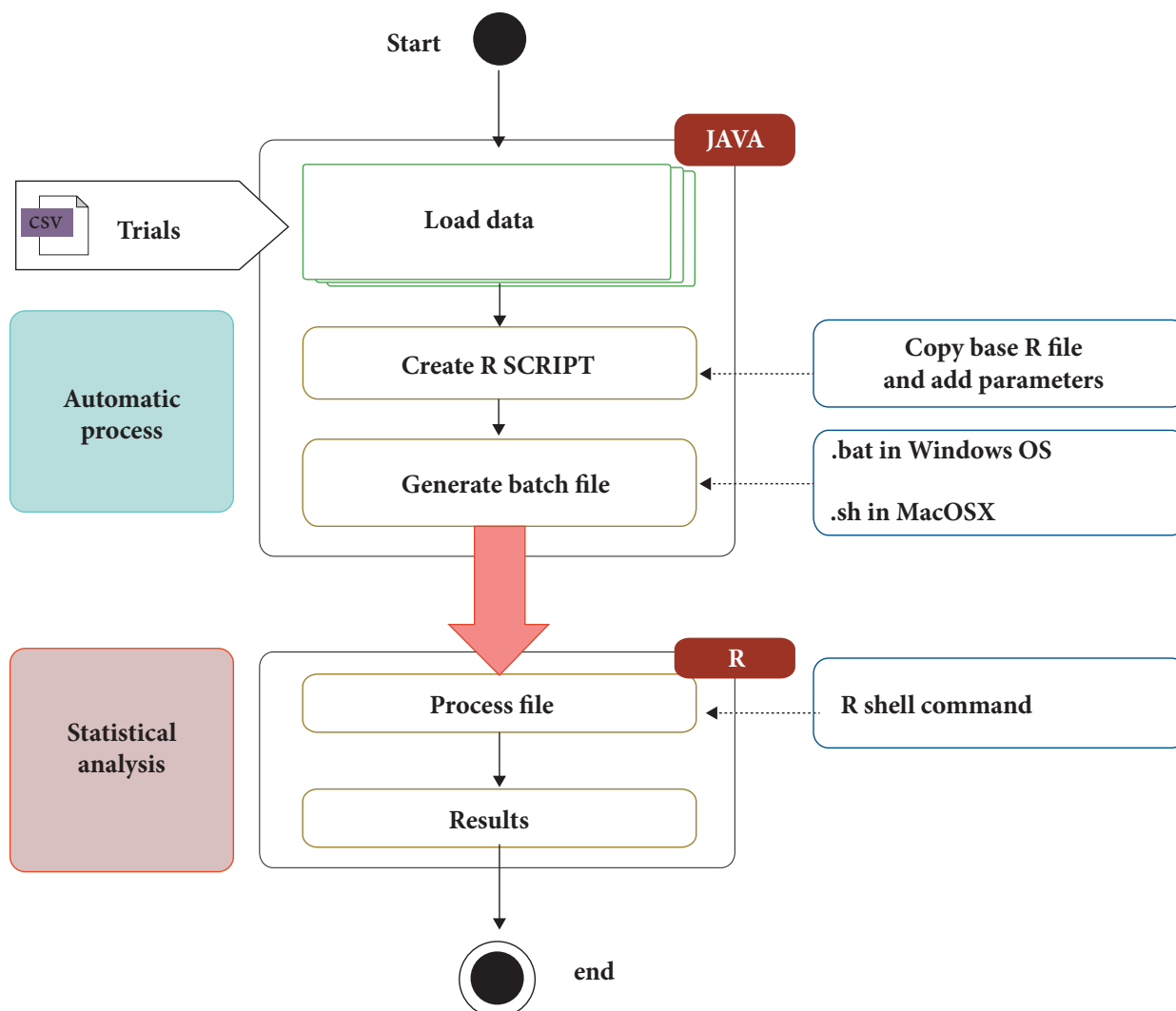
**Figure 2. Automation of data analysis using Java© and R. The processes used for loading and generating the input files used for R package are shown in the JAVA square, and the R square shows the process and result files generated during the data analysis.**

components analysis by considering the model $\bar{y}_{ij} = \mu + \tau_i + \lambda \alpha_i \gamma_j + \bar{\epsilon}_{ij}$, where $\lambda$ is the largest singular value of $ZZ'$ and $Z'Z$ (for $Z = \bar{y}_{ij} - \bar{y}_i$) and $\alpha_i$ and $\gamma_j$ are the corresponding eigenvectors. An extension to Williams work was done by considering the bilinear GEI term as $(\tau\delta)_{ij} = \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk}$. Thus, the general formulation of this linear-bilinear model is

$$\bar{y}_{ij} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij} \qquad \text{(Eqn. 2)}$$

Where the constant $\lambda_k$ is the singular value of the $k_{ith}$ multiplicative component that is ordered $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_t$; the terms $\alpha_{ik}$ are elements of the $k_{ith}$ left singular vector of the true interaction and represents the genotypic sensitivity to hypothetical environmental factors represented by the $k_{ith}$ right singular vector with elements $\gamma_{jk}$. The $\alpha_{ik}$ and $\gamma_{jk}$

satisfy the ortho-normalization constraints $\Sigma_i \alpha_{ik} \alpha_{ik'} = \Sigma_j \gamma_{jk} \gamma_{jk'} = 0$ for $k \neq k'$ and $\Sigma_i \alpha^2_{ik} = \Sigma_j \gamma^2_{jk} = 1$. This model was called "additive main effects and multiplicative interaction (AMMI)" model by Zobel *et al.* (1988) and Gauch (1988).

A biplot is a simultaneous representation of the rows and columns of a two-way array of data in terms of directions and projections Gabriel (1978). The biplot is designed to operate with one matrix (two-mode), and the singular vector of the genotypes and environmental scores for the first two components are drawn from the origin to the end of the vector. Vectors of genotypes (or environments) that are related are parallel (or near parallel), whereas vectors with angles greater than 90 % represent negligible or negative relationship of genotypes (or environments). Furthermore, vectors of genotypes and environments pointing

**Figure 3. Trial field book used for collecting phenotypic information. Experimental design is defined on the entries section and germplasm names are used for the field book. The traits section show the phenotypic values collected from plant height, senescence, and ear color.**

on the same direction denote positive GE between those genotypes and those environments, whereas vectors of genotypes and environments pointing in opposite directions denote negative GE.

## DISCUSSION

There are several packages that can be used to generate field books for breeding germplasm, as well as to generate trials for evaluation, with exporting and importing functions, using several statistical packages to evaluate information recorded in the field or in the lab with different types of devices. However, these packages are very expensive, require a lot of training, and are not user friendly. Therefore we programmed a package that can handle several crops, is comprehensive and user friendly, at no cost for the final user (Table 2).

The IBFieldbook was designed to meet the challenges of collecting, managing, and analyzing phenotypic information for different crops, in different experimental stations around the world. This tool has been used in nine crops (beans, cassava, chickpeas, cowpea, groundnuts, maize, rice, sorghum, and wheat) handling more than 150 000 pedigrees; more than 11 000 different field trials in different locations worldwide for the Consultative Group on International Agricultural Research (CGIAR) centers. Related information can be downloaded at: https://www.integratedbreeding.net/ib-tools/project-planning-and-queries/crop-databases-ib-workbench.

The IBFieldbook is a complete system for breeding management, and the software is freely available to the public. The system includes the trait management for all the variables that are measured in the field, export and import functions to work with the information into hand held devices or Microsoft Excel© software. The system also allows direct modification or addition of information using a programmed graphical user interface, and to run statistical analysis with the information collected interacting directly with R package. It has printing facilities and stores the information into a relational database management system.

## CONCLUSION

The IBFieldBook is a freely available open source software tool that helps breeders, technicians, and molecular breeders to generate their field books, collect phenotypic information in the field, analyze their collected data for single-environment and multi-environment statistical analyses, store the data in a comprehensive database, and report their results, thus helping users improve their germplasm. It was successfully tested in Africa, Asia, and Latin America. Free downloadable copies of the application and examples of databases for the nine crops can be obtained by filling out the form at: http://www.g2apps.net/index.php?consulta=PRODUCTOS&principal=IB-FIELDBOOK.

**Table 2. Feature comparisons of different software packages used to generate field books.**

| Legend n/l – Not listed in article or on software website | IBFieldbook | Prism Central Software Solutions, Inc. (2012) | AGROBASE (2012) | IMIS-Fieldbook (Banziger and Vivek, 2007)[1] |
|---|---|---|---|---|
| Freely available to public | Y | N | N | Y |
| Handle multiple crops | Y | Y | Y | N |
| Validation and verification of scores from traits | Y | Y | Y | Y |
| Capability to add traits, scales, methods | Y | Y | Y | Y |
| Capability to generate statistical analysis | Y | Y | Y | Y |
| Easy user friendly interface | Y | Y | Y | N |
| Work in multiple operating system | Y | N | N | N |

[1]**Banziger M, B S Vivek (2007)** IMIS-Fieldbook Software for managing a maize breeding program. CIMMYT.

## ACKNOWLEDMENTS

## BIBLIOGRAPHY

**AGROBASE (2012)** Generation II® Plant and Crop Breeding Software. http://www.agronomix.com.

**Banziger M, G O Edmeades, D Beck, M Bellon (2000)** Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. CIMMYT.

**Central Software Solutions, Inc. (2012)** PRISM® http://www.teamcssi.com.

**Delannay X, G McLaren, J M Ribaut (2011)** Fostering molecular breeding in developing countries. Mol. Breed. 29:857-873.

**Eberhart S A, W A Russell (1966)** Stability parameters for comparing varieties. Crop Sci. 6:36-40.

**FAO Statistical Yearbook (2013)** World food and agriculture. Food And Agriculture Organization Of The United Nations, Rome, Part 3, Feeding the world. http://www.fao.org/docrep/018/i3107e/i3107e00.htm.

**Finlay K W, A A Wilkinson (1963)** The analysis of adaptation in a plant breeding program. Austr. J. Agric. Res. 14:742-754.

**Gabriel K R (1978)** Least Squares Approximation of Matrices by Additive and Multiplicative Models. J. Royal Stat. Soc., Series B 40:186-196.

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

**McLaren C G, R M Bruskiewich, A M Portugal, B Cosico (2005)** The international rice information system. A platform for meta-analysis of rice crop data. Plant Physiol. 139:637-642.

**Ortiz R, R Trethowan, G O Ferrara, M Iwanaga, J H Dodds, J H Crouch, J Crossa, H J Braun (2007)** High yield potential, shuttle breeding, genetic diversity, and a new international wheat improvement strategy. Euphytica 157:365-384.

**Ribaut J M, de M C Vicente, X Delannay (2010)** Molecular breeding in developing countries: challenges and perspectives. Curr. Opin. Plant Biol. 13:1–6.

**Williams E J (1952)** The interpretation of interactions in factorial experiments. Biometrika 39:65-81.

**Yates F, W G Cochran (1938)** The analysis of groups of experiments. J. Agric. Sci. 28:556–580.

**Zobel R W, M J Wright, H G (1988)** Gauch Jr Statistical analysis of a yield trial. Agron. J. 80:388-393.