Article

Research on Hybrid Crop Breeding Information Management System Based on Combining Ability Analysis

Yan-yun Han [1,2,3,4]**, Kai-yi Wang** [1,2,3,4,]*****, Zhong-qiang Liu** [1,2,3,4]**, Shou-hui Pan** [1,2,3,4]**,**
Xiang-yu Zhao [1,2,3,4]**, Qi Zhang** [1,2,3,4] **and Shu-feng Wang** [1,2,3,4]

[1] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China;
hanyy@nercita.org.cn (Y.-y.H.); liuzq@nercita.org.cn (Z.-q.L.); pansh@nercita.org.cn (S.-h.P.);
zhaoxy@nercita.org.cn (X.-y.Z.); zhangqi_lt@nercita.org.cn (Q.Z.); wangsf@nercita.org.cn (S.-f.W.)
[2] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
[3] Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100097, China
[4] Beijing Engineering Research Center of Agricultural Internet of Things, Beijing 100097, China
***** Correspondence: wangky@nercita.org.cn

check for updates

Abstract: Combining ability analysis can be used to preliminarily identify the advantages and disadvantages of combinations and parents in earlier generations, enabling breeders to reduce the range of material, save breeding time, and improve breeding efficiency. An approach for combining ability analysis through the hybrid crop breeding information management system is presented. The general combining ability prediction effect of parents and the specific combining ability prediction effect of combinations are calculated to analyze hybrid combinations using the hybrid crop breeding information management system. The results provide the basis for parent selection and combination selection. The plant breeding trial management function of the system can provide convenient diallel crossing trial design, field planting plan, and combining ability analysis. In the system, the genealogy of breeding materials is traced with the combining ability test crosses. The selection of high-generation breeding materials can be performed in accordance with the combining ability test results of early generation materials. The system has been successfully applied to a large Chinese seed company. The combining ability test function automates data analysis and eliminates days in the decision-making process. The efficiency of the combining ability test analysis and test report generation has improved to more than double by using the system.

Keywords: combining ability analysis; breeding informatization; crop breeding information management system

1. Introduction

Plant breeding programs can develop genotypes adapted to specific agricultural environments and can result in lower inputs to achieve higher and more sustainable productions with lower energy costs to accommodate the growing population [1]. With the expansion of the breeding scale, the data generated in the crop breeding process have accumulated and are enriched. Based on the experience of breeders, selecting new varieties with high yield, multi-resistance, and high quality from a massive breeding population is difficult. Therefore, using agricultural information technology to assist traditional breeding is effective [2]. The modern seed industry is a highly information-intensive industry. Information technology is an important part of the modern breeding technology system and provides important technical support to improve breeding efficiency. Using a crop breeding information management system to manage and maintain breeding information can clearly establish

the relationship among various data, and the system can be used to analyze these data. Many advanced foreign seed enterprises—such as Pioneer, Syngenta, and Monsanto—have established modern breeding information management systems based on information technology [3]. These systems provide strong technical support for the efficient statistical analysis of massive breeding data. Moreover, the pedigree and genetic relationship of breeding materials can be obtained clearly, thereby improving breeding efficiency. The structure, function, and technical content of these proprietary breeding information management systems are confidential, but such systems certainly have huge storage capacity for massive data and can efficiently analyze complex data. In addition to these private breeding information management systems, a foreign piece of commercial breeding information management software is also being developed and improved. The *AGROBASE* breeding information management software developed by *Agronomix*, a leading Canadian breeding software developer and service company, is used by breeders in more than 70 countries around the world [4]. *PRISM*, which is an acronym for Plant Research Information Sharing Manager, is a breeding information management system developed by the *CSSI* company of America; it provides information management support for grain and vegetable crops [5]. The *Phenom-networks* breeding information management system developed by the bioinformatics company founded by Hebrew University of Israel allows breeding data management, the statistical analysis of phenotypic data, and the genetic analysis of quantitative trait locus location through web pages [6]. The *LABKEY* software developed by the French company *DORIANE* is a research process and data management piece of software particularly designed for agricultural research, such as seed selection and breeding, agronomic experiments and tests, biotechnology laboratory research, and food processing. The *RnDExperience* software is designed for agronomic research and biological research; it supports the creation of new plant varieties and management of technical information in biotechnology laboratories [7]. *E-Brida* of the Netherlands is a breeding information system that analyzes and explores data by using a pedigree tree. Information on mothers, fathers, crossings, self-pollination, or selection is provided with a full survey of origin, including photos and specific features of pedigree items [8]. The *PLABSTAT* software from Germany is a computer program for the statistical analysis of plant breeding experiments; it can analyze diallel and factorial crosses [9].

China's breeding research institutions and number of seed industry practitioners rank first in the world. However, most seed enterprises and breeding research institutes remain at a low level of informatization; they still use manual measurement, paper records, empirical decision-making, and other working methods, with a serious imbalance in input and output [10,11]. Traditional breeding data management and analysis methods have gradually become one of the primary bottlenecks restricting the development of the seed industry. With the vigorous development of seed industry informatization, some key enterprises are learning from foreign experience and from their own innovations; moreover, such enterprises introduce and develop a series of information management systems in the key links of seed industry education, breeding, and promotion [12]. Certain companies introduce foreign plant breeding software with poor application effects, because the foreign breeding software-advocated technique differs from that used in the domestic seed industry. Furthermore, foreign breeding software lacks good technological support, particularly in the stages of personalized customization, and system implementation lacks a positive response to requirements [3]. In recent years, China's independently developed commercial breeding information management system, such as the gold seed breeding platform, has been gradually applied to breeding enterprises and breeding research institutions. However, the system lacks the application and analysis of combining ability.

In crop breeding, the utilization of heterosis and the cultivation of conventional varieties are generally inseparable in various hybridization methods. Hybridization is an important means of crop breeding, and the key to successful hybridization is parent selection. Considering that the performance of the parents is not necessarily the same as that of the hybrid offspring, the quality of the hybrid combination can only be identified after several generations. Thus, the selection of parents is difficult. If the quality of the combination can be preliminarily identified in the early generations,

emphasizing the importance of correct parent selection, then the quality of the combination can be identified as early as possible, and the breeding effect can be improved. Combining ability analysis can be used to preliminarily identify the advantages and disadvantages of combinations and parents in early generations, thereby enabling breeders to reduce the range of material, save breeding time, and improve breeding efficiency. Therefore, we construct a crop breeding information management system, i.e., the gold seed breeding cloud platform, based on combining ability analysis and application. The system is based on a cloud computing platform. It can improve flexibility, reduce infrastructure requirements, improve accessibility, and efficiently handle large data sets [13]. The business flow of the system is shown in Figure 1. Breeders select parent materials annually on the basis of breeding objectives. An intelligent retrieval function is provided in the system to rapidly search for parent breeding materials [11]. The incomplete, complete, and partial diallel crossing methods are built into the system, and these help with the design of a crossbreeding scheme of the combining ability test. Moreover, previous hybridized combinations, the orthogonal combination, and the reciprocal hybrid are detected automatically. Many trial design methods, such as randomized blocks and completely random, are provided for breeders based on the crossbreeding scheme. An assistant designs the planting planning in the field based on the experimental plan developed by the breeder. The workers sow the seeds based on the planting plan. During crop growth, an assistant collects the trait data. The breeder evaluates the materials and analyzes the combining ability of breeding materials in the system with one click. This feature helps breeders to identify superior parents and hybrids on the basis of general combining ability (GCA) and specific combining ability (SCA) performances, which can be used in a future breeding program.
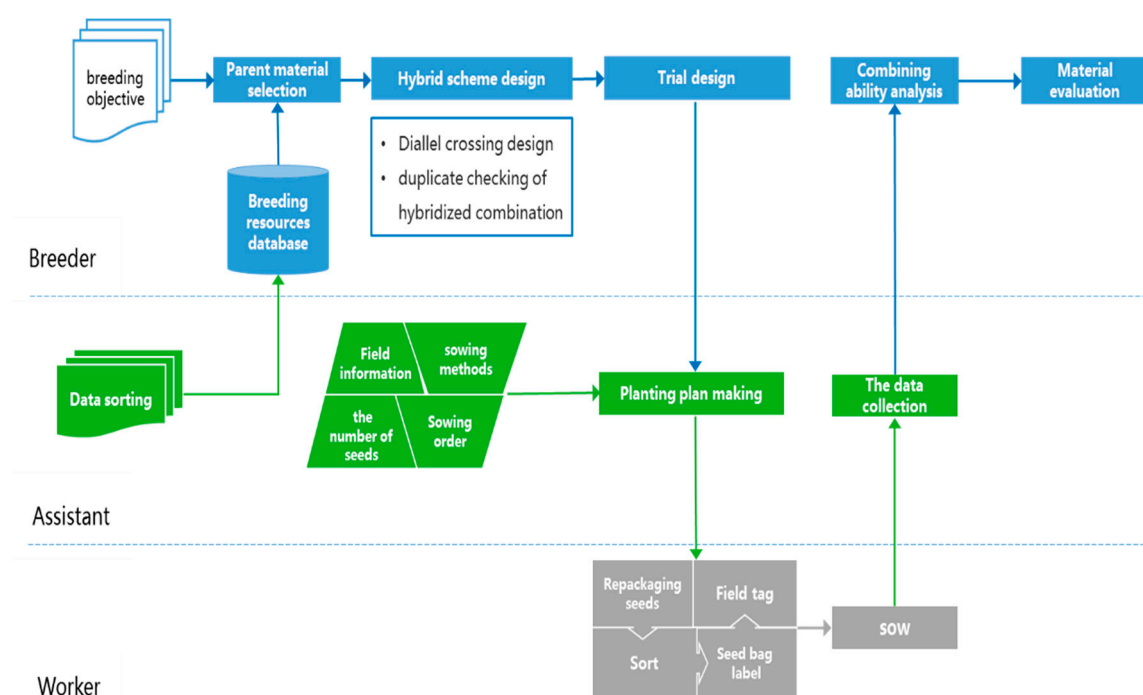

Figure 1. System business flow chart.

2. Materials and Methods

2.1. Diallel Crossing Trial

The practice of plant breeding shows that the performance of the parents is often inconsistent with that of the hybrid offspring. On the one hand, some parent materials perform well, but their hybrid offspring do not. On the other hand, some parent materials do not perform particularly well on their own, but good individuals or combinations can be separated from their hybrids. This difference in

offspring due to different combinations of parents indicates that different combinations of parents have different abilities, and such a phenomenon is known as combining ability. Combining ability analysis is used to identify better combiners, which can then be hybridized to exploit heterosis and to select better crosses for direct use or for further breeding work [14]. Studies on heterosis and combining ability analyses are indispensable tools in any breeding program; they provide the desired genetic description that is needed to improve crop variety or heterotic exploitation [15]. The selection of parent materials based on the combining ability is important to obtain offspring with hybrid vigor. Combining ability is an index to measure the compatibility of various traits of hybrid parents and is used as a basis for selecting parents. The GCA is an important indirect criterion in choosing parents [16]. Parents with high GCA values can be used in the creation of improved inbred lines [17]. SCA can be used as an index to determine the importance of a particular cross combination for exploitation through heterosis breeding and hybridization followed by recombination [18].

Diallel cross design is an important program of hybrid crop breeding. Good combinations and parents can be selected by analyzing diallel cross trials that can be used to estimate heterosis and combining ability. Diallel cross is used to study the genetic diversity of quantitative traits and estimate combining ability [19]. Diallel cross methods include complete, incomplete, and partial diallel crosses. Complete diallel crosses are the total hybridization among a group of parents, which may also include self-crossing and reciprocal cross combinations. In a previous study, Griffing (1956) divided complete diallel crosses into four types according to whether parent lines or the reciprocal cross combinations are included in the diallel crossing design (Table 1; the number of inbred lines or varieties is p.). Since Griffing proposed the four methods for the estimation of the combining ability of different homozygous lines, many researchers have used such methods in their breeding programs [20].

(1) The first type includes parents and the whole F_1 materials, with a total of P^2 combinations.
(2) The second type includes a set of F_1 materials, which include parents and reciprocal cross combinations. The number of hybridized combinations is $P + \frac{1}{2}P(P-1) = \frac{1}{2}P(P+1)$.
(3) The third type does not include parent materials, only reciprocal cross F_1 materials. The number of hybridized combinations is $\frac{1}{2}P(P-1) + \frac{1}{2}P(P-1) = P(P-1)$.
(4) The fourth type does not include parent materials, only a set of reciprocal cross F_1 materials. The number of hybridized combinations is $\frac{1}{2}P(P-1)$.

Table 1. Four types of complete diallel crosses.

I	A	B	C	D	II	A	B	C	D	III	A	B	C	D	IV	A	B	C	D
A	×	×	×	×	A	×	×	×	×	A		×	×	×	A		×	×	×
B	×	×	×	×	B		×	×	×	B	×		×	×	B			×	×
C	×	×	×	×	C			×	×	C	×	×		×	C				×
D	×	×	×	×	D				×	D	×	×	×		D				

Incomplete diallel cross is the mutual hybridization between two sets of parents, which does not include reciprocal cross (Table 2). The characteristic of this hybrid design is the division of the pure-breeding parents into two groups, and all kinds of possible intergroup hybridization are achieved.

Table 2. 3×5 incomplete diallel cross.

P_2 \ P_1	D	E	F	G	H
A	×	×	×	×	×
B	×	×	×	×	×
C	×	×	×	×	×

2.2. Combining Ability Analysis

A 3×5 incomplete diallel cross scheme is shown in Table 2. This scheme has two sets of parents. One set (P_1) has three parents—namely, *A*, *B*, and *C* ($n_1 = 3$)—which are precocious and have weaker disease resistance. The other set (P_2) has five parents—namely, *D*, *E*, *F*, *G*, and *H* ($n_2 = 5$)—which are more resistant to disease but mature later. Fifteen combinations are produced in the scheme. A completely randomized block design of three replications ($b = 3$) is used in the trial on the basis of the scheme. By taking plot yield as the unit of calculation, the current paper aims to estimate the combining ability of each parent's yield. Trial data are shown in Table 3. The area of each plot is 25 m². The plot yield is measured in kg. A total of 45 observed values are found. The plot of the A × D cross combination has the highest average yield. The plot of the B × E cross combination has the lowest average yield. Moreover, the combining ability is analyzed.

Table 3. Combining ability trial data.

Combination \ Block	I	II	III	Sum Total of Combinations X_{ij} (X_a)	Mean Value
A × D	30	30	27	87	29
A × E	22	19	22	63	21
A × F	23	26	26	75	25
A × G	25	23	24	72	24
A × H	20	25	24	69	23
B × D	23	26	23	72	24
B × E	18	17	16	51	17
B × F	19	16	19	54	18
B × G	21	22	20	63	21
B × H	18	21	21	60	20
C × D	20	20	17	57	19
C × E	20	24	22	66	22
C × F	19	18	20	57	19
C × G	21	21	18	60	20
C × H	17	19	21	57	19
sum total of blocks (X_b)	316	327	320	$\sum X = 963$	$\overline{X} = 21.4$

Combining ability analysis is divided into the following processes. The first step is variance analysis, which tests whether remarkable differences are found among the combinations. The degrees of freedom are decomposed, and then the sum of squares is calculated to obtain the ANOVA table. The degrees of freedom among the blocks are calculated on the basis of Equation (1). The degrees of freedom among the combinations are calculated on the basis of Equation (2). The degrees of freedom for random error are calculated on the basis of Equation (3). The sum of squares can be calculated on the basis of Equations (4)–(8). The result of the variance analysis is shown in Table 4. The difference among the combinations is remarkable. The variance among the combinations is composed of the GCA variance and the SCA variance components of the parents. Therefore, the variance among the combinations is further decomposed to analyze the difference between the two components.

$$d_{f(\text{block})} = b - 1 \tag{1}$$

$$d_{f(\text{combination})} = n_1 \cdot n_2 - 1 \tag{2}$$

$$d_{f(\text{error})} = (b - 1)(n_1 \cdot n_2 - 1) \tag{3}$$

$$S_{\text{total}} = \sum x^2 - C \tag{4}$$

$$C = \frac{\left(\sum x\right)^2}{n_1 n_2 b} \tag{5}$$

$$S_b = \frac{1}{n_1 n_2}\left(\sum x_b^2\right) - C \tag{6}$$

$$S_v = \frac{1}{b}\left(\sum x_a^2\right) - C \tag{7}$$

$$S_e = S_{total} - S_v - S_b \tag{8}$$

Table 4. Results of variance analysis.

Sources of Variation	Degrees of Freedom	Sum of Squares	Variance	F Value
blocks	2	$S_b = 4.13$	$V_b = 2.07$	$V_b/V_e = 0.709$
combinations	14	$S_v = 418.8$	$V_v = 29.9$	$V_v/V_e = 10.2$ **
random error	28	$S_e = 81.87$	$V_e = 2.92$	

** means the difference is extremely significant.

The bidirectional classification table (Table 5) of two sets of parents (P_1 and P_2) is obtained on the basis of Table 3. The degrees of freedom of P_1 are calculated on the basis of Equation (9). The degrees of freedom of P_2 are calculated on the basis of Equation (10). The degrees of freedom for female parent × male parent are calculated on the basis of Equation (11). The sum of squares of the two sets of parents and that of the interaction of the two sets of parents are calculated on the basis of Equations (12)–(14). The results of ANOVA among the combinations are shown in Table 6.

$$d_{f(P_1)} = n_1 - 1 \tag{9}$$

$$d_{f(p_2)} = n_2 - 1 \tag{10}$$

$$d_{f(p_{1,2})} = (n_1 - 1) \times (n_2 - 1) \tag{11}$$

$$S_{p1} = \frac{1}{n_2 b}\sum_{i=1}^{n_1} x_{i.}^2 - C \tag{12}$$

$$S_{p2} = \frac{1}{n_1 b}\sum_{i=1}^{n_2} x_j^2 - C \tag{13}$$

$$S_{p12} = S_v - S_{p1} - S_{p2} \tag{14}$$

Table 5. Bidirectional classification table of two sets of parents P_1 and P_2.

$P_{1(i)}$ \ $P_{2(j)}$	D	E	F	G	H	$X_{i.}$
A	87	63	75	72	69	366
B	72	51	54	63	60	300
C	57	66	57	60	57	297
$X_{.j}$	216	180	186	195	186	

Table 6. Results of ANOVA of the combinations.

Sources of Variation	Degrees of Freedom	Sum of Squares	Variance	F Value
g_i (P_1)	2	$S_{p1} = 202.8$	$V_{p1} = 101.4$	$V_{p1}/V_e = 34.73$ **
g_j (P_2)	4	$S_{p2} = 88.8$	$V_{p2} = 22.2$	$V_{p2}/V_e = 7.60$ **
S_{ij} ($P_{1,2}$)	8	$S_{p1,2} = 127.2$	$V_{p1,2} = 15.9$	$V_{p1,2}/V_e = 5.45$ **
random error	28	$S_e = 81.87$	$V_e = 2.92$	

** means the difference is extremely significant.

The second step is estimation of the combining ability effect value. The GCA is estimated on the basis of Equations (15) to (16). $\hat{g}_{i.}$ expresses the GCA effect of the *i*th parent in the set of P_1. $\hat{g}_{.j}$ expresses the GCA effect of the *j*th parent in the set of P_2. The analysis results for the GCA effect are shown in Table 7. The SCA is estimated on the basis of Equation (17). The GCA of parents' plot yield and SCA of combinations are shown in Table 8.

$$\hat{g}_{i.} = \frac{\sum_{j=1}^{n_2}\left(X_{ij} - \overline{X}_{..}\right)}{n_2} \tag{15}$$

$$\hat{g}_{.j} = \frac{\sum_{i=1}^{n_1}\left(X_{ij} - \overline{X}_{..}\right)}{n_1} \tag{16}$$

$$\hat{S}_{ij} = X_{ij} - \overline{X}_{..} - \hat{g}_{i.} - \hat{g}_{.j} \tag{17}$$

Table 7. Analysis results of the general combining ability (GCA) effect.

P_1(i) \ P_2(j)	D(1)	E(2)	F(3)	G(4)	H(5)	$X_i.$	$\hat{g}_{i.}$
A(1)	29	21	25	24	23	24.4	3.0
B(2)	24	17	18	21	20	20	−1.4
C(3)	19	22	19	20	19	19.8	−1.6
$X_{.j}$	24	20	20.7	21.7	20.7	$X... = 21.4$	
$\hat{g}_{.j}$	2.6	−1.4	−0.7	0.3	−0.7		

Table 8. GCA of parents' plot yield and specific combining ability (SCA) of combinations.

$\hat{S}_{ij}$	D(1)	E(2)	F(3)	G(4)	H(5)	$\hat{g}_{i.}$
A(1)	2.0	−2.0	1.3	−0.7	−0.7	3.0
B(2)	1.4	−1.6	−1.3	0.7	0.7	−1.4
C(3)	−3.4	3.6	−0.1	−0.1	−0.1	−1.6
$\hat{g}_{.j}$	2.6	−1.4	−0.7	0.3	−0.7	

As shown in Table 8, in group P_1, the yield of parent A has the largest GCA effect. In group P_2, the yield of parent D has the largest GCA effect. In addition, the A × D combination has higher SCA. Therefore, the average yield of this combination is higher. Although the SCA of C × E is the highest, the GCA of its two parents is lower. Comparative analysis shows that the GCA effect of the parents and the SCA effect of the combination do not correspond. Even if the GCA of the parents is high, the SCA of the combination is not always high. The parents with low GCA can also produce a combination with high SCA.
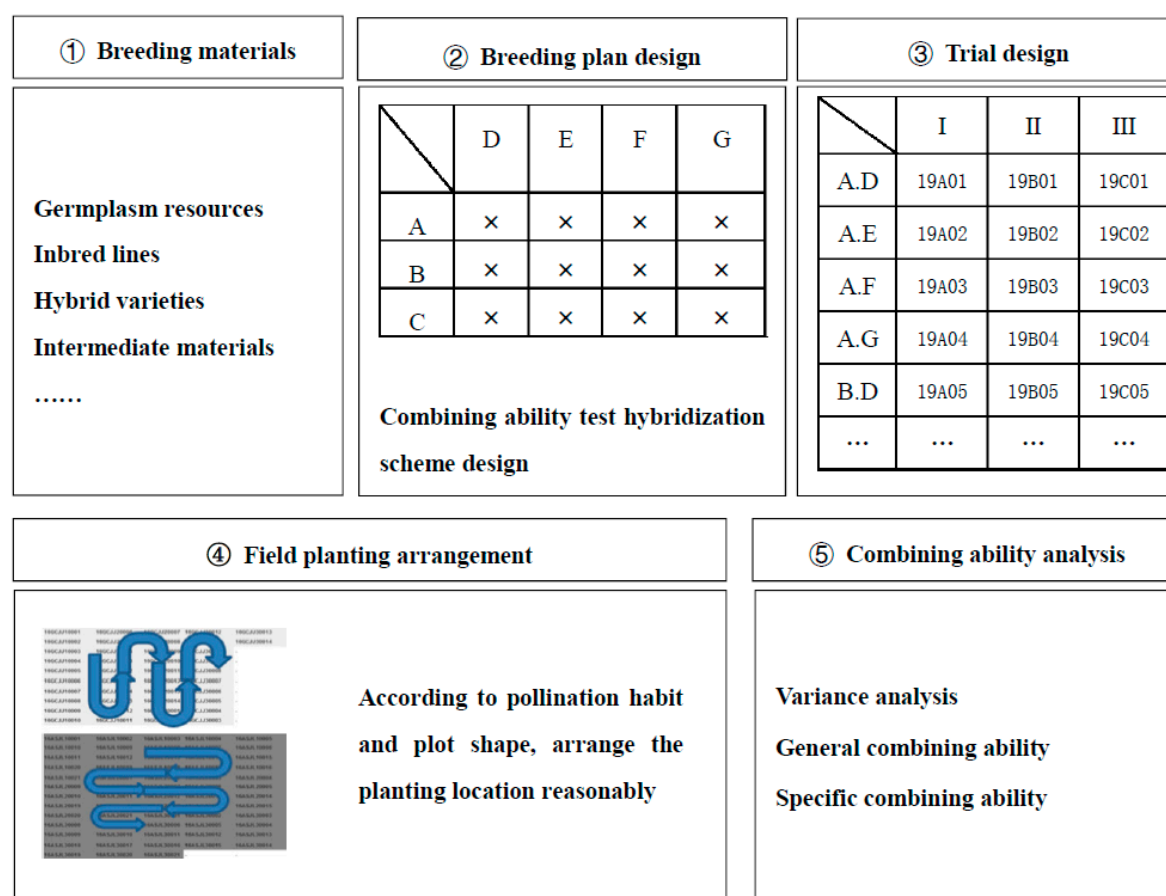
In addition, the GCA and SCA effects estimated by this method are based on the total average. For a multi-point test, the same test will have different total averages in different environments. To unify and compare the results of the multi-point test, the percentage of the combining ability effect value in terms of the total mean is used as the standard in measuring the combining ability.

2.3. Design of Crop Breeding Trial Management

The diallel cross trial is important, but its analysis and interpretation are difficult, and the calculation is complicated. Furthermore, certain difficulties are found in the implementation of a diallel cross trial. The number of combinations needed for the trial increases sharply with the number of the selected parents. Combining ability analysis has various formulas, and the process is complex.

Therefore, the whole process of the breeding trial management function is designed and developed on the basis of the combining ability test design characteristics (Figure 2).

① Breeding materials	② Breeding plan design	③ Trial design

① Breeding materials

Germplasm resources

Inbred lines

Hybrid varieties

Intermediate materials

......

② Breeding plan design

	D	E	F	G
A	×	×	×	×
B	×	×	×	×
C	×	×	×	×

Combining ability test hybridization scheme design

③ Trial design

	I	II	III
A.D	19A01	19B01	19C01
A.E	19A02	19B02	19C02
A.F	19A03	19B03	19C03
A.G	19A04	19B04	19C04
B.D	19A05	19B05	19C05
...	...	...	...

④ Field planting arrangement

According to pollination habit and plot shape, arrange the planting location reasonably

⑤ Combining ability analysis

Variance analysis

General combining ability

Specific combining ability

Figure 2. Process of plant breeding trial management function.

The whole process of the plant breeding trial management function has five parts. The first part is the crop breeding material management module. Crop breeders have hundreds of thousands of breeding materials, such as germplasm resources, inbred lines, hybrid varieties, and intermediate materials, and these are the bases of crop breeding. Crop breeding material management provides an intelligent retrieval function for the multidimensional data query, helping breeders efficiently select the parents to be used for crossing [11].

The second part is the crop breeding plan design module. In this module, breeders can efficiently design the hybridization scheme of the combining ability test. According to the breeding objectives, the hybrid combination, hybrid way, and hybrid quantity can be determined, and these support complete, incomplete, and partial diallel crosses. After many years of accumulation, the number and kind of hybrid combinations increase remarkably. Without the management of the information system, certain combinations are inevitably repeated. In particular, the repetitive useless combinations are a waste of labor and resources. Therefore, in this module, repeated combinations, orthogonal combinations, and reciprocal hybrids are validated automatically. Furthermore, the results of the combining ability test of a sister line can be provided for scientific hybridization design.

The third part is the trial design module. In this module, many trial design methods are provided, such as contrast, interval contrast, completely random, randomized block, Latin square, and split-plot designs. With one click, crop breeders can take a hybrid trial design and expand it to many years. In addition, when analyzing the data, all the means are stored for every location and every year. Breeders can see all trial data from all years across all locations with all traits.

The fourth part is the field planting arrangement module. In this module, breeding materials' planting location can be arranged reasonably with one click based on breeders' pollination habit and plot shape. Moreover, the same field can support the automatic arrangement of different planting specifications, thereby effectively solving the problem of the diversification of planting specifications.

The fifth part is the combining ability analysis module. The prompt analysis of field trial data is difficult for breeders because of the many barriers involved when using professional statistical software, including the statistical analysis system. Some breeding teams or breeding research institutions employ statistical analysis professionals, which requires additional cost. Many breeding teams lack statistical analysis professionals, and breeders only use Excel for simple data summaries. As such, data resources are often wasted. Our system is equipped with built-in breeding statistical analytical methods to facilitate the analysis of field trial data. This system does not require the use of a licensed statistics package. In the system, the operation is simple. Combining ability analysis is conducted by extracting trial data. Breeders only need to select the combining ability trial and analytical method to produce trial data results, such as variance analysis, GCA, and SCA.

2.4. Combining Ability Analysis Based on Incomplete Diallel Hybridization

An algorithm of the system for combining ability analysis based on an incomplete diallel hybridization is proposed. The input variables of the algorithm are two data sets for female and male seeds. The output variables of the algorithm are the GCA and SCA matrix. The algorithm consists of five steps, as follows. (1) An incomplete diallel cross matrix and hybridized combination are produced on the basis of female and male seed information. (2) Trial design is finished based on hybridized combination information and the field trial design method (e.g., randomized block and others.). (3) The trait data of the plot are extracted (e.g., plot yield, plant height, grain weight, and days to maturity). (4) The results of variance analysis are obtained (e.g., degrees of freedom, sum of squares, mean square, F value, random error, and total variation). (5) The combining ability analysis result (GCA and SCA matrix) is obtained.

The algorithm detail information as follows:

Algorithm name: combining ability analysis based on incomplete diallel hybridization

Input:

female seeds: $F = \{ \begin{array}{cccc} F_1 & F_2 & \cdots & F_k \end{array} \}$; male seeds: $M = \{ \begin{array}{cccc} M_1 & M_2 & \cdots & M_n \end{array} \}$;

Output:

GCA and SCA matrix result is as follows:

$$
R = \begin{pmatrix}
\hat{S}_{ij} & F_1 & F_2 & \cdots & F_k & \hat{g}_{i.} \\
M_1 & \hat{S}_{11} & \hat{S}_{12} & \cdots & \hat{S}_{1k} & \hat{g}_{1.} \\
M_2 & \hat{S}_{21} & \hat{S}_{22} & \cdots & \hat{S}_{2k} & \hat{g}_{2.} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
M_n & \hat{S}_{n1} & \hat{S}_{n2} & \cdots & \hat{S}_{nk} & \hat{g}_{n.} \\
\hat{g}_{.j} & \hat{g}_{.1} & \hat{g}_{.2} & \cdots & \hat{g}_{.k} & /
\end{pmatrix}
$$

Implementation steps:

1. An incomplete diallel cross matrix and hybridized combination are produced on the basis of female seed information ($F = \{ \begin{array}{cccc} F_1 & F_2 & \cdots & F_k \end{array} \}$;) and male seed information ($M = \{ \begin{array}{cccc} M_1 & M_2 & \cdots & M_n \end{array} \}$;).

2. The trial is designed based on hybridized combinations ($C = \{ \begin{array}{cccc} F_1M_1 & F_1M_2 & \cdots & F_kM_n \end{array} \}$) and on the trial design method (e.g., randomized blocks method ∈ {sequential method | completely random method | randomized blocks method | interval contrast method | … }).

3. The trait data of the plot are extracted (e.g., plot yield, plant height, grain weight, and days to maturity).

4. The results of variance analysis are obtained (e.g., degrees of freedom, sum of squares, mean square, F value, random error, and total variation).

5. The combining ability analysis results are obtained (GCA of female seeds and male seeds), and the SCA of hybridized combinations,

$$R = \begin{pmatrix} \hat{S}_{ij} & F_1 & F_2 & \cdots & F_k & \hat{g}_{i.} \\ M_1 & \hat{S}_{11} & \hat{S}_{12} & \cdots & \hat{S}_{1k} & \hat{g}_{1.} \\ M_2 & \hat{S}_{21} & \hat{S}_{22} & \cdots & \hat{S}_{2k} & \hat{g}_{2.} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ M_n & \hat{S}_{n1} & \hat{S}_{n2} & \cdots & \hat{S}_{nk} & \hat{g}_{n.} \\ \hat{g}_{.j} & \hat{g}_{.1} & \hat{g}_{.2} & \cdots & \hat{g}_{.k} & / \end{pmatrix}).$$

2.5. Genealogy Tracing Based on Hybrid Combinations

The core of cross breeding, particularly maize breeding, involves the selection and breeding of inbred lines with good combining ability and good characteristics and the use of parent materials to rapidly produce good hybrids. Furthermore, the complete pedigree information of a breeding variety can be useful in determining its parent source. The favorable genes of a breeding variety can be estimated on the basis of the favorable genes of its parents or distant generation parents [21,22]. With maize breeding as an example in the construction of a genealogy, the pedigree relationship between parents and offspring should be considered, and the hybrid combination and combining ability analysis results that have participated in the combining ability test should be added.

The populations of inbred line plants show a considerable difference in combining ability, and the inbred lines in the early generation show a more stable state in combining ability. Therefore, both inbred lines can pass their advantages in combining ability to their offspring. The self-crossed offspring will not have much impact on the combining ability. Thus, combining ability analysis can be used to preliminarily identify the advantages and disadvantages of a combination and the parents in earlier generations, thereby enabling breeders to reduce the range of treatment materials and to improve the breeding efficiency. However, the combining ability of the early generation cannot represent the combining ability of the later generation. In general, maize will become stable after four or five generations of self-crossing, during which the combining ability test results can better reflect the situation of its combining ability. Therefore, the combining ability determination of an early generation cannot completely replace the determination of a later generation. Although the determination of an early generation can reduce the workload of subsequent selection, the determination of combining ability in the later generation is necessary to select excellent inbred lines with high combining ability.

We constructed a genealogy of breeding materials based on the combining ability test's hybrid combinations (as shown in Figure 3). The combining ability of the third and fifth generation breeding materials was tested, and the results were added to the genealogy construction. Based on the GCA effect, the breeding potential of new inbred lines was predicted, thereby providing a basis for the decision in the selection and breeding of good inbred lines. The probability of selecting high combining ability inbred lines from third lines with high combining ability was higher than that of third lines with low combining ability. Therefore, the selection of the fourth, fifth, sixth, and seventh generations can be based on the combining ability test results of the third-generation materials. Furthermore, the fifth-generation combining ability results can be an important reference for hybrid seed production.
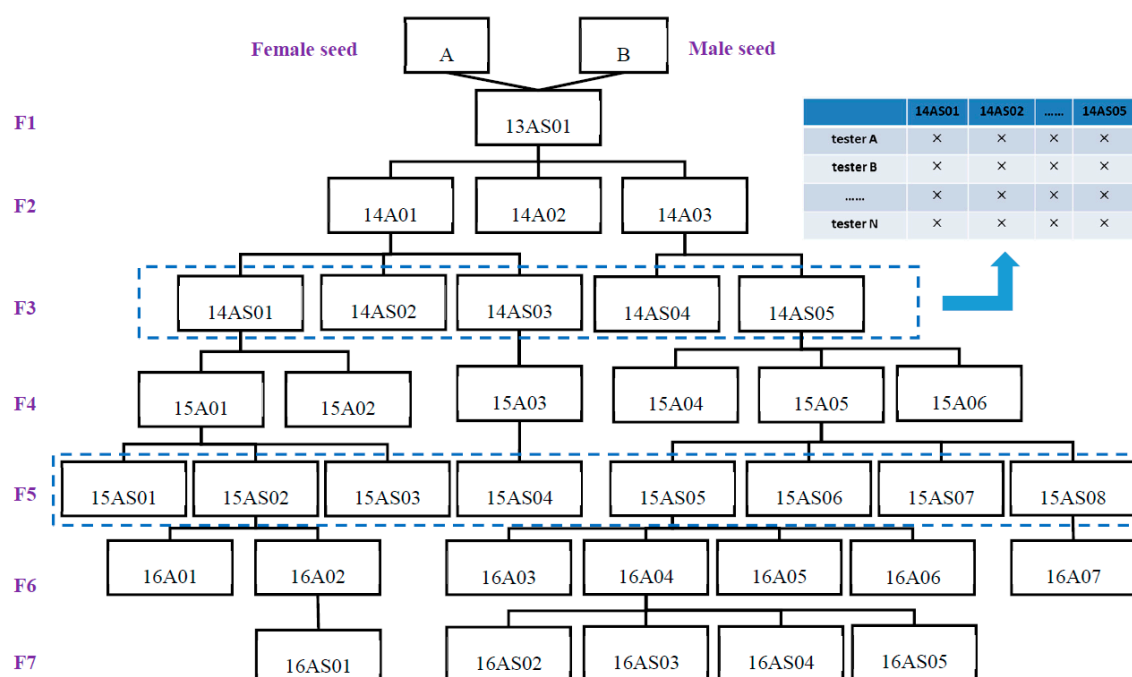
Figure 3. Genealogy tracing based on the combining ability test's hybrid combinations.

The algorithm of genealogy tracing based on combining ability test's hybrid combinations is presented, as follows.

Algorithm name: Genealogy tracing based on combining ability test's hybrid combinations

Input: Parent materials of combining ability test

Output: Genealogical tree

Implementation steps:

1. A class, a list, and a stack are created for storing nodes.

2. Crossbred F_1 information of the parent material, which participates in the combining ability test, is found and assigned to the tree node.

3. All children of F_1 are found and added into the list.

4. If the list size is over 0, and the children list is pushed into stack, then go to Step 5. Otherwise, jump to Step 7.

5. Element in the stack is popped and assigned to the tree node.

6. If all children of the tree node are found and added to the list, then jump to Step 4.

7. The tree node is emptied and added into the tree node list. If the stack is empty, then go to Step 8. Otherwise, jump to Step 5.

8. The tree node list is returned, and the genealogy tree is created.

9. All children of the parent material are found and associated with the combining ability test results.

3. Results

3.1. Hybridization Design and Data Analysis

In practical production, combining ability can be used to evaluate the value of a parent material in heterotic utilization or cross breeding. Combining ability is important to successfully predict the genetic capability of parent lines and crosses [15]. Moreover, we have developed a crop breeding information management system with an incomplete diallel cross scheme design function and incomplete diallel cross trial analysis function. Figure 4 shows a part of the system interface, and the primary functions are implemented as follows:

(1) This function efficiently supports the design diallel cross scheme. Female seeds and male seeds are selected by an intelligent retrieval function based on breeding objective. An incomplete diallel cross scheme can be produced with one click (Figure 4a).

(2) An incomplete diallel cross trial analysis function supports the one-key extraction of trial data (Figure 4b) and one-key generation of combining ability analysis data (Figure 4c).
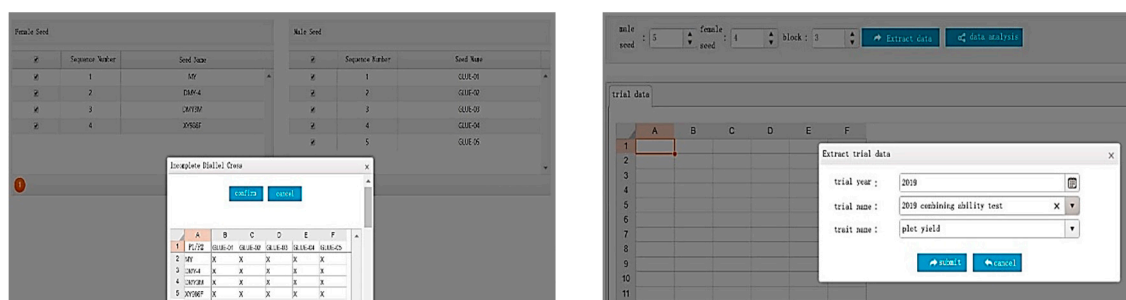
(**a**) Incomplete diallel cross scheme design.

(**b**) Incomplete diallel cross trial analysis.

Analysis of Variance:

Sources of Variation	Degree of Freedom	Sum of Squares	Mean Square	F Value
blocks	2	0.367	0.184	0.240
male seeds	4	16.380	4.095	3.262*
female seeds	3	302.183	100.728	80.240**
combinations	12	15.064	1.255	1.643
random error	38	29.026	0.764	
total variation	59	363.020		

The gca of parents' plot yield and sca of combinations:

	MY	DMY-4	DMY3M	XY986F	g(i.)
GLUE-01	-0.375	0.972	2.991	-3.588	-0.516
GLUE-02	1.227	0.945	-0.402	-1.770	2.117
GLUE-03	-1.135	-0.331	-0.049	1.515	-2.470
GLUE-04	1.906	-0.983	-1.786	0.863	-0.841
GLUE-05	-1.623	-0.603	-0.755	2.980	1.710
g(.j)	-9.859	-4.148	6.211	7.796	

(**c**) Analysis result.

Figure 4. Hybridization design and data analysis interfaces.

3.2. Genealogy and Detailed Information of the Breeding Material

In the system, the genealogy of breeding materials is constructed in association with the combining ability testing combinations and combining ability analysis results. The whole genealogy tree of each breeding material can be traced. When the breeder clicks any node of the genealogical tree, detailed information, images, and data analysis reports pertaining to a certain material can be viewed, as shown in Figure 5. The function facilitates the elucidation of the parental source and main characteristics of breeding materials. Furthermore, early or late generation combining ability test results can be viewed. Inbred line selection and hybrid seed production can reference the combining ability test results, because the combining ability analysis can be used to preliminarily identify the advantages and disadvantages of the combination and parents in earlier generations. Thus, breeders can reduce the range of treatment materials and improve the breeding efficiency.
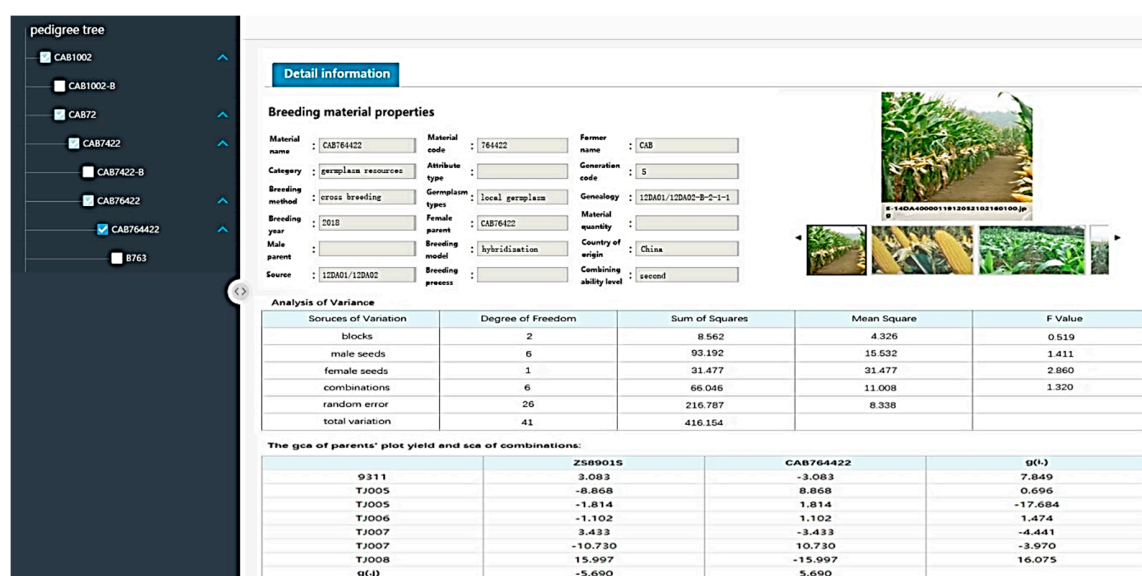
Detail information

Breeding material properties

Material name	: CAB764422	Material code	: 764422	Former name	: CAB
Category	: germplasm resources	Attribute type	:	Generation code	: S
Breeding method	: cross breeding	Germplasm types	: local germplasm	Genealogy	: 12DA01/12DA02-B-2-1-1
Breeding year	: 2018	Female parent	: CAB76422	Material quantity	:
Male parent	:	Breeding model	: hybridization	Country of origin	: China
Source	: 12DA01/12DA02	Breeding process	:	Combining ability level	: second

Analysis of Variance

Sources of Variation	Degree of Freedom	Sum of Squares	Mean Square	F Value
blocks	2	8.562	4.326	0.519
male seeds	6	93.192	15.532	1.411
female seeds	1	31.477	31.477	2.860
combinations	6	66.046	11.008	1.320
random error	26	216.787	8.338	
total variation	41	416.154		

The gca of parents' plot yield and sca of combinations:

	ZS8901S	CAB764422	g(i.)
9311	3.083	-3.083	7.849
TJ005	-8.868	8.868	0.696
TJ005	-1.814	1.814	-17.684
TJ006	-1.102	1.102	1.474
TJ007	3.433	-3.433	-4.441
TJ007	-10.730	10.730	-3.970
TJ008	15.997	-15.997	16.075
g(.j)	-5.690	5.690	

Figure 5. Genealogy and detailed information of breeding material interface.

4. Conclusions

Parent evaluation and combination selection are important problems in conventional crop breeding. Given the complex genetic background, large breeding population, and different environment and selection criteria, the evaluation results are often blind and one-sided. Moreover, the expression of parent-related traits provides an important basis for the evaluation of hybrid combination. Although conventional field phenotype evaluation based on material phenotype traits has a good evaluation function, the evaluation work for an increasing breeding material scale takes a long time and requires a large amount of work. The current research used information technology and a breeding information management system. The GCA prediction effect of parents and the SCA prediction effect of combinations were calculated to analyze the hybrid combination, thereby providing the basis for parent and combination selection. The best possible combination model that can be used as a reference for the effective utilization or further improvement of a crop population is determined. The application value and utilization potential of the population can be evaluated on the basis of the GCA effect, and the genetic relationship of the population is deduced on the basis of SCA [23].

Combining ability information and gene action that governs the inheritance of yield and quality-related traits can help breeders to select suitable parents and develop an appropriate breeding strategy [18]. Combining ability is widely used to screen the best parents in hybrid breeding [24]. In addition, combining ability can be partitioned into GCA and SCA. GCA is the average performance of a parent genotype over several cross combinations and is associated with additive gene effects, whereas SCA reveals the average performance of crosses that perform better and is typically associated with non-additive gene effects (dominance) [25]. Superior parents and hybrids are identified on the basis of GCA and SCA performance, which can be used in a future breeding program and to suggest suitable breeding approaches. In general, high GCA is one of the key parameters for selecting parent lines to develop hybrids with strong heterosis [26].

Diallel cross design is an important crop breeding scheme. The analysis of a diallel cross test can be used to estimate heterosis and combining ability to select the best combination and parents. However, the implementation of diallel cross breeding has two challenges. First, the number of combinations required for the experiment increases sharply with the number of test parents, and the scale of the experiment is relatively large. Second, the process of combining ability analysis is complex, and the equations involved are many. Every year, a finite window of time is available to plant breeders, during which they must work to analyze many months' worth of data that they have been collecting and worked on, including the analysis of combining ability test data. Data analysis results can provide

a basis for making decisions, which affect the plant breeding program remarkably. However, many breeding teams lack statistical analysis professionals, and breeders tend to agonize over breeding data statistical analysis. Our system is equipped with built-in breeding statistical analytical methods, including combining ability analysis methods. This system does not require the use of a licensed statistics package. Moreover, this system has solved the practical problem that users do not understand regarding the importance of biometric statistics. The system's operation is simple. Breeders only need to select the combining ability test trial and analytical method to generate the trial data analysis results with one click.

The plant breeding trial management function of the system can provide convenient diallel crossing trial design, field planting plan, and combining ability analysis. The system offers a complete solution for all stages of a combining ability test, from breeding material management and breeding plan design to trial design, field planting arrangement, and combining ability analysis. Furthermore, we trace the genealogy of breeding materials with the combining ability test's crosses. The selection of high-generation breeding materials can be used as a reference for the combining ability test results of early-generation materials.

The system has been applied to the *Beidahuang Kenfeng* Seed Co., Ltd. in 2019, which is a large-scale, modern, and comprehensive state-holding seed company with an integrated R&D, production, processing, sales, service, and import and export business in China. Combining ability test program formulation, parent material selection, trial design, field planting arrangement, and combining ability data processing and analysis are applied to many hybrid breeding crops, such as hybrid rice, maize, and soybean. More than 80,000 hybrid species are managed, and over 300,000 trait data are collected. The efficiency of combining ability test analysis and test report generation has increased to more than double.

Combining ability analysis can be used to improve breeding efficiency by providing the vital parameters needed in hybrid breeding programs [27]. The combining ability test can be easily analyzed, allowing breeders to evaluate and compare the data from different trials. The combining ability test function automates data analysis and saves days in the decision-making process. The system will not make the decision for breeders. However, the system will give breeders confidence to make their decisions with all the data needed to support such decisions. The combining ability analysis method built into the system is suitable for the cross breeding of many crops, such as corn and hybrid rice, as well as trees, such as Chinese fir and pine. However, our systematic process applies to crop breeding. The system applies to the breeding of many crops, such as hybrid rice, maize, wheat, and soybean. Moreover, the informatization needs of the breeders at the forefront are satisfied, which is our goal. We hope that the system can be used in forest breeding in the future.

Author Contributions: Y.-y.H. conceived research and wrote the manuscript; K.-y.W. provided important advice in the development of this study and manuscript; Z.-q.L. and S.-h.P. designed this research; X.-y.Z., Q.Z. and S.-f.W. developed the information system and discussed the content of this study. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fess, T.F.; Kotcon, J.B.; Benedito, V.A. Crop Breeding for Low Input Agriculture: A Sustainable Response to Feed a Growing World Population. *Sustainability* **2011**, *3*, 1742–1772. [CrossRef]
2. Wang, J.C.; Gao, Z.F.; Li, D.S.; Zhu, D.; Wu, H. The Application of Agricultural Information Technology in Wheat Breeding. *Crops* **2018**, *184*, 43–49. (In Chinese)
3. Han, Y.Y.; Wang, K.Y.; Liu, Z.Q.; Zhang, Q.; Zhao, X.; Wang, S. A crop trait information acquisition system with multitag-based identification technologies for breeding precision management. *Comput. Electron. Agric.* **2017**, *135*, 71–80. [CrossRef]
4. Mulitze, D.K. AGROBASE/4: A microcomputer database management and analysis system for plant breeding and agronomy. *Agron. J.* **1990**, *82*, 1016–1021. [CrossRef]
5. Central Software Solutions, Inc. PRISM Plant Breeding Software. 2013. Available online: http://www.teamcssi.com/index.html (accessed on 17 January 2020).
6. Phenome Networks Ltd. Phenome Networks. 2019. Available online: http://phenome-networks.com/ (accessed on 17 January 2020).
7. Doriane. 2019. Available online: www.doriane.com (accessed on 18 January 2020).
8. EBrida. 2019. Available online: https://www.e-brida.com/ (accessed on 19 January 2020).
9. Utz, H.F. *Analysing Diallel and Factorial Designs by PLABSTAT*; University of Hohenheim: Stuttgart, Germany, 2001.
10. Liu, Z.Q.; Wang, K.Y.; Zhao, X.Y.; Li, M. Research on Information Model of Plant Breeding under Cloud Environment. *J. Agric. Mech. Res.* **2017**, *39*, 7–11. (In Chinese)
11. Han, Y.Y.; Wang, K.Y.; Liu, Z.Q.; Pan, S.; Zhao, X.; Wang, S. Golden seed breeding cloud platform for the management of crop breeding material and genealogical tracking. *Comput. Electron. Agric.* **2018**, *152*, 206–214. [CrossRef]
12. Sun, H.Y.; Li, J.M.; Liu, Z.; Fang, W.; Ma, S.; Han, G.R.; Wan, C.; Li, J.W.; Wang, Y.X. Thoughts on the construction of seed industry informatization in China. *China Seed Ind.* **2014**, *6*, 1–3. (In Chinese)
13. Yang, F.; Wang, K.; Han, Y.; Qiao, S. A Cloud-Based Digital Farm Management System for Vegetable Production Process Management and Quality Traceability. *Sustainability* **2018**, *10*, 4007. [CrossRef]
14. Murtadha, M.A.; Ariyo, O.J.; Alghamdi, S.S. Analysis of combining ability over environments in diallel crosses of maize (Zea mays). *J. Saudi Soc. Agric. Sci.* **2018**, *17*, 69–78. [CrossRef]
15. Ene, C.O.; Ogbonna, P.E.; Agbo, C.U.; Chuakwudi, U.P. Heterosis and combining ability in cucumber (*Cucumis sativus* L.). *Inf. Process. Agric.* **2019**, *6*, 150–157. [CrossRef]
16. Bouchetat, F.; Aissat, A. Evaluation of the genetic determinism of an F1 generation of barley resulting from a complete diallel cross between autochthones and introduced cultivars. *Heliyon* **2019**, *5*, e02744. [CrossRef] [PubMed]
17. Longin, C.F.H.; Gowda, M.; Mühleisen, J.; Ebmeyer, E.; Kazman, E.; Schachschneider, R.; Schacht, J.; Kirchhoff, M.; Zhao, Y.; Reif, J.C. Hybrid wheat: Quantitative genetic parameters and consequences for the design of breeding programs. *Theor. Appl. Genet.* **2013**, *126*, 2791–2801. [CrossRef] [PubMed]
18. Kumari, J.; Dikshit, H.K.; Singh, B.; Singh, D. Combining ability and character association of agronomic and biochemical traits in pea (*Pisum sativum* L.). *Sci. Hortic.* **2015**, *181*, 26–33. [CrossRef]
19. Aghamiri, S.M.M.S.; Mostafavi, K.; Mohammadi, A. Genetic study of agronomic traits in barley based diallel cross analysis. *Adv. Environ. Biol.* **2012**, *6*, 62–68.
20. Balliu, A.; Hallidri, M. Combining Ability Test between Some Tomato Genotypes. In *Balkan Symp on Veg & Potatoes*; ISHS: Thessaloniki, Greece, 2002.
21. E, Z.G.; Wang, L. Construction of Database for Chinese Rice Varieties and Their Genealogy. *Chin. J. Rice Sci.* **2011**, *25*, 565–566. (In Chinese)
22. Yin, Z.; Guang-Lian, X.; Xin, Y. The Construction of Jiangsu Province Crop Variety Pedigree Traceability System. *Jiangsu Agric. Sci.* **2013**, 412–414. (In Chinese)
23. Tan, H.; Zheng, D.B.; Huang, A.H.; Chen, G.; Wu, Y.; Huang, K. Evaluation on combining ability and heterosis pattern of ten maize populations. *Guangxi Agric. Sci.* **2009**, *40*, 7–11. (In Chinese)
24. Rahimi, M.; Rabiei, B.; Samizadeh, H.; Ghasemi, A.K. Combining ability and heterosis in rice (*Oryza sativa* L.) cultivars. *J. Agric. Sci. Technol.* **2010**, *12*, 223–231.

25. Sibiya, J.; Tongoona, P.; Derera, J.; Rij, V.N. Genetic analysis and genotype x environment (G x E) for grey leaf spot disease resistance in elite African maize (*Zea mays* L.) germplasm. *Euphytica* **2013**, *185*, 349–362. [CrossRef]

26. Romanus, K.G.; Hussein, S.; Mashela, W.P. Combining ability analysis and association of yield and yield components among selected cowpea lines. *Euphytica* **2008**, *162*, 205–210. [CrossRef]

27. Xiang, C.; Zhang, H.; Wang, H.; Wang, J.; Wang, W.; Xia, J.; Gao, Y.; Ye, G. Dissection of combining ability for yield and related traits using introgression lines in the background of a key restorer line in rice (*Oryza sativa* L.). *Field Crop. Res.* **2016**, *193*, 154–163. [CrossRef]