

User manual for

Predhy.GUI

Performs Genomic Prediction of Hybrid Performance

With

Graphical User Interface

(Vision 2.1)

Yang Xu, Guangning Yu, Yuxiang Zhang, Yanru Cui,

Shizhong Xu, Chenwu Xu

(xuyang_89@126.com)

Last updated on April, 2025

contents

1. Getting started	3
1.1 installation	3
1.2 Run predhy.GUI	3
2. Dataset input	3
2.1 Genotype datasets	3
2.1.1 Input_genotype dataset	3
2.1.2 Inbred_gene dataset (*.csv format file)	4
2.2 Phenotype datasets (*.csv format file)	5
2.2.1 Hybrid phenotype	5
2.2.2 Parent phenotype	5
2.3 Parent names dataset (*.csv format file)	6
3. Operation process	7
3.1 convertgen	7
Dataset Input	7
Method select & Parameter setting	7
Run the software	8
3.2 cv	9
Dataset Input	9
Method select & Parameter setting	10
3.3 predhy.predict	13
Dataset Input	13
Method select & Parameter setting	13
Run the software	14
3.4 predhy.predict_NCII	15
Dataset Input	15
Method select & Parameter setting	16
Run the software	17
3.5 crodesign	19
Dataset Input	19
Method selection & Parameter setting	19
Run the software	20

1. Getting started

The software package `predhy.GUI` runs only in the R software environment and can be freely downloaded from the R website (<https://cran.r-project.org>).

1.1 installation

Within R environment, the `predhy.GUI` software can be installed online using the below command:

```
install.packages("predhy.GUI")
```

1.2 Run `predhy.GUI`

Once the software `predhy.GUI` is installed, users may run the software using two commands:

```
library("predhy.GUI")  
predhy.GUI()
```

2. Dataset input

2.1 Genotype datasets

2.1.1 Input_genotype dataset

Numeric format for Genotypic dataset (*.csv format file)

The first column stands for marker ID. Among the remaining columns, each column lists all the genotypes for one individual while the first row shows the individual names. For each marker, homozygous genotypes are expressed by 1 and -1, respectively, and the heterozygous genotypes are indicated by zero, missing values are indicated by NA.

	R001	R002	R003	R004	R005	R006	R007	R008
SNP1	-1	1	1	1	-1	1	-1	-1
SNP2	-1	1	1	1	-1	1	-1	-1
SNP3	-1	1	1	1	-1	1	-1	-1
SNP4	-1	1	1	1	-1	1	-1	-1
SNP5	-1	1	1	1	-1	1	-1	-1
SNP6	-1	1	1	1	-1	1	-1	-1
SNP7	-1	1	1	NA	-1	1	-1	-1
SNP8	-1	1	1	1	-1	1	-1	-1
SNP9	-1	1	1	1	-1	1	-1	-1
SNP10	-1	1	1	1	-1	1	-1	-1

Hapmap format for Genotypic dataset (*.txt format file)

Please see the TASSEL software in details. Here we introduce simply. The first eleven columns describe the specific information of markers and individuals, and their column names must be "rs#", "alleles", "chrom", "pos", "strand", "assembly#", "center", "protLSID", "assayLSID", "panel" and "QCcode".

The values for marker genotypes should be character, such as AA, TT, CC, GG, NN, AC and AG, where the "NN" indicates missing or unknown genotypes. In the 2 and 5 to 11 columns, "NA" indicates **no information** available. All the individual genotypic information will be showed from the 12 to last columns. In each column, individual name is listed in the first row, i.e., "A002", and the others are the genotypes (character).

rs	alleles	chrom	pos	strand	assembly	center	protLSID	assayLSID	panel	QCcode	A002	A003	A004	A005	A006
SNP_1_14068	T/C	1	14068	NA	NA	NA	NA	NA	NA	NA	NA	TT	TT	NA	TT
SNP_1_338176	G/T	1	338176	NA	NA	NA	NA	NA	NA	NA	NA	NA	GG	NA	GG
SNP_1_703171	G/A	1	703171	NA	NA	NA	NA	NA	NA	NA	GG	GA	GG	GA	GA
SNP_1_1033512	C/T	1	1033512	NA	NA	NA	NA	NA	NA	NA	TT	TT	CC	NA	TT
SNP_1_1401306	A/C	1	1401306	NA	NA	NA	NA	NA	NA	NA	CC	CC	CC	NA	CC
SNP_1_1465404	C/T	1	1465404	NA	NA	NA	NA	NA	NA	NA	CC	CC	CC	CC	CT
SNP_1_1725463	C/T	1	1725463	NA	NA	NA	NA	NA	NA	NA	CT	CT	CC	CT	CT
SNP_1_1866006	C/T	1	1866006	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SNP_1_2045326	G/A	1	2045326	NA	NA	NA	NA	NA	NA	NA	GG	AA	AA	GG	GG
SNP_1_2670571	A/G	1	2670571	NA	NA	NA	NA	NA	NA	NA	AA	AA	AA	AA	AA
SNP_1_2950255	G/C	1	2950255	NA	NA	NA	NA	NA	NA	NA	GG	GG	GG	GG	GG
SNP_1_3818861	A/T	1	3818861	NA	NA	NA	NA	NA	NA	NA	AA	AA	AA	AA	AA
SNP_1_4185501	C/G	1	4185501	NA	NA	NA	NA	NA	NA	NA	GG	CC	CC	CC	CC
SNP_1_4616639	G/T	1	4616639	NA	NA	NA	NA	NA	NA	NA	NA	GG	GG	GT	GT
SNP_1_5036129	G/A	1	5036129	NA	NA	NA	NA	NA	NA	NA	GG	GG	GG	GG	GG

2.1.2 Inbred_gene dataset (*.csv format file)

A matrix for genotypes of parental lines in numeric format, coded as 1, 0 and -1. The first columns indicates the names of inbred lines, which must be provided. Among the remaining columns, each column lists all the genotypes for a SNP while the first row shows the SNP names.

It can be obtained from the original genotype using convertgen function.

	SNP_1_14068	SNP_1_338176	SNP_1_703171	SNP_1_1033512	SNP_1_1401306	SNP_1_1465404	SNP_1_1725463	SNP_1_1866006
A002	0.521126761	0.800711744	1	-1	-1	1	0	0.580952381
A003	1	0.800711744	0	-1	-1	1	0	0.580952381
A004	1	1	1	1	-1	1	1	0.580952381
A005	0.521126761	0.800711744	0	-0.239875389	-0.865319865	1	0	0.580952381
A006	1	1	0	-1	-1	0	0	0.580952381
A007	0	-1	1	-1	-1	1	0	-1
A008	1	1	0	-1	-1	1	0	-1
A010	1	0	1	1	1	1	0	0.580952381
A011	1	1	0	-1	-1	1	0	-1
A012	1	1	1	-1	-1	1	0	1
A013	1	1	0	-1	-1	1	0	0.580952381
A014	1	1	1	-1	-0.865319865	0	0	-1
A015	-1	0.800711744	0	-0.239875389	-0.865319865	1	0	1
A016	0	0	1	-1	-1	1	0	-1
A017	-1	0	1	-1	-1	1	0	1
A018	1	0	1	-1	-1	1	0	1
A020	0.521126761	1	1	1	1	1	0	1
A021	-1	1	1	-1	-1	1	0	1
A022	1	0.800711744	1	-1	-1	1	0	-1
A023	1	1	1	1	-1	1	0	1

2.2 Phenotype datasets (*.csv format file)

2.2.1 Hybrid phenotype

A data frame with three columns. The first column and the second column are the names of male and female parents of the corresponding hybrids, respectively; the third column is the phenotypic values of hybrids. The names of male and female parents must match the rownames of inbred_gen. Missing (NA) values are not allowed.

M	F	Trait1
A002	A017	1433.745
A003	A393	1451.795
A003	A256	952.38
A003	A187	522.58
A003	A071	1457.775
A003	A439	1320.1
A005	A429	1638.91
A005	A430	1592.485
A006	A017	2050.12
A006	A021	1948.125
A006	A304	1474.83
A006	A268	1499.175
A006	A010	1010.345
A006	A030	953.685
A007	A021	1541.34

2.2.2 Parent phenotype

A matrix of phenotypic values of parent. The names of the matrix must match the rownames of inbred_gen.

	parent_phe
A002	1
A003	1
A004	1
A005	1
A006	2
A007	1
A008	1
A010	1
A011	1
A012	1

2.3 Parent names dataset(*.csv format file)

male_name: a data frame with only one column, of the names of male parents, with “M” in the first row.

female_name: a data frame with only one column, of the names of female parents, with “F” in the first row.

M	F
A002	A008
A003	A008
A003	A008
A003	A008
A003	A010
A003	A010
A005	A010
A005	A010
A006	A010
A006	A010
A006	A011
A006	A011
A006	A011
A006	A011
A007	A012

3. Operation process

3.1 convertgen

Dataset Input

Users must first click the drop-down menu to select the genotype file type, which includes “HapMap format with single bit”, “HapMap format with double bit”, “numeric format”. Then users can click the file input box to upload their data.

The screenshot shows a web browser window with the URL `http://127.0.0.1:5655`. The browser's address bar includes "Open in Browser" and "Publish" buttons. The page has a navigation bar with tabs: "Predhy", "convertgen" (selected), "cv", "predhy.predict", "predhy.predict_NCII", and "crodesign". On the left, a sidebar contains links: "Description", "Input files" (highlighted in blue), "Parameters", and "Results". The main content area is titled "Input genotype". It features a "file type" dropdown menu currently set to "HapMap format with double bit". Below this is an "input_geno" section with a "Browse..." button and a text input field containing "input_geno.hmp.txt". A blue "Upload complete" button is at the bottom of this section. To the right of the input fields, there is explanatory text: "the type of genotype. There are three options: 'hmp1' for genotypes in HapMap format with single bit, 'hmp2' for genotypes in HapMap format with double bit, and 'num' for genotypes in numeric format." and "genotype in HapMap format or in numeric format. The names of individuals should be provided. Missing (NA) values are allowed."

Method select & Parameter setting

missingrate: max missing percentage for each SNP, users are allowed to choose one by sliding the bottom on the sliderInput.

maf: minor allele frequency for each SNP, users are allowed to choose one by sliding the bottom on the sliderInput.

Impute: users can click on the checkbox to decide whether to impute *NA* SNP or not.

http://127.0.0.1:5655 Open in Browser Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description
Input files
Parameters
Results

Parameters for SNPs

missingrate max missing percentage for each SNP, default is 0.2.

0 0.03 0.06 0.12 0.18 0.24 0.3

maf minor allele frequency for each SNP, default is 0.05.

0 0.02 0.04 0.08 0.12 0.16 0.2

☒ Impute

logical. If TRUE, imputation. Default is TRUE.

Run the software

After uploading all the needed files and setting the parameters, users can run the Software simply by clicking “Start calculation”. When calculation is down, the result will be given in the datatable below the panel, user may download the full data by clicking at “Download Genotype” bottom.

http://127.0.0.1:5655 Open in Browser Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description
Input files
Parameters
Results

Converted genotype

▶ Start calculation

Show 10 entries Search:

	SNP_1_14068	SNP_1_338176	SNP_1_703171	SNP_1_1033512
A002	0.5211267605633803	0.800711743772242	1	-1
A003	1	0.800711743772242	0	-1
A004	1	1	1	1
A005	0.5211267605633803	0.800711743772242	0	-0.2398753894080997
A006	1	1	0	-1
A007	0	-1	1	-1
A008	1	1	0	-1
A010	1	0	1	1
A011	1	1	0	-1
A012	1	1	1	-1

Showing 1 to 10 of 348 entries

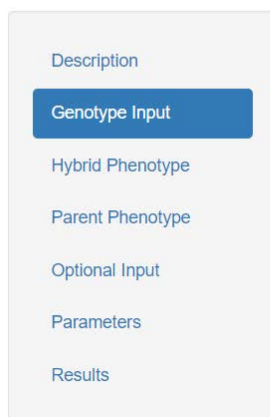
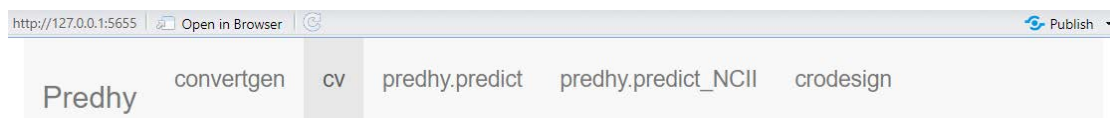
Previous 1 2 3 4 5 ... 35 Next

[Download Genotype](#)

3.2 cv

Dataset Input

Users must upload the `inbred_gen` and hybrid phenotype files, while the design matrix and the parental phenotype are optional. In design matrix module, users should upload the design matrix if you select “**Input a design matrix**”; In parent phenotype module, users should upload the parent phenotype if you select “**Input parent phenotype**”; users don’t need to upload those file, which will be ignored, if you select “**Not included**”. The dominance genotype is also optional, in dominance genotype module, if you select “**Include dominance genotypes**”; users don’t need to upload this file and the dominance genotype will be calculated automatically; if you select “**Not included**”, it will be ignored.



Input genotype

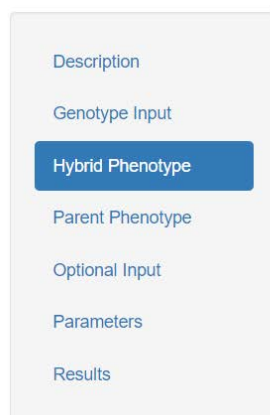
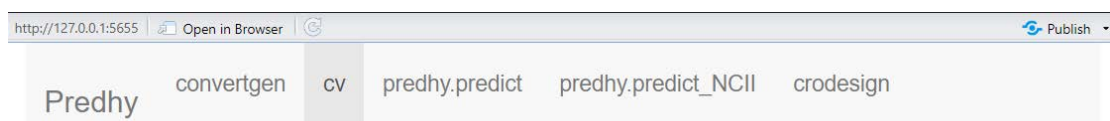
`inbred_gen`

Browse...

`inbred_gen.c`

Upload complete

A matrix for genotypes of parental lines in numeric format, coded as 1, 0 and -1. The row.names of `inbred_gen` must be provided. It can be obtained from the original genotype using `convertgen` function



Input hybrid phenotype

`hybrid_phe`

Browse...

`hybrid_phe.c`

Upload complete

A data frame with three columns. The first column and the second column are the names of male and female parents of the corresponding hybrids, respectively; the third column is the phenotypic values of hybrids. The names of male and female parents must match the rownames of `inbred_gen`. Missing (NA) values are not allowed.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description

Genotype Input

Hybrid Phenotype

Parent Phenotype

Optional Input

Parameters

Results

Input parent phenotype

A matrix of a phenotypic values of parent (Optional)

Parent Phenotype

Browse... parent_phe.csv

Upload complete

☐ Not included
☒ Input parent phenotype

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description

Genotype Input

Hybrid Phenotype

Parent Phenotype

Optional Input

Parameters

Results

Input design matrix of the fixed effects & dominance genotypes

design matrix of the fixed effects(Optional)

☒ Not included
☐ Input a design matrix

dominance genotypes(Optional)

☒ Not included
☐ Include dominance genotypes

Method select & Parameter setting

Method: There are eight GS methods in the predhy.GUI, including "GBLUP", "BayesB", "RKHS", "PLS", "LASSO", "EN", "XGBoost", "LightGBM". Users may select one of those methods or all of them simultaneously with "ALL".

Number of folds: The k for k-fold cross validation.

Replicates: Repeat number of independent replicates for the cross-validation.

The random number: The random number.

CPU: the number of CPU for parallel calculation.

The screenshot shows a web browser window with the URL `http://127.0.0.1:5655`. The browser's address bar includes "Open in Browser" and "Publish" buttons. The application's navigation bar at the top contains the "Predhy" logo and several tabs: "convertgen", "cv" (which is the active tab), "predhy.predict", "predhy.predict_NCII", and "crodesign". On the left side, there is a vertical sidebar with a list of menu items: "Description", "Genotype Input", "Hybrid Phenotype", "Parent Phenotype", "Optional Input", "Parameters" (highlighted in blue), and "Results". The main content area is titled "Select models & other parameters". It contains several input fields: a dropdown menu for "GS methods" set to "GBLUP", a text input for "the number of folds" set to "5", a text input for "replicates" set to "1", a text input for "the random number" set to "133", and a text input for "the number of CPU" set to "5".

Run the software

After uploading all the needed files and setting the parameters, users can run the Software simply by clicking “Start calculation”. The result will be print on the panel if a single method is selected. If you chose “ALL” in method, a plot of cross validation result for eight methods will be given.

Predhy

convertgen

cv

predhy.predict

predhy.predict_NCII

crodesign

Description

Genotype Input

Hybrid Phenotype

Parent Phenotype

Optional Input

Parameters

Results

Trait predictability (R_Square)

► Start calculation

GBLUP R_Square = 0.2251

Predhy

convertgen

cv

predhy.predict

predhy.predict_NCII

crodesign

Description

Genotype Input

Hybrid Phenotype

Parent Phenotype

Optional Input

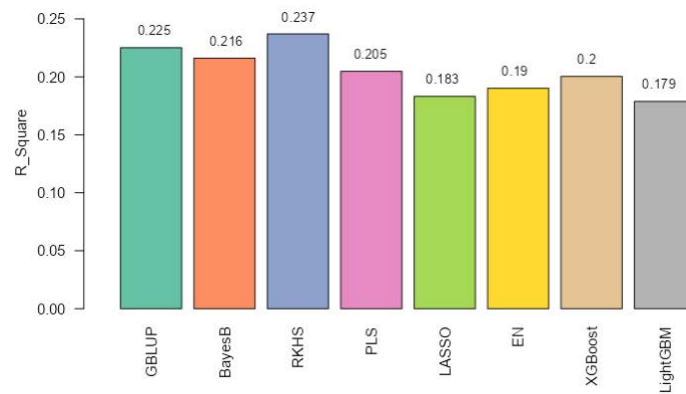
Parameters

Results

Trait predictability (R_Square)

► Start calculation

Trait predictability of 8 methods



3.3 predhy.predict

This function was designed to predict all potential crosses of a given set of parents using a subset of crosses as the training sample.

Dataset Input

Users must upload the inbred_gen and hybrid phenotype files, while the parent phenotype is optional.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv **predhy.predict** predhy.predict_NCII crodesign

Input genotype & phenotype

inbred_gen

Browse... inbred_gen.c

Upload complete

hybrid_phe

Browse... hybrid_phe.c

Upload complete

A matrix of a phenotypic values of parent (Optional)

☒ Not included

☐ Input parent phenotype

Method select & Parameter setting

Method: There are eight GS methods in the predhy.GUI for hybrid performance predicting, including "GBLUP", "BayesB", "RKHS", "PLS", "LASSO", "EN", "XGBoost", "LightGBM". Users should select one of those methods.

Prediction model: There are four options: the additive model, the additive-dominance model, the additive-phenotypic model, the additive-dominance-phenotypic model, user can choose one by select one of the choices.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv **predhy.predict** predhy.predict_NCII crodesign

Description

Input files

Methods & Models

Selection

Phenotypic values

Select methods & models

GS methods

GBLUP

GS models

the additive model

Select hybrids: Selection of hybrids based on the prediction results. There are three options: select = "all", which selects all potential crosses. select = "top", which selects the top n crosses. select = "bottom", which selects the bottom n crosses. User can decide number hybrids to select when select = "top" or select = "bottom".

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv **predhy.predict** predhy.predict_NCII crodesign

Description

Input files

Methods & Models

Selection

Phenotypic values

Select hybrids

the selection of hybrids based on the prediction results

the top n crosses

the number of selected top or bottom hybrids, only when select = "top" or select = "bottom".

100

Run the software

After uploading all the needed files and setting the parameters, users can run the Software simply by clicking "Start calculation". When calculation is down, the result will be given in the datatable below the panel, user may download the full data by clicking at "Download Results" bottom.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv **predhy.predict** predhy.predict_NCII crodesign

Description

Input files

Methods & Models

Selection

Phenotypic values

Phenotypic values of the predicted hybrids

▶ Start calculation

[Download Results](#)

Show entries Search:

top_100 ▲▼

A062/A291	1926.446905224698
A169/A291	1923.3362122002
A133/A291	1920.018562590968
A027/A291	1919.059207234987
A017/A291	1916.779254205353
A038/A291	1916.639470265751
A062/A169	1905.340900313774
A052/A291	1904.015803388827
A062/A133	1902.023250704543
A291/A398	1901.883709023961

Showing 1 to 10 of 100 entries

Previous 2 3 4 5 ... 10 Next

3.4 predhy.predict_NCII

This function was designed to predict all potential crosses of a given set of parents (usually between different heterotic groups) using a subset of crosses as the training sample, following the North Carolina mating design II.

Dataset Input

Users must upload the inbred_gen and phenotype files, along with the Heterotic group dataset (two files, one contains male_names, the other contains female_names), while the parent phenotype is optional.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description

Input files

Parent names

Methods & Models

Selection

Phenotypic values

Input genotype & phenotype

inbred_gen A matrix for genotypes of parental lines in numeric format, coded as 1, 0 and -1. The row.names of inbred_gen must be provided. It can be obtained from the original genotype using `convertgen` function

Browse... inbred_gen.c

Upload complete

hybrid_phe A data frame with three columns. The first column and the second column are the names of male and female parents of the corresponding hybrids, respectively; the third column is the phenotypic values of hybrids. The names of male and female parents must match the rownames of inbred_gen. Missing (NA) values are not allowed.

Browse... hybrid_phe.c

Upload complete

A matrix of a phenotypic values of parent (Optional)

☒ Not included

☐ Input parent phenotype

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description

Input files

Parent names

Methods & Models

Selection

Phenotypic values

Input names of parents

male_name a vector of the names of male parents.

Browse... male_name.csv

Upload complete

female_name a vector of the names of female parents.

Browse... female_name.csv

Upload complete

Method select & Parameter setting

Method: There are eight GS methods in the predhy.GUI for hybrid performance predicting, including "GBLUP", "BayesB", "RKHS", "PLS", "LASSO", "EN", "XGBoost", "LightGBM". Users should select one of those methods.

Prediction model: There are four options: the additive model, the additive-dominance model, the additive-phenotypic model, the additive-dominance-phenotypic model, user can choose one by select one of the choices.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description

Input files

Parent names

Methods & Models

Selection

Phenotypic values

Select methods & models

GS methods

PLS

GS models

the additive model

Select hybrids: Selection of hybrids based on the prediction results. There are three options: select = "all", which selects all potential crosses. select = "top", which selects the top n crosses. select = "bottom", which selects the bottom n crosses. User can decide number hybrids to select when select = "top" or select = "bottom".

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description

Input files

Parent names

Methods & Models

Selection

Phenotypic values

Select hybrids

the selection of hybrids based on the prediction results

the top n crosses

the number of selected top or bottom hybrids, only when select = "top" or select = "bottom".

100

Run the software

After uploading all the needed files and setting the parameters, users can run the Software simply by clicking "Start calculation". When calculation is down, the result will be given in the datatable below the panel, user may download the full data by clicking at "Download Results" bottom.

Predhy

convertgen

cv

predhy.predict

predhy.predict_NCII

crodesign

Description

Input files

Parent names

Methods & Models

Selection

Phenotypic values

Phenotypic values of the predicted hybrids

▶ Start calculation

[Download Results](#)

Show 10 entries

Search:

top_100

A014/A014	2171.048111502145
A014/A038	2168.236895244563
A038/A014	2168.236895244563
A038/A038	2165.425678986982
A014/A358	2156.81104166041
A358/A014	2156.81104166041
A038/A358	2153.999825402828
A358/A038	2153.999825402828
A358/A358	2142.573971818676
A014/A253	2107.792308368091

Showing 1 to 10 of 100 entries

Previous 1 2 3 4 5 ... 10 Next

3.5 crodesign

This function was designed to generate a mating design for a subset of crosses based on a balanced random partial rectangle cross-design (BRPRCD) (Xu et al. 2016).

Dataset Input

Users need to upload the Parent names dataset (two files, one contains male_names, the other contains female_names).

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description
Parent names
Input parameters
Results

Input parent names

male parent name a table containing names of male parents

Browse... male_name.
Upload complete

female parent name a table containing names of female parents

Browse... female_name.
Upload complete

Method selection & Parameter setting

percentage: User can decide the percentage of all potential hybrids to be evaluated in the field by clicking the numericInput.

seed: The random number.

http://127.0.0.1:5655 | Open in Browser | Publish

Predhy convertgen cv predhy.predict predhy.predict_NCII crodesign

Description
Parent names
Input parameters
Results

Parameters

percentage an integer denoting 1/d percentage of crosses to be evaluated in the field.

50

seed the random number

123

Run the software

After uploading all the needed files and setting the parameters, users can run the Software simply by clicking “Start calculation”. When calculation is down, the result will be given in the datatable below the panel, user may download the full data by clicking at “Download crodesign” bottom.

http://127.0.0.1:5655

Open in Browser

Publish

Predhy

convertgen

cv

predhy.predict

predhy.predict_NCII

crodesign

Description

Parent names

Input parameters

Results

Results

▶ Start calculation

Download crodesign

Show 10 entries

Search:

	crossID	male_Name	female_Name
1	1	A008	A007
2	2	A054	A007
3	3	A156	A007
4	4	A005	A007
5	5	A335	A007
6	6	A426	A007
7	7	A025	A007
8	8	A011	A007
9	9	A186	A007
10	10	A092	A030

Showing 1 to 10 of 3,370 entries

Previous12345...337Next