SoftwareX 24 (2023) 101544

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original Software Publication

RankspeQ: An R package platform for genotype characterization and performance-ranking based on MultispeQ measurements

Jonatan Soto [a,*], Johan Aparicio [a], Aquiles Darghan [b], Milan Oldřich Urban [a,*]

[a] Alliance of Bioversity International & International Center for Tropical Agriculture (CIAT), Cali, Colombia
[b] Department of Agronomy, Universidad Nacional de Colombia, Bogotá, Colombia

ARTICLE INFO

Keywords:
Ranking
Fluorescence
Photosynthetic efficiency
Confusion matrices

ABSTRACT

With the new technologies in plant phenotyping, robust and reliable tools are still required to analyze large-scale and multivariate datasets. The RankspeQ is a novel R package developed to evaluate genotype performance and to support selection-driven decisions based on leaf traits and environment-related variables measured by the MultispeQ device. The presented software consists of 3 main functions: i) data cleaning, ii) computational trait-genotype ranking, and iii) comparison of accessions against grain yield or another crop trait. Optionally, the evaluation can be performed by alternative groups which can be defined by the user, such as genepools, families, among others. The software development as well as the data evaluation was made with datasets of *Phaseolus spp*. experiments. However, R code - with easy modifications - can be used on any other crop. This valuable tool helps to understand the hidden potential of MultispeQ equipment and - most - identify crop traits useful in genotype characterization in particular environments. The tool has direct potential for physiologists, breeders etc. as it identifies the best performing accessions. However, it also targets false positive results with low yield but high photosynthetic performance. We also propose to use a new efficiency index to calculate the ratio of incoming radiation for net photosynthesis in proportion to light dissipation processes. Further updates will include new algorithms (e.g. trait heritability), generalization to other species and a shiny interface to make the software user friendly.

Required metadata
Current code version

Nr	Code metadata description	*Please fill in this column*
C1	Current code version	*V1.0*
C2	Permanent link to code/repository used of this code version	*https://github.com/jssotob/RankspeQ*
C3	Code Ocean compute capsule	*Not applicable*
C4	Legal Code License	*MIT + file LICENSE*
C5	Code versioning system used	Git
C6	Software code languages, tools, and services used	*R*
C7	Compilation requirements, operating environments & dependencies	*R ≥ 4.0* *RStudio≥1.2.5*
C8	If available Link to developer documentation/manual	*Not applicable*
C9	Support email for questions	J.S.Soto@cgiar.org; jssotob@unal.edu.co

1. Motivation and significance

In the last years, the interest and development of new tools for faster and reliable plant phenotyping techniques has tremendously increased [1]. The quantification of qualitative traits of interest for selection i.e. genomic or phenomic and prediction [2] should help the breeding programs to support decision processes for the future crop improvement strategies. MultispeQ, as a handheld, fast and multi trait device connected to the PhotosynQ platform [3] has been developed for addressing the challenges to collect phenotypic data, store it in an open-source repository and analyze it. The biggest advantage of this device is that it allows to measure different parameters related to the light/dark phase of photosynthesis (quantum yield Phi2 + other indices of photosynthetic efficiency, chlorophyll fluorescence, Linear Electron Flow (LEF), among others), traits related to morphology of the leaf structure (leaf temperature differential, leaf angle, leaf thickness, SPAD etc.) as well as environmental parameters such as air temperature and humidity, atmospheric pressure, among others at the real time of leaf

* Corresponding authors.
E-mail addresses: J.S.Soto@cgiar.org, jssotob@unal.edu.co (J. Soto), m.urban@cgiar.org (M.O. Urban).

https://doi.org/10.1016/j.softx.2023.101544
Received 19 October 2022; Received in revised form 4 August 2023; Accepted 29 September 2023
Available online 17 October 2023

Table 1
Selected traits and Pearson correlation to final yield (all significant) of common bean genotypes.

Trait	Coefficient	Trait	Coefficient
LEF	−0.68	FmPrime	0.61
NPQt	−0.43	FvP_over_FmP	0.71
PhiNPQ	−0.72	Phi_Index	0.64
PS1.Oxidized.Centers	−0.54	PS1.Active.Centers	0.35
Vh.	−0.46	PS1.Over.Reduced.Centers	0.27
v_initial_P700	−0.38	FoPrime	0.28
P700_DIRK_ampl	−0.39	Fs	0.39
gH.	−0.39	kP700	0.19
Leaf.Temperature.Differential	−0.19	tP700	0.23
Phi2	0.64	Relative.Chlorophyll	0.19
PhiNO	0.58	value1	0.33

measurements. MultispeQ generates a large-scale dataset from a single project, specially whether a time series during the crop cycle are evaluated. However, the physiological data (especially connected to photosynthetic performance) always needs to be verified for outliers. Therefore, statistical methods are required for both visualizing, analyzing and/or doing selection of genotypes based on the preferred traits (e.g. the grain yield or individual yield components) [2].

In this study, a case study plant is *Phaseolus vulgaris* - common bean. It is an important crop for human nutrition as it has a valuable content of minerals (Ca, Cu, Fe, Mg, Mn, Zn), fiber and proteins and is widely produced/used/consumed in developing countries mostly by small farmers [4]. A study conducted by Dramadri et al. in 2021 [5] on 256 common bean genotypes included measurements by MultispeQ the 7th and 9th week after sowing. Some relevant photosynthetic (PS) traits were evaluated to identify potential QTL regions for selection under drought stress (DS). The authors report that there was no evidence against the null hypothesis among the PS traits such as Relative Chlorophyll, Quantum Yield of Photosystem 2 (Phi2), and incoming light lost via non-regulated processes (PhiNO). Likewise, in the same study, the authors showed at least a weak correlation to yield components and partitioning traits under drought conditions. Another study conducted by Zhu et al. in 2020 [6] on maize and wheat integrated different tools to quantify the leaf chlorophyll content under different doses of nitrogen fertilization. The unitless relative chlorophyll obtained by MultispeQ (leaf greenness) was correlated to actual leaf chlorophyll content units (destructively) with a positive Pearson correlation (r over 0.90). These transformed values were contrasted with data obtained in laboratory and unmanned aerial vehicle (UAV) imaging. Four modeling algorithms included machine learning techniques were applied. The authors concluded that the use of both hyperspectral UAV sensors as well as ground measurements taken by MultispeQ to estimate leaf chlorophyll content is an important advance to quantify the nitrogen stress with high accuracy. Yan et al. (2020), exposed maize to Fe_3O_4 nanoparticles at levels of 50 and 500 mg.kg^{-1} which generated a morphological toxicity in leaves and roots [7]. However, traits such as LEF and Phi2 measured by MultispeQ in this study did not show evidence against the null hypothesis between the treatments.

The current state of the art regarding to MultispeQ data and analysis is not widely explored. This suggests that the potential of the device is still not fully used and awaits to be discovered. One problem can be the lack of deep understanding of photosynthesis- and fluorescence-related data. Another problem is likely connected to the lack of a quick and reliable platform, which will process data and give a user a meaningful and quick result. In our study, we tried to solve the latter problem, partially because our common bean database is large and complex, using different genepools, genotypes, treatments, and environments. The team of a breeder, crop physiologists, database specialist and statisticians were created to target this task. We constructed an R package able to clean, process and analyze data on every crop type. Our wish is that everyone can rapidly rank crop accessions evaluated (after each measurement) and thus understand its performance in a particular target population of environments understanding better the GxE. When deeper knowledge is available, in the future we will implement/explore other MultispeQ traits or their combinations into the presented package to deal with biotic and abiotic stressors of other crops.

2. Software description

RankspeQ is a package with 3 main functions written in R using object-oriented programming. The structure of each one is further explained below.

Any project conducted in “Plants” kingdom with the MultispeQ device by the commonly used protocols “Photosynthesis RIDES no open/close” or “Photosynthesis RIDES 2.0” or any other that contains the traits described in Table 1 and uploaded in the PhotosynQ Network can be passed through the functions of this package for cleaning and ranking of genotypes.

A dataset to be passed through the functions described in the following sections must include the following variables:

i) A project question related to genotypes to be compared/ranked. It is highly desirable to be named specifically as “Genotype” or any other related string. The function returns a warning message in the case the provided genotype string is not found in the raw dataset.
ii) For a given day of sampling, it is highly recommended to complete all the measurements in a single date since the genotype ranking and contrasting against yield components (or other traits) is done by different dates. To compare genotypes measured the same part of the day also is physiologically correct methodology. Likewise, a comparison for photosynthesis acclimation can be tested by repeating the measurements twice in a day as in the morning (earlier than 12:00 pm) and the afternoon (later than 12:00 pm).

Common mistakes while taking measurements with a MultispeQ include the wrong selection of a project question, mistakenly identified genotype; over and/or under samplings making an unbalanced data, among others. It is a policy of PhotosynQ that any wrong observation and/or mistake uploaded into the Network cannot be deleted. However, the user can omit them manually before applying the functions. Therefore, to avoid wrong calculations and under/over estimation, the dataset needs to be properly prepared and always checked for outliers as photosynthetic data can easily be measured out of the physiologically acceptable ranges. To simplify this, we included a cleaning function as the first step of data preparation.

2.1. The function MSPQ_tidy

This function is developed to prepare the raw dataset for the further analysis and genotype-trait classification. The return is a list that includes nine objects whether the argument plotIm = TRUE, or eight otherwise and it is the main argument for the upcoming function MSPQ_ranks.

It contains five arguments which are described as follows: i) df. A required argument with the raw MultispeQ data frame generated either by the “Photosynthesis RIDES no open/close” or “Photosynthesis RIDES 2.0” protocols. ii) genotype. A required string of length one (SL1) with the name of the genotype column in the dataset. It is case sensitive. iii) time.dif. A required logical flag, TRUE whether a date of measurement includes two repetitions as morning and afternoon and a comparison between these time intervals is desired; otherwise, FALSE whether only one repetition was done in a day regardless the time of the day. iv) data_name = NULL. A default argument with a project name defined by the user. It must be a character string of length one and is implemented in the summary object v) plotIm = FALSE. A default argument. If TRUE the boxplots of the imputed variables are generated and saved into the

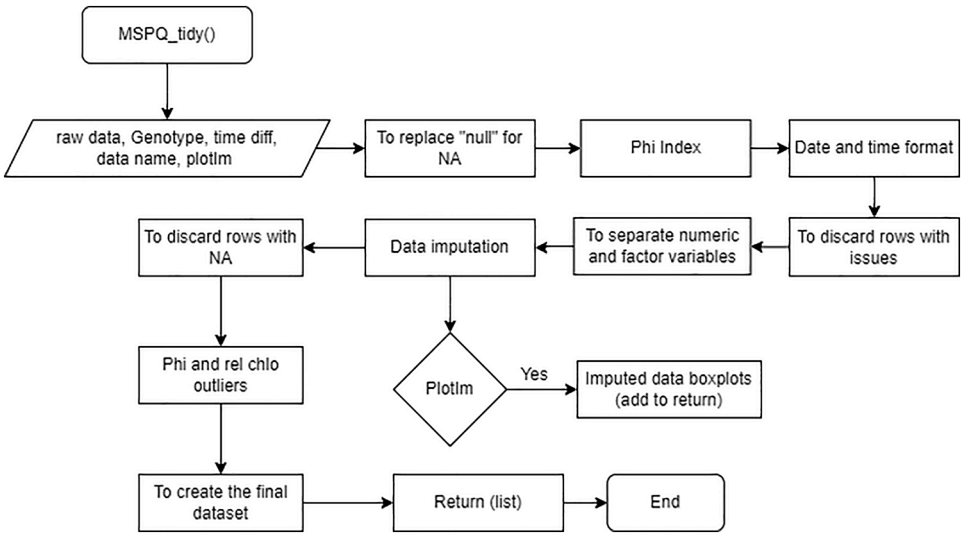

Fig 1. Flowchart of the MSPQ_tidy function.

return. All the required dependencies are automatically installed and loaded in case of missing in the R session of the user.

The function MSPQ_tidy has ten implemented steps. The algorithm flowchart and the description of each procedure is detailed below (Fig 1).

2.1.1. *Replacing "null" for NA*

Randomly, the character string "null" can be found at several cells (observations), affecting the class of a given trait i.e., numeric to character. For the MultispeQ RIDES protocols, the most affected 'replaced' traits are traits related to the absorbance. Therefore, this procedure finds this string and replaces it by missing values NA. Therefore, the class of the affected variable returns to the original.

2.1.2. *Calculating Phi_Index*

The family of the three Phi variables calculated by the MultispeQ have a compositional characteristic, providing relative information [8] as they describe the destination of the captured light. This means that their sum is always equal to 1 and each Phi variable is the proportion-of-use (ratio) of the incoming PAR light intensity for the three different photosynthesis light-phase mechanisms i.e., net photosynthesis or quantum yield (Phi2), non-photochemical quenching (PhiNPQ) and non-regulated processes (PhiNO). However, to understand the general response of the plant, we suggest to create a new index (Phi Index), which calculates the effectivity of quantum yield over both non-productive (but protective and important) values.

Phi Index is defined as the occurrence ratio of net photosynthesis divided by the sum of occurrence ratios of non-photosynthetic events (also defined as odd [13]) and calculated by the Eq. (1):

$$Phi\ Index = \frac{Phi2}{PhiNPQ + PhiNO} \tag{1}$$

In this sense, if Phi Index is higher than 1 it means that the incoming light is mostly used for net photosynthetic production. If Phi Index is equal to 1 it means that the incoming light is equally distributed into both net photosynthesis and energy dissipation/tissue protection by heat dissipation or other processes. If Phi Index is less than 1 it means that the incoming light is being mostly dissipated and quenched by chlorophyll fluorescence and other processes than used for net photosynthetic production. The minimum hypothetical value 0 means that there are no excited electrons coming into Photosystem II.

However, we understand that further exploration of the relationship between these parameters but especially its regression with the final yield or other important crop performance traits remains still an open question.

2.1.3. *Formatting dates and creating the time (AM/PM) variable*

The time variable in the MultispeQ dataset is originally formatted as MM/DD/YYYY hh:mm AM/PM. Since the hour is recorded exactly during measurements, this variable may contain a high amount of factor levels (times). Therefore, this procedure deletes the hour from each cell and two variables are created "date" formatted as MM/DD/YYYY and "time" which includes both AM and PM levels whether time.dif=TRUE and two measurements of the experiment were conducted in a day to identify any possible genotypic variability as well as to obtain information on genotypic acclimation. In case that the AM/PM strings (measurements) are missing in the original dataset, the function MSPQ_tidy will not continue and the error message "There is not AM/PM indicator in column "time" and/or hour is missing, check out first" will be displayed in the console.

2.1.4. *Discarding rows with issues*

The non-empty values in the "Issues" variable mean that those observations returned an error or were red-labeled by the equipment. In the function, these rows are discarded and separated into the removed_observations object that can be found, verified and analyzed in the output list not to lose any important data. The summary table includes the number of issues and the removed_freq collapsible tree displays dates, sources of variation and frequencies of removed observations in order to trace whether the removals are product of wrong measurements or, importantly, a genotype-response effect (GxE).

2.1.5. *Separating factors and character columns from the dataset*

The MultispeQ dataset generated by the protocols "Photosynthesis RIDES no open/close" or "Photosynthesis RIDES 2.0", besides the project questions, contains multiple factors or non-numerical variables that can be either used for another type of analyses or contains unique values. Therefore, this procedure removes the variables that meet at least one of the below statements and are not included into the final dataset: i) If the variable is non-numerical except for date, time, the project questions, and device ID (MAC address of the MultispeQ). ii) If the variable is a character string. iii) If the 50% or more of the variable is empty. iv) If the length of the unique values is less or equal than three.

Some other variables removal include: i) The class of the "ID" variable is an integer corresponding to a unique value for each observation. This variable is also removed. ii) The variable "SPAD_650" is removed if it is identical to the variable "Relative.Chlorophyll". iii) The variables "Longitude and Latitude" are removed in case of missing values.

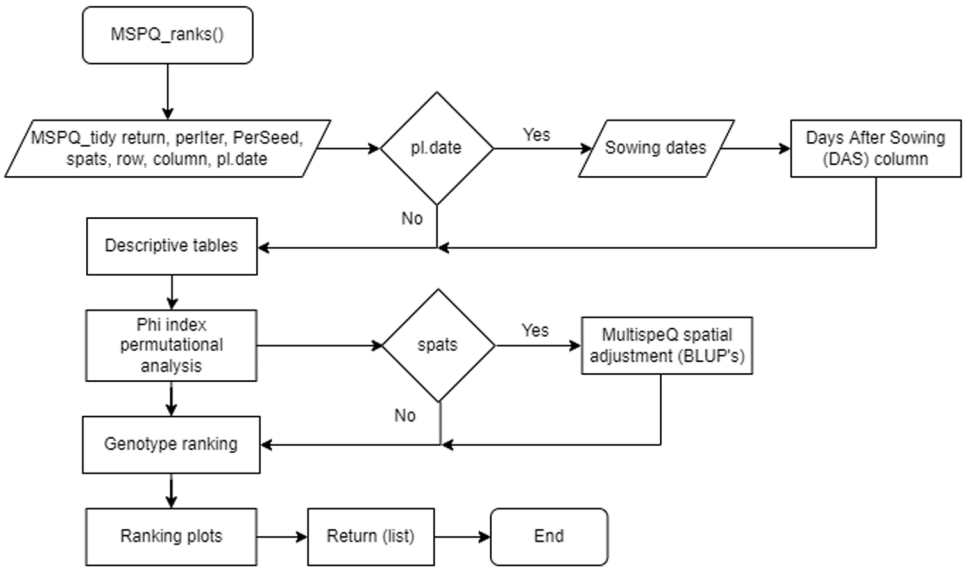

Fig 2. Flowchart of the MSPQ_ranks function.

Table 2
Error messages returned by the function target_trait_comparison.

Error message	Cause	Solution
The argument target.trait.file must be a data frame that contains the *Genotype* column and the yield component to evaluate.	target.trait.file is not a data frame	The target.trait.file must be converted to a data frame.
The variable *Genotype* was not found in the target.trait.file dataset.	Either if the genotype column is not included in the target.trait.file or its name is different than in the MultispeQ dataset.	Make sure that the *Genotype* column is included in the target.trait.file and its name is the same as the MultispeQ dataset.
The genotype(s) *Genotype ID* from the yield file not found in the ranked genotypes. Should it be named as *Genotype ID (MultispeQ genotypes)*?	Either if a genotype ID or name from the target.trait.file is not found in the MultispeQ genotypes. The function tries to suggest a correct name.	Check and change the genotype names that were not found in the target.trait.file. All must be same named as the MultispeQ dataset.
The MultispeQ ranks obtained by the function MSPQ_ranks were adjusted by SpATS. The spatial variables *row* and *column* were not found in the yield file.	If the SpATS procedure in the function MSPQ_ranks was called, at least one of the displayed columns are not included in the target.trait.file or their names are different than the ones used into the MSPQ_ranks.	The target.trait.file data frame must contain the same spatial coordinates columns as in the MultispeQ dataset if the SpATS procedure was conducted into the function MSPQ_ranks.
The variable(s) *sources of variation* do not exist in the yield file. Make sure that it/they exist and try again.	If any other source of variation different than *Genotype* was evaluated in the experiment and it is not included in the target.trait.file data frame or its name is different than in the MultispeQ dataset.	All the project questions and answers (i.e. Genotype, Treatment, Block, etc.) from the MultispeQ project must be included in the target.trait.file in order to conduct the yield contrast for each level.

2.1.6. Data imputation

As mentioned in 2.1.1, MultispeQ gives randomly several "null" observations. Once they are mutated into NA values, this procedure imputes them by using the R package mice (Multivariate Imputation by Chain Equations) [9], following the steps as follows:

i) The variables "Latitude" and "Longitude" are not imputed. ii) By the function aggr the variables-with-empty values are identified as well as the count of the empty cells. iii) The dataset is split by dates-of-measurement and the function mice is applied to each variable, including the following arguments $m = 5$ as the number of multiple imputations; method = "sample" which takes random samples from the observed values for the imputation, seed = 500 as the random seed; printFlag = FALSE as hiding the procedures in console, making the computing process more efficient. A progress bar is implemented for this step.

Since the MSPQ_ranks function includes some of the absorbance-related traits (Table 1) for conducting the computational genotype-ranks, these imputations are required in order to avoid empty values/responses from the measurements.

If the user sets the argument plotIm = TRUE into the function MSPQ_tidy. The boxplots of the imputed variables per day will be obtained in the final output. The temporal distribution of the variable "PS1.Open.Centers" and its imputed values (red dots) after the mice algorithm for the example illustrated in the Section 3.

2.1.7. Finding and discarding rows with empty values

If an observation contains missing values even after the cleaning of the variables and data imputation, the row is discarded and moved to the removed_observations object of the final output.

2.1.8. Removing PAR and Phi2 outliers

Since there is a linear but negative correlation between the observed incoming light (PAR) and Phi2, a low value of light intensity (darkness) might induce an overestimation of the photosynthesis efficiency. With that in mind, it is recommended not to take MultispeQ measurements when the observed PAR is lower than 100 μmol m^{-2} s^{-1}. Therefore, if PAR is lower than 1 μmol m^{-2} s^{-1} or higher than 2500 μmol m^{-2} s^{-1} as well as Phi2 is lower than 0.03 or higher than 0.85, the observation is discarded and moved to the removed_observations object of the final output for verification.

2.1.9. The return

The function MSPQ_tidy returns a list with the objects as follows: i) The numerical dataset generated by both variables and selected observations, including the date and time columns after formatting, the project questions as well as the device ID. ii) Non-numerical dataset with

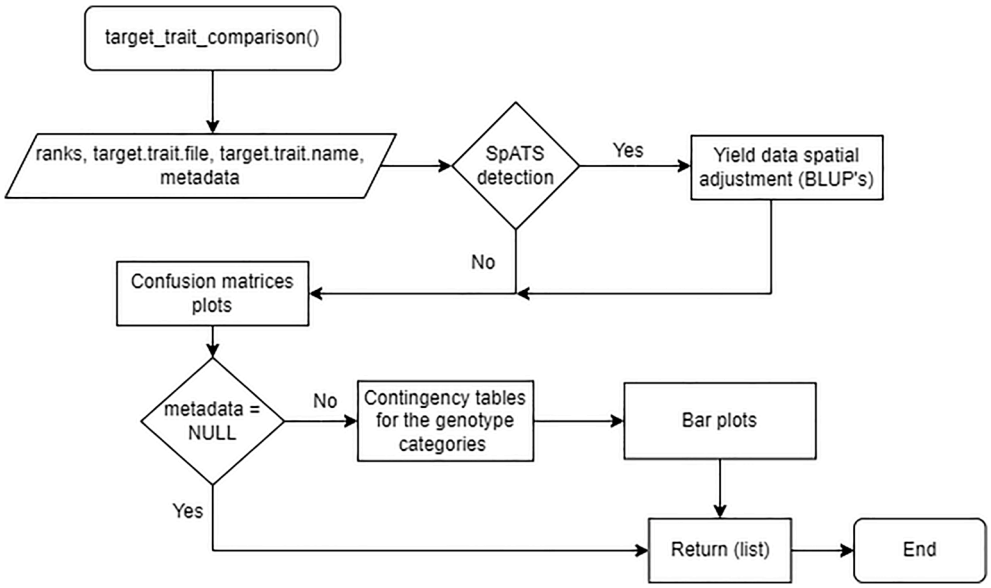

Fig 3. Flowchart of the target_trait_comparison function.

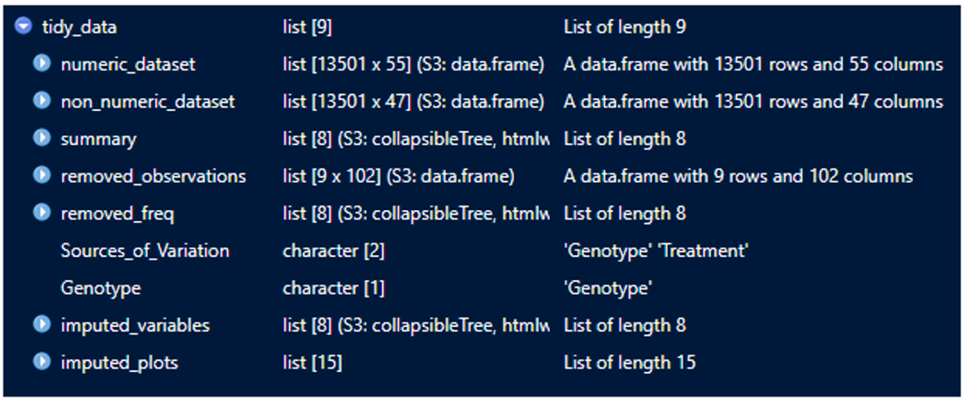

Fig 4. Return of the MSPQ_tidy function for the "19-06 BASE100" dataset.

all the removed variables. iii) A non-collapsed collapsible tree object with the summary of the process. It lists the initial dimensions of the raw dataset and calculates the proportion of removals. It includes other values such as whether "Latitude" and "Longitude" variables were removed; whether "SPAD_650" was removed, among others. iv) removed_observations data frame that contains all the observations removed due to issues, outliers in the defined reference values and/or any missing values. v) A collapsed collapsible tree with the contingency table of removed observations by date-of-sampling and sources of variation (project questions, treatments). vi) A character string with the sources of variation (project questions) and row, column names in case of spatial arrangement. vii) A character string with the name of the genotype column. viii) A non-collapsed collapsible tree with the imputed variables and the number of missing values of each of them. ix) If plotIm = TRUE, a list with the boxplots for each imputed variable.

2.2. The function MSPQ_ranks

This function is developed for a further analysis and genotype-trait classification for a series of measurements. The return is a list that includes nine objects which are detailed below. The function contains seven arguments described as follows: i) out. This is the only required argument of the function and must be in the list returned by the function MSPQ_tidy. ii) perIter = 100. An integer of length one (IL1) with the number of permutations for the Phi Index ratio analysis; 100 by default. iii) PerSeed = 123. An IL1 with the random seed for the permutational Phi Index ratio analysis; 123 by default. iv) spats. A logical flag, FALSE by default. If TRUE, the step 2.2.4 is conducted. v) row = NULL by default. If spats = TRUE, it must contain a SL1 with the name of the row variable associated to each observation. vi) column = NULL by default. If spats = TRUE, it must contain a character string with the name of the column variable associated to each observation i.e. row = "row" and column = "column". vii) pl.date = FALSE. A logical flag, FALSE by default. If TRUE, a series of questions will appear in console in case the user desires to provide the sowing date or multiple dates. The date format must be mm/dd/yyyy. The function will request to type again in case of a different format. The algorithm calculates the Days After Sowing (DAS) for each sampling date.

The function MSPQ_ranks will return different errors in case of missing row and col arguments or mismatches when spats = TRUE.

The function MSPQ_ranks has four implemented steps. The algorithm flowchart and the description of each procedure is detailed below (Fig 2).

2.2.1. Descriptive tables

The numerical dataset from the MSPQ_tidy return is grouped by date, time and the sources of variation (project questions). The mean, standard deviation, median and coefficient of variation tables are

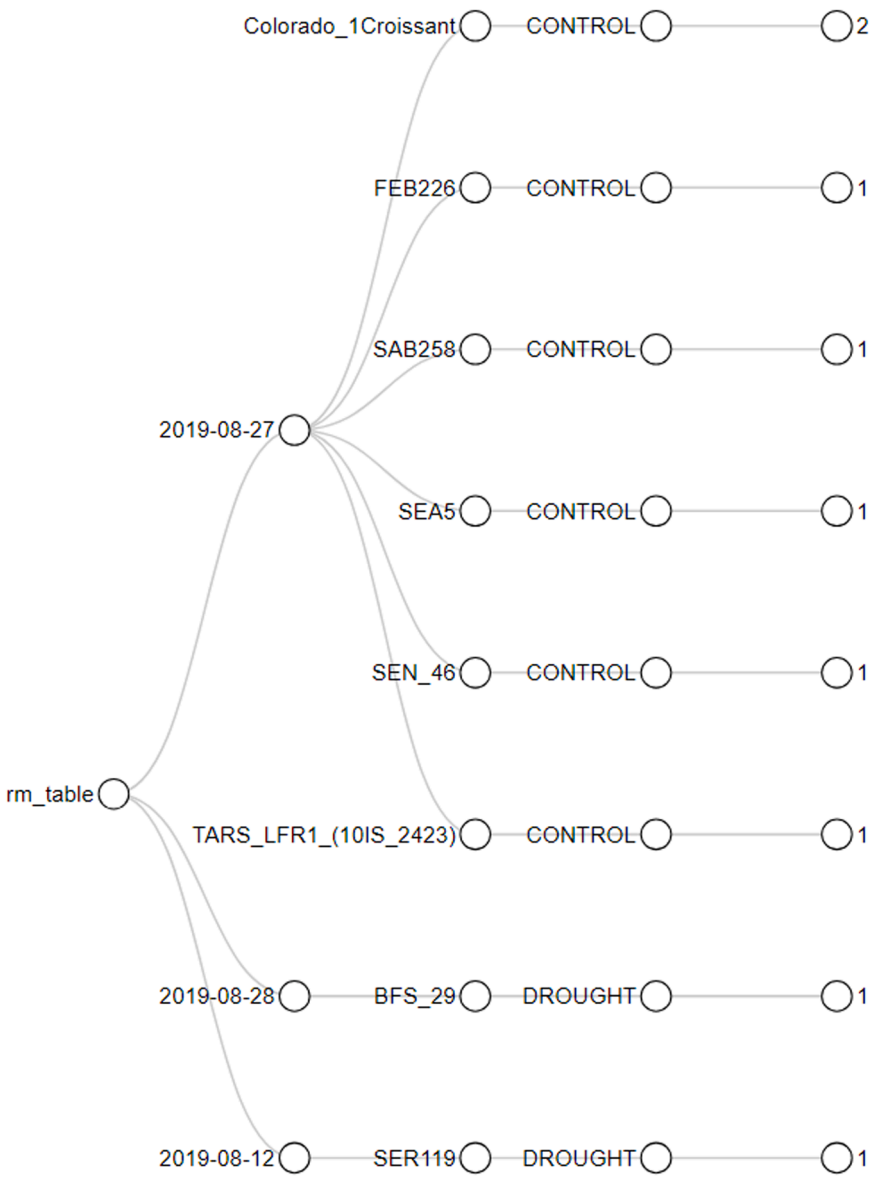

Fig 5. Interactive collapsible tree with the frequency of removed observations.

summarized for all the traits (photosynthetic, leaf and environmental). These data frames are the first 4 objects of the return list.

2.2.2. Phi Index permutational analysis

As explained in 2.1.2, the Phi Index is an odd (ratio of occurrence events). With these values calculated for every observation, the procedure conducts a permutational analysis of the morning/afternoon repeated measurements only if time.diff = TRUE in MSPQ_tidy in order to determine a ratio between quantum yield and dissipation processes in the morning and the afternoon for a given day. Physiologically speaking, there should be differences because many factors and feedback regulation loops influence the actual ratio of incoming light conversion into Photosystem II.

The steps for this analysis are: i) To group the dataset by dates and sources of variation. ii) For every single group, the Phi Index replicates are randomized as (perIter – 1) times (by default perIter = 100, thus 99 permutations are run). iii) For every permutation, the odd ratio AM/PM is calculated. If the odd ratio is less than 1, the quantum yield is higher in the afternoon than in the morning, vice versa if higher than 1. iv) The last odd ratio (100th by default) is calculated from the original dataset. v) The number of odd ratios lower than the original (100th by default) is calculated by the Eq. (2):

$$D = sum\left(odd\ ratio < odd\ ratio_{periter}\right) \tag{2}$$

vi) the permutational P-value of the test is calculated by the Eq. (3):

$$Pvalue_{perm} = \frac{D}{perIter} \tag{3}$$

vii) The evaluation of the p-value returns TRUE whether p-value is lower than 0.05 or FALSE otherwise. viii) A collapsed collapsible tree is saved with the contingency table of the number of days/samplings in which the evaluation of p-value == TRUE meaning that there is a significant difference between the Phi Index in the morning and in the afternoon by sources of variation (project questions).

A progress bar is implemented for this step.

2.2.3. Genotype-trait score computation

The selected traits included in the genotype rank-score (the traits can differ for different crops) are listed in Table 1. This selection was made by using a correlation matrix of all available MultispeQ variables (48 traits) obtained from the numerical dataset of the MSPQ_tidy function

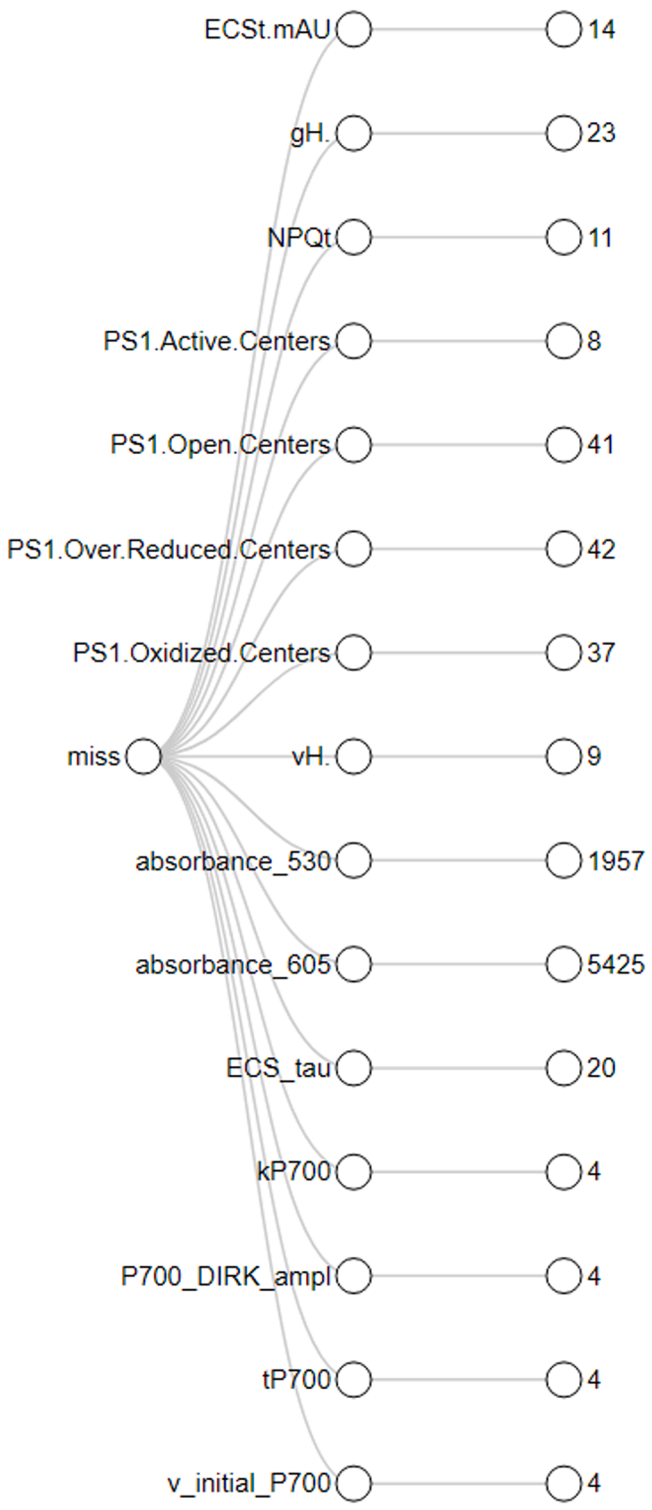

Fig 6. Interactive collapsible tree with the frequency of imputed data.

with the final yield of a common bean (*Phaseolus vulgaris*) and tepary bean (*P. acutifolius*) experiment with 13,510 datapoints taken in 13 different dates and two measurements per day (morning and afternoon) (dataset not shown).

The traits were selected by the significant positive or negative Pearson correlations. From 48 traits, 22 traits were selected important for further analyses. As mentioned above, this list of pre-selected variables can easily be changed based on different crop data, user experience or other preferences. However, based on the number of measured plants and its quality, the *Phaseolus spp*. scientists can use the set as an authoritative list.

The steps to conduct the ranking of genotypes based on their MultispeQ performance are: i) To group the dataset by sampling dates either from the medians table (2.2.1.) or the spatial adjusted dataset by the SpATS procedure (2.2.4.). ii) For every trait and every sampling date, the values are ranked by the absolute value of the correlation described in Table 1, i.e. positively correlated variables are ranked from lowest to highest and opposite for negative correlations. Therefore, negative correlations give a low score for a high trait value and vice versa. iii) A score between 1 - *n* (number of genotypes) is given to each genotype. iv) The sum of the trait scores is computed as a total score per genotype and per date-of-measurement. v) A multi-facet scatterplot for every time of the day (AM/PM) and source of variation is created where each facet is a date of measurement. The abscissa (x) axis corresponds to the genotypes and the ordinate (y) axis to the cumulated trait score. The facets are sorted accordingly to the cumulated trait score.

2.2.4. *Data adjustment by spatial arrangement*

This is an optional procedure adapted from the R-shiny package MrBean [10] which can be applied if the experiment includes a spatial factor in the design such as genotype grid arrangement in rows and columns. The function MSPQ_analysis includes the optional logical argument spats = FALSE by default. If TRUE, the dataset must contain two extra numerical columns as row and column with respective coordinates per observation. In addition, the arguments row = NULL by default and column = NULL by default must contain a character string with the name of the columns i.e. row = "row" and column = "column".

The steps to conduct the spatial adjustment are:

i) To split the numerical dataset obtained from the MSPQ_tidy function by dates of measurement; time of measurement (AM/PM) and sources of variation such as treatment i.e. irrigated, drought.
ii) For every single subset a spatial model is fitted by the function SpATS available in the R package SpATS (Rodríguez-Álvarez et al., 2018). The defined arguments are as follows: a) response: a character string with the traits to be adjusted (Table 1); genotype: a character string with the column name of the genotypes; row and column names taken from above; covariate = c("Leaf.Temperature", "Light.Intensity..PAR.") by default to include these covariates into the model formula.
iii) To obtain the BLUP's by applying the function predict to the fitted model.
iv) To build the new numerical dataset with sources of variation and the spatial adjusted BLUP's for all the traits in order to follow the step 2.2.3.

2.2.5. *The return*

The function MSPQ_ranks returns a list with the objects as follows: i) A data frame with the means of the replicates per date-of-sampling and sources of variation. ii) A data frame with the standard deviation of the replicates per date-of-sampling and sources of variation. iii) A data frame with the medians of the replicates per date-of-sampling and sources of variation. iv) A data frame with the coefficient of variation of the replicates per date-of-sampling and sources of variation. v) A data frame with the spatial adjusted BLUP's whether spats = TRUE. vi) A series of plots with the cumulated trait scores obtained from the step 2.2.3. vii) a collapsed collapsible tree with the results of the Phi2 permutational analysis. viii) A character string with the row and column arguments whether spats = TRUE. These will be used in the function target_trait_comparison.

2.3. *The function target_trait_comparison*

This function compares the previous MultispeQ ranks for all the dates

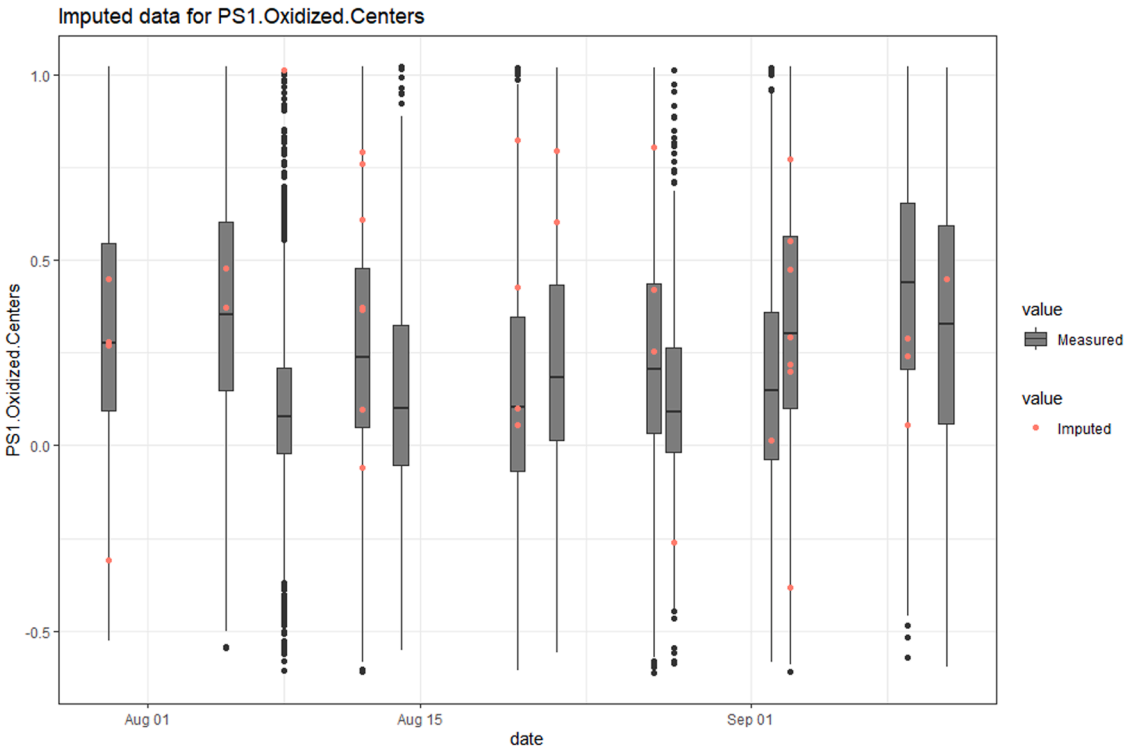

Fig 7. Distribution of the imputed variable *PS1.Oxidized.Centers* by dates-of-measurement.

with the final yield (i.e. kg ha^{-1}; g plant^{-1}). or literally any other individual target trait or yield component (i.e. Pod Harvest Index, crop biomass, seed quality related to protein or mineral content, plant vigor, vegetation or growth indices etc.). Likewise, a file with relevant information about the genotypes passport or grouping traits such as accessions; abiotic traits; among others, can be optionally included to evaluate the behavior of the MultispeQ scores with these specific groups of genotypes. The function contains 4 arguments described as follows: i) ranks. The list returned by the previous function MSPQ_ranks. ii) target.trait.file. A data frame with the actual yield or any other response trait mentioned above of the genotypes in wide data format where each column must contain and be named as the same MultispeQ project questions. The function will detect if the spatial adjustment (SpATS) of the MultispeQ traits was done. In that case, this data frame must also include the spatial columns of each observation and they must be named in the same way as in the function MSPQ_ranks. iii) target.trait.name. A character SL1 with the name of the variable that contains the yield data in the target.trait.file; it is case sensitive. iv) metadata = NULL. An optional argument. A data frame in wide format with one or more grouping traits associated to each evaluated genotype (i.e. Genepool, abiotic traits).

Before conducting the following steps, the function target_trait_comparison evaluates the class of the provided arguments and will return different errors in case of any mismatch or missing required arguments (Table 2).

The function target_trait_comparison has three steps. The algorithm flowchart and the description of each procedure is detailed below (Fig 3).

2.3.1. *Yield data adjustment by spatial arrangement*

This procedure is the same as described above (Section 2.2.4.) and is conducted on the yield data if the MultispeQ dataset was spatially adjusted by row and column coordinates. If not, the mean of the yield data for every genotype and source of variation is calculated and used for the following steps.

2.3.2. *Contrast of the MultispeQ ranks and yield data by confusion matrices*

The contrast of the MultispeQ ranks (2.2.3.) with the final yield is done for every sampling event and time separately. The steps to conduct this procedure are as follows: i) the yield data is sorted in descending order. ii) in both sorted yield data and MultispeQ cumulated trait score data frames (generated from 2.2.3.) the genotypes are clustered in groups by deciles. The first decile corresponds to the group of genotypes with the lowest values of both yield and MultispeQ score. The sorting increases up to the decile 10 for the group of genotypes with the both greatest values, accordingly. iii) for every single genotype a comparison of clusters for both variables (yield and MultispeQ trait-score) is done by a logical evaluation. TRUE if both deciles match or whether the difference between them is not bigger than ± 1 (i.e. cumulated trait score decile = 3 and yield decile =2) and FALSE if otherwise. iv) a series of interactive confusion matrices are generated where the targets (columns) correspond to the yield ranking clusters and predictions (rows) which correspond to the MultispeQ score ranking (Fig 12).

Based on the above-mentioned parameters, every confusion matrix classifies the genotypes into four different categories: a) Predicted (the matrix diagonal). If the behavior of MultispeQ trait-score is similar to the evaluated yield component (i.e. coordinates 4, 5 in the Fig. 12). b) False Positive. A group of genotypes with low yield but high MultispeQ trait-related score. c) False Negative. A group of genotypes with high yield but low MultispeQ trait-related score. d) Low prediction. The remaining group of genotypes that do not coincide in any of the previous described categories. For breeders, this last group will need some additional attention and depend on whether positive or negative selection scheme is applied in the breeding program. From the plant physiology view, groups with low prediction and “false” groups show some signs of photosynthetic acclimation in the particular environment. However, these groups earn detailed attention in any case as some of the functional traits can be of high importance (leaf trichomes, low SPAD, early/late maturity etc.) for stress resistance. The yield correlation to MultispeQ value ranking is better when the number of genotypes in the matrix diagonal is higher. The genotype names are interactively

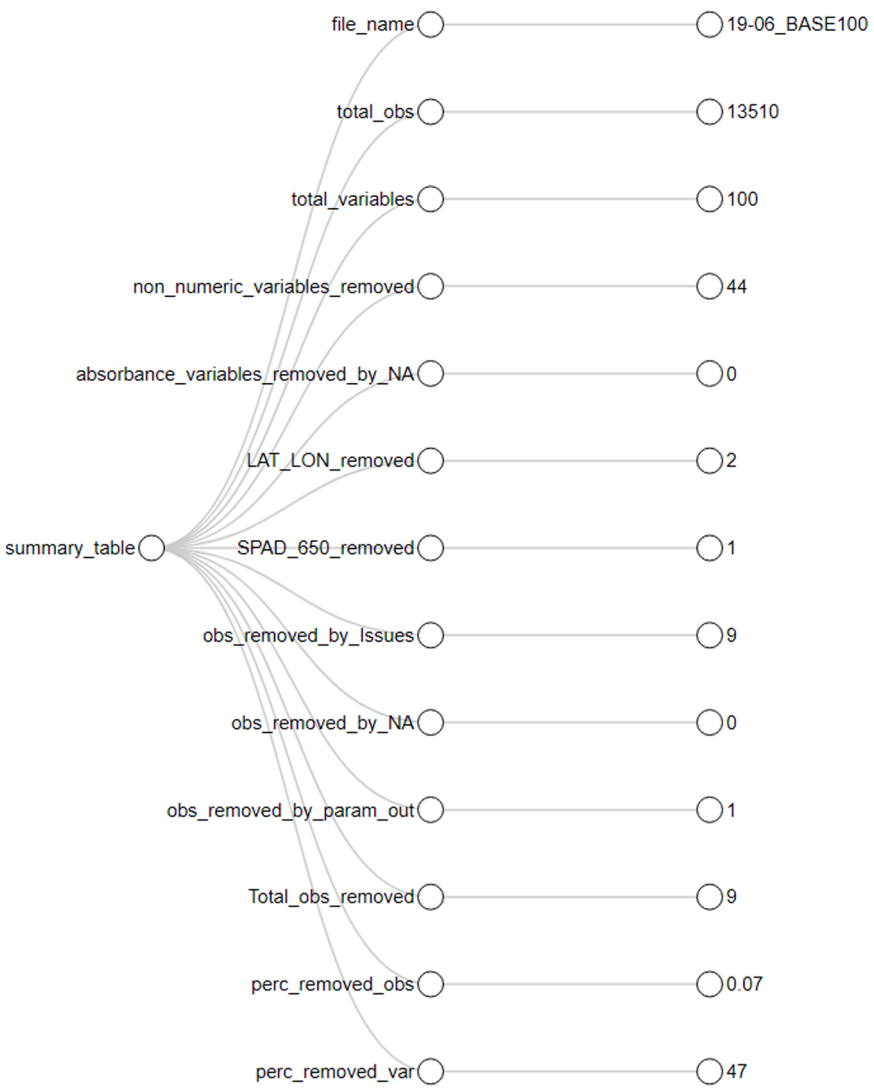

Fig 8. Interactive collapsible tree with the summary of the tidying process of the raw dataset.

ranks	list [11]	List of length 11
means	list [1768 x 55] (S3: tbl_df, tbl, da	A tibble with 1768 rows and 55 columns
std	list [1768 x 55] (S3: tbl_df, tbl, da	A tibble with 1768 rows and 55 columns
var_coef	list [1768 x 55] (S3: tbl_df, tbl, da	A tibble with 1768 rows and 55 columns
medians	list [1768 x 55] (S3: tbl_df, tbl, da	A tibble with 1768 rows and 55 columns
BLUP_df	list [1768 x 29] (S3: data.frame)	A data.frame with 1768 rows and 29 columns
rank_table	list [1768 x 28] (S3: tbl_df, tbl, da	A tibble with 1768 rows and 28 columns
rank_plots	list [4]	List of length 4
permutes	list [8] (S3: collapsibleTree, htmlw	List of length 8
Sources_of_variation	character [5]	'DAS' 'date' 'time' 'Genotype' 'Treatment'
Genotype	character [1]	'Genotype'
SPATS_variables	character [2]	'row' 'col'

Fig 9. Return of the MSPQ_ranks function for the "19-06 BASE100" dataset.

displayed when the mouse cursor steps by a pixel.

2.3.3. *Metadata evaluation*

This procedure is conducted if the argument metadata is a data frame with relevant information about the genotypes passport or grouping traits such as accessions; abiotic traits; among others. The input must be a data frame in a wide format with one or more traits related to each evaluated genotype. The steps to conduct this procedure are as follows:

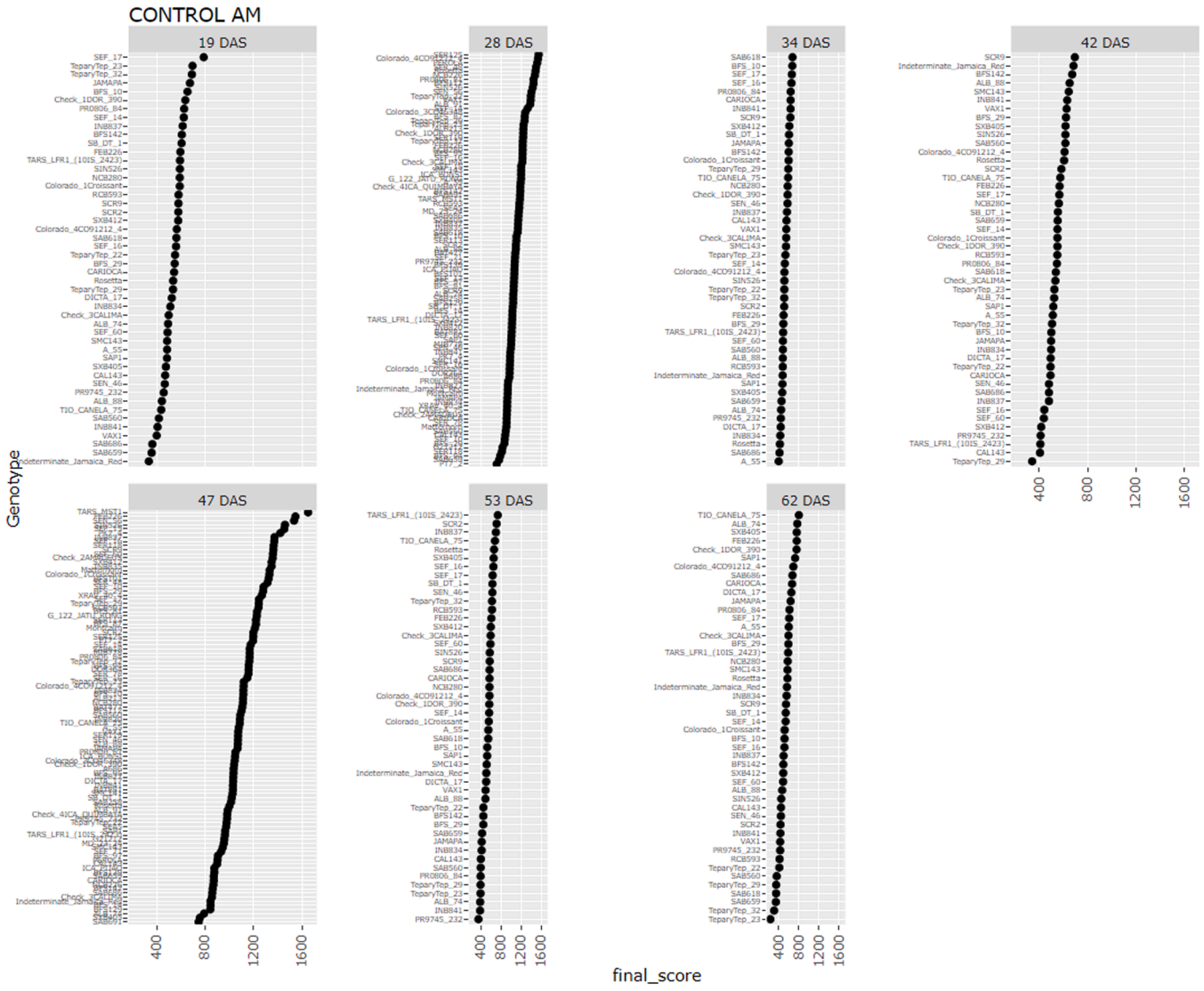

Fig 10. The sample of ranking plot of the evaluated genotypes for the morning measurements and irrigated (CONTROL) treatment. The abscissa corresponds to the cumulated trait score computed after the all-variables ranking; the ordinate the genotypes listed from the lowest to the highest score.

i) The genotypes and their related traits are joined to the previous contrast process (2.3.2.). ii) A sum of each individual group of Predicted, Low Prediction, False Positive and False Negative results is done for every sampling date. iii) Bar plots are generated from these contingency tables where the abscissa corresponds to the sampling dates and the ordinate to the frequency (number of genotypes) of each prediction category for every trait (Fig 13).

2.3.4. The return

The function target_trait_comparison returns a list with the objects as follows: i) A list with the series of interactive confusion matrices. ii) A list of character strings with the number of predicted genotypes. iii) A list with the contingency tables whether metadata was called. iv) A list with the bar plots for the metadata. v) A data frame with the spatial adjusted BLUP's whether spats = TRUE in the function MSPQ_ranks.

3. Illustrative examples

To calibrate the platform, we used the MultispeQ dataset ID 7844 [dataset] [11]. The dataset is open access and can be obtained upon user sign in. The experiment was conducted in 2019 in the Alliance Bioversity International – CIAT at Palmira, Colombia campus. A population of 100 genotypes of different *Phaseolus* species including 74 common beans (*P. vulgaris*), 4 tepary beans (*P. acutifolius*) and 22 interspecific accessions was evaluated in control and terminal drought trials (irrigation interrupted at flowering time, 26 days after sowing). Each treatment contained three spatial repetitions. The experimental unit corresponds to a plot with six rows per genotype, each of 5.4 m under both treatments. The whole population was therefore 600 plots. The MultispeQ evaluations were conducted on 13 different dates during the whole crop cycle, measuring the youngest but fully expanded and healthy leaf of 6 different plants selected randomly from the central plot rows avoiding borders. The project questions or sources of variation for this experiment were *Genotype* (as integers from 1 to 300 derived from 100 genotypes and 3 spatial repetitions) and *Treatment* (control and drought). The dataset pre-processing include: i) to discard 1 date where the measurements could not be completed due to rain. ii) to assign the genotype names to the *Genotype* column. iii) to include the *row* and *column* variables with the spatial coordinates of every plot.

3.1. Applying the MSPQ_tidy function

The function MSPQ_tidy was applied with the following arguments. i) df = df. The raw dataset already pre-processed. ii) genotype =

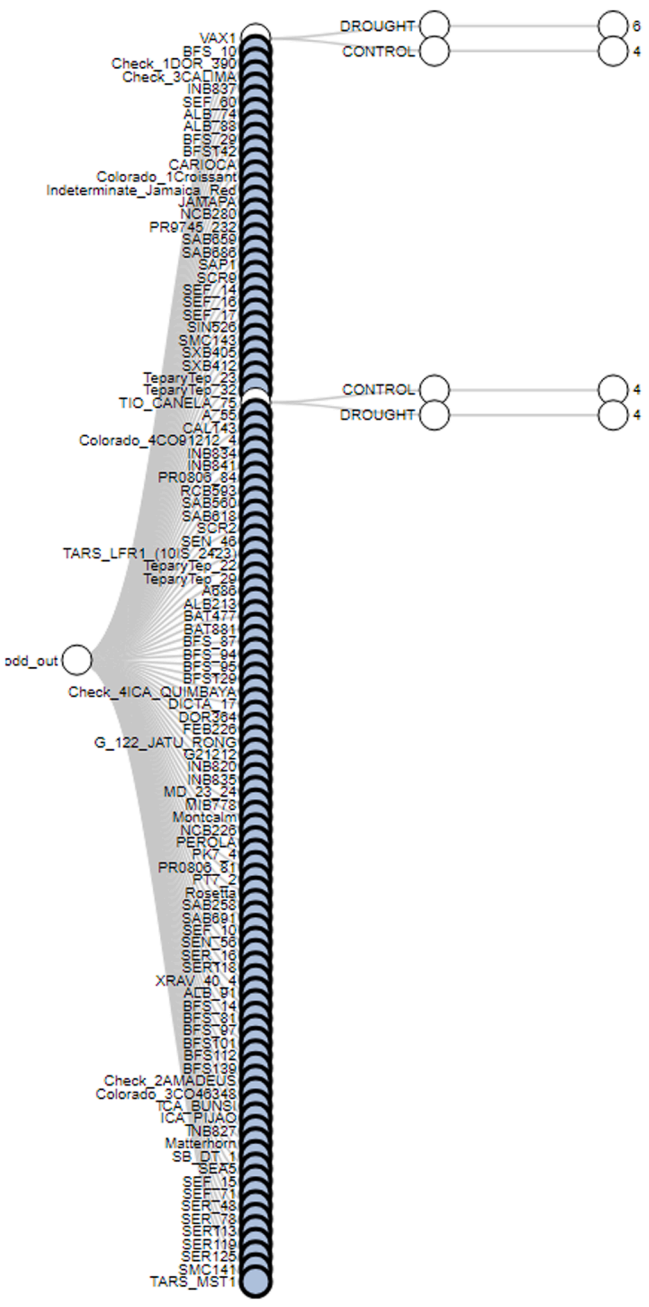

Fig 11. Interactive collapsible tree with the Phi Index permutation analysis results.

Table 3
Sample of the target.trait.file data frame for the "19-06 BASE100" experiment. *Final yield in kg ha^{-1}.

Genotype	Treatment	col	row	YDHA*
SEF_71	CONTROL	1	1	3515.78706
SER_16	CONTROL	1	2	3352.98789
BFS112	CONTROL	1	3	2923.43459
SER113	CONTROL	1	4	3164.28779
SEF_14	CONTROL	1	5	2432.53392
FEB226	CONTROL	1	6	2447.65431

"Genotype". A character SL1 with the name of the *Genotype* column. iii) time.diff = TRUE. For every single sampling date, two measurements were done as in the morning and the afternoon. iv) data_name = "19-06_BASE100". A character SL1 with the name of the experiment (optional argument). v) plotIm = TRUE. The return will include the distribution plots of the imputed values (originally NA). Therefore, the code chunk for running this function is:

```
tidy_data <- MSPQ_tidy(df = df, genotype = "Geno-
type", time.dif = TRUE, data_name = "19
-06_BASE100", plotIm = TRUE)
```

During the function execution, the following messages were printed in the console:

```
Replacing null for NA
Calculating Phi Index
formatting dates and creating the time (morning/af-
ternoon) variable
Discarding rows with issues
Applying control structures
Data Imputation
|=====================================-
==============================| 100%
Finding and removing rows with NA's
Removing PAR, Phi2 and Relative.Chlorophyll outliers
Making final df
Making summary table
Done!!!
```

A list with nine elements that were described above is the output of this function (Fig 4).

A total of nine observations were removed due to measurement issues. By calling the object removed_freq the frequency (cases) of the discarded observations by date and sources of variation is displayed to control if distribution is purely accidental (environment-driven) or genotype-specific (genotype-driven) (Fig 5).

The object imputed_variables is also an interactive collapsible tree that displays the variables and the number of imputed values (Fig 6). In our case, the largest amount of data to impute corresponds to the variable *absorbance_605* (5425 values) and the variables with the least values to impute were *kP700, P700_DIRK_ampl, tP700*, and *v_initial_P700* (4 values per variable).

The distribution of the imputed variables can be displayed by calling the object imputed_plots. Each element of this sub-list is a boxplot by dates-of-measurement of the variable and the red dots are the imputed values (Fig 7).

Finally, by calling the object summary, a new collapsible tree is displayed with a summary of the different processes conducted on the raw dataset (Fig. 8). The raw dataset dimensions were 13,510 observations (rows) and 100 variables (columns). The 0.07% of the observations were removed (9 values) due to issues. Likewise, the 47% of the variables (47 columns) were removed as 44 of them being non-numerical, *SPAD_650* as being identical to *Relative.Chlorophyll* and both *Latitude* and *Longitude* as they contained at least one or more NA and these are not imputed. The details of dataset cleaning are explained in 2.1.

3.2. Applying the MSPQ_ranks function

The function MSPQ_ranks was applied with the following arguments. i) out = tidy_data. The object returned previously by MSPQ_tidy function. ii) perIter = 100. iii) PerSeed = 123. As default arguments. iv) spats = TRUE as the dataset includes the spatial coordinates for conducting spatial analysis. v) row = "row". A character SL1 with the name of the row column in the dataset. vi) column = "col". A character SL1 with the name of the column variable in the dataset. vii) pl.date = TRUE as the planting dates will be provided. Therefore, the code chunk for running this function is:

```
ranks <- MSPQ_ranks(out = tidy_data, spats = TRUE,
row = "row", column = "col", pl.date = TRUE)
```

Since pl.date = TRUE, a series of questions appear in the console for providing the sowing date (or dates in case the experiment was sown in a wider range of time). In this case, both treatments called "CONTROL" for

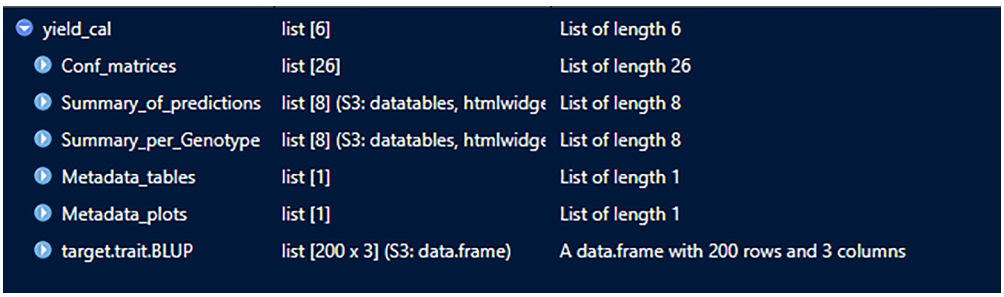

Fig 12. Return of the target_trait_comparison function for the "19-06 BASE100" experiment.

Fig 13. Interactive confusion matrix of a given sampling date for the "19-06 BASE100" experiment.

Table 4
Counts distribution of the four predicted categories for the CONTROL 28 DAS AM confusion matrix.

Conf_matrix	Variable	Count
CONTROL 28DAS AM	Predicted	33
CONTROL 28DAS AM	Low Prediction	51
CONTROL 28DAS AM	False Positive	7
CONTROL 28DAS AM	False Negative	9

irrigation and "SEQUIA" for drought were sown in two different dates. Therefore, the dates provided in the console are as follows (the values in italic are the input answers):

```
Are there multiple planting dates? Y/N: Y
Which of the 'Treatment' 'Genotype' was sown in multiple dates: Treatment
Please type the planting date for CONTROL in format mm/dd/yyyy: 07/11/2019
The planting date provided for CONTROL is: 2019-07-11
Please type the planting date for DROUGHT in format mm/dd/yyyy: 07/17/2019
The planting date provided for DROUGHT is: 2019-07-17
```

Subsequently, the execution of the further processes is displayed in console by the following messages:

```
Descriptive tables
Phi Index permutational analysis
|=======================================-================================| 100%
Adjusting variables with spatial components
|=======================================-================================| 100%
Ranking
...
Making return
Done!!!
```

A list with eleven elements that were described above is the output of this function (Fig 9).

The object Sources_of_variation includes the element days after sowing (DAS calculated using the provided dates). Likewise, the object BLUP_df is a data frame with the estimated BLUP's from the three spatial repetitions of each treatment. The ranks were calculated from this data set instead of the original data.

For the morning measurements of the "CONTROL" treatment, the behavior of the genotypes for the different phenological stages varies as well as the cumulated trait score (Fig 10). The abscissa corresponds to the cumulated trait score computed after the all-variables ranking; the

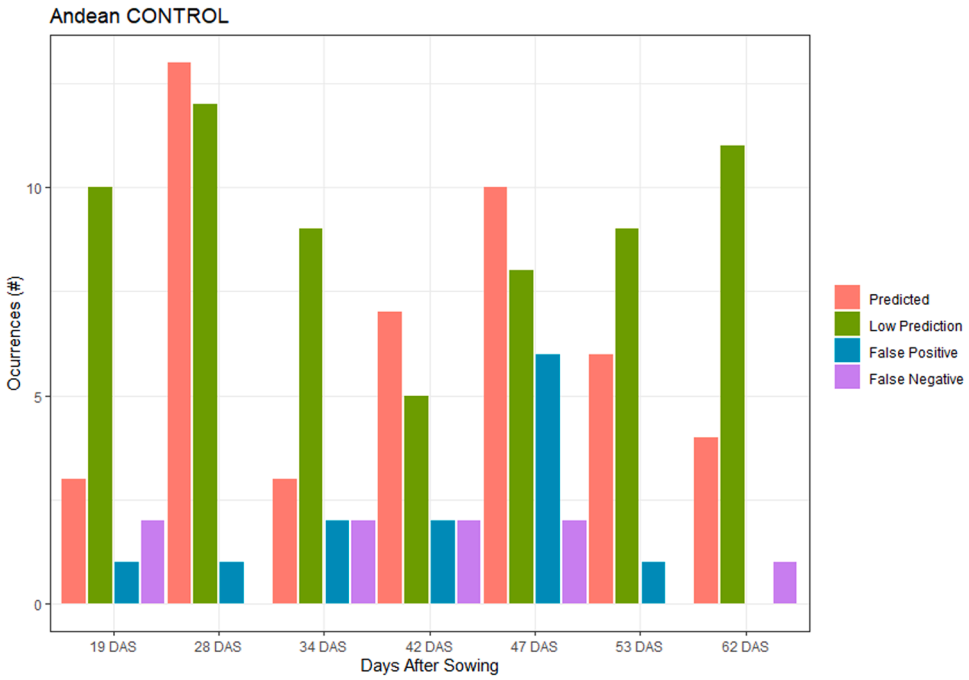

Fig 14. Distribution of four prediction categories for Andean genepool in irrigation treatment for the "19-06 BASE100" experiment.

ordinate the genotypes listed from the lowest to the highest score. The order of the genotypes will always be sorted in that way. These plots are stored into the rank_plots object from the returned list.

The object permutes is an interactive collapsible tree where the results of the Phi Index permutational analysis can be observed. The Fig. 11 shows the results for two randomly selected genotypes.

For example, for the cultivar VAX1 (the upper part of the tree, Fig. 11), in drought treatment, the Phi Index was significantly higher in the morning than in the afternoon in 6 different dates, meaning that the intercepted PAR light was mostly used in net photosynthesis in the morning than in the afternoon. Whereas in the irrigation treatment (control) index was different in four dates of measurements. Another example, for TIO CANELA 75, in both treatments, the Phi Index was higher four times in the morning hours than in the afternoon measurements. Please note, the detailed analysis of individual genotypes, their environmental, physiological and other performance is out of the scope of this study.

3.3. *Applying the target_trait_comparison function*

The function target_trait_comparison was applied with the following arguments. i) ranks = ranks. The object returned by the previous function. ii) target.trait.file = target.trait.file. A data frame with the yield information for every single plot (Genotype). Since the SpATS procedure was previously conducted, this data frame must contain the same spatial coordinates. The Table 3 is a sample of its first six observations. iii) target.trait.name = "YDHA". A character SL1 with the name of the yield variable to contrast with the previous ranks. In this case, the variable is yield in kg ha^{-1}. iv) metadata = metadata. A data frame with the corresponding genepool of each genotype. In this case, Andean gene pool with 13 genotypes, Mesoamerican gene pool with 61 genotypes, Interspecific lines with 22 genotypes and Tepary lines with four genotypes. Therefore, the code chunk for running this function is:

```
Yield_cal <- target_trait_comparison(ranks = ranks, target.trait.file = target.trait.file, target.trait.name = "YDHA", metadata = metadata)
```

During the function execution, the following messages were printed in the console:

```
Adjusting yield with spatial components
...
Plotting confusion matrices
...
The genotypes metadata to analyze is: Genepool
Making return
Done!!!
```

A list with six elements that were described above is the return of this function (Fig 12).

The object Conf_matrices is a list with 26 interactive plots (in this case), where each one corresponds to a confusion matrix for a day of sampling (DAS) as well as time of the day and treatment (Fig 13). For better orientation, the object Summary_of_predictions counts the number of the four prediction categories (Predicted, Low Prediction, False Positive, False Negative) in an interactive table, summarizing the results of the confusion matrices. This table can be sorted by confusion matrix name, prediction category or counts (Table 4).

The False Negative genotypes might surely be included into a positive breeder selection since they are clustered in the highest yield group although their MultispeQ performance is low. The opposite is true for the False Positive genotypes, which are low yielding anyway. Unfortunately, the scope of this study does not allow to discuss in detail these cases, even though they are extremely interesting from crop physiology point of view. However, probably also breeders can learn something from these cases, as yield quality is not included here and can contain/explain the important hidden part of the story. More detailed genotype x environment x temporal dynamics (e.g. photosynthetic traits) need to be conducted to verify the value of the "False" as well as "Low Prediction" genotypes.

For the whole experiment (all dates, treatments and evaluated genotypes), the potential of yield prediction by the MultispeQ (diagonal of the confusion matrices) was approximately 30% (30 up to 100 genotypes predicted well). About 50% of the population with a Low prediction and a total of 85% of the genotypes including the False Negatives was predicted using extended datasets. Likewise, we can say that the MultispeQ measurements done in the morning shown a better behavior in comparison to the final yield in both treatments (irrigated and drought stress). The 30% is actually an excellent result considering

many variables (physiological, biochemical, morphological, climatic, pedologic, agronomic), diverse portfolio of accessions (different genepools using different strategies and traits combinations), however also usually non-significant relationship of photosynthetic performance with yield performance together with very low heritability of yield caused by strong environmental effect.

The object Metadata_plots is a list that contains the barplots for every genepool (Andean, Mesoamerican, Interspecific and Tepary) with the distribution of predictions to the yield for both treatments. In the irrigated treatment (CONTROL), For example, the highest number of occurrences (measurements) that one or more of the Andean genotypes were predicted (present in the diagonal of confusion matrices) either in the morning and/or the afternoon is at 28 DAS which is close to the flowering stage and then 47 DAS which is the end of the linear seed filling growth phase (Fig 14).

Finally, the object trait.target.BLUP is a data frame with the predicted BLUPs for the yield trait (YDHA) by the SpATS procedure.

4. Impact

The MultispeQ Analyzer, called RankspeQ, is as an R package with three individual functions to help crop breeders, physiologists, agronomists, crop modelers and others to understand the dynamic of photosynthesis-related leaf traits and explore/understand it in relationship to the final seed yield or any other trait of interest. Rapidity and high precision in a genotype-performance based selection anchored on verified and well analyzed quantitative data from a set of traits evaluated under different experimental conditions is much needed. RankspeQ considers a group of traits previously identified as highly correlated to yield (kg ha^{-1} or g $plant^{-1}$) and uses them to compute and rank the behavior of individual genotypes on the background of the entire population (experiment). Therefore, a genotype selection before the harvest can be done by identifying the extremes with high and low scores (across more phenological stages). A phenological stage, able to show the highest predictability power and thus worth to reduce measurements in other stages still needs to be identified and validated across different environments. However, a trait calibration process (or calibration experiment) based on different statistical techniques is necessary and highly recommended to apply in order to use the RankspeQ in a new crop, especially if yield data from the region is not available yet. If needed, this (re)calibration ranking could be done every time after each measurement, or can be (re)run when an important breeding decision is required or the effects of some acute stress (e.g. un expected heat wave) are part of the scientific hypotheses.

Different crop performance during the crop cycle (diurnal or phenology-related responses) is analyzed and if the yield data is obtained, they can be contrasted to the MultispeQ rank-scores. Therefore, the user is encouraged to identify the individual dates (= phenological stages) where the yield prediction of individual genotypes is highly correlated to a selected list of measured traits and decrease the frequency of measurements in the upcoming trials. The grouping of genotypes can be actually based on any attribute specified earlier in the database. The attributes can be different groups (in the case of common bean Andean, Mesoamerican, interspecific etc.), genepools/species (secondary and tertiary such as *P. montanus, P. acutifolius* etc.), or resistance/character such as drought/heat and pathogen resistant accessions, among others. Other specific traits can be evaluated if genotypic data are available from previous experiments (seed mineral content, effect of soil mineral deficiency etc.).

RankspeQ will be updated according to the users' feedback and issues identification. Changes can include trait-for-ranking selection based on automatic PCA, time series approach or trait heritability (h^2) to support either negative or positive genotype selection process. Experiments conducted in other different crops can help to calibrate the package to be used in different species. Likewise, a shiny [12] interface will be implemented to make the tool easier to use, allowing to amplify the users to use it.

Since conducting measurements by MultispeQ on plant experiments under different conditions is a low cost, easy, reliable, and rapid, the implementation of our semi-automatic tool, which requires basic knowledge of R programming is a progressive step towards discovering the utility of a MultispeQ equipment, especially in the era of "black-box" high throughput phenotyping tools which require teams or specialized programs to understand them. MultispeQ is a unique device able to give us leaf-relevant data, including the immediate climatic conditions at the moment of a measurement.

5. Conclusions

RankspeQ is developed for helping the scientific community, physiologists, breeders, and agronomists to support their decision-making processes on reliable quantitative/qualitative data-driven selection. Based on a middle-throughput phenotyping tool, MultispeQ, data are easily collected by the device (around 100–300 accessions can be phenotype in one day). This software ranks genotypes and contrast results against selected/preferred yield component or another crop trait of interest. In other approached similar results (ranking) are generally available only at the end of the experiment, so if breeder is using "negative selection" the ranking can be a useful method. Here however, a preliminary dataset (first experiment) can be taken as an important data source and the most useful traits (highly regressed to the trait of interest) in the particular environment can be easily identified and used further for "environment-specific ranking." This serves to check the population behavior under particular environmental conditions and should offer/target reliable data/traits available for climate change-oriented modelers or other specific tasks.

By users' feedback and experts' criteria (we kindly ask you for your feedback), the future version(s) will include new functions and algorithms to discover the maximum potential that MultispeQ can provide as well as an easier and rapid way to process the datasets. Likewise, we expect to develop photosynthesis indices from different parameters measured by MultispeQ in order to support accurate parental selection for different breeding programs schemes. The possible use of drones/satellites in crop performance estimation should/could be then corrected by ground-truthing using MultispeQ in the near future. Modelers are encouraged to adjust their models to accept MultispeQ outputs, as they are extremely valuable (climate data + tissue data + plant data) especially if the future models will be driven by physiological/phenomics hypotheses related to "what a plant really sense and how we can use it to understand/predict" more than using complicated `black box' models with plethora of inputs.

CRediT authorship contribution statement

Jonatan Soto: Software, Data curation, Writing – original draft. **Johan Aparicio:** Software, Writing – review & editing. **Aquiles Darghan:** Software, Writing – review & editing. **Milan Oldřich Urban:** Conceptualization, Supervision, Data curation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Authors are deeply thankful to Dr. Steve Beebe for his support to develop and implement the package. Steve's questions focused on the breeder's aspect were critically important for improvement of our ideas. UMO is thankful to GIZ and PIAF (Germany) for their constant support. SJ is thankful to colleagues Diego Conejo and Sergio Cruz for their support on dataset cleaning and visualization.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.softx.2023.101544.

References

[1] Walter A, Studer B, Kölliker R. Advanced phenotyping offers opportunities for improved breeding of forage and turf species. Ann Bot 2012;110:1271–9.
[2] Schmidhalter U, Alqudah AM, Mayer JE, Bustos-Korts D, van Eeuwijk FA, Boer MP, et al. Combining crop growth modeling and statistical genetic modeling to evaluate phenotyping strategies. Front Plant Sci 2019;10. Available from, www.frontiersin.org.
[3] Kuhlgert S, Austic G, Zegarac R, Osei-Bonsu I, Hoh D, Chilvers MI, et al. MultispeQ Beta: a tool for large-scale plant phenotyping connected to the open photosynQ network. R Soc Open Sci 2016;3(10).
[4] Beebe S. Common bean breeding in the tropics. Plant Breed Rev 2012;36:357–426.
[5] Dramadri IO, Nkalubo ST, Kramer DM, Kelly JD. Genome-wide association analysis of drought adaptive traits in common bean. Crop Sci 2021 [Internet]n/a(n/a). Available from, https://acsess.onlinelibrary.wiley.com/doi/abs/10.1002/csc2.20484.
[6] Zhu W, Sun Z, Yang T, Li J, Peng J, Zhu K, et al. Estimating leaf chlorophyll content of crops via optimal unmanned aerial vehicle hyperspectral data at multi-scales. Comput Electron Agric 2020;178.
[7] Yan L, Li P, Zhao X, Ji R, Zhao L. Physiological and metabolic responses of maize (Zea mays) plants to Fe3O4 nanoparticles. Sci Total Environ 2020;718.
[8] Greenacre M. Compositional data analysis. Annu Rev Stat Appl 2021;8(1):271–99. https://doi.org/10.1146/annurev-statistics-042720-124436.
[9] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011;45. Available from, http://www.jstatsoft.org/.
[10] Aparicio J.S., Ariza-Suarez D., Aparicio J., Raatz B. Web application for spatial modelling of field trials aplicación web para la modelación espacial de ensayos de campo. 2019. [Internet] Available from: https://www.researchgate.net/publication/335206864.
[11] Urban M.O. 19-06 BASE100 [Internet]. Palmira: PhotosynQ; 2019. Available from: https://photosynq.org/projects/19-06-base-100.
[12] Wickham H. Mastering Shiny - Build interactive apps, reports, and dashboards powered by R. ShinyStuff; 2021. p. 298.
[13] Persoskie A, Ferrer RA. A most odd ratio: interpreting and describing odds ratios. Am J Prev Med 2017;52(2):224–8. https://doi.org/10.1016/j.amepre.2016.07.030.