This section of the SCIPDb provides detailed methodology followed in data mining, curation, analysis, and final integration into the database. SCIPdb hosts a collection of available literature on various stress combinations known to exist in nature. We report the presence of 100 stress combinations involving different abiotic and biotic stressors. Due to certain limitations in data analysis and for maintaining uniformity, we have considered a few exceptions. For example, while deciding the stress combinations, we have considered the combinations where the individual stresses affect the plant growth independently. The studies involving CO_2 and herbicides were not considered for data extraction as they may not be considered stresses. The studies which clearly show the combined stress imposition either simultaneously or sequentially were considered for the data extraction. We excluded studies involving the forest tree species (field experiments) from data extraction due to the complex stress response involved. It is also complicated to compare the physiological response of trees to crop plants. Since we aimed to measure the impact of combined stress on plants in their natural state, we have excluded studies related to genetically modified plants while extracting data. However, most of the combined stress studies which were not considered for data extraction (reports, thesis, reviews, etc.) were listed in the ancillary articles category in the reference file. The details about the exceptions and exclusions are provided in the end. The detailed methodology followed, starting from literature mining to integration into the database, is described below.

Deciding the Stress combinations:

Based on the knowledge of the team and the literature in plant science research, we have divided the stress combinations into three categories, namely abiotic-abiotic (consisting of two or more abiotic stress), abiotic-biotic (wherein one stressor is abiotic, and the other is either pathogen or pest) and biotic -biotic (consisting of two or more pathogens/pests). We have not included stress combinations in the present state, including elevated CO_2, herbicides, and insecticides, since these are anthropogenic. We also provide a list of several stress combinations which can exist but are not reported in the literature. The reference of articles that indicate the possible occurrence of such stress combinations is provided. Also, we plan to include many combinations other than the ones presently listed. A list of all stress combinations that are to be covered in the database is listed here.

Literature mining

To retrieve all available articles under combined stress and to have >90% literature coverage, different kind of keywords have been used along with the different types of search engines. Bibliography from each article has also been searched to make sure that no articles are left out. For a list of keywords used to retrieve articles under each stress combination, refer “List of keywords used for literature mining” file.

The list of search engines used for literature mining was listed as below,

1. Google- https://www.google.com/
2. NCBI PubMed- https://www.ncbi.nlm.nih.gov/
3. Searchit- http://searchit.libraries.wsu.edu/primo_library/libweb/action/search.do?vid=WSU&dscnt=0&dstmp=1470197360669&fromLogin=true
4. Jstor- http://www.jstor.org/
5. Webofknowledge- http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=Q25r3nQ9RdXUq7DPaGm&preferencesSaved=
6. Google scholar- https://scholar.google.co.in/
7. Sci-hub- https://sci-hub.tw/
8. Krishikosh (for agriculture-related thesis)- http://krishikosh.egranth.ac.in/
9. Shodhganga (for non-agriculture related thesis)- http://shodhganga.inflibnet.ac.in/
10. Consortium of e-resource in agriculture (CERA)- http://cera.iari.res.in/index.php/en/
11. CABI- https://www.cabi.org/

Sorting of articles

After retrieving all the articles related to particular stress combination, articles were sorted as ‘main research articles’ and ‘ancillary articles’. This was done based on the type of articles (research, review, reports, etc.) and type of data (morpho-physiological and molecular data). Main research articles with morpho-physiological data were considered for data extraction, whereas reports, thesis, book chapters, abstracts, reviews, gene overexpression, and gene-silencing studies were listed under ancillary articles and not considered for data extraction. All these types of articles were integrated into the database under the ‘phenomics’ tab. Articles with transcriptome study were considered for integration into the database under the ‘transcriptome tab.’

Some intended exclusions:

1. Plant competition, in this database, referred to intraspecific competition and constituted studies related to the effect of plant density involved with other stresses on plants. Cases of inter-specific competitions have not been discussed under this category.

3. In most cases, we have not included tree species. However, in some places where the economically important tree species have been considered, we have included studies wherein the experiments involved young saplings.

Listing out parameters and their classification

From the main research articles, parameters studied in each article were listed out and classified into type A, B, and C parameters based on their significance in reflecting the net impact of stress.

- **Type A** includes growth (plant height, biomass, leaf area, leaf number, root length, shoot weight, root weight, etc.) and yield (seed weight, seed number, test weight, etc.), attributing parameters that directly reflect the impact of stress.
- **Type B** includes physiological (photosynthesis, stomatal conductance, transpiration, chlorophyll content, etc.) and pathogenesis (disease index, pathogen load, disease score, etc.) related parameters which indirectly reflect the impact of stress.
- **Type C** includes biochemical parameters such as proline content, MDA content, nutrient composition, ROS content, etc., which also explains the impact of stress but to a lesser extent compared to the other two classes of parameters.

For the complete list of parameters hosted in the database, refer “trait ontology” file.

Data extraction and depiction

Once parameters were listed out from each article, data values were extracted into the excel file. The values from the table were directly copied into the excel sheet, whereas values from graphs were extracted using the ‘GetData Graph Digitizer’ tool for better accuracy. Since data is heterogeneous and to make it uniform, it was normalized by subjecting to calculation using the formula mentioned below.

For few parameters, such as electrolyte leakage, pathogenesis-related parameters, etc., were not subjected to calculation as they imply different meanings after subjecting to calculation. Using both the calculated and not calculated values table was prepared, reflecting the net impact of stress and interaction between the two stresses at the plant interface. This table was used in preparing the ‘data page’ file for each article and was finally represented in the database in tabular form. For easy understanding, percent change values were shown along with arrows in red and green color. A red-colored downward arrow indicates the parameter is affected under stress, and the higher the positive value greater the damage to the parameter under stress. Green-colored upward arrow indicated parameters are not affected under stress conditions as compared to control.

Data analysis

Data presented in tabular form was analyzed by comparing the individual and combined stress values of each parameter. Suppose percent change values are greater in combined stress compared to both the individual stressed. In that case, the outcome of combined stress is depicted as ‘negative.’ If percent change values are less in combined stress compared to both the individual stress, then the outcome is depicted as “positive.’ Together with the tabular part, a brief inference was written for each article which was finally checked for grammar and plagiarism. Each of the data pages starts with a brief introduction of the stress combination,

with information about the number of studies available in different crops for that particular stress combination.

Data integration into the SCIPDb

The above data was finally integrated and depicted as an HTML page specific to each plant species. The frontend user interface was implemented using HTML5, CSS and PHP (version: 7.0.12). The back-end schema was designed using MySQL, an open-source relational database management system, and stored in MySQL tables (Version: 5.7.17). To provide an interactive interface and enhanced user experience, Bootstrap 4, JavaScript, and jQuery were used. For data page format, please visit the 'phenomics' section.

Visualizations

Tableau public desktop (Version 2020.4, https://public.tableau.com/en-us/s/) was used to visualize high dimensional phenomics data in the form of an interactive Treemap. For the creation of radial trees, Flourish studio was used (https://app.flourish.studio/). An interactive geographical map was generated using Google My Maps. (https://www.google.com/maps/about/mymaps/).