

# Final Report

---

relating to the Community Plant Variety Office (CPVO) Project Grant Agreement for an Action  
(Agreement No 7517891)

"INTEGRATION OF MOLECULAR DATA INTO DUS TESTING IN DURUM WHEAT: DEVELOPMENT OF A COMMON  
ONLINE MOLECULAR DATABASE AND A GENETIC DISTANCE CALCULATION TOOL"

Project Acronym: "DURDUStools"

## *Project Leader*

Austrian Agency for Health and Food Safety GmbH (AGES), Spargelfeldstraße 191, 1220 Wien

Contact person and project manager:

Alexandra Ribarits (+43 (0)5 0555-34914, alexandra.ribarits@ages.at)

## *Project Partners*

Centro di ricerca Difesa e Certificazione (CREA DC) – Partner 1

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC) – Partner 2

Nemzeti Élelmiszerlánc-biztonsági Hivatal (NÉBIH) – Partner 3

Euroseeds – third party

## *Contributors to this report*

Alexandra Ribarits (AGES), Svenja Bomers (AGES), Tadej Zerak (AGES), Oliver Alber (AGES), Jutta Taferner-Kriegl (AGES), Jakob Seereiter (AGES), Anna Pia Maria Giulini (CREA DC), Silvio Pezzetti (CREA DC), Antonio Escolano García (INIA-CSIC), Ana Lázaro Somoza (INIA-CSIC), Ferenc Somogyi (NÉBIH), Sándor Kőrösi (NÉBIH), Barbara Fürnweiger (AGES)

## *Project duration*

January 2021 – June 2023

## *Summary of the project plan*

To utilize the results achieved in DURDUS, the aim of the follow-up project DURDUSTools was to develop an easily accessible online genetic distance (GD) calculation tool that can be used by durum wheat DUS experts for the choice of comparators in DUS tests. In order to increase the quality of DUS decisions in the entrusted EOs, DNA profiles of all varieties from the durum wheat reference collection were linked to the GD calculation tool. A prototype application was developed in the first year of the project and critically tested by all participating EOs. In parallel, the GD threshold, as defined in the DURDUS project, was evaluated and refined. In the second year of the project, the EOs tested the implementation of the final version of the GD calculation tool for the DUS field trials planning. During the final stage of the project, a draft partnership agreement was defined to ensure the maintenance, access to and the use of the GD calculation tool (DurdusTools) beyond the project duration.

## Table of Contents

Milestones in the project.....	5
Summary.....	8
Security of DurdusTools.....	9
Genotyping.....	9
SNP marker selection .....	9
Curation of database .....	13
Genetic distance calculation .....	14
Technical functionalities of DurdusTools.....	14
Login .....	14
User .....	15
Variety management .....	15
Viewing variety information .....	15
Downloading variety data .....	16
Adding new varieties.....	16
Updating and uploading new variety data.....	17
Genetic distances .....	17
Information.....	18
Notifications.....	18
Warning sign.....	18
Information sign.....	19
Administrator features .....	20
Uploading new genotypic data.....	20
Database management.....	20
Notify users.....	20
Testing of the common database and genetic distance calculation tool .....	20
Task 1: Candidate varieties .....	20
Task 2: Historical data.....	21
The use of DurdusTools by DUS experts .....	21
Austria .....	22
Hungary .....	22
Spain .....	22
Italy .....	23

Conclusions, Benefits and limitations .....	23
Workflow.....	24
Definition of a harmonized workflow.....	24
Common workflow .....	24
Draft Partnership Agreement.....	26
Project management.....	27

## Milestones in the project

The planned duration of the project was 2 years (24 months). The scheduled project start was in January 2021, and the intended project closing was in January 2023. Adjustments in the timetable of DURDUSTools were made due to the extension of the project DURDUS, as approved by the CPVO. Further amendments were necessary in January 2023 to solve the issues observed with the set of markers selected, and in April 2023 to allow the partners to finalize their conclusions and define the conditions of the partnership agreement for the maintenance of the database after the project. The final duration of the project is thus 30 months. Additional changes in time and location were caused by the current COVID-19 pandemic. For a detailed overview about the steps and the timing in the DURDUSTools project, please refer to **Table 1**.

**Table 1: Implementation of milestones.**

Milestone	Current state of implementation	Scheduled date according to work plan	Actually completed	Comment
Preliminary requirement analysis and survey for partner EOs, Euroseeds and CPVO	100%	January – March 2021	April 2021	Adjusted due to the extension of DURDUS
Sampling of varieties not yet genotyped	100%	April 2021	June 2021	Adjusted due to the extension of DURDUS
Project meeting and workshop (2 days) defining technical and legal aspects for the molecular database and genetic distance calculation tool	100%	May 2021	04 May 2021	Amendment: According to the plan, the first meeting was to be held in Vienna, Austria. Due to the COVID-19 situation, the meeting was rearranged to a one-day online meeting.
Interim report (requirement analysis)	100%	April 2021	May 2021	Adjusted due to the extension of DURDUS
Set up of the common molecular database and genetic distance calculation tool prototype version	100%	April – July 2021	July 2021	Online meeting to present the tool prototype version on 29.07.2021 and subsequent distribution of the login data
Genetic analysis of varieties	100%	May – July 2021	July 2021	By SGS Institut Fresenius GmbH – TraitGenetics Section
Integration of new molecular data	100%	July 2021	July 2021	
Testing of the genetic distance calculation tool prototype version	100%	July – September 2021	October 2021	Report was provided by the EOs on 15.10.2021

Investigation of genetic distance threshold by using historical data	100%	July – September 2021	October 2021	Report was provided by the EOs on 15.10.2021 Amendment: Additional evaluation of genetic distance threshold as defined in DURDUS
Feedback of EOs on the tested prototype version	100%	September 2021	October 2021	Report was provided by the EOs on 15.10.2021
Adjustment of the database and the genetic distance calculation tool according to the feedbacks, addition of features to the final genetic distance calculation tool	100%	October – December 2021	November 2021 to date	Continuous improvement and adjustment according to feedback of EOs
Web-meeting Presentation of the final database and the genetic distance calculation tool	100%	January 2022	25 November 2021	
Genetic analysis of varieties	100%	-	April 2022	Additional task (Amendment to the original work plan) Genotyping of current candidate varieties and remaining reference varieties
Integration of new molecular data	100%	-	April 2022	Additional task (Amendment to the original work plan)
Final test of the common database and genetic distance calculation tool	100%	January 2022	January 2022	
Practical implementation at the EOs (spring sowing)	100%	February – March 2022	February – March 2022	Genetic distance calculation results were used for the planning of the DUS trials second year
Practical implementation at the EOs (autumn sowing)	100%	September – October 2022	September – October 2022	Genetic distance calculation results were used for the planning of the DUS trials second year
Final project meeting (2 days) Feedback on practical implementation; discussion and preparation of the draft Partnership Agreement	100%	November – December 2022 (to be defined)	14 – 15 September 2022	At NÉBIH, Siofok, Hungary Additional online meeting on 12 December 2022
Meeting with GEVES for presentation of DURDUStools	100%	-	12 December 2022	Additional task (Amendment to the original work plan)
Project meeting (online) for discussion of the slightly modified genetic distance calculation	100%	-	18 January 2023	Additional task (Amendment to the original work plan)
Online meeting of AGES and service provider to discuss recently discovered challenges concerning consistency of the genotypic data	100%	-	08 February 2023	Additional task (Amendment to the original work plan)

Project meeting (online) for presentation of the new SNP marker selection by the service provider	100%	-	02 March 2023	Additional task (Amendment to the original work plan)
Project meeting (online) to discuss the new SNP marker selection presented by service provider	100%	-	03 April 2023	Additional task (Amendment to the original work plan)
Project meeting (online) for final conclusions on the final SNP marker selection	100%	-	19 April 2023	Additional task (Amendment to the original work plan)
Online meeting of AGES and service provider to discuss the possibility to incorporate control varieties	100%	-	24 April 2023	Additional task (Amendment to the original work plan)
Project meeting (online) for discussion of the first drafts of the partnership agreement	100%	-	27 April 2023	Additional task (Amendment to the original work plan)
Project meeting (online) for discussion on the control varieties and cost divisions	100%	-	09 May 2023	Additional task (Amendment to the original work plan)
Project meeting (online) for discussions on the draft partnership agreement and final report	100%	-	17 May 2023	Additional task (Amendment to the original work plan)
Submission of the draft final report including the draft Partnership Agreement	100%	January 2023	June 2023	Adjusted due to the extension of the project duration

## Summary

With the development of a common online molecular database and a genetic distance calculation tool (DurdusTools), molecular data can be integrated into DUS testing in durum wheat. Seeds of durum wheat varieties and candidate varieties were genotyped by using a commercial wheat array of an external service provider (SGS Institut Fresenius GmbH TraitGenetics Section). Each genotype is assigned a unique identifier number to anonymize the seeds for the service provider, and to assure the identification to the correct genotype in the tool. The generated genotypic data is uploaded in a molecular database with limited access. For the genetic distance calculation between each individual pair within the database, the tool selects specific SNP markers that were defined a priori based on quality parameters. Besides the genotypic data, a variety file exists that contains basic information about all varieties and candidate varieties in the molecular database. The information columns were developed together with the DUS experts (Year of genotyping, Denomination, Breeder's reference, Responsible EO, Status, Year of registration, Name of Breeder, Comment). The tool uses information from both collections, genetic distance calculation and information in the variety file to create the output. The downloadable output is an Excel file that shows information of the most similar varieties to the candidate variety of interest.

Access to the tool is only given to the five entrusted EOs for durum wheat with the purpose of DUS testing, and to the tool administrator. To this end, the EOs are provided with the information of the genetic distance of specific pairs and do not need to have access to the raw genotypic data. The downloading of the raw genotypic data is restricted only to the project coordinator to avoid unauthorized use. The confidentiality of the seeds provided to the service provider is guaranteed by anonymization. The originated raw genotypic data is stored online in the genetic distance calculation tool, which is placed on a special server using the coordinator's IT infrastructure and well-established security features. Furthermore, the variety names in the raw genotypic data are encrypted to increase the safety of the data.

The EOs use the information of the genetic distance together with the phenotypic assessment after the first year of DUS trials to select comparators to be grown in the field next to the candidate varieties under investigation. For this, the EO is responsible to send 20 seeds per candidate variety to the coordinator, encode the seeds according to the automatically assigned code and upload the information in the tool. After two years, the status of the candidate variety is updated. The Status can either be C (= candidate variety; genotype is still in the application process), V (= variety; genotype is registered/protected in the EU) or nV (= no registered/protected variety; candidate variety was not registered/protected as variety). The defined threshold in DURDUS (0.32, with a safety margin up to 0.40) is used as general guidance for distinctness between two individuals. However, the DUS experts concluded to retain



flexibility in the application of this threshold in their countries, always in combination with morphological distance. After testing the practical application of DurdusTools, the experts confirmed that the tool is user-friendly and helps to improve the quality of the DUS tests.

## Security of DurdusTools

During the project the following safety measures were followed:

- The genotypic data is kept secure and confidential to the highest possible extent.
- The genotypic data is stored on a safe server.
- Only the coordinator staff working in the project have access to it.
- The variety names in the genotypic data were encrypted to increase the safety of the data.
- No sharing of the genotypic data is foreseen.
- Access to the genetic distance calculation tool was only given to the four entrusted EOs involved in the project and the project coordinator.

These provisions will be continued after the end of the project.

## Genotyping

For collecting the genetic data, two Illumina Infinium SNP marker micro arrays were used, due to an add-on of 5.000 markers by the service provider. Concretely, a 20K SNP micro array (in 2018) and an updated 25K SNP micro array (in 2019, 2020, 2021, 2022) were used (Table 2). Altogether, 896 genotypes were genotyped in the course of the projects DURDUS und DURDUS tools.

Year	Number of genotypes	SNP micro array
2018	583	20K
2019	5	25K
2020	94	25K
2021	94	25K
2022	120	25K

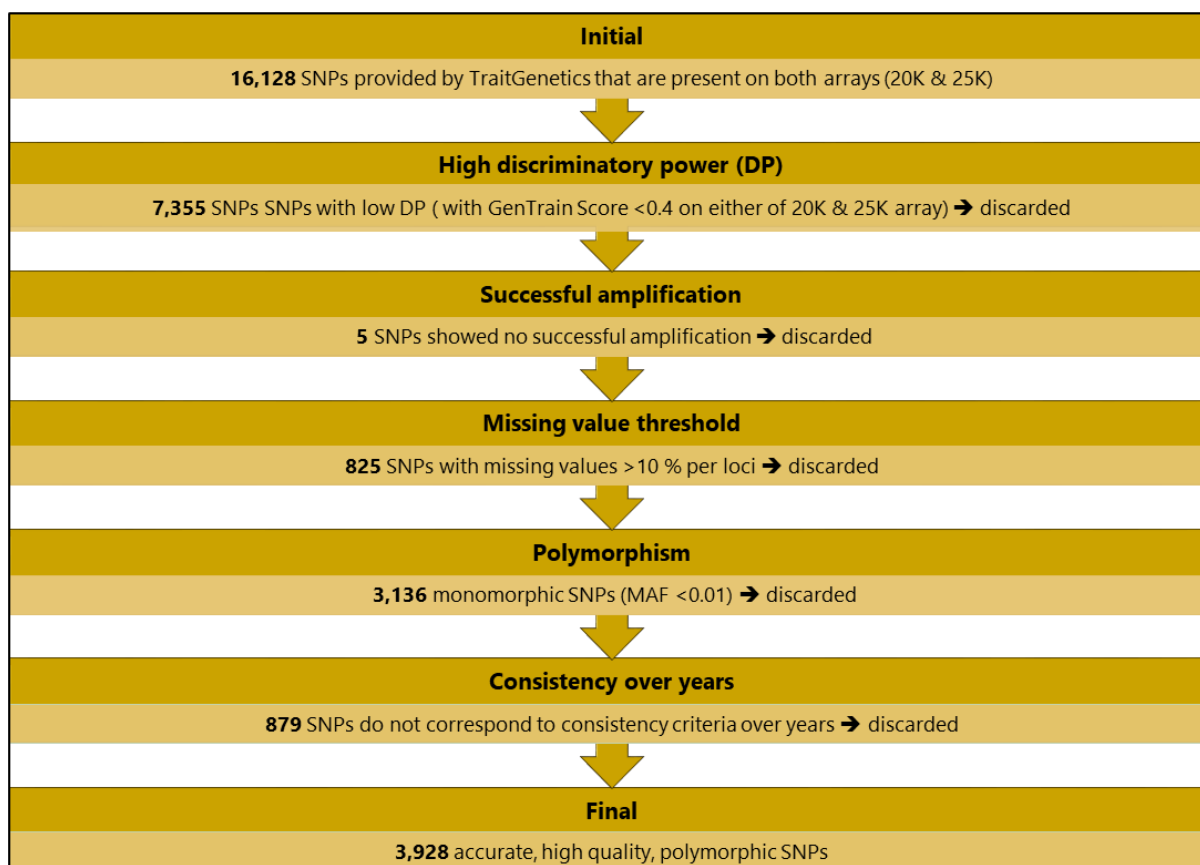
**Table 2: Overview of the different genotyping processes grouped by year.**

## SNP marker selection

Based on the available genotypic data, the SNP marker selection was adjusted three times during the DURDUS and DURDUS tools projects. After the first genotyping process in 2018, 8,804 high quality SNP markers were selected. In 2019, the number of selected high quality

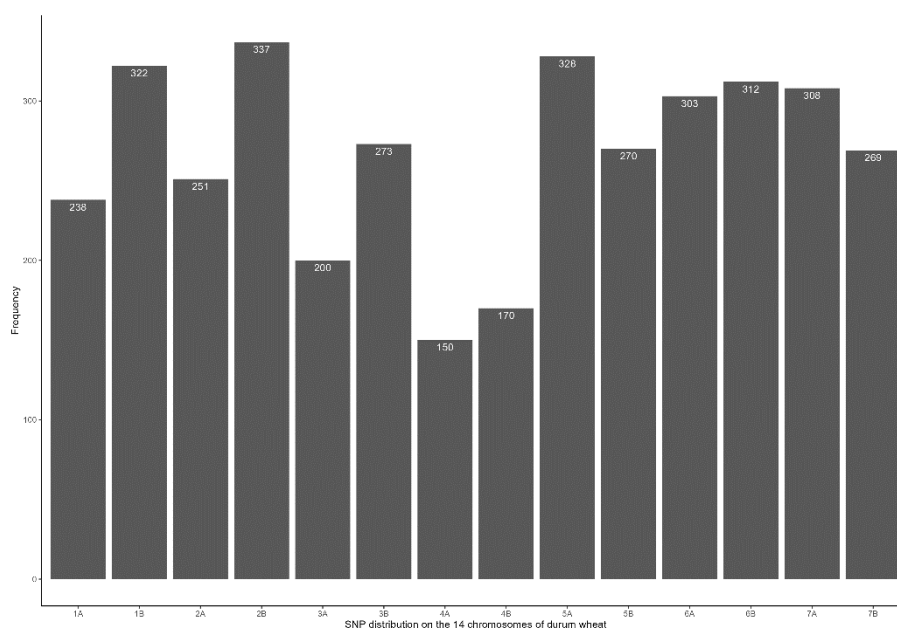
SNP markers was slightly reduced, from originally 8,804 selected SNP markers to 8,194 SNP markers that were common to both 20K and 25K arrays. After the genotyping process in 2020, it was decided to further increase the quality of the selected SNP markers by considering the value of the GenTrain Score – an Illumina quality parameter for reliable clustering and allele calling, which further reduced the number of SNP markers to 4,807. Towards the end of the year 2022, consortium members found discrepancies in the genetic distances of selected varieties with the same denomination that had been genotyped in both 2018 and 2022. Extensive testing allowed to exclude technical and calculation errors. Further investigations thus focussed on the stability of the selected SNP markers, and some SNP markers were found to show different genotyping results for varieties with the same denomination depending on the SNP micro array used. To explore potential solutions, the service provider was contacted. It was concluded that the requirement to obtain stable genotypic results over several years was a specific challenge of DUS testing, and particular attention should be given to ensure the consistency of data independent of the specific technology used to obtain them. The service provider agreed to examine in detail the genotyped collection of durum wheat varieties in DURDUS and DURDUSTools to select the most stable and reproducible SNP markers across the different years and micro arrays. The following quality criteria were used for this purpose: Minor Allele Frequency (MAF  $<0.03$ ), Frequency of heterozygous genotypes (AB  $<0.1$ ), Call Frequency (CF  $>0.9$ ) and GenTrainScore (GTS  $>0.2$ ). These selection criteria resulted in 6,295 high quality SNP markers that are supposed to be easy to score and to show stable results over the years.

The final DURDUS/DURDUSTools selection of SNP markers comprises the intersection of the 6,295 stable SNP markers as selected by the service provider with the 4,807 high-quality polymorphic SNP markers that had been in use during variety assessment and selection by the DUS experts. As a consequence, 879 SNP markers that were not considered stable and reproducible by the service provider were excluded from the previous SNP marker selection. The consortium thus agreed to follow the service provider's recommendations but also to consider the results of the DUS expert evaluation. The final SNP marker selection for the genetic distance calculation was thus reduced to 3,928 high quality, polymorphic, and stable SNP markers with high discriminatory power (Figure 1). The reduction of the number of SNP markers based on various quality parameters for the final data set is expected to ensure a reliable and sustainable characterization of the material.



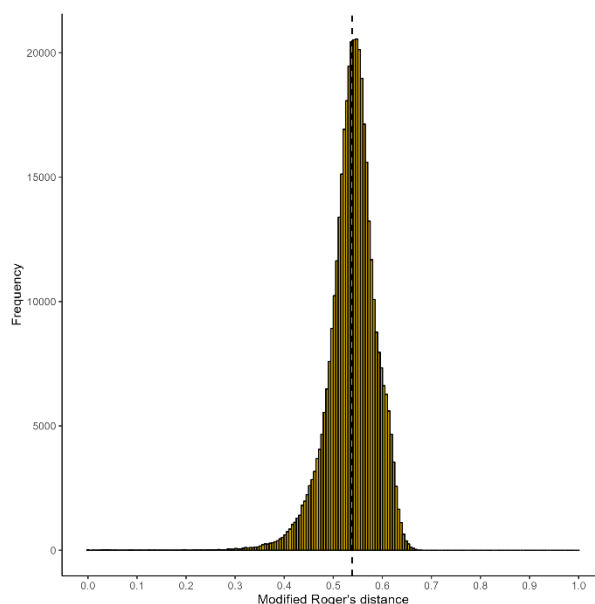
**Figure 1: Flowchart showing the SNP marker selection process from initial 16,128 SNP markers that are present on both arrays (20K & 25K) provided by SGS Institut Fresenius GmbH - TraitGenetics Section to 3,928 high quality, polymorphic and consistent SNP markers with high discriminatory power. Selection was performed based on genotyping results of the DURDUS/DURDUSTools collection of durum wheat genotypic data (high-quality data of 893 genotypes).**

The 3,928 SNP markers are evenly distributed across all 14 durum wheat chromosomes as shown in Figure 2.



**Figure 2: Frequency distribution of the 3,928 selected SNP markers on the 14 chromosomes of durum wheat provided by SGS Institut Fresenius GmbH - TraitGenetics Section. For 197 SNP markers no genetic position could be determined.**

The genetic distance within the DURDUSTools high-quality genotypic data collection of 893 durum wheat genotypes displayed a normal distribution with an average MRD among pairs of 0.54 (minimum = 0.00, maximum = 0.68; Figure 3).



**Figure 3: Frequency distribution of the modified Roger's distance (MRD) within the DURDUSTools collection of durum wheat genotypic data (893 genotypes) calculated by the use of the selected 3,928 SNP markers and the pairwise deletion method.**

The new SNP marker selection of 3,928 SNP markers slightly increased the genetic distance results by an average of 0.03 compared to the previously used 4,807 SNP markers, with maximum changes of +0.05 and -0.12, respectively. The genetic distance discrepancies detected in December 2022 between selected varieties with the same denomination genotyped in both 2018 and 2022 were reduced. We expect that this SNP marker selection provides stable results in the calculation of genetic distance over years.

### Curation of database

For further curation of the database and to support the aim to obtain stable genotypic results over years, the service provider created a specific cluster file for reliable and sustainable clustering and allele calling based on the genotyped collection of durum wheat varieties in DURDUS and DURDUStools. The specific cluster file will be used exclusively for all further consortium analyses using the 25K Wheat Illumina Infinium Micro-Array.

In addition, it was decided to include two defined control samples in each genetic analysis to ensure technical consistency and reproducibility of the data between years. These controls are carried out by the service provider. For this purpose, seeds of selected varieties (Sherekhan, Spineto) are sent to the service provider, who extracts the DNA and keeps it over the years to use it every time that the genotyping is carried out. Control sample genotype data is routinely checked for consistency in each analysis. In general, the reproducibility error as a parameter for the consistency of duplicate DNA samples as controls must be less than 0.1%. This parameter indicates the deviation between the replicates of a control in the database, and is checked based on all SNPs relevant for durum wheat on the micro array. In case of discrepancies that cannot be solved by the service provider and that are not due to technical reasons, the service provider informs the consortium that will initiate further investigations.

Furthermore, each EO submits seeds of two varieties for internal control of the entire genotyping process. These seeds must be collected each year and for every new genotyping process from the same reference sample stored at the respective EO. In accordance with the procedure for the candidate varieties, the information of the control varieties is also uploaded each year and the seeds are sent anonymously to the coordinator. After having received the results of the genotyping, each EO checks the genetic distances of the control varieties for plausibility. If the genetic distance is not plausible, the consortium is contacted. A warning appears indicating that there are duplicate denominations in the database. Thus, after having validated the identity of the sample based on the genetic distance, each EO must decide which genotype to keep as the reference sample.

## Genetic distance calculation

The genetic distance is calculated using the modified Roger's distance (MRD). The MRD calculation was automated by adapting the R script developed by the Beissinger lab (<https://www.uni-goettingen.de/en/603177.html>, accessed on June 05, 2023), using the sum of differences of allele counts for all markers. As the MRD calculation cannot handle missing values (non-available values, NAs), the method of pairwise deletion is used for the calculation of the genetic distance. As the number of SNP markers varies between each pairwise comparison, the final number of SNP markers that are used for the calculation of the genetic distance between two varieties are indicated in the output of genetic distance calculation (SNP markers used). Please see the DURDUS final report for further details.

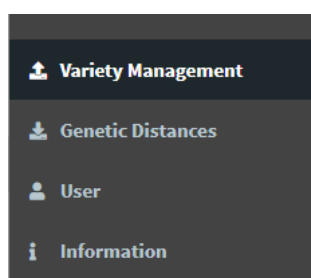
For the calculation of the genetic distances within the DURDUStools collection of 893 durum wheat genotypes an average amount of 3,864 SNP markers was used. The minimum amount of SNP markers used for the calculation of the genetic distance between two varieties was 2,944 SNP markers.

## Technical functionalities of DurdusTools

DurdusTools was developed to integrate the use of molecular data into DUS testing in durum wheat by the development of an easily accessible genetic distance calculation tool to be used by DUS experts. The focus was especially on ease of operation and efficiency through many automated steps. At the same time, the security of the genetic data was ensured.

### Login

The tool can be accessed via the web address <https://shiny.ages.at/DurdusTools/>. The access to the tool is password protected and access is given only to DUS experts of entrusted examination offices for durum wheat.



After a successful login, a toolbar is presented on the left side of the screen. It consists of the following panels designed to navigate through the tool: Variety Management, Genetic Distances, User and Information.

## User

In the User panel, information about the currently logged in user is presented: username, EO, and email. Here, the email and password associated with user currently logged in can be changed.

By default, each EO had access to the tool via two e-mail addresses chosen by the EO. More users can be implemented by the administrator. The tool administrator is contacted for access to the tool or for changing of the permission rights. Permission rights include the following options: (1) the download and upload (update), or (2) only the download of data.

## Variety management

In the variety management panel the user has the following three features:

- Viewing variety information
- Downloading variety data
- Adding new varieties
- Updating and uploading new variety data

### Viewing variety information

At the top of the page there is a table with all the variety information with the following columns:

Variety Management									
Select columns ▾	Delete selected rows	Show 10 ▾ entries	Search: <input type="text"/>						
Individual	Year of genotyping	Denomination	Breeder's reference / Synonym	Responsible EO	Status	Year of registration	Name of Breeder	Comment	Excluded from genetic distance calculation

Column	Description
Individual	ID that is used for the genotype as a code for the service provider.
Year of genotyping	Year in which the genotyping was performed.
Denomination	Denomination of the genotype (if Status is V).
Breeder's Reference / Synonym	Breeder's reference or any synonym used for the genotype.
Responsible EO	Entrusted EO that has sent the genotype to the coordinator. Possible values: AT, ES, HU, IT, FR.
Status	Current Status of the genotype. Three possible values: C = candidate variety; genotype is still in the application process. V = variety; genotype is registered/protected in the EU. nV = no registered/protected variety; candidate variety was not registered/protected as variety.
Year of registration	Year in that the genotype got the registration/protection (if Status is V).

Name of breeder	Name of the breeder from which the genotype originated.
Comment	Additional comment. The possible content will be defined.
Excluded from genetic distance calculation	Information why genotype is excluded from genetic distance calculation.

In addition, the columns ID, username, Inserted and Updated are hidden and can be selected by clicking on the "Select columns" button. The columns ID represents a unique identifier (number), which is used to distinguish varieties internally. The column username shows the user who last updated/uploaded the variety information. The columns Inserted and Updated represent date and time when the information of the genotype was inserted and last updated respectively.

The table can be filtered by using the search textbox on top right. Any value from any column can be typed in the search textbox and the table view will filter for respective rows. For example, specific values like Denomination or Breeder's reference can be searched by using the search field. When using this function, capital or small letters, special characters, specific letters like á or ó or spaces are ignored.

### Downloading variety data

The variety data can be downloaded for individual use.

Download variety data

Download variety list of selected rows
Download variety list of my EO
Download full variety list

Three buttons for downloading variety data are available: (1) Download variety list of selected rows, (2) Download variety list of my EO, and (3) Download full variety list.

- (1) Information from selected rows is downloaded. Specific rows can be selected from the table by clicking on them. The search feature can also be used for selecting rows.
- (2) Information for all varieties for which user EO is responsible is downloaded.
- (3) Information for all varieties is downloaded.

### Adding new varieties

Add new varieties

Add new varieties

The "Add new varieties" function is only available to users with the appropriate permission rights. This allows the EOs to upload information on new candidate varieties each year before genotyping and to obtain a corresponding ind\_number for each individual for anonymization of the seeds before submission to the coordinator. In this way, the genetic data is protected, but can be



assigned to the corresponding genotypes in the tool. The numbers are filled in ascending order and at the time of uploading, so that the samples added in the same year show consecutive numbers. The automatically assigned identification number is used by the EOs to label the seed packages to be submitted for genotyping.

## Updating and uploading new variety data

### Upload new variety data

Choose a changed varieties file to upload

Browse...

No file selected

The "Upload new variety data" function is only available to users with the appropriate permission rights. This allows the EOs to fulfill their responsibilities to update the Status and add information about existing varieties.

It is the responsibility of each EO to keep the variety list up-to-date. When the EO has received the results of the past DUS tests, the user changes the status of the candidate varieties to either variety (V) if it has been registered or to no variety (nV) if it has not been registered. The user also adds the name of the breeder and the year of registration. It is only possible to change the variety information if the user belongs to the responsible EO or if no responsible EO is specified (denoted by "-"). A user cannot delete varieties. Some columns (ID, Individual, Responsible EO, Excluded from genetic distance calculation, Username, Inserted, Updated) cannot be modified by the user as they are crucial for the functionality of the tool. In case of desired changes, the administrator needs to be contacted.

## Genetic distances

In the genetic distance panel, the user is able to download the results of the genetic distance calculation. The tool uses information from both collections, genetic distance calculation and information in the variety file to create the output.

Genetic distance filter

from

0

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

to

1

Individuals filter (names, separated with comma)

Status filter

V

Download genetic distances

Before the download, a simple list of filters is available: (1) Genetic distance filter, (2) Individuals filter and (3) Status filter.

- (1) In the genetic distance filter, minimum and maximum genetic distance between varieties that are to be included in the output can be set.

- (2) In the individual filter, the following options can be inserted in the text field for selecting the output: Individual, Denomination or Breeder's reference. Those values need to be separated with comma (","), and they can be mixed.
- (3) In the Status filter, the Status of the varieties displayed in the output can be selected. The value can either be C (=candidate variety) or V (=reference variety). Genotypes with the Status nV (=no registered/protected variety) cannot be chosen for genetic distance calculation as they are not of common knowledge.

The output is an Excel workbook with the result of the genetic distance calculation for each individual in a separate sheet. For example, if the individual filter is filled with "ind\_701,Don Pedro,05e2848", an Excel file with 3 sheets will be the output. It contains the genetic distances, the number of SNP markers used in the pairwise calculation and relevant information about the individual that has been used for the comparison. The individuals are sorted by their genetic distances, so that the most similar varieties are located at the top (Figure 4).

Individual	Distance	SNPs used	Year of genotyping	Denomination	Breeder's reference / Synonym	Responsible EO	Status	Year of registration	Name of Breeder	Comment
ind_189	0.20	4342	2018	Variety 1	abc	ES	V		Breeder 1	
ind_59	0.22	4793	2018	Variety 2		ES	V		Breeder 2	Deletion from EU plant variety database 10.07.2020
ind_177	0.22	4796	2018	Variety 3		ES	V		Breeder 3	Deletion from EU plant variety database 22.11.2018
ind_51	0.24	4777	2018	Variety 4		IT	V		Breeder 4	Deletion from EU plant variety database 15.03.2017
ind_284	0.25	4225	2018	Variety 5		FR	V	1999	Breeder 5	Deletion from EU plant variety database 07.05.2020
ind_366	0.25	4178	2018	Variety 6		IT	V		Breeder 6	
ind_417	0.26	4539	2018	Variety 7		ES	V		Breeder 7	Deletion from EU plant variety database 22.11.2018
ind_396	0.26	4797	2018	Variety 8		ES	V		Breeder 8	
ind_481	0.26	4707	2018	Variety 9	def	IT	V	2015	Breeder 9	
ind_161	0.27	4801	2018	Variety 10		ES	V		Breeder 10	Deletion from EU plant variety database 22.11.2018
ind_169	0.27	4800	2018	Variety 11		ES	V		Breeder 11	
ind_401	0.28	4795	2018	Variety 12		IT	V		Breeder 12	
ind_225	0.28	4786	2018	Variety 13		IT	V		Breeder 13	
ind_29	0.28	4801	2018	Variety 14		ES	V		Breeder 14	
ind_294	0.28	4801	2018	Variety 15	ghi	IT	V	2011	Breeder 15	

**Figure 4: Example output of genetic distance calculation**

If no filters are set, the output contains information for all available pairwise comparisons in one sheet.

## Information

In the information panel, instructions for using the tool, contact information in case of technical or data difficulties, as well as background information is presented and can be downloaded.

## Notifications

## Warning sign



On the upper right corner of the app there is a little triangle in case there is a warning that needs a response.

Possible warnings are:

- (1) No information about new candidate varieties was uploaded in current calendar year.
- (2) Duplicated denominations are present between varieties that are included in genetic distance calculation. The denominations are clickable for presenting in the variety table. The user will see duplicated denominations for varieties the EO is responsible for. To solve the issue and to make the notification disappear, the responsible EO that is displayed for the other individual is contacted. The EOs communicate and decide bilaterally which individual to keep as reference variety. Both EOs are able to change the column "Excluded for genetic distance calculation" for this respective variety. The notification will disappear when the responsible EO adds something (e.g. "doubled genotype") to the column "Excluded for genetic distance calculation" for the individual that is excluded.
- (3) Similar denominations are present between varieties that are included in genetic distance calculation. Two denominations are similar if they only differ in special characters, specific letters like á or ó, spaces and capital lettering, for example prospero and Próspero. To solve the issue and to make the notification disappear, the same procedure as described in (2) has to be applied.
- (4) Information on the candidate varieties needs to be updated. User will receive the notification when they have genotypes with the Status "C" that were not updated two years after they were inserted. To solve the issue and make the notification to disappear, the Status (V or nV) is updated and the Name of Breeder, Year or registration and Denomination is added to the respective genotype.

#### Information sign



On the upper right corner of the app there is a little information sign displaying an "i" in case new information is available.

The user receives an information if the Breeder's reference of one of his candidate varieties is similar to that of another variety from another EO. The Breeder's reference is clickable for

presenting in the variety table. This information appears only when the Status is candidate variety (C) and does not appear when the Status is variety (V) or no variety (nV). The individuals are similar if they differ only in special characters, certain letters such as á and ó, spaces and capital letters. These individuals are clickable for easier navigation. All messages in the panel can be dismissed.

Information notifications may include miscellaneous notifications, for example updates of the tool.

## **Administrator features**

For proper functionality and security of the tool, some features are only available for the administrator.

### **Uploading new genotypic data**

The Genotypic data panel consists of simple file selection feature, where the xlsx file with genotypic data can be uploaded. The variety names in the raw genotypic data file are encrypted to increase the safety of the data. After uploading, the recalculation of the genetic distances will begin immediately.

### **Database management**

This panel consists of table view of (up to) last 10 changes to database. The administrator can see who changed the database, what was the action and when exactly it was executed. In case of error, each action can be reverted by a corresponding button.

### **Notify users**

After inserting the password, the administrator can click a button to send a standard email to all users with the purpose to (1) inform that new genotypic data is available or (2) to remind to upload information of new candidate variety data or (3) to send out new instructions for the workflow of the current year.

## **Testing of the common database and genetic distance calculation tool**

In the course of the project, the EOs were asked to get familiar with the tool, as well as to test and sharpen the genetic distance as defined in DURDUS. To this end, different tasks were established to be completed by the EOs during August to October 2021. The EOs provided an evaluation report on 15 October 2021.

### **Task 1: Candidate varieties**

This task was designed to help the EOs to get familiar with the tool (feature: GD calculation) and evaluate its usability.

The EOs were asked to work on the following objectives:

- Get the genetic distances of the candidate varieties that were DUS tested in 2020/2021 to their potential comparators.
- Use your standard (GAIA) method to find the comparators to be grown in the field in the second year of DUS tests for the same candidate varieties.
- Compare the two methods.

The EOs were asked to provide a report containing:

- Compare the results of the two methods: How useful is the DURDUSTools approach?
- Did you have any problems using the tool? Please specify in detail.

## Task 2: Historical data

This task was designed to help the EOs to get familiar with the tool (feature: GD calculation) and sharpen the genetic distance.

The EOs were asked to work on the following objectives:

- Select 20 random varieties and use the GD calculation tool.
- Select 5 varieties that cover the whole GD range (0 – 0.6) to the most possible extent.
- From the GD output, select representative varieties for each GD range (minimum number)
  - GD 0 – 0.1: 1 variety
  - GD 0.1 – 0.2: 1 variety
  - GD 0.2 – 0.3: 1 variety
  - GD 0.3 – 0.4: 2 varieties
  - GD 0.4 – 0.5: 1 variety
  - GD 0.5 – 0.6: 1 variety
- Compare the phenotypic distances of the 35 pairs using your standard (GAIA) method.

The EOs were asked to provide a report containing:

- The genetic distance of the selected pairs
- Taking into account the phenotypic distances, would you grow the pairs in the field? (YES or NO)

## The use of DurdusTools by DUS experts

Currently, the EOs may use the information of the genetic distance together with the phenotypic assessment after the first year of DUS trials to select comparators to be grown in the field next to the candidate varieties under investigation.

All partner EOs reported on their experience how they use the genetic distance (GD) calculation tool and which threshold they apply when planning their field trials. Generally, it was concluded that the tool is user-friendly and helps to improve the quality of the DUS tests. The previously defined threshold (0.32, with a safety margin up to 0.40) is used as general guidance, and the DUS experts explained how they concretely apply this threshold in their countries and for their material.

## **Austria**

Reference varieties with a genetic distance of less than 0.35-0.40 to the candidate variety are included in the second year trials.

Varieties that would not be considered for side-by-side comparisons based on the GAIA results are included. The phenotypic assessment reflected the GD results by showing that some specific pairs were indeed very similar but differed in minor details. GD calculation increased the safety by considering similar varieties and minimized the risk of missing relevant reference varieties.

## **Hungary**

The Hungarian EO included reference varieties with a genetic distance of less than 0.32 - 0.40 to the candidate variety. This range was defined for safety reasons, whereas the GAIA between varieties is usually high. Defining the GAIA distance is crucial, therefore the GD should currently be used after the first year of DUS testing. Further evaluation of the threshold is needed.

Increase of safety and quality.

## **Spain**

Varieties with a GD lower than 0.2 should be included in the trial, regardless of the GAIA value. Varieties with a low GAIA distance value (here: 0-2) should be included in the trial, regardless of the GD. Varieties with a low GAIA distance (here: 3-5) could be excluded if the GD is more than 0.45. Varieties with a medium GAIA distance (here: 6-9) could be excluded if the GD is more than 0.32.

Decisions should be made case-by-case basis and always include the morphological data as indicated by UPOV. The GD calculation can support the decision. The tool is very useful to increase the quality on the assessment of distinctness because it allows including varieties in the DUS trial which might not have been detected based on morphological data. A completely novel benefit of the GD calculation is the recognition of similar candidate varieties under evaluation at another EO. Sharing experiences between EOs increases security.

## Italy

A threshold of 0.32-0.35 was applied. For the selected varieties with a GD around 0.32-0.35 a morphological comparison was performed to exclude them from the side-by-side comparison. The field planning has been done using both genomic and morphological information.

It was emphasized that the aim of using the tool is to reduce the number of reference varieties, and thus the work and costs. A further benefit of the GD calculation is the knowledge if a breeder submitted two identical samples, one for registration and one for protection.

## Conclusions, Benefits and limitations

From the reports of the EOs it became clear, that the tool was very useful for the planning of the DUS trials. GAIA and the genetic distance came to different conclusions in some cases. It was concluded that the genetic distance (GD) should not be the sole basis for selection of reference varieties and the phenotypic evaluations should always be considered. A common threshold applied for all EOs could not be found.

### Benefits

Increases the security of assessing distinctness in the field comparisons by detecting varieties with a very low GD (i.e.  $<0,2$ ), including varieties from the reference collection not found to be similar based on morphological characteristics, and candidate varieties under evaluation in other examination offices.

Refines the number of varieties needed to be grown side by side in the second year of testing by combining GD with the GAIA value following the UPOV model for the combination of morphological and molecular distances.

Each EO can compare its own candidate varieties with the candidate varieties that are in the registration procedure in another country in Europe at the same time.

### Limitations

The GD between a candidate variety and the reference collection cannot be used as the only method to assess distinctness and to exclude varieties from field trials. GDs calculated by DurdusTools should always be considered together with morphological differences.

Discrepancies in the genetic distances of varieties with the same denomination were detected in December 2022. As a consequence, the functionality of the tool and the GD calculation were thoroughly checked. After having excluded technical errors, the issue was further investigated and the service provider was contacted. It was concluded that having consistent genotypic data across multiple year and SNP marker micro arrays has so far not been given enough attention

as this is a specificity of our requirements in DUS testing. To secure the quality and the consistency of the genotyping, a specific cluster file for our analyses was created and it was decided to include control varieties in each genotyping process. This procedure is expected to provide stable and reproducible data over different years.

## Workflow

### Definition of a harmonized workflow

In the definition to create a common workflow for all participating EOs, crucial factors are:

- timeline of applications
- receipt of seed material
- planning of the field trials: September/October – winter trials, February/March – spring trials
- sowing of seeds
- availability of the results of the DUS tests

The following challenges apply:

- (1) The time between the receipt of the seed from the applicant and the planning of the growing trials/sowing is not sufficient to finish genotyping. Genotyping is expected to take 4-6 weeks, whereas the EOs typically only have about 3-4 weeks between receiving the seed and planning the trials.
- (2) The schedules of the EOs differ to a considerable extent.

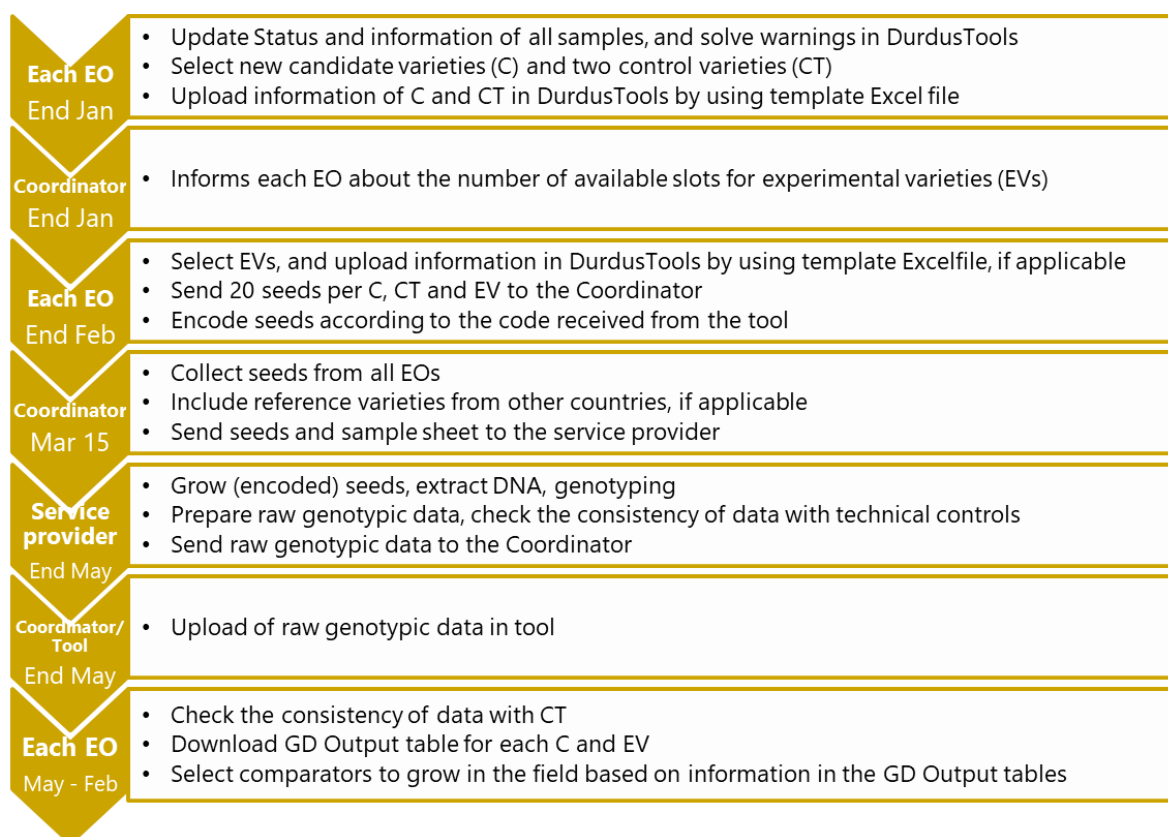
Outstanding effort would be needed to fully harmonize timelines. The consortium came to the conclusion that this effort outweighs the benefits. For this reason, GD calculation is not available in the first year of DUS tests, except in Austria due to different timelines. Here, the VCU tests start one year earlier than the DUS tests, and seeds are available before the first year of DUS tests. Therefore, for the Austrian EO, it is possible to consider the GD already in the first year trials. This is especially useful to grow closely related genotypes next to each other and not to miss relevant reference varieties. Exclusion of reference varieties only on the basis of GD without morphological consideration is currently not aimed at and needs further evaluation.

### Common workflow

The existing workflow from the DURDUS final report was updated and adjusted according to the presented timelines. It was agreed that at the beginning of the year an automatic email will be sent to all partners to present the different steps during the started year.



- (1) End of January: Each EO shall ensure that the status of and the information on all varieties and candidate varieties for which it is responsible for is up to date. Each EO has received the results of the past DUS tests and can update the status of the candidate varieties (Cs), either to variety (V) if it has been registered or to no variety (nV) if it has not been registered. The EOs are responsible for solving any issues indicated by the warnings in the tool. This includes the removal of duplicated denominations, e.g. in case of control varieties or experimental varieties. Each EO has updated the variety list in DurdusTools with the information on the new candidate varieties and selected two control varieties (CTs) per EO to be genotyped.
- (2) End of January: AGES, the Coordinator, informs the EOs about the number of available slots for experimental varieties (EVs) per EO. Accordingly, the EOs may add EVs to be genotyped to the variety list in DurdusTools.
- (3) End of February: AGES has received 20 seeds of each candidate variety (Cs), control varieties (CTs) and experimental varieties (EVs) from the EOs. The seeds are anonymized by the EOs. To do this, EOs receive a corresponding ind number when uploading the information into DurdusTools. In this way, the genetic data can be assigned to the corresponding genotypes. Anonymized seeds can be sent at any timepoint until the end of February and will be collected at AGES.
- (4) March 15: The Coordinator sends all seeds of Cs, CTs and EVs from the EOs to the service provider. The Coordinator seeks to include missing reference varieties that are listed in the EUPVP - Common Catalogue but are from countries that are not part of this agreement. The Coordinator adds all samples to the electronic sample sheet that is sent to the service provider by e-mail. AGES (the Coordinator) is the single point of contact for the service provider and serves as an intermediary between the EOs and the service provider. This will also help to minimize potential sources of error.
- (5) March 15 to end of May: The service provider receives the seeds, grows the seeds, extracts the DNA from several individual plants, and performs genotyping with an Illumina Infinium 25K SNP marker micro array. The DNA of two defined control samples (varieties Sherekhan and Spineto) is included in each analysis to ensure the technical consistency of the data between years. These controls are done by the service provider. Raw genotypic data is expected to be delivered 6-8 weeks after the seeds are shipped.
- (6) End of May: An Excel file containing the raw genotypic data is sent to the Coordinator who uploads it to the tool. All EOs will be automatically notified by e-mail that the genotypic data is now available for GD calculation.
- (7) End of May: The EOs can use the results of the GD calculation of their CTs to check the consistency of their control varieties as quality control. The results of the GD calculation of their Cs can be used for planning their DUS trials. The EOs can use the results of the GD calculation of their EVs for their own purposes.



## Draft Partnership Agreement

The drafting of the partnership agreement commenced after the final adjustment of the SNP marker selection during the extension period of DURDUSTools. The consortium wanted more clarity on the way forward, in particular what further testing would be needed in the routine use of DurdusTools. These aspects are included in the draft partnership agreement, which is provided as an Annex to this report. It covers all technically relevant aspects for routine use and maintenance of DurdusTools and shall be checked and finalized together with the legal departments of the participating parties. Thus, the draft partnership agreement will be finalized after the end of the project, together with GEVES that has not been part of the DURDUSTools project but has been invited to participate in the routine use of the genetic distance calculation tool DurdusTools. All five entrusted EOs for durum wheat have agreed to and will participate in the routine use of DurdusTools in DUS testing.

## Project management

In the course of the project, three amendments were made to Grant Agreement No 7517891:

- 1) To extend the date of submission of the interim technical report of the “DURDUSTools” project from 30 April 2021 until 31 May 2021;
- 2) To extend the duration of Grant Agreement No 7517891 until 31 March 2023 and to allow the consortium to progress in solving the issues observed with the set of markers selected;
- 3) To extend the duration of Grant Agreement No 7517891 from 31 March 2023 until 16 June 2023, in order for the partners to finalize their conclusions and define the conditions of the partnership agreement for the maintenance of the database after the project;

The amendments did not entail additional costs for the CPVO.

The project plan was amended accordingly as shown in Table 1. Smooth implementation of the work plan was ensured by regular communication via email and a number of online meetings as detailed below. In addition, one on-site meeting of the project team was organized.

A first ad-hoc online meeting was performed with all project partners on March, 24, 2021 to explain the proposed features of the genetic distance calculation tool following the requirement analysis and to discuss related aspects.

The requirement analysis was distributed among the partners on March, 24, 2021, critically reflected and discussed in the offices and returned to AGES on April, 14, 2021.

The first project meeting as scheduled was held online on May, 04, 2021 to agree upon the features of the DURDUSTools prototype. All partners attended the meeting with exception of the third party (Euroseeds), representatives of which were excused due to conflicting schedules. In addition to the partners, a representative of CPVO was present in the first meeting. An additional online project meeting will be held on June, 09, 2021 between AGES and representatives of Euroseeds and CPVO to focus on the security and access rights of the genetic distance calculation tool, as well as the information of breeders about the project.

Following Article II.9 of the Grant Agreement for an Action with CPVO, a Consortium Agreement was concluded between the project leader AGES and the project partners INIA-CSIC, CREA DC and NÉBIH, as well as the third party Euroseeds.

For the presentation of the DURDUSTools prototype version and the tasks for the testing phase an online meeting was performed on July, 29, 2021. All partners attended the meeting with

exception of the third party (Euroseeds) and the representatives of the Austrian EO. The meeting was recorded and sent to the missing partners. The testing phase of the prototype started with the distribution of the login data on August, 03, 2021 and ended with the due date of the report on October, 15, 2021.

A follow-up to the first meeting of DURDUSTools was organized online between representatives of AGES, CPVO and Euroseeds on June 09, 2021. The objective was to discuss, in particular, the confidentiality of data and security aspects, the access to and the use of the genotypic data that is stored in the DURDUSTools database, and informing the breeders about the project.

The first version of the interim report was submitted in May 2021 and complemented with an annex after the follow-up meeting on 09 June 2021.

An online project meeting with all project partners was organized on November 25, 2021 to collect the feedback and the experience concerning the use of the DURDUSTools prototype.

A physical DURDUSTools meeting was organized on September 14 and 15, 2022 in Siófok, Hungary to exchange and discuss about the past experiences with DURDUSTools in planning the DUS field trials. The aim was to create a common workflow and to prepare the basis for a partnership agreement. All project partners, as well as CPVO were involved.

An online meeting was performed on November 30, 2022 to discuss the summary of the meeting in Siófok and to prepare for the meeting with GEVES on December 12, 2022.

In the morning of December 12, 2022 GEVES, an entrusted EO for durum wheat not involved in DURDUSTools, was invited to an online meeting to discuss the routine application of DURDUSTools in DUS testing. All DURDUSTools project partners and representatives of CPVO were present. GEVES concluded that it would be part of the routine application of DURDUSTools for DUS testing.

On the afternoon of December 12, 2022, the project partners of DURDUSTools, together with CPVO and Euroseeds had another online meeting to discuss remaining open questions and to decide on the next steps concerning the report and the draft partnership agreement.

On January 18, 2023, the project partners of DURDUSTools had another online meeting to discuss the (draft) Partnership Agreement and present a slightly modified version of the genetic distance calculation in DURDUSTools with mathematical modification of the calculation and removal of presumably problematic SNP markers. Afterwards, the EOs were send a link to a copy version of the tool to test the new calculation of genetic distances.

On February 08, 2023, AGES and the service provider had a meeting to discuss the recently discovered challenges concerning the consistency of the genotypic data across multiple years and SNP marker micro arrays. It was agreed that the service provider will investigate the genotyping data created in DURDUS and DURDUStools in detail and select for the most stable and reproducible SNP markers over different years.

On March 02, 2023, the service provider presented the conclusions of the new 6,295 SNP marker selection to all project partners. On March 14, 2023, AGES created a copy of the tool to test the results of the genetic distance calculation with the new SNP marker selection.

On April 03, 2023, the project partners had an online meeting to discuss the conclusions of the genetic distance calculation with the new 6,295 SNP marker selection. It was agreed to additionally test another reduced SNP marker selection based on the DURDUS/DURDUStools SNP marker selection. On the next day, a copy version of the app was created to test the results of this new 3,928 SNP selection.

On April 19, 2023, the conclusions of the results of the genetic distance calculation with the new SNP marker selection were shared among all project partners in an online meeting. The final conclusions on the 3,928 SNP marker selection was made and integrated in the tool.

On April 24, 2023, the project coordinator had an online meeting with the service provider to discuss the possibility to incorporate control varieties.

On April 27, 2023, the first drafts of the partnership agreement were shared and discussed with all project partners in an online meeting.

On May, 09, 2023, the procedure with control varieties and the division of the costs were discussed between all EOs.

On May 17, 2023, the draft partnership agreement and draft final report were discussed between all EOs.