



# Root2Res

Root phenotyping and genetic improvement for rotational crops resilient to environmental change

## Genotypic assembly

Kelly Houston, HUTTON




Funded by  
the European Union



UK Research  
and Innovation

Project funded by

 Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra  
Swiss Confederation

Federal Department of Economic Affairs,  
Education and Research EAER  
State Secretariat for Education,  
Research and Innovation SERI

Root2Res has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101060124. Its work is supported by Innovate UK through the Horizon Europe Guarantee scheme Grant Agreement No. 101060124 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) grant No. 23.00050.

This deliverable is the culmination of activity in Task 3.2: *Provide comprehensive genotypic data for all the study crops*. This task had the objective of collating genotypic data for all crops. The aim was to assemble, and where necessary supplement, all the available genotypic data for the germplasm to be used in the project.

Deliverable Number	Work Package / Task
D 3.1	WP3 T1
Lead Beneficiary	Deliverable Author (S)
HUTTON	Kelly Houston (HUTTON)
Planned Delivery Date	Actual Delivery Date
31.08.2025	29.08.2025

Type of deliverable			
R	Document, report (excluding periodic and final reports)		✗
DEM	Demonstrator, pilot, prototype, plan designs		
DEC	Websites, patents filing, press & media actions, videos, etc.		
DATA	Data sets, microdata, etc.		x
OTHER	Software, technical diagram, algorithms, models, etc.		

Dissemination level			
PU	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page))		
SEN	Sensitive, limited under the conditions of the Grant Agreement		✗

Root2Res has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101060124. Its work is supported by Innovate UK through the Horizon Europe Guarantee scheme Grant Agreement No. 101060124 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) grant No. 23.00050.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, UK Research and Innovation (UKRI), European Research Executive Agency (REA) or Swiss State Secretariat for Education, Research and Innovation (SERI). Neither the European Union nor any other granting authority can be held responsible for them.

## Contents

<b>1. Details of the populations and genotyping used for each of the different species.....</b>	<b>4</b>
1.1. Barley North:.....	4
1.2. Barley South:.....	4
1.3. Durum wheat:.....	4
1.4. Faba bean:.....	4
1.5. Potato:.....	5
1.6. Sweet Potato:.....	5
1.7. Lentils:.....	5
1.8. Peas:.....	6
<b>2. Plans for how the data will be used within the project .....</b>	<b>7</b>
<b>3. Legacy of the data for Root2Res and wider community .....</b>	<b>7</b>
<b>4. References .....</b>	<b>7</b>

## Tables

Table 1. Details of the data, and where to find them open source and links where possible.....	6
--	---

# 1. Details of the populations and genotyping used for each of the different species.

## 1.1. Barley North

A nested associated mapping (NAM) population was developed by JHI and KWS, backcrossing eight landraces to KWS Irina (a spring two-rowed barley). The landraces used as donors are Afghan, China Huang Yen, Cornish, Donegal, Eire 6-row, Floye, Golden Archer, Hen Gymro, Millenium, Morayshire Gold, Webbs Naked, Zephyr. This population was genotyped using the barley 50k SNP chip (Mayer et al., 2017). The full population of 396 of BC<sub>1</sub>F<sub>4</sub> lines were analysed with the corehunter R package to reduce the size of the population to 144 (approximately 40% of the original population), while maintaining genetic diversity. To validate whether we had maintained genetic diversity in the core population of 144 lines, GWAS analysis was carried out using traits including heading date and observed that the same associations for these traits regardless of if the analysis was carried out with the full or core population.

## 1.2. Barley South

The CBBT collection used for the GWAS analysis to study root traits in droughted conditions comprised of 260 varieties and advanced lines assembled by ICARDA Association panel. These lines were selected as they are more suited to the conditions in Spain and Morocco, where these field trials took place and potentially represent genetic diversity not captured in the NAM population used in the barley north field trials. This collection has been genotyped with the barley 50k SNP chip (Mayer et al., 2017).

## 1.3. Durum wheat

The CERREALMED population of 300 accessions including cultivars and landraces was used for this species. This population has been genotyped with the 90K genotyping array to provide 16,634 SNPs. This population is combination of lines identified by ARVALIS and ICARDA and is part of the CERREALMED project.

## 1.4. Faba bean

Data for the NORFAB panel of faba bean accessions using single primer enrichment technology (SPET) was generated. This included 196 faba bean accessions for 323007 biallelic SNP markers. The data set was obtained after filtering out markers from the raw VCF file containing 1.86 M variants according to the following criteria: positions with overall quality (QUAL) score and mean mapping quality (MQ) lower than 20 were discarded, genotypes supported by 3 or fewer reads were set to missing following a filter for missingness tolerating up to 50% of missing values. Finally, multiallelic and

monomorphic sites as well as indels and variants in close proximity to indels (within 10 bp) were discarded following a filter based on minor allele frequency of 0.01. VCF handling and all filtering steps were performed using BCFtools 1.15.

## 1.5. Potato

The germplasm to be used for the potato aspect of this project is the diploid 06h1 *Solanum phureja* population, which consists of 100 lines. These lines have been genotyped using genotyping by sequencing and has been published in Gartner et al., 2024 <https://doi.org/10.1007/s00122-024-04605-0>.

## 1.6. Sweet Potato

A collection of 200 diverse sweet potato lines has been genotyped with DarTSeq to provide genetic data. The data is undergoing validation and approval before it will be made available open access.

## 1.7. Lentils

A total of 220 lentil accessions were selected among the Training-core collection developed within the INCREASE H2020 project (Bellucci et al., 2021; Guerra-Garcia et al., 2022). The lentil diversity panel for the Root2Res project was selected based on the following criteria: sampling a diverse and representative subset of the T-core based on genetic diversity (<https://www.corehunter.org/>) and geographic distribution; seed availability for GrowScreen-PaGe experiment (Gioia et al., 2016); and availability of other phenotyping data coming from other project/experiments.

Lentil ( $2n=2x=14$ ; 3.7 Gbp) accessions were sequenced using the CRISPR-Cas9-based depletion protocol (Jumpcode Genomics) developed by Rossato et al. (2023) within the INCREASE H2020 project and made available to the Root2Res project. By using a custom set of guide-RNAs, this method enabled targeting of 2.9 Gbp of repeats and concentrating the sequencing on coding and regulatory regions of the genome, significantly minimizing sample preparation procedure, sequencing costs, and genomic redundant information.

The sequencing data were pre-processed for quality check and aligned to the *L. culinaris* CDC Redberry v2.0 reference genome (Ramsay et al., 2021). Variants calling was performed using GATK HaplotypeCaller v4.1.7.0.

As a preliminary step, single nucleotide polymorphisms (SNPs) found within identified repeated regions were excluded from the dataset, subsequently we retained only polymorphic and biallelic sites with a mean depth value within 3 and 15, missingness less than 0.7 and minor allele frequency higher than 0.01. The study of individual missingness led to the removal of 9 accessions with more than 50% missing data resulting in 211 accessions. The filtering pipeline produced a final dataset of 765,453 single nucleotide polymorphisms.

## 1.8. Peas

The pea population to be used is USDA core collection (254 lines), which has been genotyped using GBS and the data has been published online (<https://doi.org/10.3929/ethz-b-000681298>). The population was genotyped-by-sequencing (GBS) following the protocol proposed by Poland et al. (2012; <https://doi.org/10.1371/journal.pone.0032253>) using a combination of PstI and MspI as restriction enzymes.

Table 1. Details of the data, and where to find them open source and links where possible.

Dataset	Genotyping method	Germplasm	Reference if available
<b>Barley North</b>	50K SNP array	KWS Irana landrace NAM population n = 396	
<b>Barley South</b>	50K SNP array	260 varieties and advanced lines assembled by ICARDA Association panel	
<b>Durum wheat</b>	90K SNP array	A combination of lines identified by ARVALIS and ICARDA and part of CerealMed project and mainly elite and advanced lines	
<b>Faba bean</b>	SPET sequencing	196 faba bean accessions for 323007 biallelic SNP markers	
<b>Potato</b>	GBS	100 lines of the diploid 06h1 Solanum phureja population	Gartner et al., 2024 <a href="https://doi.org/10.1007/s00122-024-04605-0">https://doi.org/10.1007/s00122-024-04605-0</a>
<b>Sweet potato</b>	DArTseq array		
<b>Lentils</b>	WGS using a CRISPR-Cas9-based depletion protocol	A total of 220 lines, 107 landraces, 24 breeding lines, 25 cultivars, rest unassigned	

Dataset	Genotyping method	Germplasm	Reference if available
<b>Peas</b>	GBS	USDA core collection (320 lines)	<a href="https://doi.org/10.3929/ethz-b-000681298">https://doi.org/10.3929/ethz-b-000681298</a>

## 2. Plans for how the data will be used within the project

The data generated and collated within WP3 of Roots2Res will primarily be used for Genome Wide Association Studies or Quantitative Trait Loci mapping experiments conducted in the project. The aim of these experiments is to identify loci, or ideally genes, linked to variation in root traits that underpin the crops ability to be resilient under abiotic stress conditions (including drought). The genotypic data will also be used to explore population structure within the populations of the different crops used in this WP to help to interpret the results from the mapping analysis.

## 3. Legacy of the data for Root2Res and wider community

The genotypic data generated by this WP will be used for other analysis in the project. For example, we could use this data to determine how genetically distinct smaller numbers of individual lines are from each other to complement experiments carried out other WPs, that are focused on specific root related traits. We could also use this data to guide our selection of genetically diverse or similar individuals for additional experiments in other WP. The legacy of these data is to help breeders to consider roots in their breeding programs and be part of a genetic toolbox for the future for these crops. Another legacy of this data for the wider community is that these genotyped populations can be used to map other traits to address other questions in future project. As genotyping can be expensive, these populations are an excellent resource for the crop research community.

## 4. References

Bellucci, E., Mario Aguilar, O., Alseekh, S., Bett, K., Brezeanu, C., Cook, D., De la Rosa, L., Delledonne, M., Dostatny, D. F., Ferreira, J. J., Geffroy, V., Ghitarrini, S., Kroc, M., Kumar Agrawal, S., Logozzo, G., Marino, M., Mary-Huard, T., McClean, P., Meglič, V., ... Papa, R. (2021). The INCREASE project: Intelligent Collections of food-legume genetic resources for European agrofood systems. *Plant Journal*, 108(3), 646–660. <https://doi.org/10.1111/tpj.15472>

Guerra-García, A., Gioia, T., von Wettberg, E., Logozzo, G., Papa, R., Bitocchi, E., & Bett, K. E. (2021). Intelligent Characterization of Lentil Genetic Resources: Evolutionary History, Genetic Diversity of Germplasm, and the Need for Well-Represented Collections. *Current Protocols*, 1(5). <https://doi.org/10.1002/cpz1.134>

Rossato, M., Marcolungo, L., De Antoni, L., Lopatriello, G., Bellucci, E., Cortinovia, G., Frascarelli, G., Nanni, L., Bitocchi, E., Di Vittori, V., Vincenzi, L., Lucchini, F., Bett, K. E., Ramsay, L., Konkin, D. J., Delledonne, M., & Papa, R. (2023). CRISPR-Cas9-based repeat depletion for high throughput genotyping of complex plant genomes. *Genome Research*, 33(5), 787–797. <https://doi.org/10.1101/GR.277628.122/-/DC1>

Ramsay, L., Koh, C. S., Kagale, S., Gao, D., Kaur, S., Haile, T., Gela, T. S., Chen, L.-A., Cao, Z., Konkin, D. J., Toegelová, H., Doležel, J., Rosen, B. D., Stonehouse, R., Humann, J. L., Main, D., Coyne, C. J., McGee, R. J., Cook, D. R., Bett, K. E. (2021). Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. *BioRxiv*, 2021.07.23.453237. <https://doi.org/10.1101/2021.07.23.453237>

Gioia T, Galinski A, Lenz H, Müller C, Lentz J, Heinz K, Briese C, Putz A, Fiorani F, Watt M, Schurr U, Nagel KA. (2016) GrowScreen-PaGe, a non-invasive, high-throughput phenotyping system based on germination paper to quantify crop phenotypic diversity and plasticity of root traits under varying nutrient supply. *Funct Plant Biol*, 44(1):76-93. <https://doi.org/10.1071/FP16128>