Allele mining

Plant Genetic Resources (3502-470)

23 May 2025

Table of contents

1	Motivation	1				
2	Learning goals	2				
3	The use of PGR for breeding of improved varieties	2				
	3.1 Diversity in countries of origin	2				
4	Approaches to allele mining	3				
5	Query of genebank databases	4				
	5.1 Identification of new disease resistance alleles in wheat	5				
6	Focused identification of germplasm strategy (FIGS)	5				
7	Examples of FIGS analyses	6				
	7.1 General structure of a FIGS analysis	6				
	7.2 Drought-tolerant faba beans	8				
	7.3 Using FIGS to identify new sources of resistance genes	8				
8	Allele mining based on population genetic analysis	10				
9	Key concepts	11				
10	Summary	11				
11	Further reading	13				
12	2 Study questions					
Ref	ferences	13				

1 Motivation

The identification of new and useful genetic variation is one of the primary goals and functions of both *ex situ* and *in situ* gene banks, as well as the diversity that is present in world wide breeding programs. The identification and subsequent utilitation of genetic variation is frequently called **allele mining**, which is defined as the identification of useful genetic variation ind exotic germplasm (wild relatives and landraces) as resource for improving modern elite varieties.

The purpose of this lesson is to demonstrate the application of allele mining. A key goal is to raise the awareness that knowledge of the population genetic history of a crop and of its geographic distribution is highly useful for allele mining because in combination with environmental parameters it allows to identify adaptive variation in known genes or novel genes of interest.

2 Learning goals

- 1. Understand how allele mining is used to identify and utilize useful genetic variation from exotic germplasm (wild relatives and landraces) for improving modern elite varieties .
- Describe and contrast core-collection sampling, environmental association mapping, targeted allele mining of known genes, tests of selection, and population-genetic modelling as strategies for uncovering adaptive variation.
- Outline the Focused Identification of Germplasm Strategy (FIGS), including the integration of phenotypic, passport (geographic/ecological), and climatic data, and the use of multivariate analyses to predict adaptive traits.
- Describe how sequence-based diversity analyses can reveal signatures of positive selection and prioritize candidate resistance alleles in plant genetic resources.

3 The use of PGR for breeding of improved varieties

We define genetic resources as exotic (non-native) or old landraces and the crop wild relatives of modern crops that evolved in and adapted to specific environments and agricultural systems. The resources can be crossed into current breeding populations for the continuous improvement of elite material. Genetic resources can may also be useful as sources for new traits to meet new challenges and requirements in crop cultivation:

- Adaptation to global change Create varieties that tolerate higher temperatures and longer periods of low precipitation or periods of flooding, for example
- **Opening of new regions for cultivation** Expand the cultivation range, for example in cooler regions such as Northern Europe by improving cold tolerance (of maize, for example)
- **New uses** Develop new uses such as bioenergy crops for biogas production by selecting varieties with a later flowering time or a larger growth potential for higher biomass.

The main challenge is to identify useful genetic variation that allows to achieve these goals and to separate it from disadvantageous variation during the introgression into elite material by classical breeding or genetic engineering.

3.1 Diversity in countries of origin

In many cases the major crop plants originated in different countries than those in which they are mainly cultivated (Figure 1). Since crop genetic diversity in the areas of domestication tends to be high than in other regions, chances that useful alleles can be found there are also higher, and therefor such PGR are of particular interest. On the other hand, if a crop was successful after domestication and spread rapidly into new areas, local adaptation in combination with human artificial selection may have led to the



Figure 1 – Centers of domestication and centers of current cultivation of major crops. Source: Diamond (2002)

origin of new, locally adapted varieties *outside* the center of origin of a crop.

This fact, which applies to many crop species, is a good motivation for allele mining by utilizing the *whole* diversity present in genetic resources. One example is the improved cold tolerance of a Swiss maize landrace, which originated over the last few hundred years of cultivation in the cool Rhine valley of Switzerland (Peter et al., 2009) and is now evaluated for introgression into modern breeding material (Hölker et al., 2019).

4 Approaches to allele mining

Several methods are available for allele mining using genetic resources. They are mainly based on the concept of selecting a subset (core collection) of accessions from a large collection of genebank accessions using criteria that either enrich or identify useful genetic variation for a trait or a certain environment. Methods also differ with respect to which type of data are used to investigate subsets of accessions: phenotypic information, genetic information, passport data (geographic coordinates of original sites, ecological information) and climatic data at the original site.

Some approaches are:

- **Core collections** Maximization of genetic, ecological or geographic diversity **Environmental association mapping** Identification of association between environmental and phenotypic variation
- Allele mining of known genes Utilize geographic distribution or annotation (functional description) and genes and genetic variants to identify novel alleles
- **Tests of selection** Discovery of new genes with a footprint of selection reflecting local adaptation
- **Population genetic modeling** to identify a large number of new alleles at candidate genes

The collection and storage of PGR *ex situ* is far advanced and reasonably comprehensive even though significant gaps remain (Ramirez-Villegas et al., 2022). Currently, there is a shift towards a characterization and utilization of these resources. Several projects attempt to characterize genetic variation in

ex situ genebanks using field trials and analysis of genetic variation. One prominent example is the Seeds of Discovery project at CIMMYT in Mexiko¹, which involves the phenotypic analysis of thousands of gene bank accessions of wheat and maize for numerous agronomic traits. An important outcome of these projects are publicly available and digitized resources that can be used by other researchers to identify useful genetic variation in PGR. The underlying accessions can then be obtained from the genebanks and used for further breeding approaches.

In the following some examples and approaches for allele mining that are currently available in the scientific literature are presented.

5 Query of genebank databases

A very simple approach to increase the chance for identifying useful genetic variation in genetic resources is to query the passport information of gene bank databases. One example is the GENESYS database² which allows the construction of very specific search queries of publicly available genetic resources. How it works is explained in a short video.³ The ordering information and a KML file⁴ with the geographic coordinates can be downloaded and used for later analyses.

As an example, one may want to find *Aegilops tauschii* from Armenia with frost tolerance.⁵ There are 14 accessions that satisfy those requirements, and they are distributed in the the range shown in Figure 2.



Figure 2 – Example of a Genesys query for frost tolerant wheat relative Aegilops tauschii from Armenia. The KML file with geographic coordinates is imported into Google Earth. Source: http://agro.biodiver.se/2014/07/ searching-genesys-the-video/

¹ http://seedsofdiscovery.org

² https://www.genesys-pgr.org/explore

³ http://www.screencast.com/t/ 0pdvMs0pc81r

⁴ XML-formatted file with geographic coordinates

⁵ Example taken from http://agro.biodiver. se/2014/07/searching-genesys-the-video/

5.1 Identification of new disease resistance alleles in wheat

In an application of an allele mining strategy, 733 wheat accessions obtained from the German genebank at IPK Gatersleben, which represent a broad geographic distribution because they originate from 20 countries, were investigated by DNA sequencing and functional assays (classical pathogenicity tests) for new alleles at the *Pm3* powdery mildew resistance genes (Bhullar et al., 2010).

Out of 8 cloned sequences of the *Pm3* gene from this sample, two new alleles were identified, which can be further tested by backcrossing into elite material. The new alleles were found in accessions from China and Nepal. Figure 3 shows the schematic sequence alignment of previously known and new alleles of the *Pm3* gene, of which two were found to be functional in pathogenicity tests.

	cc	NBS	IS		LRR		
Pm3CS							
Pm3a		4	And William Contract of		<u>t</u> , 4		
Pm3m			A (1.1.1 *** ***	\			
Pm3_4650			1. 199. March 199. Mar				-
Pm3_8152					1	11,000	75
Pm3_15011			1 I I I	1		1.1.1.1.1.1.1.1.1.1	
Pm3_7524				<mark>и</mark> в			<u> </u>
Pm3_11150				1		C	
Pm3_2816				1	1	<u> </u>	
Pm3_2616				1	1		F
Pm3_3220				1			17. * •
'	500	1000	1500	2000	2500 300	0 3500	4000

Figure 3 – Schematic representation of newly identified Pm3 alleles. The first three alleles are previously known alleles, the alleles below are eight new alleles of which the two blue ones are functional resistance allele. Red bars indicate amino acid polymorphisms relative to the Pm3CS allele, and black bars synonymous changes. Source: Bhullar et al. (2010)

6 Focused identification of germplasm strategy (FIGS)

A more formalized approach that was shown to work for the identification or new resistance alleles is **Focused Identification of Germplasm Strategy** (**FIGS**). This method is based on the hypothesis tha adaptive traits expressed in a genotype (i.e., a genebank accession) reflect the selection pressures of the environment in which it was collected. For this reason, both trait data and environmental data (mostly climate data) are combined to produce an *a priori* information or hypothesis that reflects a trait-environment relationship. The trait data are obtained from field trials at multiple locations. The key requirement is that the ecogeographic variables act in a similar fashion (i.e., exert the same selection pressure) at each of the source locations of the plant genetic resources studied.

In a second step, a set of accessions is used to define a set of accessions that contain the trait with a high probability. This set is then investigated using more elaborated analyses using a modified version of principal component analysis. Since these additional investigation can be time-consuming and expensive, FIGS is to make allele mining more efficient.

The key outcome of FIGS is a quantitative trait-environment relationship, which is based on the assumption that agro-climate parameters influence

(via evolutionary adaptation) adaptive traits that are favorable for agriculture (i.e., time to flowering, drought tolerance).

7 Examples of FIGS analyses

7.1 General structure of a FIGS analysis

In the following one FIGS analysis is described. It is important to note that only the key concept is similar over FIGS studies, but that the data and statistical analyses used may differ between studies. The task in our example was to select a subset of accessions from Nordic (i.e., Scandinavian) barley accessions for further studies.

Based on earlier studies in ecology and plant breeding it was established that *ecological niche modeling*⁶ can be used to predict the distribution of wild relatives of crop plants. Other early studies also attempted to correlate phenotypic with climatic variation, but most associations were weak. In the following an outline of a FIGS analysis is shown (Figure 4) as described by Endresen (2010):

- 1. Identify the geographic origin of a large collection of barley landraces from plant genetic resource databases
- 2. Produce field trial data of a selected set of accessions representing a broad and climatically diverse range of regions-of-origin of accessions. Conduct field trials at multiple locations.
- 3. Measure a substantial number of traits (> 10) in the field trials.
- Receive climatic data from climate databases such as the WorldClim database⁷
- 5. Organise the data into a three-dimensional array with three dimensions: Genotype, trait, as well as year with location.
- 6. Create a second three-dimensional array with the dimensions: Genotype, 12 monthly means of climate data, and a set of climate variables.
- Analyse with the PARAFAC (Generalized PCA for higher order arrays) and N-PLS method (N-way partial least squares for multiway regression).

i Some mathematical background on multivariate analysis

This is a short (advanced) description of the to multivariate analyses previously used for FIGS.

PARAFAC (Parallel Factor Analysis)

PARAFAC is based on a tensor decomposition model, which seeks to approximate a three-way (or higher) data array $\mathcal{X}_{i,j,k}$ as a sum of *F* rank-one components:

$$\mathcal{X}_{i,j,k} \approx \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f}, \qquad (1)$$

where $A = [a_{i,f}]$, $B = [b_{j,f}]$, and $C = [c_{k,f}]$ are the factor (loading) matrices for each mode (e.g. germplasm accessions, markers/genes, environments).

The method provides uniqueness without rotation under certain "mild"

⁶ It describes the range of environmental parameters within which a species can exist. See Wikipedia

⁷ http://www.worldclim.org



Figure 4 – Multiway data design of a FIGS analysis. The crop trait data (a) was organized as tri-linear (three-way) cube with 28 landrace measurements (14 samples, 2 replications) by 6 crop trait properties by 6 experimental conditions (2 yr, 3 locations). The environmental data (b) was organized as a trilinear (three-way) cube with 14 landrace origin locations by 12 monthly means by 3 climate variables. Source: Endresen (2010)

conditions (e.g. factors being linearly independent), the CP/PARAFAC decomposition is essentially unique, avoiding ambiguous rotations common in matrix PCA. An important characteristic is a **multilinear structure**, in which each component *f* captures a simultaneous pattern across all modes—e.g. a specific allele–environment combination that differentiates a subset of accessions.

The expected output of a PARAFAC analyses are the following:

- 1. Factor matrices A, B, C:
 - A scores reveal which accessions carry each latent allele– environment pattern.
 - *B* loadings show which genes or markers contribute to each pattern.
 - C weights indicate which environmental variables (e.g. climatic descriptors) drive each component.
- 2. Core consistency diagnostics: Measures (e.g. CORCONDIA) to select the optimal number of components *F*.
- Explained variation: Percent variance in the original tensor explained by the chosen PARAFAC model.

N-PLS (N-way Partial Least Squares Regression)

N-PLS is an **extension by multiway regression** and generalisation of partial least square regression (PLS). It models the relationship between an *N*-way predictor tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ and a response matrix $Y \in \mathbb{R}^{I \times M}$ (e.g. trait values) via latent (hidden) variables. This is achieved by **simultaneous decomposition**: At each component *h*, N-PLS finds weight vectors $w_h^{(X)}$ (one per mode of \mathcal{X}) and q_h for *Y* such that the covariance between the projected scores

$$t_h = \mathcal{X} \times_2 w_h^{(X,2)} \times_3 w_h^{(X,3)}, \quad u_h = Y q_h$$
(2)

is maximized. Here, \times_n denotes the *n*-mode tensor-vector product. It also results in *orthogonal deflation** after extracting each component,

both \mathcal{X} and Y are deflated and remove the effect of t_h and u_h to ensure subsequent components capture new covariance. The expected outputs of N-PLS are

- 1. **X-mode weight vectors** $\{w^{(X,n)}\}$: One set per mode, indicating how to project each dimension (accession, marker, environment) onto latent scores.
- 2. **Y-weights** *q*: How each response (trait) loads onto latent variables.
- 3. Scores $T = [t_h]$ and $U = [u_h]$: Latent representations of accessions in predictor and response spaces, respectively.
- 4. **Regression coefficients**: A multiway coefficient array *B* that can predict *Y* from new \mathcal{X} data.
- 5. **Model diagnostics**: Cross-validated R^2 , RMSEP (root-meansquare error of prediction) per trait, and the number of significant components.

In a FIGS analyses, these methods allow one to:

- **PARAFAC**: uncover simultaneous "hotspots" of allelic variation and environmental gradients that define subsets of germplasm with putatively adaptive alleles.
- N-PLS: directly link multi-environmental genotypic profiles to measured trait variation (e.g. drought tolerance), facilitating the prediction of promising accessions from the broader collection.

The results of this study showed that for traits such as heading dates, ripening days and harvest index, the missing phenotypes in landraces can be predicted with an acceptable degree of accuracy. However, the number of accessions in the landraces and the extent of field trials used to calibrate the predictive models are very crucial (more trials are better).

7.2 Drought-tolerant faba beans

Another study attempted to identify climatic variables that are associated with the drought tolerance in faba beans, *Vicia faba* Khazaei et al. (2013). Based on genebank passport data (Figure 5), two large sets of differentially adapted faba bean accessions were defined, which were then evaluated for a set of phenotypic traits.

Machine learning algorithms split the accessions into two groups based on phenotypic traits. A comparison of the *a priori* defined groups based on climatic data from the site of origin and the phenotypic trait distribution indicated a almost perfect match of those groups, which suggested that the climatic information of the site of origin can be used to identify differentially adapted genotypes with a high confidence (Figure 6).

7.3 Using FIGS to identify new sources of resistance genes

The study by Bari et al. (2012) investigated the distribution against yellow rust resistance alleles in wheat. Using machine learning approaches such as Support Vector Machines (SVM) to establish a relationship between climate



Figure 5 – Geographic distribution of the faba bean accessions used in the study. Green accessions are from moisture-limited (dry) areas, blue accessions from wetter environments. Source: Khazaei et al. (2013)

Rank	drought related parameter	model					
		rpart-caret	RF				
			mean decrease	e accuracy mean decrease Gini			
1	Leaflet temperature	34.91	0.26	55.41			
2	Canopy temperature	13.68	0.10	31.64			
3	Relative water content	12.46	0.02	9.64			
\$	Leaflet area	9.95	0.01	6.39			
5	Stomatal length	6.70	0.01	2.30			
i	Fertile tillers	4.72	0.00	0.47			
,	stomatal area	4.13	0.00	1.01			
1	Transpiration rate	3.61	0.03	6.94			
•	Stomatal area per unit area of leaflet	2.75	0.01	3.52			
0	Photosynthetic rate	2.34	0.01	4.25			
1	Days to flowering	2.21	0.00	1.95			
2	Intercellular CO ₂	1.64	0.00	0.93			
3	Stomatal density	1.26	0.01	2.42			
4	Water use efficiency	1.21	0.01	1.89			
15	Stomatal conductance	0.86	0.01	1.83			
16	Stomatal width	0.14	0.00	0.82			

doi:10.1371/journal.pone.0063107.t005

Figure 6 – Potential predictors of climate variables identified in faba beans. Source: Khazaei et al. (2013)

variables and phenotypic data on resistance, a better match than with regression approaches was found. There was a relationship between ecographic variation and the occurrence of yellow rust resistance genes among accessions from different regions. The average success rate of prediction whether a site of geographic origin contained a resistant accession was 76%.

Other examples are the detection of wheat accessions resistant to the Russian wheat aphid (*Diuraphis noxia*)(El Bouhssini et al., 2011), and stem rust (*Puccinia graminis* Pers.) in wheat (*Triticum aestivum* L. and *Triticum turgidum* L.) and net blotch (*Pyrenophora teres* Drechs.) in barley (*Hordeum vulgare* L.) (Filip Endresen et al., 2011). Another study identified new potential sources of resistance genes for the Ug99 stem rust pathogen (Endresen et al., 2012).

Careful reading of these studies, however, indicates that the overall rate of hits is relatively low and that further studies in each crop are required to verify that the observed resistances are real. Furthermore, FIGS is a pure pattern matching approach and does not include information about the genetic history of a crop and the causes of the association between geographic distribution of phenotypes and climatic parameters. For this reason, there is much room for improvement and further research.

8 Allele mining based on population genetic analysis

Simular to the analysis of geographic and population genetic patterns, the analysis of genetic variation can also be used to identify relevant candidate genes.

One example is the identification of new alleles of a disease resistance gene against barley yellow mosaic virus (Hofinger et al., 2011). The eukaryotic translation initiation factor 4E (*eIF4E*) gene located on chromosome 3H is a source of resistance to the bymoviruses Barley yellow mosaic virus and Barley mild mosaic virus. Nearly all modern barley cultivars have two recessive *eIF4E* alleles, *rym4* and *rym5*, that confer isolate-specific resistances.

The DNA sequence of *eIF4E* was analysed in 1,090 barley landraces and noncurrent cultivars originating from 84 countries. A very high nucleotide diversity was found in the coding sequence of *eIF4E*, and all nucleotide polymorphisms in the coding sequence were amino acid polymorphisms, which is very unusual. A total of 47 different *eIF4E* haplotypes were identified, and phylogenetic analysis provided evidence of strong positive selection acting on this barley gene.

The *el4FE* haplotype diversity was higher in East Asia, whereas SNP genotyping identified a comparatively low degree of genome-wide genetic diversity in 16 of 17 tested accessions (each carrying a different elF4E haplotype) from this same region. In addition, the calculation of selection statistics using coalescent simulations showed evidence of non-neutral variation for *elF4E* in several geographic regions, including East Asia, the region with a long history of the bymovirus-induced yellow mosaic disease.

The results indicate that *eIF4E* may have played a role in local adaptation to different habitats. However, no functional studies of the new resistance



Figure 7 – Geographic distribution of the eIF4E haplotypes. Source: Hofinger et al. (2011)

alleles were conducted, and for this reason, it was not clear whether the new alleles are really of functional importance or effect.

9 Key concepts

- □ Allele mining
- Focused identification of germplasm strategy

10 Summary

- Allele mining is the identification and utilization of useful genetic variation in plant genetic resources.
- The targeted identification of alleles contributes to adapt crops to climate change, allows new regions of cultivation and new uses such as biomass production
- Allele mining is based on the concept that genetic diversity is highest in the center of domestication
- Mining of plant genetic resources includes different types of information: Genotypic and phenotpic data, passport information and climatic data from the site of origin.
- Different methods for the identification of useful genetic variation exist: Core collections, Environmental association mapping to identify novel genes, allele mining of known genes, tests of selection
- The focused identification of germplasm strategy (FIGS) is a multivariate approach for the identification of novel plant genetic resources, that combines data on genetic variation, phenotypic variation and environmental variation.



Figure 8 – Distribution of diversity. The numbers indicate the different haplotypes identified at the eIF4E locus and their geographic distribution. Source: Hofinger et al. (2011)

11 Further reading

Here are a few further studies on allele mining in plants:

- sharwood_mining_2022 Mining for alleles in photosynthetic traits of plants
- Haupt and Schmid (2020) Example of the combination of FIGS and the generation of core collection
- rebetzke_review_2019 A review on the role of phenotyping in allele mining
- wambugu_role_2018 A review on the role of genomics in allele mining

12 Study questions

- 1. What are the different approaches for the identification of alleles for the purpose of allele mining?
- 2. Can you identify bottlenecks and limitations in the different approaches for allele mining?
- 3. Why is the resolution and the prediction ability of FIGS limited, and how could it be improved?
- 4. Which traits are more suiof allele mining methods like FIGS or population genetics approaches?
- 5. Why are genetic resources originating from *outside* the center of domestication of a crop of interest for allele mining?
- 6. Can you think of a further improvement of allele mining approaches with genomics approaches (e.g., genome sequencing or the analysis of gene expression) or functional approaches such as genome editing?
- 7. The *elF4E* gene is a highly conserved (i.e., slowly evolving) gene but harbors a large number of amino acid polymorphisms in barley landraces. Can you come up with some explanations for the large number of amino acid polymorphisms? How would you test whether a large number of polymorphisms is unusual or not as evidence that the polymorphisms reflect different resistance alleles agains the barley yellow mosaic virus?

References

- Bari A, Street K, Mackay M, Endresen DTF, Pauw E, Amri A. 2012. Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genetic Resources and Crop Evolution* 59:1465–1481. doi:10.1007/s10722-011-9775-5
- Bhullar NK, Zhang Z, Wicker T, Keller B. 2010. Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene Pm3: A large scale allele mining project. *BMC plant biology* **10**:88.
- Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature* **418**:700–707. doi:10.1038/nature01019

- El Bouhssini M, Street K, Amri A, Mackay M, Ogbonnaya FC, Omran A, Abdalla O, Baum M, Dabbous A, Rihawi F. 2011. Sources of resistance in bread wheat to Russian wheat aphid (Diuraphis noxia) in Syria identified using the Focused Identification of Germplasm Strategy (FIGS). *Plant Breeding* **130**:96–97. doi:10.1111/j.1439-0523.2010.01814.x
- Endresen DTF. 2010. Predictive association between trait data and ecogeographic data for nordic barley landraces. *Crop Science* **50**:2418. doi:10.2135/cropsci2010.03.0174
- Endresen DTF, Street K, Mackay M, Bari A, Amri A, De Pauw E, Nazari K, Yahyaoui A. 2012. Sources of Resistance to Stem Rust (Ug99) in Bread Wheat and Durum Wheat Identified Using Focused Identification of Germplasm Strategy. *Crop Science* **52**:764. doi:10.2135/cropsci2011.08.0427
- Filip Endresen DT, Street K, Mackay M, Bari A, De Pauw E. 2011. Predictive Association between Biotic Stress Traits and Eco-Geographic Data for Wheat and Barley Landraces. *Crop Science* **51**:2036. doi:10.2135/cropsci2010.12.0717
- Haupt M, Schmid K. 2020. Combining focused identification of germplasm and core collection strategies to identify genebank accessions for central European soybean breeding. *Plant, Cell & Environment* n/a. doi:10.1111/pce.13761
- Hofinger BJ, Russell JR, Bass CG, Baldwin T, Dos REIS M, Hedley PE, Li Y, Macaulay M, Waugh R, Hammond-Kosack KE, Kanyuka K. 2011. An exceptionally high nucleotide and haplotype diversity and a signature of positive selection for the eIF4E resistance gene in barley are revealed by allele mining and phylogenetic analyses of natural populations. *Molecular Ecology* 20:3653–3668. doi:10.1111/j.1365-294X.2011.05201.x
- Hölker AC, Mayer M, Presterl T, Bolduan T, Bauer E, Ordas B, Brauner PC, Ouzunova M, Melchinger AE, Schön C-C. 2019. European maize landraces made accessible for plant breeding and genome-based studies. *Theoretical and Applied Genetics* **132**:3333–3345. doi:10.1007/s00122-019-03428-8
- Khazaei H, Street K, Bari A, Mackay M, Stoddard FL. 2013. The FIGS (Focused Identification of Germplasm Strategy) Approach Identifies Traits Related to Drought Adaptation in Vicia faba Genetic Resources. *PLoS ONE* 8:e63107. doi:10.1371/journal.pone.0063107
- Peter R, Eschholz TW, Stamp P, Liedgens M. 2009. Swiss Flint maize landraces—A rich pool of variability for early vigour in cool environments. *Field Crops Research* **110**:157–166. doi:10.1016/j.fcr.2008.07.015
- Ramirez-Villegas J, Khoury CK, Achicanoy HA, Diaz MV, Mendez AC, Sosa CC, Kehel Z, Guarino L, Abberton M, Aunario J, Awar BA, Alarcon JC, Amri A, Anglin NL, Azevedo V, Aziz K, Capilit GL, Chavez O, Chebotarov D, Costich DE, Debouck DG, Ellis D, Falalou H, Fiu A, Ghanem ME, Giovannini P, Goungoulou AJ, Gueye B, Hobyb AIE, Jamnadass R, Jones CS, Kpeki B, Lee J-S, McNally KL, Muchugi A, Ndjiondjop M-N, Oyatomi O, Payne TS, Ramachandran S, Rossel G, Roux N, Ruas M, Sansaloni C, Sardos J, Setiyono TD, Tchamba M, van den Houwe I, Velazquez JA, Venuprasad R, Wenzl P, Yazbek M, Zavala C. 2022. State of ex situ conservation of landrace groups of 25 major crops. *Nature Plants* 1–9. doi:10.1038/s41477-022-01144-8