

# The FIGS (Focused Identification of Germplasm Strategy) Approach Identifies Traits Related to Drought Adaptation in *Vicia faba* Genetic Resources

Hamid Khazaei<sup>1\*</sup>, Kenneth Street<sup>2</sup>, Abdallah Bari<sup>2</sup>, Michael Mackay<sup>3</sup>, Frederick L. Stoddard<sup>1</sup>

<sup>1</sup> Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland, <sup>2</sup> International Centre for Agricultural Research in the Dry Areas, Aleppo, Syria, <sup>3</sup> Queensland Alliance for Agricultural and Food Innovation, The University of Queensland, St Lucia, Queensland, Australia

## Abstract

Efficient methods to explore plant agro-biodiversity for climate change adaptive traits are urgently required. The focused identification of germplasm strategy (FIGS) is one such approach. FIGS works on the premise that germplasm is likely to reflect the selection pressures of the environment in which it developed. Environmental parameters describing plant germplasm collection sites are used as selection criteria to improve the probability of uncovering useful variation. This study was designed to test the effectiveness of FIGS to search a large faba bean (*Vicia faba* L.) collection for traits related to drought adaptation. Two sets of faba bean accessions were created, one from moisture-limited environments, and the other from wetter sites. The two sets were grown under well watered conditions and leaf morpho-physiological traits related to plant water use were measured. Machine-learning algorithms split the accessions into two groups based on the evaluation data and the groups created by this process were compared to the original climate-based FIGS sets. The sets defined by trait data were in almost perfect agreement to the FIGS sets, demonstrating that ecotypic differentiation driven by moisture availability has occurred within the faba bean genepool. Leaflet and canopy temperature as well as relative water content contributed more than other traits to the discrimination between sets, indicating that their utility as drought-tolerance selection criteria for faba bean germplasm. This study supports the assertion that FIGS could be an effective tool to enhance the discovery of new genes for abiotic stress adaptation.

**Citation:** 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(5): e63107. doi:10.1371/journal.pone.0063107

**Editor:** Jean-Marc Lacape, CIRAD, France

**Received:** November 17, 2012; **Accepted:** March 28, 2013; **Published:** May 8, 2013

**Copyright:** © 2013 Khazaei et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** HK was supported by CIMO (Centre for International Mobility) and is currently supported by Emil Aaltonen Foundation (Emil Aaltosen Säätiö) fellowship. The project was also supported by the University of Helsinki and Legume Futures (KBBE 2009-1-2-01, contract 245216 CP-FP 'Legume Futures: Legume supported cropping systems for Europe'). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: hamid.khazaei@helsinki.fi

## Introduction

Drought coupled with heat stress, expected to increase in frequency and intensity is likely to expand due to climate change [1,2]. Faba bean (*Vicia faba* L.) is an important source of protein, often referred to as poor man's meat, in those dry areas of developing countries most likely to be impacted by climate change [3,4]. This has significant food security implications because faba bean is relatively sensitive to terminal moisture stress when compared to other temperate-season grain legumes [5–7] so drought is a major constraint to its production and yield stability. Therefore it is imperative that natural variation for traits related to drought adaptation be identified from the faba bean genepool and introduced into improved cultivars. Economic analysis of cultivar development showed that the identification of a desirable trait is of equal importance to the process of transferring it into improved backgrounds because it reduces the time taken to develop improved cultivars [8].

Genetic resource collections conserved in genebanks are the most obvious place to look for useful traits, but given the size of these collections, searching for specific and often rare traits has been likened to searching for a needle in a haystack. Further,

evaluating large collections for some parameters can be extremely costly. For example, the International Center for Agricultural Research in the Dry Areas (ICARDA) houses a globally important collection of over 9500 faba bean accessions. It would be beyond the resources of most research programs to evaluate this entire collection for variation in leaf morpho-physiological traits related to plant moisture stress. What is needed therefore is a means of wisely selecting an economically feasible set size that has a better probability of capturing useful variation than if material was selected at random or through the use of other techniques that do not focus on the sought-after trait.

The core collection was proposed as a way to work with fewer accessions that would represent, "with a minimum of repetitiveness, the genetic diversity of a crop species and its relatives" [9]. There are numerous examples of methodologies to develop core collections (*see* Hodgkin et al. [10] for examples), which in practice tend towards limiting the size of the sub-set to around 10% [11,12] of the original collection size. Although one of the stated purposes of core collections is to improve utilization, the vast majority of reported research seems to focus more on methods (or sampling strategies) to establish core collections [13–16] and the analysis of the diversity held within core collections [17–20]. A number of

references suggest alternative types of collections, or sets of collections, to enhance the efficiency of capturing diversity or addressing utilization, including specialized core collections [21], mini core sets [22], nested core collections [23] and composite collections [24]. Despite this diversity of core collection methodology, there seems to be a lack of literature that demonstrates that core collections have had a significant impact on the utilization of genetic resources. Rare and adaptive alleles, most of which are thought to be functional, may even be missed from a core collections [21,25–29].

The Focused Identification of Germplasm Strategy (FIGS) was designed to improve the efficiency with which specific adaptive traits are identified from genetic resource collections. FIGS is based on the premise that adaptive traits displayed by an accession will reflect the selection pressures of the environment from which it was originally sampled [30–33]. The FIGS approach uses both trait and environmental (climate) data to develop *a priori* information or specialized knowledge as per Gollin et al. [8] based on a quantification of the trait-environment relationship [32,34,35]. This *a priori* information is then used to define a set of accessions with a high probability of containing the desired traits.

Many adaptive traits can be linked to agro-climatic parameters. For example, using monthly values for a range of climatic variables, FIGS detected sources of resistance in wheat (*Triticum* ssp.) for biotic stresses such as powdery mildew (*Blumeria graminis* (DC) Speer f.sp. *tritici*) [36,37], Sunn pest (*Eurygaster intergriceps* Put.) [38], Russian wheat aphid (*Diuraphis noxia* Kurdj.) [39] and stem rust (*Puccinia graminis* Pers.) [34,35] as well as net blotch (*Pyrenophora teres* Drechs.) in barley (*Hordeum vulgare* L.) [34]. Further, Endresen et al. [40] demonstrated how eco-geographic data from the collection sites of 14 Nordic barley landraces (*Hordeum vulgare* L.) was successfully correlated to morphological traits using multi-linear data modelling techniques and conclude that the FIGS approach can be used efficiently as a targeted germplasm selection method.

However, so far studies are few on the effectiveness of FIGS to detect traits that impart tolerance to abiotic stresses such as moisture availability, and there are certainly none for faba bean.

The aim of this study was to compare the leaf morphology and phenology of two sets of faba bean accessions originating from environments with contrasting seasonal moisture availabilities. The underlying hypothesis is that ecotypic (climatic) differentiation occurred so traits associated with plant moisture regulation and lifecycle will differ between the two sets. From this, we would further expect that set membership based on collection site environmental descriptors would be the same when the accessions are classified using trait measurements.

## Materials and Methods

### Construction of FIGS Sets

Two sets containing landrace accessions of faba bean were selected from the collection conserved by ICARDA that contains 9545 entries, representing 21% of the worldwide germplasm collection [41]. One set was chosen to maximize the probability of having drought-related adaptive traits, the “dry set”, and the other was constructed as a control from accessions originating from environments with higher moisture profiles - the “wet set”. The origin of the selected accessions is presented in Figure 1, and the ICARDA accession numbers are given in Table S1.

The dry set (201) was constructed as follows. Accessions from collection sites where the annual rainfall was below 300 mm/year or greater than 550 mm/year were not considered. Of the remaining accessions, one accession per collection site was chosen

at random. A hierarchical cluster analysis was performed using the following collection site agro-climatic parameters: *precyr*, *ariry*, *tmnyr*, *tmaxy*, *bio4*, *bio15*, *bio16*, and *bio19* extracted from ICARDA and Worldclim-databases and Hijmans et al. [42] (Table 1). The climate variables were chosen to combine temperature and precipitation factors that would influence the length of growing season and seasonal moisture availability. The between-groups linkage option was set as the clustering algorithm, using squared euclidian distances as the distance measure. The procedure created 20 clusters. Accessions contained in 6 clusters were dropped because the average aridity index for the cluster was above 0.6 or below 0.1 (indicating irrigated sites). For each of the remaining clusters the accessions were sorted according to the *bio15* climate variable (a measure of the variation in seasonal moisture availability) for their respective collection sites. Any accession with a score of 50 or lower was discarded. The remaining accessions within each cluster were ranked based on collection site long-term yearly precipitation. A set of 201 accessions was chosen by selecting the lowest ranked accession in each cluster and repeating the process until the set size was achieved.

The wet set was chosen from sites that receive over 800 mm/year of rainfall (long-term average). One accession per site was chosen at random. The remaining accessions were sorted according to collection site yearly average aridity index and 201 accessions were chosen from sites with the highest aridity indices.

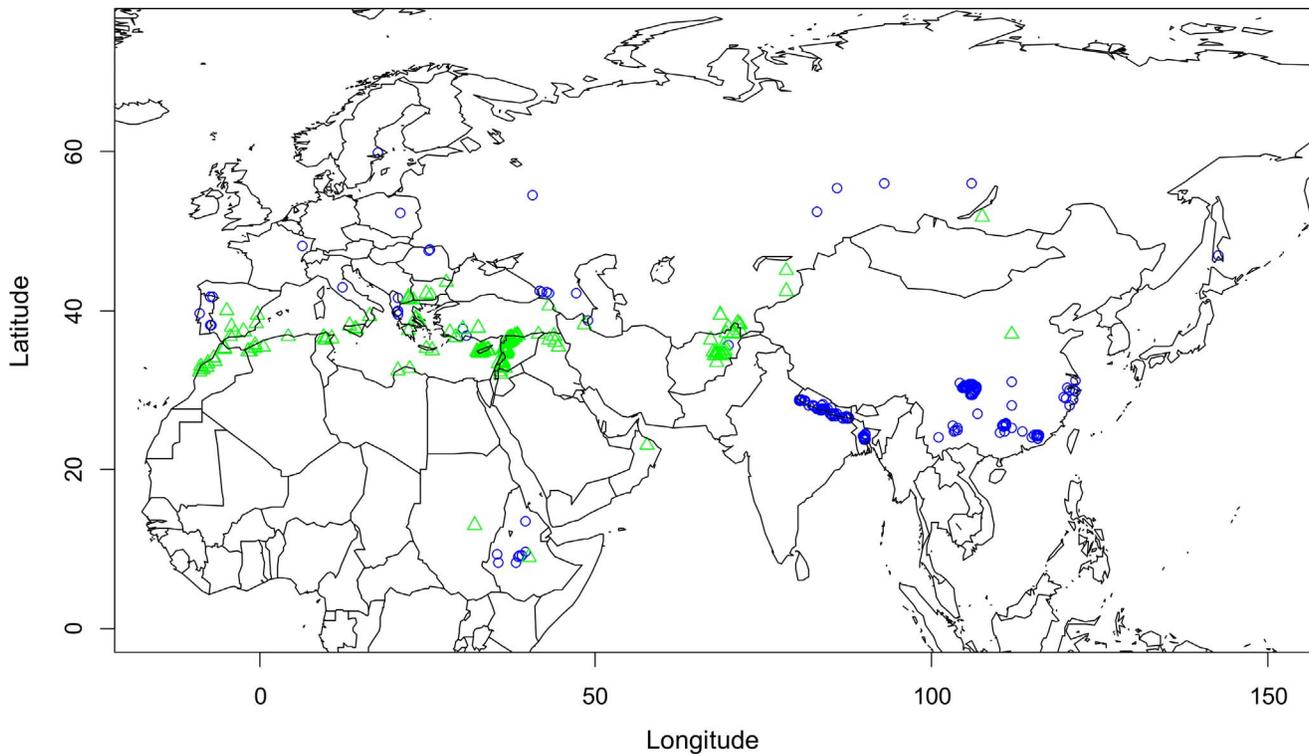
### Growth Conditions

Accessions were planted in a randomized complete block design (RCBD) with 4 replicates in a climate-controlled greenhouse of the University of Helsinki, Finland during 2010–2011, giving a total of 1608 pots. Before sowing, seeds were inoculated with *Rhizobium leguminosarum* biovar. *viciae* (faba bean strain, Elomestari Oy, Tornio, Finland). Three seeds were sown per 2 L plastic pot, which held a mixture of sand and peat (White 420 W, Kekkila Oy, Vantaa, Finland) (3:1 v/v). After 10 days, the seedlings were thinned to one per pot. Soil moisture levels were maintained at field capacity with an automatic irrigation system to ensure that each plant received the same amount of water during the experiment. At three and five weeks after sowing, 70 ml of fertilizer solution (equivalent to 20 kg of P and 24 kg of K per hectare) was added to each pot. The photoperiod was adjusted to 14 h light and 10 h dark, and the temperature was set to 21°C day/15°C night  $\pm 2^\circ\text{C}$ . Photosynthetic photon flux density (PPFD) was about 300  $\mu\text{mol m}^{-2} \text{s}^{-1}$  at the canopy level. The relative humidity was maintained at  $60 \pm 5\%$ .

**Pest control.** Thrips were controlled biologically using *Amblyseius cucumeris*, especially at seedling and flowering stages.

### Morphological Measurements

**Stomatal density and morphology.** Stomatal density (SD), length (SL) and width (SW) were measured on the middle part of the abaxial surface of the youngest, fully expanded leaflet of 8-week-old plants using the impression method [43]. The number of stomata was counted from ten different microscopic fields of view at 250 $\times$  magnification. To estimate SD, the number of stomata per field of view was converted to the number of stomata per  $\text{mm}^2$  of leaf using a standard scale. SL and SW were measured on ten stomata from the impressions using a scaled 500 $\times$  eyepiece of microscope and converted to  $\mu\text{m}$ . Stomatal area (SA) was calculated as  $\text{SA} = \text{SL} \times \text{SW}$ . Stomatal area per unit area of leaflet (SAAL) was calculated as the product of SA and SD.



**Figure 1. Geographical distribution of the two sets (wet set, blue circle and dry set, green triangle).**  
doi:10.1371/journal.pone.0063107.g001

**Leaflet area.** Leaflet area was measured using a LI-6200 leaf area meter (LI-COR, Inc., Lincoln, NE, USA). Means of four leaflets per plant were used for analysis.

**Fertile tillers.** The number of fertile tillers was counted at 16 weeks after sowing.

**Seed size.** Ten seeds from each accession were measured in order to classify them to the traditional seed size class, *minor*, *equina* and *major* according to seed length and mass [44,45].

### Physiological Measurements

**Gas exchange traits.** Gas exchange was measured on each plant at 6 weeks and 8 weeks after sowing, using a LI-6400 portable photosynthesis system (LI-COR, Inc.) equipped with a

2×3 cm leaf chamber with a LED light source (6400-02B, 90% red and 10% blue). Photosynthesis photon flux density (PPFD) was 1000  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . A  $\text{CO}_2$ -injecting cartridge was attached to the system to control reference  $\text{CO}_2$  concentration at 400  $\mu\text{mol mol}^{-1}$ , a value close to that during plant growth. The flow rate was 400  $\mu\text{mol s}^{-1}$ . All the gas exchange measurements were done between 9 and 11 am using the youngest, fully expanded leaflet which was also used for stomatal morphology and leaflet area measurements. Measurements were logged only when the stability criteria were met, according to the manufacturer's instructions. For logistical reason, each replicate was measured on a separate day. The gas exchange measurements taken were: photosynthetic rate ( $A_{\text{net}}$ ), stomatal conductance ( $g_s$ ), transpiration rate ( $E$ ), and intercellular  $\text{CO}_2$  ( $C_i$ ). Intrinsic water use efficiency (WUE) was calculated as gas exchange rate divided by stomatal conductance ( $A_{\text{net}}/g_s$ ) [46].

**Leaflet and canopy temperatures.** Leaflet temperature was measured along with gas exchange on the LI-6400. Canopy temperature was measured using a FLUKE® 574 thermometer gun (FLUKE, Everett, WA, USA) from the fully expanded leaves used for the other measurements. Canopy temperature was measured at 6 weeks and 8 weeks after sowing. Air temperature was recorded at the time of measuring leaf temperature. Leaflet temperature is presented as: Leaflet temperature – air temperature and canopy temperature as: canopy temperature – air temperature.

**Relative water content.** Five leaflets were used for determining leaf relative water content (RWC%) according to the initial principles by Barrs and Weatherley [47]. First, fresh weight (FW) was determined. Turgid weight (TW) was measured after floating the sample on distilled water in Petri dishes in darkness at 4°C for 24 h. Dry weight (DW) was calculated by putting the samples for 48h in a 60°C oven.  $\text{RWC} (\%) = (\text{FW} - \text{DW}) / (\text{TW} - \text{DW}) \times 100$ .

**Table 1.** The climatic variables used in the selection of FIGS sets.

Code	Description
<i>precyr</i>	Long term yearly precipitation
<i>ariyr</i>	Long term yearly aridity index
<i>tminyr</i>	Long term yearly minimum temperature
<i>tmaxyr</i>	Long term yearly maximum temperature
<i>Bio4</i>	Temperature seasonality (standard deviation × 100)
<i>Bio15</i>	Precipitation seasonality (coefficient of variation)
<i>Bio16</i>	Precipitation of wettest quarter
<i>Bio19</i>	Precipitation of coldest quarter

Source: ICARDA GIS (Geographic information system) unit and world global climate data (<http://www.worldclim.org/bioclim>).  
doi:10.1371/journal.pone.0063107.t001

## Phenological Measurement

The number of days to the onset of flowering was recorded.

## Statistical Analysis

The membership of the two contrasting FIGS sets was based on *a priori* information, namely the long-term climatic conditions of the sites from which the accessions were collected. The underlying assumption was that morpho-physiological traits related to moisture stress adaptation would differ between two sets of selected germplasm. Two methods were used to determine whether the two sets are different in terms of morpho-physiological phenotypic expression.

To determine if there were differences between the sets, they were subjected to a t-test, using means across replicates for each accession, with the R statistical package [48] after testing for normality.

Multivariate analysis was employed for deeper investigation because the relationships between the collection site agro-climatic conditions and trait expression are likely to be non-linear and multi-dimensional and thus not captured in a linear framework. When trait expression differs between the two sets, this should be reflected in how the classification algorithms discriminate between accessions. Thus, we would expect the algorithms to correctly assign accessions into the sets created on climatic descriptors. Three models (Table 2) were used to classify accessions, discriminate between sets and to highlight those traits that contributed most to the discrimination. The algorithms used a learning-based approach, in which they were “trained” on a set of accessions whose set membership (wet or dry) was made “known” to the algorithm. The trained algorithm was then used to classify the accessions whose set membership was “unknown” to the algorithm into two sets (wet or dry). This is an iterative process where the model that is finally chosen by the algorithm is based on the “best” values for accuracy parameters that measure the model’s ability to classify the unknown accessions into their respective climate-based sets. These learning-based techniques need fewer assumptions and thus are more suitable when highly complex non-linear relationships are expected among input variables. They were used to overcome the problem of restrictive parametric paradigms on one hand and the prerequisite distribution assumptions on the other [56,57].

The parameters used to measure the accuracy of these models are the AUC and Kappa values. The AUC refers to the area under the curve (AUC) of the Receiver Operating Characteristics (ROC) [58,59], which is a plot of true positive rate versus false positive [60]. An AUC value of 0.5 represents randomness and would indicate that the FIGS sets are no different from randomly chosen sets. An AUC value of 0.7 and above represents high model performance [59] indicating that the wet and dry sets are

highly distinguishable and that the dry set is more prone to harbor traits that favor drought adaptation. Similar to the AUC, Kappa is a measure of agreement, where a value of 0.4 and above is an indication of good agreement between the model’s prediction and the trait measurements [61].

The datasets were presented to the algorithms as follows: the mean value for each variable was calculated over the replicates for each accession. This accession level data was combined (wet+dry sets) and standardized so that the dataset mean was zero with standard deviation of 1. The algorithms split the combined data into 2 datasets containing 2/3 and 1/3 of the accessions on a random basis. The larger dataset was used to “train” the models and quantify the association between the membership (wet/dry) and the drought-related attributes. The association was then used in turn (in reverse) to classify the “unknown” accessions of the smaller dataset. This process was performed 10 times and the results were averaged.

**Selection of important parameters.** Some of the parameters used to differentiate between two sets are expected to have more influence on the classification defined by the algorithms (Table 2). The importance of each variable was calculated based on the Gini, or impurity index, where a split node that has a mixture of both tolerance and susceptible membership (wet and dry set) is less pure.

## Results

Eleven of the 16 parameters measured differed between the sets. The members of the dry set had 21% fewer fertile tillers, flowered 2.4 days earlier, had longer stomata (4%), greater stomatal area (4%), more stomatal area per unit of leaflet (3%), 48% more leaflet area, 5% higher transpiration rate, 5% higher RWC, and cooler leaves than the wet set. The transpiration rate was 9% higher in the wet set while leaflet and canopy temperatures were lower in the dry set (Table 3). Furthermore, three quarters of the material from the dry set were large-seeded (*major* type) compared to only 20% in the wet set, whose remaining seeds were distributed equally between the *minor* and *equina* classes (Figure S1). The two sets thus contained accessions that, on average, differed morphologically and physiologically.

This assertion is supported by all 3 models used to classify the accessions based on the trait data; the accessions were placed into sets that agreed with the original climate-based classifications. The Kappa scores were all close to one, which demonstrates a high degree of accuracy given that an acceptable score is above 0.4. Likewise the AUC values were well in excess of the acceptable value of 0.7. Thus the models classified the accessions into their climate-based sets with accuracies approaching 100% (Table 4).

**Table 2.** Models used in the study to test the difference between the two sets and to select the best splitters.

Model	Tuning parameters	Library (R language)	References
Classification and Regression Training (CARET)*		caret	[49,50]
Random Forests (RF)*	Number of trees (n.tree) Number of predictors chosen at each node (mtry)	randomForest	[51–53]
Support Vector Machines (SVM)	gamma/sigma, cost (C)	svm (e1071) ksvm (kernalab)	[49,54,55]

\*Variable importance is available for these models.

doi:10.1371/journal.pone.0063107.t002

**Table 3.** The mean ( $\pm$  standard deviation) of morphological, physiological and phenological measurements on sets of 201 wet adapted and 201 dry adapted faba bean accessions, along with the difference between the set means and the value of the t-test.

Parameters	means		difference	t (df=400)
	Wet	Dry		
<b>Morphology</b>				
Stomatal density (No. mm <sup>-2</sup> )	49.6 $\pm$ 12.4	48.3 $\pm$ 6.1	-1.3	1.29 <sup>ns</sup>
Stomatal length ( $\mu$ m)	53.2 $\pm$ 3.9	55.4 $\pm$ 2.6	+2.2	6.63 <sup>***</sup>
Stomatal width ( $\mu$ m)	30.4 $\pm$ 1.61	30.3 $\pm$ 0.92	-0.1	0.20 <sup>ns</sup>
Stomatal area ( $\mu$ m <sup>2</sup> )	1622 $\pm$ 184	1685 $\pm$ 119	+63	4.07 <sup>***</sup>
Stomatal area per unit area of leaflet $\times 10^{-3}$ ( $\mu$ m <sup>2</sup> .mm <sup>-2</sup> )	77.9 $\pm$ 12.4	80.6 $\pm$ 6.8	+2.7	2.67 <sup>**</sup>
Leaflet area (cm <sup>2</sup> )	11.4 $\pm$ 5.9	16.9 $\pm$ 5.2	+5.5	9.86 <sup>***</sup>
Number of fertile tillers	2.85 $\pm$ 1.52	2.25 $\pm$ 1.1	-0.60	4.54 <sup>***</sup>
<b>Physiology</b>				
Photosynthetic rate ( $\mu$ mol m <sup>-2</sup> s <sup>-1</sup> )	7.8 $\pm$ 2.0	8.2 $\pm$ 1.1	+0.4	2.25 <sup>*</sup>
Intercellular CO <sub>2</sub> (ppm)	322.4 $\pm$ 22.3	325.2 $\pm$ 10.5	+2.8	1.64 <sup>ns</sup>
Stomatal conductance (mol m <sup>-2</sup> s <sup>-1</sup> )	0.329 $\pm$ 0.119	0.320 $\pm$ 0.072	-0.009	0.92 <sup>ns</sup>
Water use efficiency ( $\mu$ mol mol <sup>-1</sup> )	25.6 $\pm$ 7.8	26.4 $\pm$ 4.2	+0.8	1.24 <sup>ns</sup>
Transpiration rate (mmol m <sup>-2</sup> s <sup>-1</sup> )	3.26 $\pm$ 0.9	3.42 $\pm$ 0.5	+0.16	2.15 <sup>*</sup>
Leaflet temperature (°C)	-0.17 $\pm$ 0.31	-0.72 $\pm$ 0.19	-0.55	34.98 <sup>***</sup>
Canopy temperature (°C)	-0.80 $\pm$ 1.21	-1.59 $\pm$ 0.48	-0.79	13.77 <sup>***</sup>
Relative water content (%)	82.3 $\pm$ 3.9	86.3 $\pm$ 2.4	+4.0	12.59 <sup>***</sup>
<b>Phenology</b>				
Days to flowering	50.7 $\pm$ 13.1	48.3 $\pm$ 7.4	-2.4	2.24 <sup>*</sup>

df: degrees of freedom; ns: non significant; \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .  
doi:10.1371/journal.pone.0063107.t003

The accuracy of the models is also illustrated by the ROC plots (Figure 2), where displacement above the diagonal indicates non-random assignment of accessions to the correct subset. In the rpart-carec plot, there is some overlap between the sets, but both RF and SVM show mutual exclusivity of the two sets. The prediction

density plots to the right of the ROC plots demonstrate that the wet and dry sets include accessions which, in a multivariate sense, are different and that the basis for the difference will be related to the selection criteria, in this case the seasonal moisture availability at collection sites.

**Table 4.** Model accuracy values for learning-based techniques used on test data (1/3) of faba bean over 10 runs of the algorithms.

Model		AUC	omission rate	sensitivity	specificity	correct classification rate	Kappa
caret-rpart	<b>Mean</b>	<b>0.97</b>	<b>0.03</b>	<b>0.97</b>	<b>0.97</b>	<b>0.97</b>	<b>0.93</b>
	Lower	0.96	0.01	0.95	0.95	0.96	0.92
	Upper	0.98	0.05	0.99	0.98	0.97	0.95
RF	<b>Mean</b>	<b>0.99</b>	<b>0.01</b>	<b>0.99</b>	<b>1.00</b>	<b>0.99</b>	<b>0.99</b>
	Lower	0.99	0.00	0.98	1.00	0.99	0.98
	Upper	1.00	0.02	1.00	1.00	1.00	1.00
SVM	<b>Mean</b>	<b>0.99</b>	<b>0.00</b>	<b>1.00</b>	<b>0.98</b>	<b>0.99</b>	<b>0.98</b>
	Lower	0.99	0.00	1.00	0.97	0.99	0.97
	Upper	1.00	0.00	1.00	0.99	1.00	0.99

AUC: Area under the ROC curve.

RF: Random Forest.

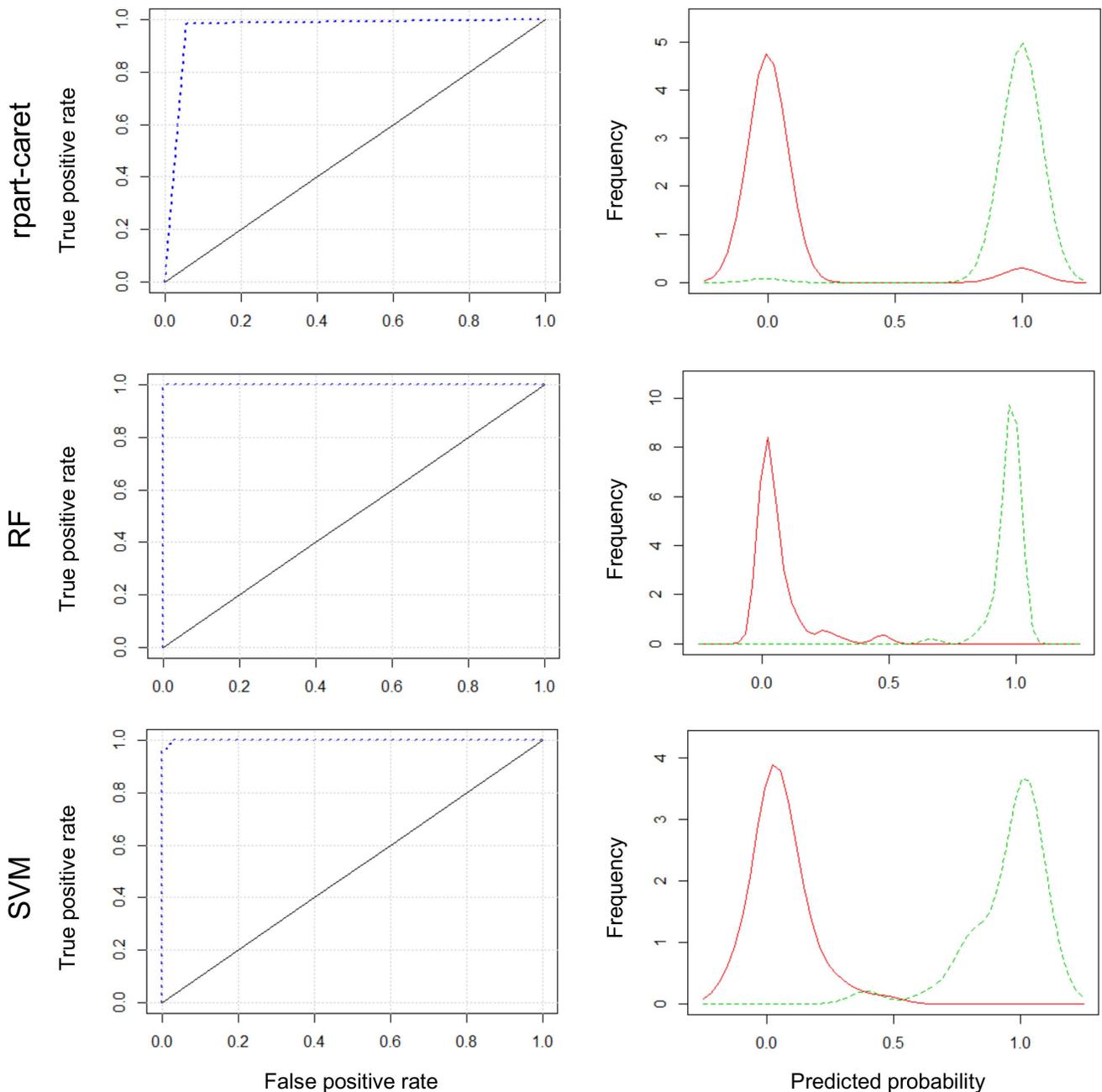
caret-rpart: Classification and Regression Training.

SVM: Support vector machine.

Correct classification rate: the overall classification of both wet and dry accessions to their respective membership group. It is the total of both correctly classified accessions as either wet or dry divided by the total of all the accessions (402).

Omission rates: the opposite of correctly classified accessions with drought-related traits in this case, which is the number of accessions lacking the traits yet they have been classified (incorrectly) as having the.

doi:10.1371/journal.pone.0063107.t004



**Figure 2. ROC plots (left) and density plots class prediction (right) for dry and wet sets using the three models; The class predictions fall out of range (0, 1) as a result of linearity/interpolation in some of the models.**  
doi:10.1371/journal.pone.0063107.g002

### Variable Importance

Of the 16 variables, leaflet temperature depression was the most informative, followed by canopy temperature, RWC, leaflet area and stomatal length (Table 5). The relative importance of the other variables differed between the three assessment methods, with transpiration rate being the third most important in RF mean decrease accuracy and fourth in RF mean decrease Gini, for example.

### Discussion

While other studies have shown that the FIGS approach was effective when employed in the search for resistance to pests and diseases (*e.g.* [34,35,37,38]), this study demonstrates its effectiveness as a method to search for adaptive traits associated with abiotic constraints. The set selection process, based on indicators of moisture availability, yielded sets whose morpho-physiology and phenology were significantly different.

This result is not all that surprising, since it has been comprehensively shown that the environment strongly influences gene flow, natural selection and thus spatial/geographic differen-

**Table 5.** Potential climate predictors based on caret R and RF packages.

Rank	drought related parameter	model		
		rpart-caret	RF	
			mean decrease 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
4	Leaflet area	9.95	0.01	6.39
5	Stomatal length	6.70	0.01	2.30
6	Fertile tillers	4.72	0.00	0.47
7	stomatal area	4.13	0.00	1.01
8	Transpiration rate	3.61	0.03	6.94
9	Stomatal area per unit area of leaflet	2.75	0.01	3.52
10	Photosynthetic rate	2.34	0.01	4.25
11	Days to flowering	2.21	0.00	1.95
12	Intercellular CO <sub>2</sub>	1.64	0.00	0.93
13	Stomatal density	1.26	0.01	2.42
14	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

tiation [62–64]. Numerous studies have documented eco-geographic variation for drought-related traits linked to environmental parameters such as phenology and carbon isotope discrimination in *Triticum turgidum* spp. *dicoccoides* (Körn.) Thell [65], as well as leaf area, electrolyte linkage and RWC in *Arabidopsis thaliana* [66]. In this context, FIGS represents a logical extension of N. I. Vavilov's work that by the 1920s had developed and illustrated the concept of centres of diversity that established the association between diversity and eco-geographic distribution [67].

Despite the above, using an eco-geographic approach to select germplasm for utilization has not been industry standard in genetic resource conservation circles. Rather, there has been a focus on the core collection concept (*e.g.* [21]). In fact, the Food and Agriculture Organization of the United Nations (FAO), in its global strategy for plant genetic resources (PGR) conservation, called for and financially supported the development of core collections as a standard and recommended practice.

However, the authors of this paper have determined that a large percentage of germplasm requests from the ICARDA genetic resources database are for specific adaptive traits. Thus it is argued that, in contrast to core collections, FIGS represents a dynamic, direct and practical approach that focuses on specific adaptive traits rather than on generalized measures of diversity, and as such could be of considerable value to the genebank user community if deployed on a regular basis. It is further suggested that as the plant breeding community prepare to tackle climate change, the efficient utilization of genetic resource collections will become increasingly important [68]. In this context, it is argued that the FIGS approach can reduce the cost and effectiveness of evaluation by reducing the number of accessions screened while providing a higher probability of identifying sought-after traits.

While this study supports the assertion that FIGS is an effective way to search for adaptive traits, there is considerable room for improvement in the approach. Since FIGS is still in its infancy, it is acknowledged that the procedure used to select the sets in this

study was more a common sense process rather one based on previous research. The rationale behind the selection of the dry set was to select material from environments that were most likely to impose relatively dry conditions during the growing season whilst not so dry that a crop would need irrigation. Faba bean is unlikely to be planted to rain fed conditions much below the 300 mm/year limit. Further the criteria on narrow range in rainfall and low aridity index were selected to favour environments where there is more likely to be higher seasonal variation for moisture availability, low rainfall tending to be coupled with high variability. The rationale here is that higher seasonal moisture variation is likely to push populations towards physiological adaptation to dry conditions rather than drought avoidance strategies (earliness, for example). The *bio15* parameter, a measure of seasonal variation in rainfall, was then used to select high variation environments. The *tm1nyr*, *tm1xyr*, *bio4*, *bio16*, and *bio19* parameters were included in the clustering procedure because they all represent factors that influence growth conditions and it was desirable to include a range of different low-moisture environments. The approach outlined above to select the dry set could have been done in different ways and further experimentation is needed to determine the optimal strategy.

Different approaches could also have been used to define the set of material originating in environments with higher seasonal moisture profiles. In this case it was considered desirable to include a wide range of environments provided they received over 800 mm of precipitation, which is considered to be favourable for faba bean cultivation.

Both sets were chosen by applying selection criteria to long-term average yearly data. However, these data do not necessarily reflect the conditions within the growing season. A more effective approach would be to use climatic data presented on the basis of growing season or different crop development phases rather than calendar year. To do this effectively there is a need for accurate continuous surface maps detailing the onset

of the growing season for different crop species. Additionally, the machine-learning algorithms used in this study could be used to create the FIGS sets using climatic variables as the input data.

While this study demonstrates that there is a difference in leaf morphology and physiology associated with water use between the two sets, it was performed under well watered conditions and thus we cannot firmly conclude that the dry set is in fact more drought tolerant. Nevertheless, the existence of a difference indicates that eco-typic differentiation has occurred in faba bean accessions from dryer environments, so we can infer that differentiation is in some way associated with adaptation to dryer seasonal moisture profiles. Indeed, eco-geographic differentiation has been found for leaf morphology in other species. For example leaf area was found to be negatively correlated with altitude (and by inference the probability of chilling stress) for *Dodonaea viscosa* subsp. *angustissima* [69]. It would appear that the same holds true for faba bean, since leaf width in this study was linked to maximum temperature regionally (latitude gradient) and leaf area to minimum temperature locally (altitude gradient).

While leaf area and RWC were positively correlated in *Quercus acutissima* [70], as found in this study ( $R^2 = 0.29$ ,  $P < 0.001$ ,  $n = 402$ ), leaf area and size diminished with declining water availability, in contrast to this study. The present results may be seen as somewhat counter-intuitive if one expects reduced leaf areas to present less evaporative surface, thus favouring tighter control on water use, which is certainly the case in xerophytic perennials. However, large leaf areas cover the soil surface more effectively, minimizing unproductive evaporation. Furthermore, 75% of the dry set accessions belong to the *major* seed type of faba bean (Figure S1) and these larger seeds tend to produce bigger seedlings with larger leaflets and more extensive root systems, which bestow the adaptive advantage of rapidly exploiting available soil moisture earlier in the season. In *Panicum virgatum* L., for example, larger seeds were linked to higher seedling vigour and better root establishment in dry environments [71], while in oat (*Avena sativa* L.) larger seeds lead to better germination under osmotic stress [72], and in faba bean larger seeds were related to higher transpiration efficiency and lower transpiration rates [73]. Furthermore, in some legume species seed size was found to be an indicator of abiotic adaptation [74].

RWC has been recognized as a reliable indicator of plant water status, and thus has been widely used as a screening parameter for drought adaptation in crop plants [47,75]. Nevertheless, screening large quantities of germplasm using RWC measurements is costly and time consuming. Since lower RWC in this study was associated with lower canopy temperatures ( $R^2 = 0.54$ ,  $P < 0.001$ ,  $n = 402$ ), it supports the assertion of Blum [76] that leaf temperature can be used as a rapid and economical phenotyping method to screen germplasm for drought adaptation. The slightly earlier flowering in the dry set is in line with expectations that earlier flowering is part of drought escape in faba bean as in many other species [6].

The current work involved the aerial part of the plant. Nevertheless, for drought adaptation, root morphology and function also play a significant role [75,77]. For example, the roots of sorghum genotypes from dry African environments were found to be deeper and more highly branched than US-derived genotypes [78]. Variation for root traits linked to drought adaptation is of particular interest, especially if they can be linked to more easily evaluated above-ground marker. A logical extension of the work reported here would be to assess differences in root morphology between the two sets.

Many genetic diversity studies still use linear based approaches such as principal component analysis (PCA). The machine learning/recursive algorithms used here represent a novel approach deserving some comment. This study demonstrates that the RF and SMV approaches are suited to studies such as this, since they can detect patterns or relationships between a dependent variable (trait data) and a set of independent variables (climate data) in large datasets [79]. They can also identify parameters that have the greatest impact on the discrimination. Used in this context, the algorithms can point to which trait or combination of traits confers the adaptation.

Further, the use of recursive partitioning is gaining momentum in areas where the data are too highly dimensional for standard regression methods such as PCA in which the decomposition of variables into reduced components leads to the loss of their individual effects, thus rendering the important variable unidentifiable in the interpretation [80]. In the present algorithms, the variables that have a strong relationship to the trait would be those that split the accessions correctly [81]. At the split, the variable that produces less entropy measured using either information theory (Shannon index) or Gini index (known as impurity measure) is ranked first. A reduction in the impurity is a prerequisite for the variable ranking/importance which can be best visualised in the graphs generated by these algorithms [82].

A further advantage of the algorithms used here is that the input data does not have to be normally distributed or conform to other assumptions related to linear models and thus do not require the tedious and time consuming pre-analysis required for linear models to ensure that the assumptions are not violated.

## Conclusions

The methods used were effective at creating sets that were different in terms of leaf morphology, physiology and phenology. This demonstrates that eco-geographic differentiation in faba beans has occurred and is related, in part, to moisture availability. Thus the underlying premise upon which FIGS is based was supported, indicating that it can be an effective tool to enhance the discovery and deployment of new genes, although the FIGS process can be improved to select for drought-adapted genetic resources. Further, the use of machine-learning algorithms was demonstrated here as an effective tool to investigate datasets that are complex and highly dimensional, so it is suggested that they are particularly suited to eco-geographic diversity studies. The results also indicate that leaf and canopy temperature could be an economical way to screen for potentially drought-adapted material as has been suggested by other authors.

## Supporting Information

**Figure S1 Distribution of seed size classes (*minor*, *equina* and *major*) among wet and dry set germplasm.** (TIF)

**Table S1 List of ICARDA accession number (wet and dry sets) used in this study.** (XLSX)

## Acknowledgments

We would like to thank Markku Tykkyläinen, Sini Lindstrom and Sanna Peltola, technical assistants of the glasshouse of the University of Helsinki for their kind assistance during the experiments.

## Author Contributions

Conceived and designed the experiments: HK KS FLS. Performed the experiments: HK. Analyzed the data: AB HK. Contributed reagents/

materials/analysis tools: HK KS AB MM FLS. Wrote the paper: HK KS AB FLS.

## References

- Barnabás B, Jäger K, Fehér A (2008) The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ* 31: 11–38.
- IPCC (2012) Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change [Field CB, Barros V, Stocker TF, Qin D, Dokken DG, et al. editors]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.
- Duc G (1997) Faba bean (*Vicia faba* L.). *Field Crops Res* 53: 99–109.
- Crépon K, Marget P, Peyronnet C, Carrouée B, Arese P, et al. (2010) Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crops Res* 115: 329–339.
- Khan HR, Link W, Hocking TJH, Stoddard FL (2007) Evaluation of physiological traits for improving drought tolerance in faba bean (*Vicia faba* L.). *Plant Soil* 292: 205–217.
- Khan HR, Paull JG, Siddique KHM, Stoddard FL (2010) Faba bean breeding for drought-affected environments: A physiological and agronomic perspective. *Field Crops Res* 115: 279–286.
- Duc G, Link W, Marget P, Redden RJ, Stoddard FL, et al. (2011) Genetic adjustment to changing climates: faba bean. In: Yadav SS, Redden RJ, Hatfield JL, Lotze-Campen H, Hall AE, editors. *Crop adaptation to climate change*, 1st edn. John Wiley & Sons, 269–286.
- Gollin D, Smale M, Skovmand B (2000) Searching an *ex situ* collection of wheat genetic resources. *Am J Agric Econ* 82: 812–827.
- Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Llimense K, Peacock WJ, Starlinger P, editors. *Genetic manipulation: Impact on man and society*, Cambridge University Press, Cambridge, 161–170.
- Hodgkin T, Brown ADH, van Hintum Th JL, Morales EAV (1995) Core collections of plant genetic resources. John Wiley & Sons, Chichester UK, 265 p.
- Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* 31: 818–824.
- Brown AHD (1989) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT, editors. *The use of plant genetic resources*. Cambridge, Cambridge University Press, 136–156.
- Holbrook CC, Anderson WF, Pittman RN (1993) Selection of a core collection from the United-States germplasm collection of peanut. *Crop Sci* 33: 859–861.
- Ortiz R, Ruiz-Tapia EN, Mujica-Sanchez A (1998) Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theor Appl Genet* 96: 475–483.
- Hu J, Zhu J, Xu HM (2000) Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor Appl Genet* 101: 264–268.
- Malosetti M, Abadie T (2001) Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. *Genet Resour Crop Evol* 48: 381–390.
- Casler MD (1995) Patterns of variation in a collection of perennial ryegrass accessions. *Crop Sci* 35: 1169–1177.
- Tohme J, Gonzalez DO, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. *Crop Sci* 36: 1375–1384.
- Bartish GI, Jeppsson N, Bartish IV, Nybom H (2000) Assessment of genetic diversity using RAPD analysis in a germplasm collection of sea buckthorn. *Agric Food Sci* 9: 279–289.
- Fu YB, Peterson GW, Williams D, Richards KW, Fetch JM (2005) Patterns of AFLP variation in a core subset of cultivated hexaploid oat germplasm. *Theor Appl Genet* 111: 530–539.
- Brown AHD, Spillane C (1999) Implementing core collections - principles, procedures, progress, problems and promise. Johnson RC, Hodgkin T, editors. *Core collections for today and tomorrow*. Rome, International Plant Genetic Resources Institute, 1–9.
- Upadhyaya HD, Ortiz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor Appl Genet* 102: 1292–1298.
- McKhann HI, Camilleri C, Bérard A, Bataillon T, David JL, et al. (2004) Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J* 38: 193–202.
- Furman BJ (2006) Methodology to establish a composite collection: case study in lentil. *Plant Genet Resour* 4: 2–12.
- Polignano GB, Uggetti P, Scippa G (2001) Diversity analysis and core collection formation in Bari faba bean germplasm. *Plant Genet Resour Newsl* 125: 33–38.
- Gepts P (2006) Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Sci* 46: 2278–2292.
- Dwivedi SL, Upadhyaya HD, Stalker HT, Blair MW, Bertoli DJ, et al. (2008) Enhancing crop gene pools with beneficial traits using wild relatives. *Plant Breed Rev* 30: 180–230.
- Pessoa-Filho M, Rangel PHN, Ferreira ME (2010) Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. *BMC Plant Biol* 10: 127. doi:10.1186/1471-2229-10-127
- Xu Y (2010) Plant genetic resources: Management, evaluation and enhancement. In: *Molecular plant breeding*. Wallingford, UK, CAB International, 15–194.
- Mackay M (1990) Strategic planning for effective evaluation of plant germplasm. In: Srivastava JP, Damania AB, editors. *Wheat genetic resources: meeting diverse needs*. John Wiley & Sons, Chichester, 21–25.
- Mackay M (1995) One core collection or many? In: Hodgkin T, Brown AHD, Van Hintum TJL, Morales EAV, editors. *Core collections of plant genetic resources*. John Wiley & Sons Ltd., Chichester, 199–210.
- Mackay M, Street K (2004) Focused identification of germplasm strategy – FIGS. In: Proceedings of the 54th Australian Cereal Chemistry Conference and the 11th Wheat Breeders' Assembly, , editors. Cereal Chemistry Division, Royal Australian Chemical Institute (RACI), Melbourne, Victoria, Australia. 138–141.
- Mackay M, von Bothmer R, Skovmand B (2005) Conservation and utilization of plant genetic resources – future directions. *Czech J Genet Plant Breed* 41: 335–344.
- Endresen DTF, Street K, Mackay M, Bari A, De Pauw E (2011) Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces. *Crop Sci* 51: 2036–2055.
- Bari A, Street K, Mackay M, Endresen DTF, De Pauw E, et al. (2012) Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genet Resour Crop Evol* 59: 1465–1481.
- Kaur N, Street K, Mackay M, Yahiaoui N, Keller B (2008) Molecular approaches for characterization and use of natural disease resistance in wheat. *Eur J Plant Pathol* 121: 387–397.
- Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus. *Proc Natl Acad Sci USA* 106: 9519–9524.
- El Bouhssini M, Street K, Joubi A, Ibrahim Z, Rihawi F (2009) Sources of wheat resistance to sunn pest, *Eurygaster integriceps* Puton, in Syria. *Genet Resour Crop Evol* 56: 1065–1069.
- El Bouhssini M, Street K, Amri A, Mackay M, Ogbonnaya FC, et al. (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 Breed* 130: 96–97.
- Endresen DTF (2010) Predictive association between trait data and ecogeographic data for Nordic barley landraces. *Crop Sci* 50: 2418–2430.
- FAO (2010) The second report on the state of the world's plant genetic resources for food and agriculture. Rome, Italy, 398 p.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25: 1965–1978.
- Wang H, Clarke JM (1993) Genotypic, intraplant, and environmental variation in stomatal frequency and size in wheat. *Can J Plant Sci* 73: 671–678.
- Muratova V (1991) Common Beans (*Vicia faba*). *Bull Appl Bot Genet Plant Breed* 50: 1–298.
- Cubero J (1974) On the evolution of *Vicia faba* L. *Theor Appl Genet* 45: 47–51.
- Fischer RA, Rees D, Sayre KD, Lu Z-M, Condon AG, et al. (1998) Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. *Crop Sci* 38: 1467–1475.
- Barrs HD, Weatherley PE (1962) A re-examination of the relative turgidity technique for estimating water deficit in leaves. *Aust J Biol Sci* 15: 413–428.
- R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available: <http://www.R-project.org>.
- Kuhn M (2008) Building predictive models in R using the caret package. *J Stat Softw* 28 (5).
- Steinberg D (2009) CART: Classification and Regression Trees. Taylor & Francis Group, LLC.
- Breiman L (2001) Random forests. *Mach Learn* 45: 5–32.
- Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, et al. (2007) Random forests for classification in ecology. *Ecology* 88: 2783–2792.
- Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9: 181–199.
- Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A (2010) R library (e1071). The R foundation for statistical computing. ISBN: 3-900051-07-0.
- Karatzoglou A, Meyer D, Hornik K (2006) Support vector machines in R. *J Stat Softw* 15 (9).
- Tirelli T, Pozzi L, Pessani D (2009) Use of different approaches to model presence/absence of *Salmo marmoratus* in Piedmont (Northwestern Italy). *Ecol Inform* 4: 234–242.

57. Drake JM, Randin C, Guisan A (2006) Modelling ecological niches with support vector machines. *J Appl Ecol* 43: 424–432.
58. Swets JA, Dawes RM, Monahan J (2000) Better decisions through science. *Sci Am* 283: 82–87.
59. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27: 861–874.
60. Freeman EA, Moisen GG (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecol Modell* 217: 48–58.
61. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
62. Lin W, Bradshaw AD, Thurman DA (1975) The potential for evolution of heavy metal tolerance in plants. III. The rapid evolution of copper tolerance in *Agrostis stolonifera*. *Heredity* 34: 165–187.
63. Spieth PT (1979) Environmental heterogeneity: a problem of contradictory selection pressures, gene flow, and local polymorphism. *Am Nat* 113: 247–260.
64. Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics* 124: 757–771.
65. Peleg Z, Fahima T, Abbo S, Krugman T, Nevo E, et al. (2005) Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ* 28: 176–191.
66. Bouchabke O, Chang F, Simon M, Voisin R, Pelletire G, et al. (2008) Natural variation in *Arabidopsis thaliana* as a tool for highlighting differential drought response. *PLoS ONE* 3: e1705. doi:10.1371/journal.pone.0001705
67. Vavilov NI (1926) Tzentry proiskhozheniya kulturnykh rastenii. (Studies on the origin of cultivated plants). *Trudy Byuro prikl Bot [in Russian]* 16: 139–248.
68. Mba C, Guimaraes EP, Ghosh K (2012) Re-orienting crop improvement for the changing climate conditions for the 21st century. *Agric Food Secur* 1: 7. doi:10.1186/2048-7010-1-7
69. Guerin GR, Wen H, Lowe AJ (2012) Leaf morphology shift linked to climate change. *Biol Lett* 8: 882–886.
70. Xu F, Guo W, Xu W, Wei Y, Wang R (2009) Leaf morphology correlates with water and light availability: what consequences for simple and compound leaves? *Prog Nat Sci* 19: 1789–1798.
71. Fan JW, Du YL, Turner NC, Li FM, He J (2012) Germination characteristics and seedling emergence of switchgrass with different agricultural practices under arid conditions in China. *Crop Sci* 52: 2341–2350.
72. Mut Z, Akay H (2010) Effect of seed size and drought stress on germination and seedling growth of naked oat (*Avena sativa* L.). *Bulg J Agric Sci* 16: 459–467.
73. Avola G, Cavallaro V, Patanè C, Riggi E (2008) Gas exchange and photosynthetic water use efficiency in response to light, CO<sub>2</sub> concentration and temperature in *Vicia faba*. *J Plant Physiol* 165: 796–804.
74. Parra-Quijano M, Iriando JM, Torres E (2012) Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. *Genet Resour Crop Evol* 59: 205–217.
75. Blum A (2011) Plant breeding for water limited environments. Springer-Verlag, New York, 255 p.
76. Blum A (2009) Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Res* 112: 119–123.
77. Blum A (2011) Drought resistance – is it really a complex trait? *Funct Plant Biol* 38: 753–757.
78. Masi CEA, Maranville JW (1998) Evaluation of sorghum root branching using fractals. *J Agric Sci* 131: 259–265.
79. Therneau TM, Atkinson EJ (1997) An introduction to recursive partitioning using the rpart routine. Technical Report 61, Section of Biostatistics, Mayo Clinic, Rochester. Available: <http://www.mayo.edu/hsr/techrpt/61.pdf>.
80. Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14: 323–348.
81. Liknes GC, Woodall CW, Perry CH (2009) Predicting forest attributes from climate data using a recursive partitioning and regression tree algorithm. In: McWilliams W, Moisen G, Czaplewski R, editors. *Forest Inventory and Analysis (FIA) Symposium 2008*; Park City, UT. Proc. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 7 p.
82. Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: A conditional inference framework. *J Comput Graph Stat* 15: 651–674.