CIHEAM
Options Méditerranéennes

Searching for climate change related traits in plant genetic resources collections using Focused Identification of Germplasm Strategy (FIGS)

Bari A., Street K., Amri A., Nachit M., Mackay M., Ouabbou H., Kahel Z., Ghanem M. E., De Pauw E., Nazari K., Alo F., El Bouhssini M., Tsivelikas A., Humeid B.

in

Porceddu E. (ed.), Damania A.B. (ed.), Qualset C.O. (ed.).
Proceedings of the International Symposium on Genetics and breeding of durum wheat

Bari : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110

2014
pages 87-94

To cite this article / Pour citer cet article

Bari A., Street K., Amri A., Nachit M., Mackay M., Ouabbou H., Kahel Z., Ghanem M. E., De Pauw E., Nazari K., Alo F., El Bouhssini M., Tsivelikas A., Humeid B. **Searching for climate change related traits in plant genetic resources collections using Focused Identification of Germplasm Strategy (FIGS).** In : Porceddu E. (ed.), Damania A.B. (ed.), Qualset C.O. (ed.). *Proceedings of the International Symposium on Genetics and breeding of durum wheat.* Bari : CIHEAM, 2014. p. 87-94 (Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110)

Searching for climate change related traits in plant genetic resources collections using Focused Identification of Germplasm Strategy (FIGS)

Abdallah Bari[1], Kenneth Street[1], Ahmed Amri[1], Miloudi M. Nachit[1], Michael Mackay[2], Hassan Ouabbou[3], Zakaria Kehel[4], Michel E. Ghanem[1], Eddy De Pauw[1], Kumarse Nazari[1], Fida Alo[1], Mustapha El Bouhssini[1], Athanasios Tsivelikas[1], Bilal Humeid[1]

[1] International Centre for Agricultural Research in the Dry Areas (ICARDA), Morocco
[2] Queensland Alliance for Agricultural and Food Innovation (QAAFI), Australia
[3] Institut National de la Recherche Agronomique (INRA), Morocco
[4] Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico

---

Abstract. Prospects to assess and explore largely untapped plant genetic resources (PGR) collections to search for climate change related traits, such as drought and heat tolerance, as well as pest and disease resistance, are possible through new approaches such as the focused identification of germplasm strategy (FIGS). FIGS approach is based on the paradigm that any germplasm is likely to reflect the selection pressures of the environment under which it evolved. The approach uses trait and environmental data (climate data including phenology data) to develop *a priori* information based on the quantification of the trait-environment relationship. If a dependency between the trait and the environment is detected, the *a priori* information is then used to define subsets of accessions with a high probability of containing the sought after traits. The subsets of accessions are then used for *a posteriori* evaluation. Recent research comparing *a priori* and *a posteriori* information supports the assertion that FIGS can be used as an effective tool to search for traits of resistance to pests and diseases as well as traits to adapt to climate change. This paper presents and discusses some of the recent results where FIGS was used to develop subsets with high probability of finding desirable traits, such as resistance to stripe (yellow) rust, in durum wheat. It also addresses ways in which current FIGS based models could be further enhanced by working the ways in which the environmental data is presented to the models, thereby improving the detection of traits associated with climate change adaptation.

Keywords. Genetic resources – FIGS – Accessions – Pests – Diseases – Resistance – Climate change.

Recherche pour des caractères liés au changement climatique dans des collections de ressources phytogénétiques en utilisant la stratégie d'identification ciblée du matériel génétique (FIGS)

Résumé. *L'évaluation et l'utilisation des collections de ressources phytogénétiques largement inexploitées pour rechercher des caractères liés au changement climatique, comme la sécheresse et la tolérance à la chaleur, ainsi que la résistance aux organismes nuisibles et aux maladies, sont aujourd'hui possibles grâce à de nouvelles approches telles la stratégie d'identification ciblée du matériel génétique (FIGS). L'approche FIGS repose sur le paradigme que tout matériel génétique est susceptible de refléter les pressions de sélection de l'environnement dans lequel il a évolué. Cette stratégie utilise des caractères et des données environnementales (données climatiques, y compris les données phénologiques) pour développer une information a priori basée sur la quantification de la relation caractère-environnement. Si une dépendance entre le caractère et l'environnement est détectée, l'information a priori est alors utilisée pour définir des sous-ensembles d'accessions ayant une forte probabilité de porter les caractères cherchés. Les sous-ensembles d'accessions sont ensuite utilisés pour une évaluation a posteriori. Des recherches récentes comparant les informations a priori et a posteriori permettent d'affirmer que la FIGS peut être utilisée comme un outil efficace pour la recherche de caractères de résistance aux organismes nuisibles et aux maladies tout comme aux caractères d'adaptation au changement climatique. Dans cet article, on présente et on discute des résultats récents de l'application de la FIGS pour développer des sous-ensembles avec une haute probabilité de trouver les caractères cherchés, comme la résistance à la rouille jaune, chez le blé dur. On discute également les possibilités d'améliorer les modèles sur lesquels est basée actuellement la FIGS, en travaillant sur la façon dont les données environnementales sont intégrées dans les modèles, améliorant ainsi la détection des caractères associés à l'adaptation au changement climatique.*

Mots-clés. Ressources génétiques – FIGS – Accessions – Organismes nuisibles – Maladies – Résistance – Changement climatique.

I – Introduction

Prospects to assess and explore largely untapped plant genetic resources (PGR) collections for agronomically important traits, particularly those linked to climate change-adaptation, are possible through new approaches such as the Focused Identification of Germplasm Strategy (FIGS). Climate change, which is the result of greenhouse gas (GHG) emissions, is causing the atmosphere to heat up (Mendelsohn & Dinar 2009). Crops such as wheat are reported to be more vulnerable to heat stress than drought (Semenov & Shewry 2011). High temperatures during the reproductive phase can reduce the number of kernels per spike, which is an important component of yield (Semenov & Shewry 2011). Both heat and drought stresses are expected to increase in their frequency and intensity in dry areas (IPCC 2012) such as central North America, Northern Africa, Central Asia, West Asia and Western Australia. Although the global climate models (GCMs) differ substantially, they all tend to indicate significant temperature increases in these areas (Girvetz *et al*. 2009). Further, this increase in temperature as a result of GHG emissions is expected to increase depending on emissions scenarios and the extent of mitigation implemented measures to curb their effects (Howden *et al.* 2007, Mendelsohn and Dinar 2009).

Plant genetic resources have contributed enormously towards increased yield in crops (Hoisington 1999) and are a ready source of trait's variation (Qualset 1975). For example, a wheat landrace from Turkey that was conserved in a genebank in 1948 was later discovered, (in the 1980s) to carry genes that are resistant to a range of fungal diseases, and are still in use in current breeding programs (Atalan-Helicke 2012, FAO). However, searching for such traits can be a daunting and costly process given that PGR consists of large collections and populations maintained *in situ* or *on-farm* that are also more prone to yield climate change related traits but yet to sampled and collected. What is required therefore is an efficient method to select material from these genetic resources so that the probability of finding and locating the required variation is maximized while reducing the number of accessions evaluated and the onverall cost implications (Gollin *et al*. 2000). The FIGS approach represents one such method.

The FIGS approach is based on the paradigm that adaptive traits exhibited by germplasm are likely to reflect the selection pressures of the environment from which the germplasm was originally sampled (Mackay and Street, 2004). For example, if a plant population is exposed over a significant period of time to weather conditions that are favourable to consistently high pathogen populations then it is likely that a selection pressure will be imposed on the plant population for the emergence of resistance genes. Paillard *et al.* (2000) found this to be the case for the evolution of powdery mildew resistance in wheat and barley landraces. Thus if a dependency between a given trait and environmental parameters can be defined then the relationship can be used to predict the likelihood of finding a desired trait in a given environment (Mackay and Street 2004, Bari *et al.* 2012, Endresen *et al.* 2012). In this context information about the environmental origins of accessions are used to define trait specific subsets of germplasm with a higher probability of containing the sought-after traits.

This paper presents and discusses how FIGS has been applied to the search for resistance to stripe (yellow) rust in durum wheat. In previous FIGS studies predictive models were applied to historic climate data to search for traits of interest. In this study the models were tested with future climate change scenarios. However, adjustments may be required in the models for change climate scenarios as well as improvement by working ways in which environmental data is presented to the models to improve the search for traits to cope with climate change adaptation. The modelling process is considering separating the induced-shift climate change variation from the overall variation for better prediction.

II – Methodology

1. Data

This study was based on a field evaluation of durum wheat accessions for response to a naturally occurring yellow rust infection at ICARDA during the 2011/2012 season. The environmental data consisted of long-term climate monthly average data for the sites from which the accessions were originally sampled. The study also consists of projected climatic data extracted from three future climate scenarios based on the Canadian Climate Centre (CCC) global circulation model (Boer *et al*. 2000).

All the climate variables were extracted from a grid cell of 1 square km (Table 1) as monthly data (De Pauw 2008). Monthly data are coarse grained and thus more prone to be out of phase in relation to critical stages of crop development (Coops *et al*. 2001), which would be further amplified by climate change effects. Thus the study also used daily data which were derived from the monthly values using models proposed by Epstein (1991) (Hofstra *et al*. 2008).

To better capture the climate change induced-shifts the predictive models were applied to climatic conditions within the growing period. Thus data averages for stages in a crops development where compared to long term climatic averages expressed as monthly values alone. Thus in the modelling process the noise created by differences in phenology between sites and climate change induced-shifts would be eliminated facilitating higher resolutions to detect environment – trait linkages.

To estimate the crop development phases a day-degree accumulation model was used from an estimated onset date for each site. The onset date was estimated using a method which determines when neither moisture nor temperature would limit plant growth. The method is based on a modification of a model developed by the Food and Agriculture Organization of the United Nations (FAO 1978, De Pauw 1982).

Table 1. The environment variables used in the study.

Variable Type	Variable Name	Variable Description	Unit	Number
Climatic	*Tmin*	Monthly minimum temperature	℃	12
	tmax	Monthly maximum temperature	℃	12
	prec	Monthly precipitation	mm	12
	tmind	Daily minimum temperature	℃	365
	tmaxd	Daily maximum temperature	℃	365
	precd	Daily precipitation	mm	365
Phenology	*Onset*	Date of sowing	day	1

All variables were standardized to a mean of zero and a standard deviation of 1. After the transformation the data was standardized and a comparison made between the transformed and non-transformed data. This data pre-processing was systematically and automatically carried out through the different models.

2. Modelling

In previous models the predictions were limited to past climate data while here the modelling was also carried out on projected future scenarios. The stripe (yellow) rust disease evaluation scores of the growing season 2011/2012 were presented to the models to detect the trait by collection site environment dependency, if it exists. The models were then run on all the durum collection data held at ICARDA using current climate set of data as well as 2 other sets of future climate data/maps reflecting two emission scenarios: *a* and *b* scenarios. The *b* scenario projects a doubling of CO2 relative to its preindustrial level (Franklin *et al*. 2013).

The modelling procedure was based on running two models using the current climate variables and then re-run the models using CC variables. The models used are SVM (Support Vector Machine) and RF (Random Forest). RF is a clustering algorithm developed to act like an ensemble classifier where the best splitters are randomly selected at each node among subset predictors ((Breiman 2001; Liaw and Wiener 2002). It is a procedure used in gene selection and classification of microarray data and genome-wide association studies for complex human diseases (Díaz-Uriarte and Alvarez de Andrés 2006; Lunetta *et al*. 2004). Support Vector Machines (SVM), on the other hand, maps input data to a more high-dimensional space that would lead to a better separation of data into respective classes by isolating those inputs which fall nearby the data boundaries (Cortes and Vapnik 1995; Principe *et al*. 2000). The mapping of input data to high-dimensional space is carried out through processes called kernel functions such as radial basis function (RBF) which is the kernel function used in this study for the SVM model. SVM models have been found to distinguish optimally between groups with minimum loss of information (Guo *et al*. 2004; Karatzoglou *et al*. 2006).

The predicted probabilities were then used to delimit areas where the conditions are conducive to occur. After defining the appropriate variograms, the maps were generated using kriging techniques (Cressie 1993). To create maps the R module was applied to irregularly spaced data (Figure 2) where the correlation between sites is (assumed) to be an exponential function of the distance.

III – Results

The results show the presence of relationship between the current or past climate data and the resistance to stripe rust. Both Receiver Operating Characteristics (ROC) values as well as Kappa values are all highly above acceptable values of 0.5 and 0.4 respectively. The ROC plots illustrate also that the curve for the two models were well above the diagonal line, which is expected when the model is different from random. The vertical trend towards the left-hand side is also an indication that the models classified the resistant accessions more correctly with fewer false positive errors (Fawcett 2006). The histograms (right side) illustrate further the extent of separation between the two trait states, resistant on one hand and tolerant on the other, with limited overlapping between the two states. The models were also able to correctly classify sites that yield either resistant or susceptible genotypes with a high correct classification when compared to the previous studies (Table 2). The accuracy of prediction as well as kappa increase reaching up to 0.83 and 0.70 respectively as we move from monthly data to aligned daily data based on onset data (Table 4).

Table 2. Accuracy and agreement parameters of daily two fit functions to generate daily data.

fit function	Stat	AUC	OR	SE	SU	CC	Kappa
Spline	mean	**0.80**	**0.32**	**0.68**	**0.92**	**0.87**	**0.61**
	upper CI	0.81	0.35	0.70	0.93	0.87	0.63
	lower CI	0.79	0.30	0.65	0.91	0.86	0.59
Loess	mean	**0.79**	**0.34**	**0.66**	**0.92**	**0.86**	**0.59**
	upper CI	0.80	0.37	0.68	0.92	0.86	0.60
	lower CI	0.78	0.32	0.63	0.91	0.85	0.57

Table 3. Accuracy and agreement parameters of daily data (spline fit function) and monthly data.

Data type		AUC	OR	SE	SU	CC	Kappa
daily data	**Mean**	**0.80**	**0.33**	**0.67**	**0.93**	**0.87**	**0.62**
	Upper CI	0.81	0.34	0.69	0.93	0.87	0.63
	Lower CI	0.79	0.31	0.66	0.92	0.87	0.61
monthly data	**Mean**	**0.79**	**0.31**	**0.69**	**0.90**	**0.85**	**0.58**
	Upper CI	0.80	0.33	0.71	0.90	0.86	0.60
	Lower CI	0.78	0.29	0.67	0.89	0.84	0.56

Table 4. Accuracy and agreement parameters of aligned data.

Data type		AUC	OR	SE	SU	CC	Kappa	
monthly	Max	0.81	0.28	0.72	0.90	0.86	0.61	
daily data	Max	0.82	0.30	0.70	0.93	0.88	0.64	
aligned daily data	Max	0.83	0.28	0.72	0.95	0.90	0.70	210 days

The presence of the existence of the dependency between climate data and the trait of resistance to stripe rust was used as a *priori* information for the prediction of stripe resistance in independent data. The results are shown in the maps for different CC scenarios. In terms of areas that might yield stem rust variation, Ethiopia was highest followed by India and Turkey. This is also similar to the results that have been reported on the regions that might yield resistance (Singh *et al*. 2006).

IV – Discussion

Recent findings on a study conducted to search for climate change traits such as traits of tolerance of drought where a comparison was made between *a priori* and *a posteriori* information supports the assertion that FIGS can be used as an effective tool to search for traits of adaptation to climate change (Khazaei *et al.* 2013). Similar comparison was also made recently for stripe rust resistance in durum wheat confirming also that FIGS is tool with potential to not only find the sought after traits but on a limited number of accessions (Bari *et al*. *in press*).

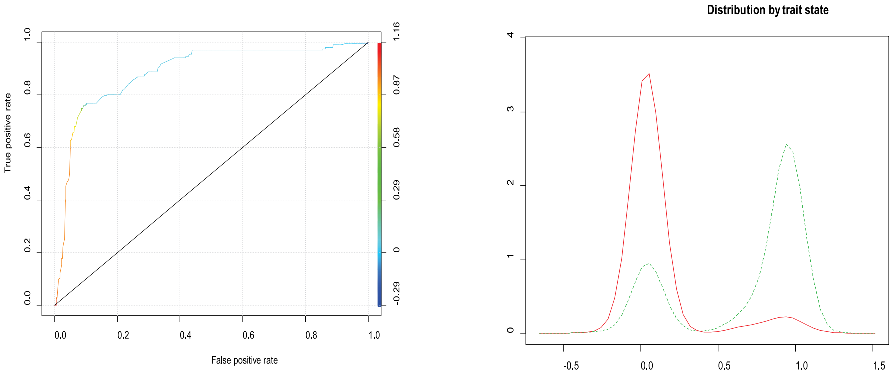

Figure 1. ROC plots for the RF and SVM models applied to the training set of accessions evaluated for Yr disease in 2011/2012 growing season at ICARDA (Left hand side). The ROC curve to the left of the diagonal plot is the true positive rate versus false positive rate. Density plots for prediction of resistance and susceptibility for the RF and SVM models (Right hand side).
[Green line indicates the probability density distribution for resistance and red line indicates susceptibility]

Acknowledgments

Our extended special thanks to Climate Change, Agriculture and Food Security (CCAFS) program and the Australian Grain Research & Development Cooperation (GRDC) for their valuable and continuous support. We thank also all colleagues involved in the newly established genetic resources and climate change platform to bring forward the tremendous potential of genetic resources not only to adapt but also to mitigate climate change.

References

Atalan-Helicke N., 2012. Conserving diversity at the dinner table: plants, food security and gene banks. Origins: *Current Events in Historical Perspective* The Ohio State University, USA, 5 (4).

Bari A., Amri A., Street K., Mackay M., De Pauw E., Sanders R., Nazari K., Humeid B., Konopka K., Alo F., 2014. Predicting resistance to stripe (yellow) rust in plant genetic resources using Focused Identification of Germplasm Strategy (FIGS). *The J. Agricultural Sci.,*152(6), pp. 906-916.

Bari A., Street K., Mackay M., Endresen D.T.F., De 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, pp. 1465-1481.

Boer G.J., Flato G.M., Reader M.C., Ramsden D., 2000a. A transient climate change simulation with historical and projected greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the 20th century. *Climate Dynamics*, 16, pp. 405-425.

Boer G.J., Flato G.M., Ramsden D., 2000b. A transient climate change simulation with historical and projected greenhouse gas and aerosol forcing: projected climate for the 21st century. *Climate Dynamics*, 16, pp. 427-450.

Breiman L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5–32.

Coops N.C., Loughhead A., Ryan P., Hutton R., 2001. Development of daily spatial heat unit mapping from monthly climatic surfaces for the Australian continent. *Internat. J. Geogr. Inform. Sci.,* 15(4), pp. 345-361.

Cressie N.A.C., 1993. Statistics for Spatial Data (Revised Edition). Wiley: New York, pp. 928.

Cortes C., Vapnik V., 1995. Support-Vector Networks. *Machine Learning*, 20(3), pp. 273-297.

De Pauw E., 1982. The concept of dependable growing period and its modelling as a tool for land evaluation and agricultural planning in the wet and dry tropics. *Pédologie*, 1982(3), pp. 329-348.

De Pauw E., 2008. Climatic and soil datasets for the ICARDA wheat genetic resource collections of the Eurasia region: Explanatory notes. *Technical Note, ICARDA GIS Unit*, Aleppo, Syria.

Díaz-Uriarte E., Alvarez de Andrés S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics,* 7, pp. 3.

Endresen D.T.F., Street K., Mackay M., Bari 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 (FIGS). *Crop Sci.,* 52(2), pp. 764-773.

Epstein E.S., 1991. On obtaining daily climatological values from monthly means. *J. Climate*, 4, pp. 365-368.

Fawcett T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, 27, pp. 861–874.

Franklin J., Davis F.W., Ikegami M., Syphard A.D., Flint L.E., Flint A.L., Hannah L., 2013. Modelling plant species distributions under future climates: how fine scale do climate projections need to be?. *Global Change Biology*, 19, pp. 473–483.

FAO, 2013. Plant genetic resources: use them or lose them. http://www.fao.org/docrep/012/ al384e/al384e00.pdf.

FAO., 1978. Report on the Agro-ecological zones project. Vol. I. Methodology and results for Africa.World Soil Resources Report, FAO, 48, pp. 158.

Gollin D., Smale M., Skovmand B., 2000. Searching an *ex situ* collection of wheat genetic resources. *American J. Agricultural Economics,* 82 (4), pp. 812–27.

Girvetz E.H., Zganjar C., Raber G.T., Maurer E.P., Kareiva P., Lawler J.J., 2009 Applied Climate-Change Analysis: The Climate Wizard Tool. *PLoS ONE*, 4(12), pp. e8320.

Guo Q., Kelly M., Graham C.H., 2004. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling*, 182(1), pp. 75-90.

Hoisington D., Khairallah M., Reeves T., Ribaut J.M., Skovmand B., Taba S., Warburton M., 1999. Plant genetic resources: What can they contribute toward increased crop productivity? *Proc. Natl. Acad. Sci. USA,* 96, pp. 5937–5943.

Howden S.M., Soussana J.F., Tubiello F.N.,Chhetri N., Dunlop M., Meinke H., 2007. Adapting agriculture to climate change. *Proc. Natl. Acad. Sci. USA,* 104, pp. 19691-19696.

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 C.B. *et al*. (eds)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 582

Jenouvrier S., Visser M.E., 2011. Climate change, phenological shifts, eco-evolutionary responses and population viability: toward a unifying predictive approach. *Int. Journal of Biometeorology*, 55, pp.: 905–919.

Karatzoglou A., Meyer D., Hornik K., 2006. Support Vector Machines in R. *J. Statistical Software* 15(9), pp. 1-28.

Khazaei H., Street K., Bari A., Mackay M., Stoddard F.L., 2013. The FIGS (Focused Identification of Germplasm Strategy) Approach Identifies Traits Related to Drought Adaptation in *Vicia faba* Genetic Resources. *PLoS ONE* 8(5), pp. e63107.

Liaw A., Wiener M., 2002. Classification and Regression by random Forest. *R. News,* 2(3), pp. 18-22.

Lunetta K.L., Hayward L.B., Segal J. van Eerdewegh P., 2004. Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet.,* 5(1), pp. 32.

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,* 21st to 24th September 2004, Canberra, ACT, Australia. Black C.K. *et al*. (eds). Cereal Chemestry Division, Royal Australian Chemical Institute (RACI), Melbourne, Australia, pp. 138–141.

Mendelsohn R., Dinar A., 2009. Climate Change and Agriculture: An Economic Analysis of Global Impacts, Adaptation, and Distributional Effects. Edward Elgar Publishing, England.

Hofstra N., Haylock M., New M., Jones P., Frei C., 2008. Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research*, 113(D21110), doi:10.1029/2008 JD010100.

Olfert O.O., Weiss R.M., Kriticos D.J., 2011. Application of General Circulation Models to Assess the Potential Impact of Climate Change on Potential Distribution and Relative Abundance of *Melanoplus sanguinipes* (Fabricius) (Orthoptera: Acrididae) in North America. *Psyche*, ID 980372).

Paillard S., Goldringer I., Enjalbert J., Trottet M., David J., de Vallavieille-Pope C., Brabant P., 2000. Evolution of resistance against powdery mildew in winter wheat populations conducted under dynamic management. II. Adult plant resistance. *Theor. Appl. Genetics,* 101, pp. 457-462.

Patterson D.T., Westbrook J.K., Joyce R.J.V., Lingren P.D., Rogasik J., 1999. Weeds, insects, and diseases; *Climatic Change,* 43(4), pp. 711-727.

Qualset C.O., 1975. Sampling germplasm in a center of diversity: an example of disease resistance in Ethiopian Barley. In: *Crop genetic resources today and tomorrow*, Frankel H., Hawkes J.D (eds), Cambridge University Press, Cambridge, pp. 81-96.

Semenov M.A., Shewry P.R., 2011. Modelling predicts that heat stress and not drought will limit wheat yield in Europe. *Sci. Rep*., 1, pp. 66.