CORDOBA UNIVERSITY
ETSIAM, INGENERIA GRAFICA y GEOMATICA
&
ICARDA DURUM BREEDING PROGRAM

Ph.D. Thesis

USE OF GEOGRAPHIC INFORMATION SYSTEMS (GIS), MORPHO-PHYSIOLOGY AND MOLECULAR MARKER TOOLS TO MODEL ADAPTATION MECHANISMS OF DURUM IN DIFFERENT ENVIRONMENTS IN MOROCCO & SYRIA

Presented by:
Zakaria KEHEL

Supervisors:
Dr. Miloudi M. Nachit
Dr. Alfonso GARCIA-FERRER PORRAS

CORDOBA (SPAIN), ----- 2015

TITULO: *USE OF GEOGRAPHIC INFORMATION SYSTEMS (GIS), MORPHO-PHYSIOLOGY AND MOLECULAR MARKER TOOLS TO MODEL ADAPTATION MECHANISMS OF DURUM IN DIFFERENT ENVIRONMENTS IN MOROCCO & SYRIA*

AUTOR: *Kehel Zakaria*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

www.uco.es/publicaciones
publicaciones@uco.es

CORDOBA UNIVERSITY
ETSIAM, INGENERIA GRAFICA y GEOMATICA
&
ICARDA DURUM BREEDING PROGRAM

Ph.D. Thesis

USE OF GEOGRAPHIC INFORMATION SYSTEMS (GIS), MORPHO-PHYSIOLOGY AND MOLECULAR MARKER TOOLS TO MODEL ADAPTATION MECHANISMS OF DURUM IN DIFFERENT ENVIRONMENTS IN MOROCCO & SYRIA

Presented by:
Zakaria KEHEL

Supervisors:
Dr. Miloudi M. Nachit
Dr. Alfonso Garcia-Ferrer Porras

CORDOBA (SPAIN), ----- 2015

Los Drs. **Miloudi M. Nachit,** Investigador del ICARDA, **Alfonso García-Ferrer Porras**, Profesor del departamento Ingeniería Geográfica y Geomática de la UCO,

INFORMAN:

Que el trabajo titulado **"Use of Geographic Information Systems (GIS), morpho-physiology and molecular marker tools to model adaptation mechanisms of durum in different environments in Morocco and Syria"**, realizado por D. Zakaria KEHEL, bajo su dirección, se considera ya finalizado y puede ser presentado para su exposición y defensa como Tesis Doctoral en el Departamento de Ingeniería Geográfica y Geomática de la Universidad de Córdoba.

M. M. Nachit

Rabat, de 2015

A. P. Garcia-Ferrer

Cordoba, de 2015

Los Drs. **Miloudi M. Nachit,** Investigador del ICARDA, **Alfonso García-Ferrer Porras**, Profesor del departamento Ingeniería Geográfica y Geomática de la UCO,

INFORMAN:

Que el trabajo titulado **"Use of Geographic Information Systems (GIS), morpho-physiology and molecular marker tools to model adaptation mechanisms of durum in different environments in Morocco and Syria"**, realizado por D. Zakaria KEHEL, bajo su dirección, se considera ya finalizado y puede ser presentado para su exposición y defensa como Tesis Doctoral en el Departamento de Ingeniería Geográfica y Geomática de la Universidad de Córdoba.

Nachit 29/04/2015

M. M. Nachit

Rabat, de 2015

A. P. Garcia-Ferrer

Cordoba, de 2015

1

TÍTULO DE LA TESIS:
USE OF GEOGRAPHIC INFORMATION SYSTEMS (GIS), MORPHO-PHYSIOLOGY AND MOLECULAR MARKER TOOLS TO MODEL ADAPTATION MECHANISMS OF DURUM IN DIFFERENT ENVIRONMENTS IN MOROCCO & SYRIA

DOCTORANDO/A:
KEHEL, ZAKARIA

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS
(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

El proyecto de investigación realizado, dentro del programa de mejora del ICARDA, supone un importante aporte en el estudio de adaptación de las distintas variedades de trigo duro en la cuenca mediterránea.
Se ha analizado la diversidad genética y su adaptación a distintas condiciones ambientales estudiando las barreras fenotípicas para optimizar la producción de las mismas.
Se han determinado las regiones mas adecuadas a los distintos patrones.

Un primer resultado de la investigación se publicó en el American Journal of Molecular Biology en 2012. "Using Bayesian and Eigen approaches to study spatial genetic structure of Moroccan and Syrian durum wheat landraces" Zakaria Kehel, Alfonso García-Ferrer, Miloudi M. Nachit

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 20 de mayo de 2015

Firma del/de los director/es

Fdo.: Alfonso García-Ferrer

Fdo.: Miloudi M. Nachit

1

Contents

List of tables 9
List of figures 11
Summary 13
Resumen 14
Acknowledgment 15
Acronyms 16
1. Preface 17
2. Introduction 19
2. 1. Durum wheat 19
2. 1. 1. Genome, origin, use and economy 19
2. 1. 2. Triticum durum 20
2. 1. 3. Western Mediterranean 20
2. 1. 4. Economy, cultivation, and use of durum wheat 20
2. 1. 5. Wheat growth development stages 22
2. 1. 6. Quantification of wheat development 24
2. 1. 7. Potential yield 25
2. 1. 8. ICARDA durum breeding program 26
2. 1. 9. Landraces and genetic resources 27
2. 2. Genotyping 29
2. 2. 1. Polymerase Chain Reaction 29
2. 2. 2. Microsatellite markers 29
2. 2. 3. Alleles 30
2. 2. 4. Linkage disequilibrium 30
2. 2. 5. Association mapping 30
2. 2. 5. Genetic diversity and structure 31
2. 2. 6. Spatial and non-spatial models in populations genetic 33
2 .3 . Geographic Information Systems (GIS) 33
3. Material and Methods 37
3. 1. Durum wheat collection 37
3. 2. Field evaluation 40
3. 2. 1. Phenotypic traits 40
3. 2. 2. Physiological traits 41
3. 2. 3. Growing environment 42

3. 3. Genotyping 45
3. 4. The GIS interface 46
3. 4. 1. The ArcMap9.2 and VBA 48
3. 4. 2. Environmental data in our GIS interface 50
3. 4. 3. Shape files in the GIS GUI 51
3. 4. 4. Methods developed within the GUI 53
3. 5. Statistical methods 54
3. 5. 1. Correlation 54
3. 5. 2. Regression 55
3. 5. 3. ANOVA 55
3. 5. 4. Path analysis 55
3. 5. 5. Stability 56
3. 5. 6. PCA 56
3. 5. 7. K-mean cluster 57
3. 5. 8. Descriptive locus statistics 57
3. 5. 9. PCA for Multi-Locus data 58
3. 5. 10. PCA to correct for stratification in association studies 58
3. 6. Spatial statistics 59
3. 6. 1. Connectivity networks 59
3. 6. 2. Spatial autocorrelation and Moran's I index 60
3. 6. 3. High/Low Clustering (Getis-Ord General G)-hot spots (Getis-Ord Local G) 61
3. 6. 4. Local and global structure 62
3. 6. 5. Multispati 62
3. 6. 6. Spatial principal components analysis sPCA 63
3. 6. 7. Geographic patterns using Monmonier's algorithm 64
3. 6. 8. Interpolating surfaces 64
3. 7. Population genetic computations 66
3. 7. 1. Genetic structure and population genetics 66
3. 7. 2. Genetic distances 67
4. Results 70
4.1. Phenotypic results 70
4. 1. 1. Agronomy 70
4. 1. 2.Physiology 81
4. 1. 3. Agronomy & physiology 84

4. 1. 4. GxE analysis 91
4. 1. 5. Spatial analysis of phenotypic data 96
4. 2. Phenotypic / Climate relationships 114
4. 2. 1. Correlation 114
4 .2. 2. GxE with genotypic covariates 117
4. 2. 3. Spatial pattern of climate variables 120
4. 3. Genotyping results 122
4. 3. 1. Locus description 122
4. 3. 2. Population structure 125
4. 3. 3. Spatial entity of molecular markers 129
4. 3. 4. Multivariate analysis 133
4. 3. 5. Evaluation of populations 142
4. 4. GIS user interface for durum landraces evaluation 146
4. 4. 1. Trait analysis 147
4. 4. 2. Marker analysis 159
Conclusions 169
References 171
Glossary 185

List of tables

Table 1: Measured morphological traits 43
Table 2: Measured physiological traits 43
Table 3: List of used SSRs and chromosomes localization 45
Table 4: Bioclimatic variables extracted from Worldclim 50
Table 5: Geographic information-shape file 52
Table 6: Traits information shape file 52
Table 7: Marker information shapefile 53
Table 8: Menus developed within the Durum GIS interface 54
Table 9: Mapping examples in landraces diversity studies 64
Table 10: Descriptive statistics of measured traits 70
Table 11: Yield relationship with other traits (Agronomic, quality, yield, coordinates) using multiple regression (contribution) 72
Table 12: Path coefficients (direct and indirect effects) of yield components to grain yield 74
Table 13: Pearson correlation between grain yield and agronomic, phonologic and quality traits 75
Table 14: Path coefficients (direct in diagonal and indirect effects in column) of agronomic, phonologic and quality traits to grain yield 76
Table 15: Descriptive statistics of measured physiological traits 81
Table 16: Yield relationship with physiological traits using multiple regression 84
Table 17: Yield relationship with all measured traits using multiple regression 86
Table 18: Variance components of genotype and genotype by environment of studied traits 91
Table 19: Significant spatial autocorrelation of mean traits 98
Table 20: Significant spatial autocorrelation of measured traits in different years 99
Table 21: High and low clusters using Getis-Ord General G 100
Table 22: Coefficients of multiple regression of GY on TKW, GFD, PH and SL 111
Table 23: Descriptive statistics of the extracted long-term climatic variable for Moroccan durum wheat landraces 114
Table 24: Reduction of the genotypic effect using a long term environmental characteristic as a genotypic co-variable 117
Table 25: Locus descriptive parameters for the dataset and for the Moroccan and the Syrian durum wheat landraces populations 123
Table 26: Moroccan durum populations information 129
Table 27: T-test value for populations found for Moroccan durum landraces 138
Table 28: Agronomic and physiological traits of the Moroccan durum wheat landraces populations 142
Table 29: GxE variance components of agronomic traits per population 145
Table 30: . Linear regression output (Table of regression parameters) 150
Table 31: Linear regression output (Table of regression parameters) 151
Table 32: Multiple regression output 153
Table 33: Path analysis output (correlation matrix) 154
Table 34: Path analysis output (path coefficients) 155
Table 35: k-mean clustering output (field of groups added to shape file) 156
Table 36: Stability analysis output (Field for mean, CV and WR added to shape file) 158

Table 37: Genetic similarity output 161
Table 38: Excel output file from PCA & AF analysis (overall view) 163
Table 39: Excel output file from PCA & AF analysis (Alleles and number of alleles per loci) 164
Table 40: Excel output file from PCA & AF analysis (number of allele per locus and per population) .. 164
Table 41: Excel output file from PCA & AF analysis (locus information per population) 165
Table 42: Excel output file from PCA & AF analysis (Genetic distances) 165
Table 43: Excel output file from PCA & AF analysis (Allele frequencies, PCA input) 166
Table 44: Excel output file from PCA & AF analysis (Eigen values) 167
Table 45: Excel output file from PCA & AF analysis (PC scores) 167

List of figures

Figure 1: Major durum wheat producers (Million tons). Source IGC, CWB. 21
Figure 2: Durum wheat top 10 importer countries (source www.factfish.com) 22
Figure 3: Durum wheat top 10 exporter countries (source www.factfish.com) 22
Figure 4: Physiological development stages of wheat (Adapted from Slafer and Rawson, 1994) 24
Figure 5: Zadok's scales for wheat physiological development 25
Figure 6: Distribution of Moroccan Durum landraces 38
Figure 7: Distribution of Syrian durum landraces 39
Figure 8: Climatic profiles for Tel Hadya experimental station during the three years of evaluation 42
Figure 9: Conception of GIS application 47
Figure 10: Example of interface of classes within ArcMap 49
Figure 11: Cluster tree of GY, quality and agronomic traits 71
Figure 12: Map interpolated residuals from multiple regression of yield on its components (GY = f (TKW, KSPK, SPM2) 73
Figure 13: Map of residuals from regression of yield on TKW 78
Figure 14: Map of residuals from regression of yield on SPM2 79
Figure 15: Map of residuals from regression of yield on KSPK 80
Figure 16: Cluster tree of physiological traits with GY 83
Figure 17: Cluster tree of all measured traits 85
Figure 18: Correlation between residuals resulting from multiple regression of GY on agronomic traits (AGRO) and on physiological traits (PHYSIOLOGY) 87
Figure 19: Map of residuals of multiple regression of GY on All traits (Yellow regions with the lowest residuals) 88
Figure 20: Cluster tree of Moroccan durum landraces using agronomic and physiological traits 90
Figure 21: Map of mean sensitivity of a landrace to environment 93
Figure 22: Map of coefficient of variation (Francis and Kanenberg) of Moroccan durum landraces 95
Figure 23: Examples of different spatial connectivity network of Moroccan durum landraces 97
Figure 24: Spatial hotspots for Plant height 101
Figure 25: Spatial hotspots for days to heading 102
Figure 26: Spatial hotspots for days to maturity 104
Figure 27: Spatial hotspots for thousand kernel weight 106
Figure 28: Spatial hotspots for spike length 108
Figure 29: Spatial hotspots for number of kernel per spike 109
Figure 30: Map of interpolated GY (High "dark" to low yield "light") 110
Figure 31: Predicted map of GY using multiple regression coefficients and raster calculation (High "dark" to low yield "light") 111
Figure 32: Voronoï tessellation and Delaunay triangulation for Moroccan durum landraces 112
Figure 33: Days to heading barriers using Monmonier's Algorithm 113
Figure 34: Thousand kernel weight barriers using Monmonier's Algorithm 114

Figure 35: Correlations between phenotypic traits and long-term climate data of landrace's origin 116
Figure 36: Climatic barriers for durum landrace's long term climate (using all climatic variables) 120
Figure 37: Climatic barriers for durum landrace's long term climate (using Tmin) 121
Figure 38: Climatic barriers for durum landrace's long term climate (using Tmax) 122
Figure 39: Genetic structure of the Moroccan and Syrian durum wheat landraces 126
Figure 40: Moroccan landraces sub-populations 127
Figure 41: Moroccan landraces population probability assignments on the Morocco studied area 128
Figure 42: Spatial distribution of alleles 217 of GWM285 130
Figure 43: Hotspots for allele 217 of GWM285 131
Figure 44: Hotspots for allele 118 of GWM375 132
Figure 45: PCA plot of Moroccan and Syrian landraces (axis1 VS axis2) 133
Figure 46: PCA plot of Moroccan and Syrian landraces (axis1 VS axis3) 134
Figure 47: Histogram of sPCA eigenvalues 135
Figure 48: Plot of variance component of the sPCA Eigen values versus spatial autocorrelation 136
Figure 49: Plot of the first component (x-axis) and its lag vector (y-axis) 136
Figure 50: Plot of the last component (x-axis) and its lag vector (y-axis) 137
Figure 51: Maps of the first spatial principal (a) and principal (b) components for the Moroccan durum wheat landraces 138
Figure 52: Contribution of alleles to the first sPCA component 139
Figure 53: Contribution of alleles to the last sPCA component 140
Figure 54: Moroccan durum landraces molecular barriers 141
Figure 55: Spatial autocorrelation in relation with distance classes 142
Figure 56: Population assignment probabilities, the first spatial and non-spatial principal explained by the genotype and some environmental factors for Moroccan durum landraces 146
Figure 57: ArcGIS interface for durum wheat landraces evaluation 147
Figure 58: Descriptive statistics program's window 148
Figure 59: Pearson correlation program's window 148
Figure 60: Spearman correlation program's window 149
Figure 61: Linear regression program's window 150
Figure 62: Linear regression output (map of residuals) 152
Figure 63: Multiple regression program's window 152
Figure 64: Path analysis program's windows 154
Figure 65: k-mean clustering program's window 156
Figure 66: Stability analysis program's window 157
Figure 67: PCA analysis program's window 158
Figure 68: PCA analysis program's output 159
Figure 69: Allele frequencies and PCA program's window on individuals 160
Figure 70: Distance calculation program's window 161
Figure 71: Genetic similarity maps 162
Figure 72: Allele frequencies and PCA program's window on populations 163

Summary

Durum wheat (*Triticum turgidum L.* var *durum*) is mainly produced and consumed in the Mediterranean region; it is used to produce several specific endproducts. The Durum wheat growing areas are subjects to various biotic and a-biotic stresses. Many varieties have been developed using new breeding technologies to cope with stresses, stabilize yield and maintain grain quality. One of the most critical step in plant breeding is the selection of genetic material that will make parents for crosses. ICARDA durum wheat breeding is studying a core collection of Mediterranean durum wheat landraces that are incorporated in breeding strategies. Field trials to characterize phenotypically and physiologically a collection of landraces were run between 2004 and 2007. Microsatellites were used also to study the genetic diversity. Geographic information system, combined with biometrical and population genetics methodologies were applied to identify phenotypic and genetic diversity spatial patterns. A large phenotypic variation was found in these collections especially yield components, morphology and quality traits, and the grain yield of landraces reached 90% of potential yield. Strong spatial patterns and barriers were found for several phenotypic traits across Morocco. Most of the phenotypic barriers overlapped with Altitude or agro-climatic barriers. In general, landraces collected in close geographic regions tend to have similar phenotypic characteristics. This study helped as well identifying different strategies of landraces in forming yield in different parts of the country. Several long time climatic variables identified to be proxies for traits variation in particular for phenology, height and grain quality. This could be used later in recognizing new regions for future germplasm collections and parents for specific crosses. The collection showed as well a high allelic diversity and strong population structure which was spatially distributed. Also, a significant molecular barrier was found and coincides mainly with the Moroccan altitude pattern and fellow the delineation of the two main mountainous chains in the country. At the end of this study, A geographic information system user interface was developed to help breeders and gene-bank managers to identify landraces and geographic region of interest. The outcome of this study supports the use of geographic information systems together with existing phenotypic data and genetic markers to assess quickly and efficiently large number of genetic resources entries held by gene-banks in particular in the context of climate change.

Resumen

El trigo duro (Triticum turgidum L. var durum) se produce y consume en la región mediterránea principalmente; empleándose para producir varios productos finales específicos. Las áreas de cultivo del trigo duro están sujetas a diversos estreses bióticos y abióticos. Muchas variedades se han desarrollado utilizando las nuevas tecnologías de mejoramiento para hacer frente a los estreses, estabilizar el rendimiento y mantener la calidad del grano. Uno de los pasos más críticos en el mejoramiento es la selección de material genético para identificar padres de cruzamientos. En el ICARDA, el programa de mejoramiento de trigo duro está estudiando una colección núcleo de las variedades locales de trigo duro del Mediterráneo que se incorporan en las estrategias de mejoramiento. Los ensayos de campo, para caracterizar fenotípicamente y fisiológicamente una colección de variedades locales, se realizaron entre 2004 y 2007. Los micro-satélites se utilizaron también para estudiar la diversidad genética. Los sistemas de información geográfica (SIG), combinados con metodologías de genética y biométrica se aplicaron para identificar formas espaciales de diversidad fenotípica y genética. Se encontró una gran variación fenotípica en estas colecciones especialmente en componentes del rendimiento, la morfología y la calidad. El rendimiento de grano de las variedades locales alcanzó el 90% del rendimiento potencial. Se encontraron importantes barreras y patrones espaciales para varias características fenotípicas a través Marruecos. La mayor parte de las barreras fenotípicas coinciden con la altitud o con las barreras agroclimáticas. En general, las variedades locales recolectadas en zonas geográficas cercanas tienden a tener características fenotípicas similares. El presente estudio ayudó también a la identificación de las diferentes estrategias con variedades locales para la determinación del rendimiento en distintas partes del país. Algunas variables climáticas identificados durante mucho tiempo son indicadores de variaciones en fenología, altura de planta y la calidad del grano. Esto podría ser utilizado más adelante en el reconocimiento de nuevas regiones para las futuras colecciones de germoplasma y padres para cruzamientos específicos. La colección estudiada demostró también una alta diversidad alélica y fuerte estructura de la población distribuida espacialmente.

Se encontró también, una barrera molecular significativa y coincide principalmente con el patrón de altitud marroquí y siguió con la delimitación de las dos principales cadenas montañosas del país. Al final de este estudio, se desarrolló una interfaz de usuario del sistema de información geográfica para ayudar a los mejoradores y los administradores de los bancos genéticos a identificar las variedades locales de la región geográfica de interés. El resultado de este estudio apoya el uso de sistemas de información geográfica, junto con datos fenotípicos existentes y marcadores genéticos para evaluar de forma rápida y eficiente un gran número de recursos genéticos en poder de bancos de genes, en particular en el contexto del cambio climático.

Acknowledgment

Acronyms

ASH: Ash content
COV: Covariance matrix
DH: Days to heading
DM: Days to maturity
GD: Genetic diversity
GE: Genotype x Environment Interaction
GFD: Grain filling duration
GIS: Geographic Information System
GUI: Graphic User Interface
Gwm: Gatersleben wheat microsatellites
GY: Grain yield
He: Expected heterozygosity
Ho: Observed heterozygosity.
ICARDA: International center for Agriculture research in the dry areas
KSPK: Number of kernel per spike
MAS: Marker Assisted Selection.
MCMC: Markov Chain Monte Carlo
PC: Protein content
PCA: Principal Components Analysis
PH: Plant height
PL: Peduncle length
QTL: Quantitative trait loci
QTLxE: Interaction QTL by environnement
SAU: Spatial Autocorrelation
SD: Standard deviation
SDS: Sedimentation test
SDSI: Sedimentation index
SDSN: Sedimentation n
SL: Spike length
sPCA: spatial Principal Components Analysis
SPM2: Number of spike per square meter
TKW: Thousand kernel weight
VAR: Variance
VBA: Visual Basic for Applications
VIT: Vitreousness
YP: Yellow pigment
ZS: Zadoc's scale

1. Preface

Wheat is the cereal most consumed cereal in the world after rice and more popular because of the presence of gluten; important in making bread and pasta production process. Wheat belongs to genus Triticum. The principal center of origin and diversity is the Fertile Crescent between the Mediterranean coasts and the Plaines of the Tigris and Euphrates. For a long time, man has grown wheat and improved its productivity and quality. Durum wheat (Triticum turgidum L. var. durum), a tetra-ploid wheat, compared to bread wheat, contains a high specific gluten that its semolina can be used to make couscous, pasta, burghul, and frike. Durum is mainly grown in the dryland areas of the Mediterranean region where biotic (diseases and insects) and a biotic (drought, cold and heat) stresses and variable environmental conditions are widespread. The Mediterranean climate is characterized by low and highly erratic annual rainfall varying from 200 to 800 mm, with usually poor rainfall distribution, and periods of drought and temperature extremes (cold and heat) that can occur at any development stage of the plant. Landraces are varieties of crops that evolved and were improved by farmers over many generations, without the use of modern breeding techniques. These varieties are generally very diverse within species, because each was adapted to a specific environment. The pace of improvement accelerated as modern breeding techniques were developed that facilitated selection of specific desirable traits. Within most types of crops, breeders have crossed different parental material and selected traits resulting in high yields. Quality changes have also been the subject of breeding effort. Other goals of breeding have included rapid and simultaneous germination, flowering, and maturation of crops.

Durum wheat grown by farmers was until recently made of landraces. In the Mediterranean region, thousands of landraces were grown. Landraces are characterized by a biotic stress tolerance, particularly drought and also good grain quality. However, they lack yield potential and diseases resistance. The durum breeding program at the International Center of Agriculture Research in Dry Areas (ICARDA), have started to use intensively the landraces germplasm to improve varieties for drought tolerance, adaptation, and grain quality. To make a good use of genetic and phenotypic diversity found in landraces. Breeders have to dissect this diversity and study the adaptation of landraces. Adaptation and responsiveness of a plant to varying environments is one of the main tasks of breeding program. To cope with environmental fluctuations, durum breeding for large and/or diversified target regions may imply the definition of a breeding strategy, and possibly exploit, Genotype x Environment interactions. Multi-environment yield trials performed for genotype selection or recommendation may also provide information for defining adaptation strategies, yield stability targets, indirect selection criteria (based on morpho-physiological traits or genetic markers), and parent germplasm and selection environments. Repeatable Genotype x Location (GL) interaction effects can be either exploited, by breeding material adapted to a specific sub-region, or minimized, by breeding material widely adapted to a region. Interfacing statistical modeling of genotype responses (e.g., by joint regression, Additive Main effects and Multiplicative Interaction or factorial regression techniques) with indirect selection theory allows for comparing different adaptation strategies of germplasm.

In the case of genetic resources and since this set of plant material is collected in specific geographic region, an alternative approach is to link climatic and soil layers with a matrix of durum landraces characteristics to greatly understand the mechanisms of adaptation to difficult stressed conditions, and to improve the selection process of parental material used in the future in a Mediterranean breeding program. This will provide better decisions to develop adapted germplasm. Such linkage will reduce environmental hazards and abiotic stress risks and thus making breeding a multidisciplinary task involving breeders, physiologist, geneticists, statisticians, and agro-ecologists, etc. This methodology involves the use of the Geographic Information Systems (GIS), a family of powerful and dynamic computer software systems that manipulates and displays layers of spatially variable data. A variety of data types are used including climatic factors (precipitation, temperature, and radiation), geo-physical features (topography, soil traits) and biological characteristics (plant information and tolerance). By

integrating these individual spatial data layers in a GIS, it's possible to better understand their interrelationships and create more useful models and maps to discover and study genetic diversity, which help the breeders to select and improve varieties.

Presently, advanced molecular technologies make it possible to efficiently measure genetic information. As for geographic information, considerable advancements in computer science have led to the development of sophisticated software (GIS), in parallel with the elaboration of a wide variety of spatial analysis methods, making it possible to extract information from any environmental profile.

Specifically, the present work reveals a GIS angle on particular aspects of molecular genetics and phenotyping of genetic resources. It falls within the discipline of GIS because GIS tools have been involved in the context of a scientific approach carried out together with biologists to assess their potential usefulness in discovering genetic diversity patterns and in bearing out hypotheses suggested by population geneticists.

The use of GIS has increased in several fields that partly deal with geographically based data such as genetic resource collections. In many cases researchers are interested in using geographical/statistical data to explain facts in their research. Another alternative methodology is to compute geographical quantities in GIS and then export these quantities to a standard statistical tool (SPSS, Excel, etc.) or vicse-versa. These statistical tools cannot be integrated into a GIS so that the statistical analyses can be executed within a GIS software e.g. ArcMap. In the case of population genetics, several softwares were developed to study the effect of space or integer space into studying the dynamic of populations and distribution of alleles. However, no software permitting integration of phenotypic, genetic and GIS analysis was developed. Since such integrated GIS and statistical programs do not really exist and there is a demand for them, it is an interesting task to deal with.

The aim of this study was 1) to study the phenotypic and genotypic diversity durum wheat landraces collections (example, the Moroccan collection), 2) to discover the spatial aspects of phenotypic traits and genetic variation and 3) develop a GIS user interface to help breeders study genetic resources. To accomplish this: 1) we used three years of phenotyping a set of durum landraces, 2) we genotyped the collection with fifty microsatellites across all durum wheat chromosomes and 3) analyzed using a set of biometrical and spatial statistics methods. This thesis includes a detailed explanation of the most statistical methodology used in breeding and genetic resources studies.

2. Introduction

2. 1. Durum wheat

2. 1. 1. Genome, origin, use and economy

Domestic wheat originated in the Fertile Crescent in the Middle East. The oldest archaeological evidence for wheat cultivation comes from Syria, Jordan, Turkey, Armenia, and Iraq (ref.). Around 9000 years ago, wild einkorn (*Triticum monococcum)* wheat was harvested and domesticated in the first archaeological sedentary farming in the Fertile Crescent (Mac Key, 2005). The wild and cultivated wheats include diploid, tetraploid, and hexaploid species for which either *Triticum monococcum or Triticum urartu* was the A genome donor. The wild tetraploids *T. turgidum* ssp. *dicoccoides* arose from hybridization between *T. urartu* and the putative B donor *Ae. speltoides*. However, it remains uncertain whether *Ae. speltoides* is the sole source of the B genome or whether the genome resulted from an introgression of several parental species (Zohary and Feldman, 1962). *T. turgidum* (AABB) includes the wild ssp. *dicoccoides* and several cultivated subspecies such as *T. turgidum* ssp. *durum* grown mainly in the Mediterranean dryland. *Triticum aestivum* L. (AABBDD) is an allopolyploid of *Aegilops tauschii* (DD) cross with *Triticum turgidum* L. (AABB). *T. aestivum* (AABBDD) arose under cultivation 8,000 years ago from spontaneous hybridization between *T. turgidum* ssp. *dicoccon* and the diploid goatgrass *Aegilops tauschii* ssp. *Strangulate* (DD). The range of distribution of Triticum relatives occurs from the Canary Islands to Western China, and from Southern Russia to Northern Pakistan and India (van Slageren, 1994). The center of variation of Triticum wild relatives includes Egypt, Palestine, Jordan, Lebanon, Syria, Turkey, Armenia, Azerbaijan, Iraq, Iran, Afghanistan, and the Turkish Republics of Central Asia (van Slageren, 1994).

Acording to Harlan (1992), the origin of *T. aestivum* was a coincidental event that probably occurred as a consequence of cultivating *T. turgidum* in close proximity to populations of wild *Ae. tauschii*. Hexaploid wheat was found in the archaeological sites of Cafer Höyük, Can Hasan and Çatalhöyük in Turkey and Abu Hureya in Syria, dating to the seventh millennium before Christ (BC). Contrary to this evidence most of the evolutionary studies have placed its origin in either Transcaucasia or the south Caspian region. However, the probable absence of *Triticum turgidum* from this region until the sixth millennium BC indicates that the biological evidence is at odds with the archaeological evidence. Further, the origins of emmer wheat (AABB) demonstrated that cultivated emmer is not monophyletic, and it was domesticated on more than one occasion and at different geographic locations in the Fertile Crescent. The demonstration that cultivated emmer has diverse origins provides evidence in favor of the hypothesis that the transition to agriculture in South West Asia was a necessary response to a changing environment rather than the result of a chance discovery. Ancient or modern farmers have grown four wheat species: einkorn (*T. monococcum*), emmer (durum), *T. timopheevi*, and bread wheat. However, only durum and bread wheat are currently used for food production, accounting for 4 and 96% of the total wheat acreage, respectively. The farming communities of the Hauran plateau in Southern Syria, as in the other parts of the fertile crescent, have contributed for millennia to the evolution and in situ conservation of the durum landrace Haurani. According to Vavilov (1951) and Harlan (1992), the landrace Haurani can be considered as an evolutionary link between wild emmer wheat (*Triticum dicoccoides*), the wild progenitor of all domesticated wheats and through breeding developed cultivars. The Haurani landrace has evolved in a heterogeneous environment with large variations in rainfall (250-459 mm), altitude, temperature extremes (cold and terminal heat), drought, length of growing season, date of sowing, etc. Growing for thousands of year has led the evolution over many generations, of gene complexes providing the landrace with adaptive traits for the rainfed areas (Nachit, 1992). The Haurani landrace was continuously cropped

for millennia until the end of 1980s. It was replaced by the new productive and drought tolerant durum varieties; its cropping area has declined to less than 5% (Nachit, 1995).

2. 1. 2. Triticum durum

Durum wheat is one of the oldest cultivated plants in the world and is grown mainly in the middle and near East region and North Africa, which are considered the centers of origin and diversification of this crop (Vavilov, 1951). Based on archeological evidence it is generally accepted that durum wheat was domesticated at least 2000 years before bread wheat (Morris and Sears 1967) during the late Mesolithic period and the early Neolithic age (Harlan 1986). The adaptation of durum wheat largely overlaps that of bread wheat, but is less widely grown (Autrique et al., 1996). On the other hand, durum wheat is better adapted to Mediterranean dryland than bread wheat. This is why over 80% of the total world durum wheat area is located in the Mediterranean basin (Porceddu et al., 1990) and this is why durum has been concentrated in the driest areas of the West Asia and North Africa (WANA) region. Durum wheat is best adapted to regions having a relatively dry climate, with hot days and cool nights during the growing season, typical of Mediterranean and temperate climates. Seed germination will occur as low as 2°C, but the optimal temperature is 15°C (Bozzini, 1988). Most durum wheat produced in the world is of spring growth habit; however, durum wheat lines with winter habit (requires vernalization to initiate the transition from vegetative growth to reproductive growth) have been evaluated for production in the southern USA (Domnez et al., 2000; Schilling et al., 2003).

2. 1. 3. Western Mediterranean

The countries include Portugal, Spain, southern France, Morocco, Algeria and Tunisia. Most are accessible and harbour eight or more species. In northern Portugal, there are landraces of wheat and rye adapted to unidentified soil problems. There are also primitive wheats, such as spelt, T.d*icoccum* and T.*monococcum*, that are still grown in Spain for specific culinary or animal uses. In North Africa, there are landraces of diploid, tetraploid and hexaploid wheats that may exhibit physical environmental stress tolerances. Collections of *Ae. bicornis* from the coastal areas of Egypt and Cyprus in the eastern Mediterranean might be useful as a source of salt tolerance. In Morocco, for example, the local populations of durum wheat offers an important genepool as sources of adaptation and tolerance to many biotic and abiotic stresses. This important genetic material is continuously subject to genetic erosion and the rapid adoption of the newly released varieties has already reduced significantly the acreage grown to landraces in many parts of Morocco. In the mountain and oasis regions of Morocco, however, wheat landraces are still widely grown by farmers.

2. 1. 4. Economy, cultivation, and use of durum wheat

Wheat is the first important and strategic cereal crop for the majority of world's population. It is the most important staple food crops of about 40% (nearly half) of the world population, occupying 17% (one sixth) of crop acreage worldwide, and providing nearly 55% of the carbohydrates and 20% (nearly fifth) of the food calories consumed globally (FAO, 2003; Gupta et al., 2008; see: http://www.slideshare.net/ifad/durum-wheat-miloudi-m-nachit-icarda-4998603). Although wheat production during the last four decades has witnessed a steady significant increase, a fatigue has been observed during the last few years, leading to the lowest current global wheat stocks ever since 1948/1949. Consequently, wheat prices have also been soaring, reaching the highest level of US $ 367 a ton as against US $ 165 a year ago.

Wheat exceeds in acreage and production every other grain crops (including rice, maize, etc.); (Gupta et al., 2008) and is therefore, the most important cereal grain crop of the world, which is cultivated over a wide range of climatic conditions and the understanding of genetics and genome organization using

molecular markers is of great value for genetic and plant breeding purposes. The world durum wheat production was estimated to be 35.4 million tons in 2012-2013. Most important producers are EU, Canada, Turkey and Syria and North Africa (Figure 1). Trade is to be around 7.6 Million tons during 2011-2012 compared to 6.8 MT of 1997-2001. There is a continuous durum wheat demand in the world. In the last decades, Africa is the most demander of durum wheat mainly because of the demand of North Africa. The regions importing the most are North Africa, European Union and Latin America (Figure 2). Within North Africa, Algeria is the main importer. The top 5 exporters during 2012 (Figure 3) of wheat are Canada, India, EU, USA and Mexico

Durum (derived from the Latin word for hard) has the hardest kernel of all wheats. Durum wheat with high protein content and gluten strength is the preferred choice of processors for producing pasta products. Durum kernels are amber-colored and larger than those of other wheat classes. Also unique to durum is its yellow endosperm, which gives pasta its golden color. Durum wheat with strong gluten characteristics forms strong, non-sticky dough ideal for pasta and couscous production. Semolina with strong gluten properties also results in pasta and couscous products with superior cooking characteristics. Durum wheat kernel is normally hard and virtually all varieties have amber, vitreous, and rather large kernels. The protein content is usually about 13%, but may reach 22%. High protein content, however, does not always guarantee optimum cooking quality (Ciaffi et al., 1991; Blanco and Giovanni 1996). The principal use of durum wheat grain is the production of semolina for use in pasta products. However, in North Africa, durum is preferred for the production of couscous; and in the Middle East and burghul. Traditional breads are also made with durum flour, particularly in Morocco and South Italy.

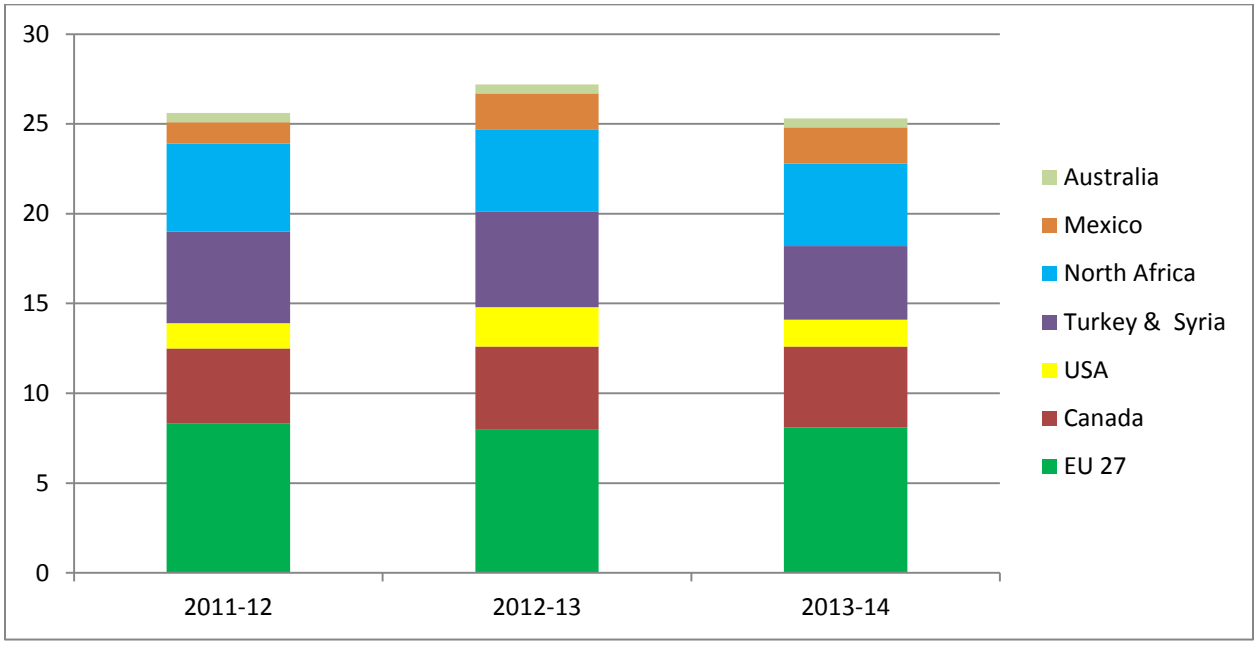


Figure 1: Major durum wheat producers (Million tons). Source IGC, CWB.

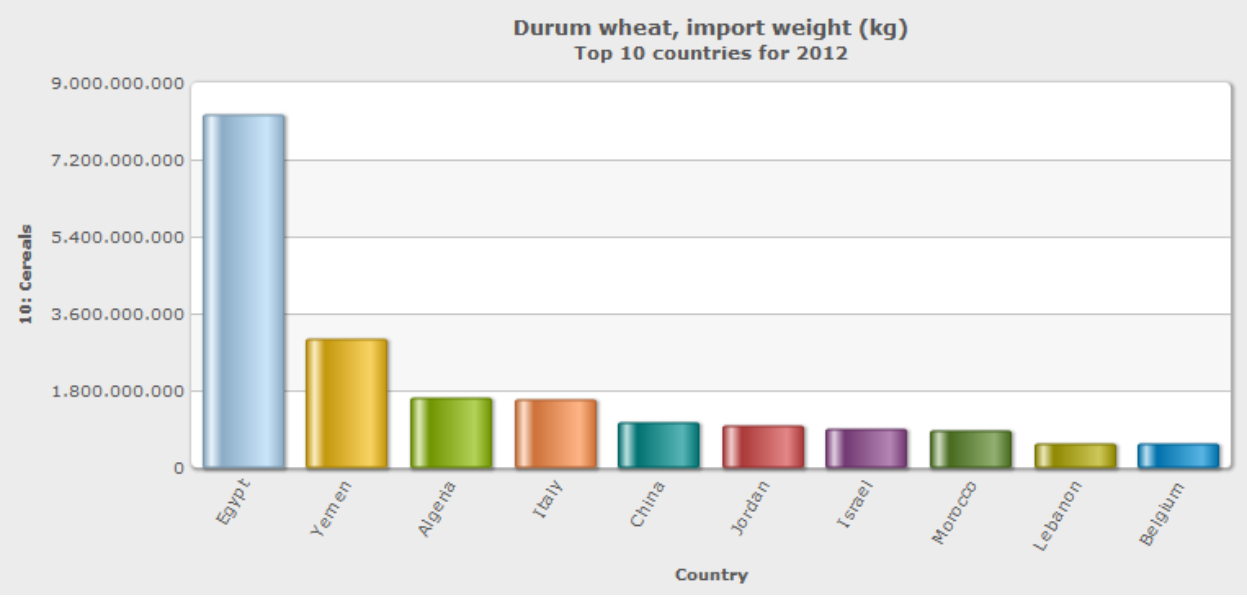


Figure 2: Durum wheat top 10 importer countries (source www.factfish.com)

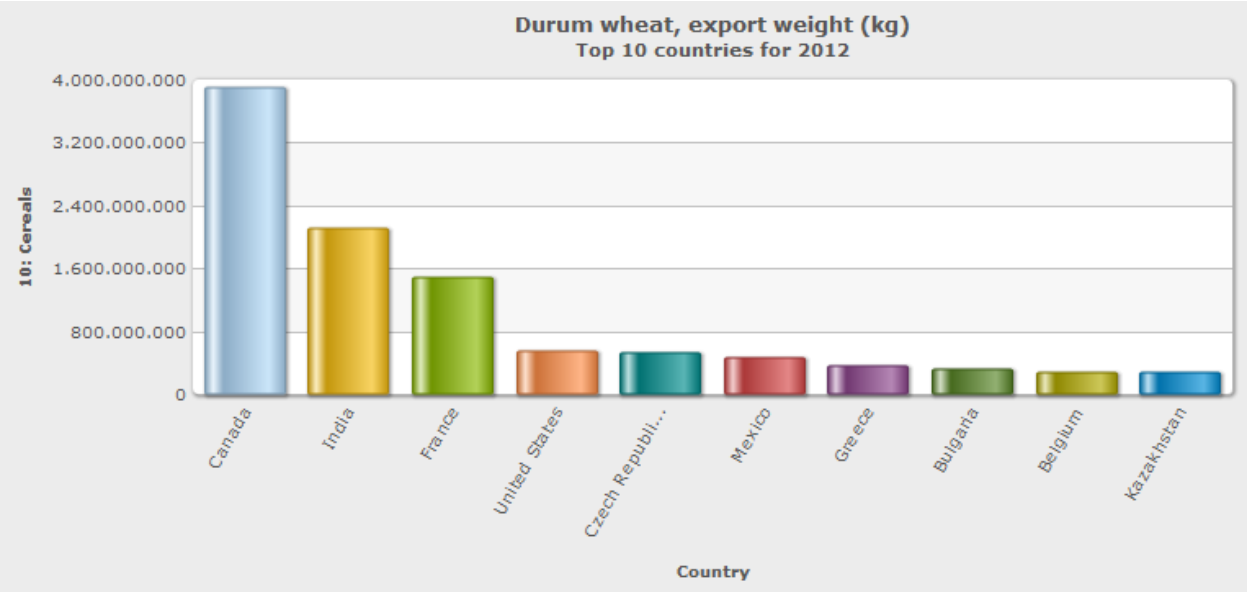


Figure 3: Durum wheat top 10 exporter countries (source www.factfish.com)

2. 1. 5. Wheat growth development stages

Physiologically: germination, emergence, tillering, floral initiation or double ridge, terminal spikelet, first node or beginning of stem elongation, boot, spike emergence, anthesis, and maturity are usually distinguished developmental stages. These stages (Figure 4) may be grouped into: germination to emergence (E); growth stage 1 (GS1) from emergence to double ridge; growth stage 2 (GS2) from double

ridge to anthesis; and growth stage 3 (GS3), which includes the grain filling period, from anthesis to maturity. Physiological maturity is usually defined as the time when the flag leaf and spikes turn yellow (Hanft and Wych, 1982). The period of each development phase depends essentially on genotype, temperature, day-length, and sowing date. Various environmental stresses may shorten the wheat growth phases.

Germination to emergence (E): Germination may occur between 4° and 37°C, optimal temperature being from 12° to 25°C. During germination, the seminal roots grow first, followed by the coleoptile, which protects the emergence of the first leaf.

Emergence to double ridge (GS1): Wheat tillers grow from the axils of the main shoot leaves. The potential number of tillers varies with genotype, particularly among flowering types, winter types having a greater number. Tillering has great agronomic importance in cereals since it may partially or totally compensate the differences in plant number after crop establishment and may allow crop recovery from early frosts. The duration of the vegetative stage (GS1) in wheat may vary from 60 to 150 days depending on sowing date and genotype. This stage has two major components: *Vernalization*, wheat flowers after the completion of a cold period. The double ridge stage is not reached until chilling requirements are met, and the vegetative phase is prolonged generating a higher number of leaves in the main shoot; Vernalization occurs at temperatures between 0° and 12°C (Ahrens and Loomis, 1963; Trione and Metzger, 1970). *Photoperiod*, after vernalization is completed, genotypes, which are sensitive to photoperiod, require a certain day-length to flower. Sensitivity to photoperiod differs among genotypes. They flower faster as the day-length increases, but they do not require a particular length of day to induce flowering (Evans et al., 1975; Major and Kiniry, 1991). The development of the inflorescence after induction occurs at a rate that is also dependent on daylength in the genotypes sensitive to photoperiod (Stefany, 1993). Vernalization and photoperiod constitute the basic processes of the adaptation of wheat to various environments. Knowledge and genetic manipulation of them should continue to provide significant tools for adaptation and yield.

Double ridge to anthesis (GS2): Wheat plants have from four to eight leaves in the main shoot when the growing apex changes from the vegetative to the reproductive stage. The length of the apex at this time is approximately 0.5 mm. Temperatures above 30°C during floret formation cause complete sterility (Owen, 1971; Saini and Aspinal, 1982). Each spikelet has from 8 to 12 floret primordia in the central part of the spike. The basal and distal spikelets have from six to eight florets. Two stages are differentiated. *Terminal spikelet*: Spikelet number per spike is already determined at this stage, varying from 20 to 30 (Allison and Daynard, 1976; Kirby and Appleyard, 1984). This stage is particularly sensitive to environmental stresses, especially nitrogen and water (Wuest and Cassman, 1992a). *Spike growth*: Once the terminal spikelet is formed, stem elongation starts and the spike begins to grow. Spike growth occurs from the appearance of the leaf prior to the flag leaf (penultimate leaf) up to ten days past anthesis (Kirby and Appleyard, 1984). Spike growth, slow in its early stages, increases greatly about the time the ligule of the flag leaf becomes visible (Krumm et al., 1990). In the wheat crop, there is a close relation between the number of kernels per unit area and the ratio between incoming radiation to the mean temperature above 4.5°C (the photothermal quotient) calculated for the 30 days preceding anthesis (Fischer, 1985a).

Anthesis to physiological maturity (GS3): The wheat spike contains only one spikelet per rachis node. Each spikelet has between three and six potentially fertile florets (Kirby and Appleyard, 1984), which are self-pollinated in 96 percent of the cases (Martin et al., 1976). Anthesis begins in the central part of the spike and continues towards the basal and apical parts during a three- to five-day period (Peterson, 1965). The proximal florets of the central spikelet are fertilized two to four days earlier than the distal florets. These grains usually have a greater weight (Simmons and Crookston, 1979). After floret fertilization, cellular division is rapid, during which the endosperm cells and amyloplasts are formed. After there is a phase of cell growth, and differentiation and starch deposition in the endosperm, which corresponds to

linear grain growth and takes from 50 to 70 percent of the grain filling period. The embryo is formed at the time of endosperm growth (Jones et al., 1985).

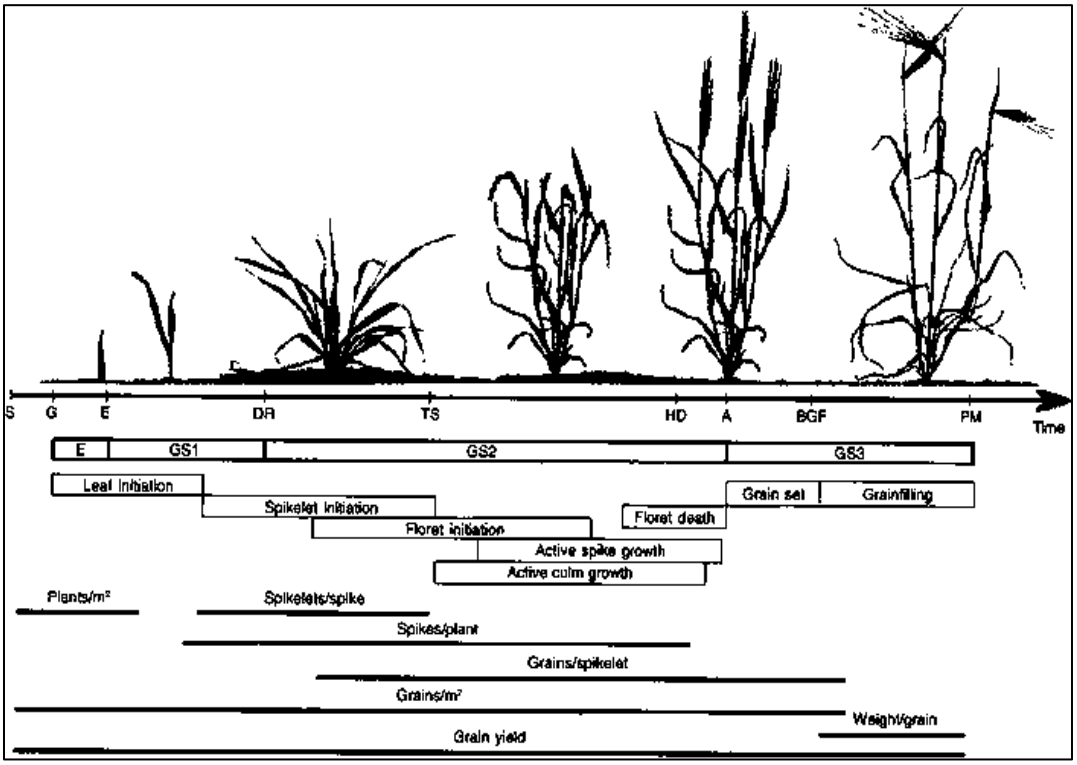


Figure 4: Physiological development stages of wheat (Adapted from Slafer and Rawson, 1994)

S=sowing; G=germination; E=emergence; DR=double ridge appearance; TS=terminal spikelet initiation; HD=heading; A=anthesis; BGF=beginning of grain filling period; PM=physiological maturity; GS=growth stage

2. 1. 6. Quantification of wheat development

There are several scales or development codes for wheat that describe visible growth stages without the need for dissection of the plant. Among these the most widely used is Haun's scale (Haun, 1973), which is most useful in defining vegetative growth stages. Feeke's scale (Large, 1954) and Zadoks' scale (Zadoks et al., 1974) provide a good description for both vegetative and reproductive stages (Figure 5). Crop development stages are determined in representative plants in the field, avoiding the borders of plots and any interfering material. Zadoks' scale is the most comprehensive and easiest to use. It describes all stages of the cereal growth cycle, incorporating characteristics not considered in other scales. It is based on a decimal code, which incorporates various aspects of plant development. The main growth stages are self-explanatory. A second digit, values from 0 to 9, gives more detail for each main growth stage using the position 5 as the middle value. Leaf numbers, for example, have decimal codes from 11 to 19 and the tillers in the main shoot from 21 to 29. After emergence, all development stages are based on observations on the main shoot, usually the tallest and thickest. After stage 40 (at 39 the flag leaf ligule is just visible and at 41, the early boot stage, the spike is beginning to swell), the stages of the main shoot and tillers become similar, and the stages are determined by viewing the whole plant. Stages 70 to 93 are determined by the development stage of individual kernels or grain in the middle of average spikes. At the ICARDA

durum program we are using this scale: see physiological data (spectral reflectance for example): ZS45 (booting stage); ZS70 (after anthesis 100% ~Milk stage)

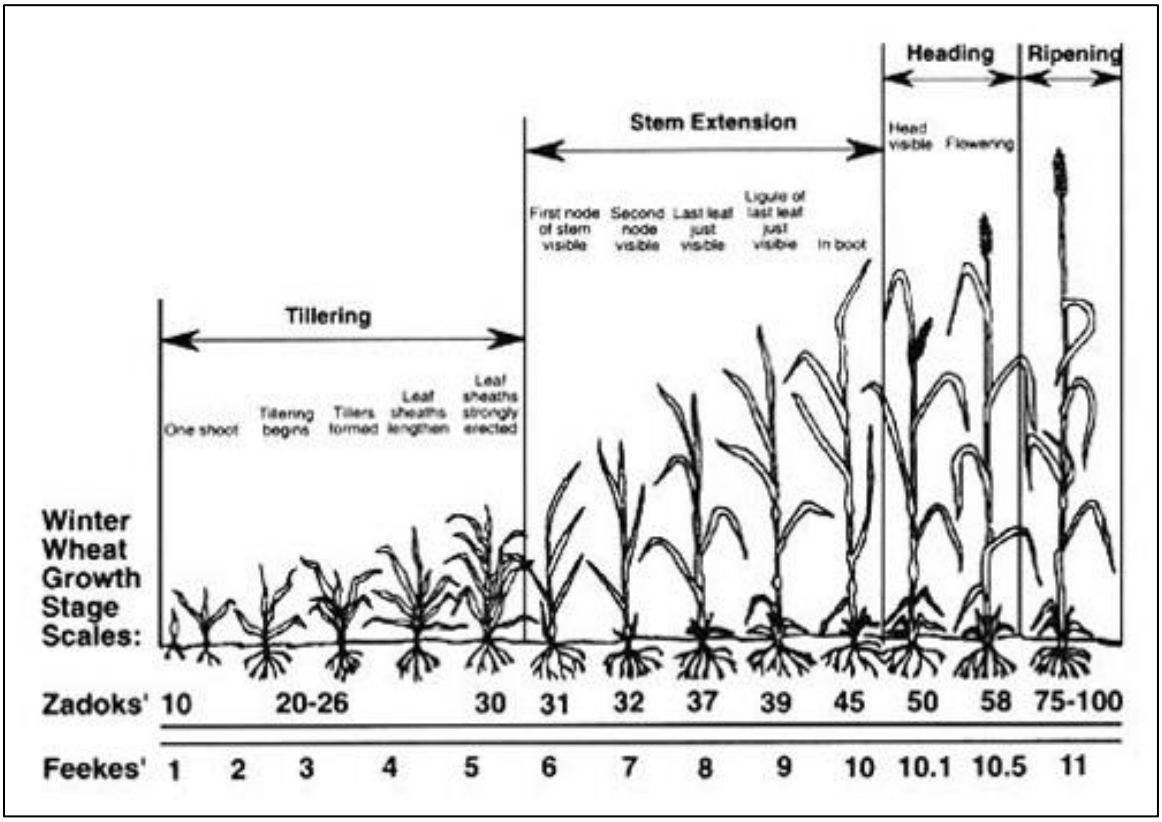


Figure 5: Zadok's scales for wheat physiological development

2. 1. 7. Potential yield

Yield potential, defined as the yield of an adapted genotype grown under optimal management and in the absence of biotic and abiotic stresses, has been found to be a very useful concept since progress in yield potential usually leads to progress in wheat yield in farmers' fields, particularly if stresses are mild. The yield of a wheat crop can be expressed as the product of two components: $GY = KM2 * KW$

where GY is grain yield (g/m^2); KM2 is the kernel number (m^{-2}); and KW is the kernel weight (g). The KM2 can be also expressed as the number of kernel per spike (KSPK) and the number of spike per meter square (SPM2). It can be also explained as the product of plants per meter square, spikes per plant, spikelets per spike, florets per spikelet and grains per floret. It follows the GY equation that changes in wheat yield potential could be achieved through changes in KM2 and/or KW. Strong associations with yield have been found with KM2 for sets of wheat genotypes (Austin et al., 1980; Slafer et al., 1990; Slafer et al., 1996). KM2 is established in the period between 20 and 30 days before flowering and 10 days after anthesis. This period coincides with tiller and floret mortality, along with the active growth of the stem (peduncle) and spike. Gains in KM2, however, do not translate directly in yield potential gain due to partial compensation by decreased KW. Slafer et al. (1996) argue that the lower KW observed with increased KM2 is not only due to a lower amount of assimilates per grain but is the result of an increased

number of grains with a lower weight potential coming from more distal florets. It has been shown that competition for limited resources during the spike growth period, including light and nitrogen, is the major cause of KM2 potential loss.

2. 1. 8. ICARDA durum breeding program

The Mediterranean climate is characterized by low and highly unpredictable annual rainfall varying from 200 to 800 mm, with usually poor rainfall distribution, and periods of drought and temperature extremes (cold and heat) that can occur at any plant stage of development (Nachit et al., 1992a, 1992b). These environmental conditions are the main causes of yield reduction and fluctuation in the Mediterranean region. As irrigation on a large scale for durum commercial production is not available to increase production, the only applicable alternative is the improvement of drought tolerance and yield stability through genetics/ plant breeding, stress physiology, and use of molecular markers as tools by the breeders (Nachit et al., 1998). Consequently, In Syria for example, during the last 10 years the contribution made by the stress tolerant and productive durum genotypes developed by the ICARDA is reflected in the spectacular production increase, from less than 1 to 3.4 million tons, without any significant increase in the cropped area (1.2 million .ha). However, with unrelenting population growth in most countries of the West Asia and North Africa region, and virtually all show food deficit, the major challenge for researchers is to increase food- mainly wheat output.

The ICARDA dryland durum-breeding program was initiated in 1977 in northern Syria. The ICARDA main research station (Tel Hadya) and its related research sites across a rainfall gradient (Lattakia, Terbol, Kfardan, and Breda) in the Middle East region are located in the heart of the Fertile Crescent. A region where a wealth of wheat landraces is found along their wild relatives in different agro-ecological zones, from lowland plains to highland plateaus; and from favorable to stressed environmental conditions (Nachit, 1992, 1998). Along drought, cold, and heat, multiple biotic stresses (diseases, insects, and viruses) are endemic with the highest virulence of disease races and insect biotypes. Thus, these combinations of abiotic and biotic stresses, makes the breeding work in the Fertile Crescent dryland both complex and challenging.

The Mediterranean basin is rich of durum landraces and wild relatives. In the ICARDA durum breeding program at ICARDA, landraces and wild relatives possessing novel traits are evaluated and used to improve durum varieties (Nachit, 1992). Different tools are used to generate durum germplasm for immediate use by durum scientists in the region: Molecular markers, stress-resistance tools, conventional cross, etc. The objectives of the durum breeding program at ICARDA, in collaboration with NARS, advance research institutes and universities are: 1) develop productive durum genetic material combining high grain quality with resistance to the main abiotic and biotic stresses encountered in the Mediterranean region; 2) Use the available genetic variation found in the durum wheat local landraces; and 3) use of molecular assisted selection. As drought is the dominant stress factor limiting durum productivity in the Mediterranean region, the ICARDA's durum breeding program has developed with collaborators a breeding strategy to improve germplasm resistant to drought, cold and heat. The cornerstone of this strategy is the introgression of resistance genes from landraces and wild relatives to cultivated durum, and the utilization of contrasting and representative environments in the Mediterranean basin. The standard experiment in durum wheat breeding program at ICARDA is to evaluate durum cultivars under a range of environmental conditions as multi-environmental trial strategy; also a particular attention is given to detailed testing environments characterizations in terms of physiologically relevant meteorological and soil variables. The dependence of genotypic performance of durum wheat on environmental conditions is an expression of the genotype by environment interaction (GE), and breeding for abiotic stress environments means to a large extent trying to understand and overcome problems imposed by GE. For statistical analysis, durum wheat breeding program is using and developing wide range of methodologies such as ANOVA, AMMI, stability (parametric and non-parametric), Wescott. Durum breeding program at

ICARDA is using the marker assisted selection (MAS) through the identification of QTL and QTLxE pattern for yield and associated traits. That was a very effective selection method compared to traditional phenotypic (or field) selection of secondary traits influencing yield.

2. 1. 9. Landraces and genetic resources

Landraces are one important category of genetic resources which have been categorized by Frankel (1977) and the Food and Agriculture Organization of the United Nations Commission on Plant Genetic Resources (FAO, 1983). Landraces and obsolete cultivars represent a very valuable part of the genetic resources of wheat (Zou and Yang 1995) because of their characteristic features such as their tolerance to locally occurring stress (Tesemma et al., 1998). Landraces, which have been developed through a combination of a natural selection and selection performed by farmers (Belay et al., 1993) and have been selected over thousands of years by farmers and nature for characteristics related to local adaptation and yield stability. Landrace varieties are an important germplasm to move towards sustainable agricultural development. In the case of durum wheat landraces, several works have reported the presence of important features for crop improvement such as resistances, early maturity and quality (Porceddu et al., 1975; Boggini et al., 1987; Pecetti *et al.*, 1992).They have some valuable traits which can contribute significantly to improvement of new durum wheat cultivars and broaden their diversity (Biesantz et al., 1990; Tesemma et al., 1998). Durum wheat landraces are less productive, but they are more tolerant to environmental stress than the modern varieties. They are still cultivated in the remote rural areas of several Mediterranean countries for local use because of their high end-product quality (Agorastos and Goulas, 2005), especially in areas of marginal agriculture, where yield and yield stability are the most desirable characteristics. The genetic erosion of these varieties could lead to the extinction of valuable resources which have not been exploited. The protection and utilization of these materials requires their conservation, evaluation and characterization (Esquinas- Alcazar, 1987).

This category of gene pool should be a major activity of germplasm banks to identify useful genetic variation and make it available to breeders.

Variation is needed to:

increase yield potential; provide new sources of biotic resistance to maintain current yield levels;

provide adaptation to the more marginal environments (abiotic stress);

provide improved industrial quality.

Evaluation and pre-breeding should be major activities of any collection. The ICARDA durum wheat collection provides the opportunity and the responsibility to raise involvement in these activities, in addition to offering new variation to breeding programs. Information systems make it possible to estimate the degree of relatedness among wheat landraces and allow breeders to increase genetic diversity by selecting materials of divergent parentage for crosses. This can reduce wheat's vulnerability to diseases and climatic changes, and automatically updates family trees as additional ancestry is discovered. A survey of breeders indicates that 75 percent of wheat breeders acknowledge that future advances in breeding will be limited by a lack of genetic resources, though this was not considered an immediate restraint for most programs (Rejesus et al., 1996). This lack of genetic resources can be mediated by increasing knowledge about the value of genetic resources and through the identification of new and novel sources of traits, both in the existing *ex situ* collections and *in situ* collections yet to be collected.

No insightful study has been done to estimate the contribution of collections to wheat improvement. Chapman (1986) examined the role of genetic resources (defined as wild materials and landraces) in wheat breeding and found it difficult to assess. He concluded that genetic resources are used in about 10% of crosses, based on the occurrence of genetic resources in pedigrees of recently released cultivars and the frequency of references to genetic resources in the *Annual Wheat Newsletter*. An example of the utility of

genetic resources is their contribution to improving wheat resistance to the rust diseases. One of the stem rust resistance genes, *Sr2*, originally transferred to hexaploid wheat from Yaroslav emmer by McFadden in 1923 (Stakman and Harrar, 1957), has provided durable resistance to the disease. Cultivars possessing *Sr2* in combination with other genes have been grown without stem rust losses on millions of hectares in North America over the last 30 years (Roelfs, 1988b). The tremendous gains in wheat production associated with the so-called green revolution in India and Pakistan would probably not have been realized without the protection from stem rust provided by *Sr2* in combination with other genes. The narrow genetic basic of durum and common wheat is a major constraint for the improvement of these crops (Feldman and Sears 1981). Therefore, it is of great importance to widen the genetic variation of desirable traits, particularly, those affecting yield and quality (Nachit 1998, 2000). Wild relatives of wheat, having a much wider range of genetic variation, could serve as an excellent source for improvement of such desirable traits. In fact, wild relatives hold rich pools of genetic variation and carry many genes of great economic potential (Feldman and Sears 1981). For this reason, many programs are now carrying out hybridization programs, based on interspecific or intraspecific crosses between wild species and cultivated wheats. For instance, the ICARDA durum-breeding project has mainly based its hybridization program on crosses between improved genotypes, Mediterranean landraces and wild relatives to improve and broaden the genetic base for resistance to biotic and abiotic stresses. Thus, landraces and wild relatives from the Middle East have been used to enhance drought tolerance, from Turkey and Algeria to incorporate cold resistance and from Morocco-Iberia region to improve resistance to root rot and Hessian fly (Nachit 1989, Nachit et al. 1995b).

To reach their research goals, many research projects either on genome sequencing or in population genetics and conservation biology have been run; and scientists have to analyze the dramatically growing amount of genetic data gradually produced by molecular techniques in the context of these projects. These works are generating a huge quantity of biological data, most of which are spatially located within a geographic context. In parallel with molecular approaches, it is highly desirable to apply a diversity of interdisciplinary (statistics, spatial statistics, molecular biology, phenotyping) approaches to understand such complex information. Geographic Information Systems (GIS) holds promise for being one of the appropriate ways to investigate genetic data from a point of view, which is somewhat unique to the traditional field of life sciences. The geographic attributes of molecular data are worthy of attention and consist of an alternative means of studying the variation of genetic diversity and of analyzing natural selection processes. Combining GIS with molecular genetics technologies will increase the power of the latter by exploiting the spatial dimension of the information they provide, proposing an alternative perspective that may lead to improved understanding of genomic functions. The visualization (exploratory spatial analysis) and the representation (cartography or thematic mapping) of spatially distributed genetic data are likely to highlight patterns of diversity and thus offer additional concrete support for interpretation. Furthermore, spatial analysis may allow the discovery of relationships between genome regions and properties of the environmental surroundings for the examined populations of plants.

Several and different works had been done to study the phenotypic diversity of the durum wheat landraces. Ahmadizadeh et al. (2010) studied the genetic diversity of 37 durum wheat landraces from Iran and Azerbaijan using multivariate analysis under stress and irrigation conditions. This study showed that under irrigated conditions biological yield, awn length and harvest index showed more direct positive effects on yield. In drought stress condition, biological yield, spike length, number of grains per spike and harvest index showed more direct positive effects on yield. Araus et al. (2007) found a significant relationships between phenotypic variation among landraces from the Middle Euphrates and both minimum temperatures and the ratio of precipitation to potential evapotranspiration of the sites of origin. In addition, consistent differences in grain yield, plant structure, and water status were found among genotypes following both north–south and east–west gradients across the Mediterranean. Moraguees et al. (2006) demonstrated that the origin of landraces influenced biomass production. Landraces from the north side of the Mediterranean basin produced 19% more tillers than those from the south, resulting in larger

biomass and leaf area allocation on tillers at anthesis. Southern landraces showed a better adaptation to drought environments. Also, the yield components differed also depending to the origin of the durum landraces and that yield components had a strong or weak correlation depending if the landraces are originated from the north or the south. The ecological and anthropological causes may have played a role in the creation of the observed variation using 11 spike characteristics using durum wheat landraces from Algeria, Ethiopia and Italy (Spagnoletti et al., 1984).

2. 2. Genotyping

2. 2. 1. Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a technique used to amplify small segments of DNA. This molecular biology method was developed in 1985 by Kary Mullis. Small single-stranded segments of DNA made of 20-30 nucleotide bases (oligonucleotides) are synthesized in vitro in order to be correctly bound to opposite strands of the DNA segment it is wished to replicate. At the points of contact an added enzyme(DNA polymerase) can start to read off the nucleotide sequence and, through bases complementarily, synthesizes a new sequence until two new double strands of DNA are formed. The sample is then heated, which makes the strands separate so that they can be read off again. The procedure is continuously repeated, doubling at each step the number of copies of the desired DNA segment. Through such repetitive cycles, it is possible to reach millions of copies of the desired DNA segment within a few hours. According to the common approach, nucleotides provided to start the reaction are radioactive to make it possible to distinguish the different alleles by autoradiography after electrophoresis. Since a few years, radioactivity is progressively replaced by fluorescent labeling. The PCR technique is presently used in numerous molecular genetics applications: Random Amplified Random DNA (RAPD), Amplified fragment length polymorphisms (AFLP), Sequence Specific PCR Based Markers, Microsatellites or Simple Sequence Repeats (SSR) and Single Nucleotide Polymorphisms (SNPs). Molecular markers are subject of continuous technical advancement and evolution. Most of markers are used in genetic diversity studies and the assessment and maintenance of genetic diversity, through the use of molecular markers is crucial as it provides a repository of adaptability to environmental and other changes (Mondini et al. 2009).

2. 2. 2. Microsatellite markers

Microsatellites are stretches of DNA that consist of tandem repeats of sequences of *mono*, *di* or *tri* nucleotides which are repeated between 10 and 20 times (for example, the frequent TG di nucleotide repeated 15 times in succession) and have no known coding function. These sequences are numerous, regularly distributed over the genome and characterized by an important polymorphism due to the variation of the number of repeats from an allele to the other. Using PCR, these repeats can be easily amplified. The number of repeat units that an individual has at a given locus can be resolved using a polyacrylamide gel whose high resolution permits a distinction of alleles whose size is one base pair different. From the gels, it is generally possible to perceive two genetic marks (alleles) for individuals as each one is inheriting one length of nucleotide repeats from his mother and one from his father and are thus considered co-dominant. Individuals with only one band have in fact received the same allele from both their mother and father. An important condition to use microsatellites in an efficient way is to make sure that the considered locus is unique. To check for it, flanking sequences on both sides of the locus have to be the same. Microsatellites are highly variable. In a population, many alleles of a single microsatellite locus, different in the number of repeats, may exist (up to 70 at a single locus). Moreover, microsatellite alleles change rapidly over time (Smith and Gaffney, 2000), evolving over time, from generation to generation. That is a reason why they are used to detect recent changes in population like effects of population fragmentation. Microsatellites are also useful for the identification of incipient differentiation of populations.

2. 2. 3. Alleles

An allele is likely to play a marker role only if it can be distinguished from other alleles. Moreover, within a population, a marker is likely to be useful only if the variety is heterozygote at the location of this marker. This is of course because for a homozygote variety, the marker provides no information to distinguish two types of descendants. And even in the case for which the father, the mother and the offspring are heterozygote (A/a), the marker is not providing information. The efficiency of a marker is assessed according to its unambiguous ability to distinguish two descendants groups according to a marker allele. A co-dominant (SSR for example) marker is a marker for which all alleles can be merely deducted from the observation of the phenotype. It is providing more information than a *dominant* (AFLP for example) marker whose recessive allele can be observed only when homozygote. A marker is providing the more information when the number of alleles is high and their frequencies are balanced. This is why highly polymorphic co-dominant markers are checked. A system to increase the information provided is to consider a group of narrowly bound markers as a unique marker called haplotype, and whose polymorphism is the result of the allelic combination of each basic marker (Crow, 1986; Suzuki, 1991).

2. 2. 4. Linkage disequilibrium

Alleles are said to be in linkage equilibrium if the frequency of a particular genotype is equal to the product of the frequencies of the individual alleles that make up the genotype. A natural way to measure the deviation from linkage equilibrium is to compare the observed and expected genotype frequencies and this is what is called linkage disequilibrium (LD):

$$D_{AB} = p_{AB} - p_A p_B$$

where D_{AB} is the coefficient of LD, p_{AB}, p_A and p_B are the frequencies the haplotype AB, allele A and allele B respectively.

The term linkage disequilibrium is actually an inappropriate name for deviations from this expectation as physical linkage between loci is neither necessary, nor sufficient to generate associations. LD is often due to the fact that a genetic link exists, but the reverse is not true and the existence of a genetic link doesn't imply LD. Linkage equilibrium is generally admitted as working hypothesis when considering a large closed population. Indeed, LD generally occurs through selection, migration, mutation, or genetic drift, and is gradually replaced by successive recombination in the course of generations. Consequently, each global linkage disequilibrium within a population is not stable and is existing only in the case of recent evolutionary processes (selection, mutation, migration, drift and admixture) or if loci are physically very close to one another. In this case, markers efficiency would be weak as the association between two alleles at two loci detected on a population's sample could not be generalized on the level of the whole population.

2. 2. 5. Association mapping

The phenotypic variation (observed) of many complex traits of many crops is influenced by multiple quantitative trait loci (QTLs), their interactions (epistasis; QTL.QTL), the environment (E), the environmental effect on QTLs (QTL.E) and on their interactions (QTL.QTL.E). Linkage analysis and association mapping are the most used methods to dissect complex traits. The traditional method to identify QTL in plants involves developing a segregating population from two genotypes (parents) varying in phenotypic values from a trait of interest, following extensive genotyping and phenotyping, significant marker-trait associations are identified. Although this method identifies genomic regions associated with traits for which the populations were developed. Furthermore, QTL identification is

limited to loci that differ between the parents, and unless large populations are used, QTL with small effects are not detected (Reimer et al., 2008). Different methods are used to identify QTL using bi-parental populations: Approximate methods including markers regression, Haley-Knott and its extended version regressions and composite interval mapping. Exact methods such as interval mapping, multiple interval mapping, multiple imputations and Bayesian interval mapping.

Association mapping (AM) is a complementary strategy to QTL mapping to identify associations between genotype and phenotype (Yu and Buckler 2006), and takes advantage of this "historical" LD to identify marker-trait relationships. The basic objective of AM is to detect correlations between genotypes and phenotypes in a sample of unrelated individuals. This technique has been successfully employed in human and animal genetics (DeWan et al. 2006; Karlsson et al. 2007) where creating large populations of segregating individuals is not practical or feasible. Compared to linkage mapping in traditional bi-parental populations, AM offers several advantages: increased sampling of allelic variation, increased mapping resolution, and reduced research time (Buckler and Thornsberry 2002; Flint-Garcia et al. 2003; Kraakman et al., 2004; Aranzana et al., 2005). The majority of studies have found that simple sequence repeats (SSRs) or single nucleotide polymorphism (SNPs) are the markers of choice when performing association studies, as a result of their ability to detect genetic variability (Eujayl et al., 2001; Stich et al., 2006a). The high level of polymorphism that SSRs provide increases the power to detect LD and facilitates higher resolution mapping (Stich et al., 2006a). Under ideal situation, the basic statistics for association analysis would be ANOVA, t-test, chi-square test and linear regression. However, as the population structure can affect the association between a trait and a marker (or a phenotype and a genotype); different methods have been developed to deal with this important factor. Bradbury et al. (2007) implemented a general linear model (TASSEL) using population structure (Q) estimated using random markers. A unified mixed model analysis for association mapping accounting for different level of relatedness between used cultivars was developed by Yu et al. (2006). Patterson et al. (2007) and Price et al. (2006) proposed a fast effective way to diagnostic population structure and used it further as a correction for association studies using chi-square test.

2. 2. 5. Genetic diversity and structure

The amount and distribution of genetic diversity (GD) affect the evolutionary potential of species and populations (Futuyma 1998) which makes genetic diversity in natural populations of great interest. Genetic structure of a species can be applied to preservation of the evolutionary potential of the species, which is one of the goals of conservation (Godt & Hamrick 1998). Genetic diversity, including the variability of alleles and genotypes, is commonly used to describe the heritable variation in a population or species. The genetic diversity of plant species reflects their breeding systems. Also, fluctuations in the number and size of populations and their bio-geographic history may play critical roles in determining the current genetic composition of species (Hamrick & Godt 1996). GD of a population can be structured by spatial factors and by the genetic backgrounds of species. Structuring can exist at different scales, for example, among populations, subpopulations or neighboring individuals (Escudero et al., 2005). The spatial distribution of plants is a product of environmental influences, including human activities, life-history traits and past demographic histories of species (Knowles et al., 1992, Frankham et al., 2002). The genetic structure of plant populations is largely shaped by factors such as selection, spatial habitat structure, isolation by distance, social organization, mating system, gene flow, genetic drift, evolutionary history, life history, and other ecological and evolutionary factors at a wide variety of spatial and temporal scales (Loveless & Hamrick 1984, Avise 2004). When dispersal between populations is restricted, gene flow between them is reduced, resulting in high genetic structuring at the population level. Populations of nearly all species exhibit at least some degree of genetic differentiation across geography (Ehrlich & Raven 1969). It is a continuing challenge for scientists to describe population genetic architectures within species and identify the biological forces responsible for them. Considering genetics only, the study of spatial structures exist since a long time. Indeed, in 1931 Sewall Wright developed adaptation and

evolution models which were incorporating spatial distribution and distance considerations (Epperson, 2003). Distance between populations or habitats remains a central issue in spatial genetics as the main reference models in this discipline directly refer to, or are controlled by it (genetic isolation by distance, stepping-stone model and infinite- island model) (Epperson, 2003; MacArthur & Wilson, 1967). A lot of different statistics in which distance is playing a role were developed within spatial genetics (Epperson, 2003). For instance, the well-known Mantel test, developed in 1967, allows testing the association of one set of pairwise measures with another. This was applied to compare geographical with genetic distances (Epperson, 2003) to find out if distance from a source was likely to explain genetic diversity gradients defined by genetic distance.

Population structure results from selection and high levels of admixture (individual accession membership proportion found in multiple sub-populations) in a population and results in increased LD between unlinked markers (Nordborg and Tavare 2002; Cardon and Palmer 2003; Farnir et al., 2000; Rostoks et al. 2006). Population structure is often used in genetic studies to summarize relationships between Individuals within and among populations, and can provide insight into evolutionary relationships. The probability of a Type I error increases in AM studies if population structure is not accounted for (Flint-Garcia et al. 2003; Gupta et al., 2005). Several methods have been proposed for estimating population structure and modeling population structure in AM studies, including distance- and model-based methods (Pritchard et al., 2000a; Ahmad 2002; Lu et al., 2005; Yu et al., 2006; Camus-Kulandaivelu et al., 2007; Peleg et al., 2008). Distance-based estimates of population structure are generally based on clustering of individuals using pairwise genetic distance estimates between individuals (Nei 1972; Rogers 1972; Nei 1978). In contrast, model-based methods assign individuals probabilistically to one or more sub-populations (Pritchard et al., 2000a). The most common model-based approach is Bayesian modeling where allele frequencies are used to estimate the likelihood of an individual belonging to a particular subpopulation. This approach allows assignment of individuals to respective populations that can be integrated into statistical models to account for population structure in AM studies (Pritchard et al. 2000a). With Bayesian modeling, the number of sub-populations is usually estimated *a priori*. Often, known relationships (pedigree, origin of the individual) and/or genetic distance methods are used to estimate a realistic number of sub-populations for calculation of model-based assignments (Liu et al. 2003; Lu et al. 2005; Agrama et al. 2007; Chao et al. 2007; Hai et al. 2007).

Few studies were conducted on detailed population structure of durum wheat landraces. Earlier work of the ICARDA durum breeding program (Autrique et al. 1996) studied genetic diversity and measured genetic distance between durum wheat cultivars and some landraces of diverse eco-geographical origin using restriction fragment length polymorphism markers (RFLP). Maccaferri et al. (2005) studied the structure and Linkage Disequilibrium (LD) of an elite collection of durum wheat using STRUCTURE and TASSEL programs. High and low molecular weight glutenin and clustering method were used by Moraguees et al. (2006) to study the genetic diversity between 63 Mediterranean durum wheat landraces. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was used to separate white glumes, black awned, black glumes, and white awned, and classified wheat-like accessions among 56 accessions of durum wheat using SSRs (Duwayri et al., 2007). Wang et al. 2007 found 87 alleles in 25 primer SSRs using 60 durum wheat accessions from seven countries. They found more alleles were identified on the B genome than A genome. Zarkti et al. (2010) measured genetic distance and diversity of 23 Moroccan durum wheat accessions of which 17 were landraces by using only 7 SSRs; and assumed that the genetic variability found in durum wheat may be anthropogenic, geographical or environmental. Melnikova et al., 2010 studied the genetic diversity using gliadin coding loci was studied with 465 durum wheat accessions from 42 countries. This study could differentiate between three groups of accessions, south, north and advanced lines from international breeding centers (ICARDA and CYMMIT). Fifty microsatellites were used as molecular markers tool to determine the genetic structure and spatial adaptation of Moroccan (98) and Syrian (90) durum wheat landraces (Kehel et al., 2013) where Bayesian and Eigen methods were used to determine the genetic diversity and structure; and to analyze the effects of spatial factors. Neighboring

landraces tend to have close genetic profile. These results demonstrate the importance of the use of the spatial Bayesian and the easily usable Eigen methods to analyze the genetic diversity and to discriminate between the durum wheat landraces.

2. 2. 6. Spatial and non-spatial models in populations genetic

Recently, taking landscape information into account in genetic studies is of a growing interest (Manel et al., 2003). Among landscape features, space is most likely to influence the genetic structuring of a set of individuals or populations (Manel et al., 2004; Coulon et al., 2006). This structuring can exhibit different patterns, such as isolation by distance (Wright, 1943), clines (Haldane, 1948), meta-populations (Hanski and Simberloff, 1997; Kerth and Petit, 2005) and barriers to gene flow (Slatkin, 1985). There is strong concern, then, in including space in the analysis of genetic data. Spatial information, since a long time, can be used a posteriori for graphical display purposes (Bertranpetit and Cavalli-Sforza, 1991; Manel et al., 2004) or to measure spatial autocorrelation (Sokal and Wartenberg, 1983; Sokal et al., 1986; Bertorelle and Barbujani, 1995; Smouse and Peakall, 1999). Such methods are not properly designed to investigate spatial patterns of genetic data but may be useful to visualize and test for spatial structure. To investigate spatial genetic structures other than the most evident, a method should be spatially explicit. To be explicit, a method should directly take spatial information into account as a component of the model used. Such methods have been developed using different approaches. Dupanloup et al., (2002) developed the SAMOVA, the spatial analysis of molecular variance. Guillot et al., (2005) GENELAND the Bayesian clustering framework, and, François et al., (2006) a hierarchical Markov random field (HMRF) model. The last two programs were proposed as improvements of STRUCTURE (Pritchard et al., 2000; Falush et al., 2003) by integrating geographic information to infer the number of populations and detect the genetic discontinuities among these populations (Coulon et al., 2006). Manel et al. (2007) proposed a method to detect genetic boundaries among multilocus genotypes. Another, maybe more concerning, issue with these methods resides in the clustering approach itself: assigning individuals to groups is a likely inappropriate strategy when individuals are genetically structured as a cline. A last approach would be to use a Mantel correlogram (Legendre and Legendre, 1998) to assess the variation of spatial autocorrelation in allelic frequencies across scales. An alternative for exploring genetic data is offered by ordination methods (such as principal component analysis PCA) because their utilization is not contingent on a particular genetic model. Hardy–Weinberg equilibrium or linkage equilibrium are thus no longer required. Basically, these methods aim at summarizing strongly multivariate data into a few uncorrelated components, forming the so called 'reduced space'. For this summary to be meaningful, the components are chosen so as to reflect most of the variability in data. Such methods can be applied on allelic frequency data to obtain a summary of the genetic variability among individuals or populations. A great illustration of such practice was offered by Menozzi et al. (1978), who used PCA to investigate the spatial patterns of the genetic variability, obtaining the well-known synthetic maps of human gene frequencies. More recently, PCA proved useful to correct for population stratification (Price et al., 2006) in AM study and to infer and test the number of subpopulations (Patterson et al., 2006). PCA seeks genetic variability, not spatial structures; it is not a likely optimal method for revealing spatial genetic patterns. Recently, a new tool for spatial pattern of genetic variability is developed called spatial principal components sPCA (Jombart et al. 2008), it is a modified PCA to still study the genetic variance between individuals taking into account their spatial autocorrelation. Two types of patterns are discriminated at sPCA: global and local structures.

2 .3 . Geographic Information Systems (GIS)

Geographic information systems (GIS) or geographic information science designates system that can store, manipulate, and analyze geo-referenced data. GIS are interdisciplinary, being a field that provides tools useful through their applications to solving problems within other disciplines. In this sense, GIS is merging cartography, spatial analysis, geostatistics, database management and software development. GIS

are considered to be applications-led technology (Longley et al., 2001). GIS consists of a two-sided field closely related to computer science, the GIS part (software, topology, databases, standards, formats, etc.), and of a collection of methods and models that explicitly use the spatial referencing of each data case, the spatial analysis (Goodchild and Haining, 2004).

Data in GIS representing real data (road, rivers, buildings, plant landraces, etc.) collected or measured directly in its environment or captured remotely (remote sensing, aerial photography). These data can be vector or raster. Vectors are the geometrical way of representing geographic features. Three different types of vectors can represent a geographical information: 1) Point (or events), which is an ordered pair (x, y) of spatial coordinates. A point indicates the place of occurrence of an event, like in the case of durum wheat landrace. 2) Polygon (or zone) which is a set of ordered pairs (x, y) of spatial coordinates, in such a way that the last point is identical to the first thus forming a closed region in the plane. It is covering an area. A country or a wheat field are polygons for example. 3) Line (or polyline) is set of ordered pairs (x, y) of spatial coordinates but representing linear features such as rivers, roads. On the other hand, raster is a matrix of rows and columns of cells where in each cell a unique value (usually between 0 and 255) is stored. Raster data can be images where each pixel (cell) has a color value. Raster can be continuous such as elevation in a digital elevation model (DEM) or discrete like for soil image or land use. These data are organized in general in database. The most used geographic database organization is the geo-relational model, that utilizes a relational database management system (DBMS) like *DBASE* or *ACCESS*, to store in its tables the attributes of the geographic objects, and separate graphic files to store the geometric representation of these objects.

GIS has been applied, for long time, for a multiplicity of uses in military, history, land survey, hydrology, archeology, anthropology, transportation, medicine, diseases surveillance, etc. For many years, GIS turned toward environmental modeling (Goodchild et al., 1993), generally concerned with explaining basic features of GIS to demonstrate how they could be efficiently applied to fields related to the natural sciences (Caloz and Collet, 1997). Because of this late reflection on what constitutes geographic information research, we are challenged by a need of integration of GIS and spatial analysis (Goodchild, 1992). This results in a gap between a trend of spatial data management for which geography is a mechanism for accessing information and whose works are technology-oriented, and a movement of spatial analysis interested in functionality and models for which geography has a fundamental role. The information management and business aspects are much more noticeable than the analysis one. The development of technologies naturally led to a GIS industry (software producers) narrowly involved together with academic GIS users. The dilemma about GIS is what to consider it as: science or business. GIS was mainly disseminated by the Environmental Science Research Institute (ESRI), which is a pure business company. ESRI contains research and science in its name but its objectives are essentially software industry. GIS science needs to empty itself from software production. The availability of GIS open source applications is probably to advance the situation that spatial analysis are more important than business. In this context, GIS are not only tools: their use belongs to a wider group of specific knowledge, which have spatial information in common and are unified within GIS.

Until now, application to genetics has been very rare. Despite its current predominance in life sciences, and its direct application to concerns of public society (health, food), genetics had until lately remained outside the scope of GIS research. In contrast, from the end of the 1960s on, biologists gradually appropriated GIS tools, mainly in ecology. Only since the mid-1990s, population geneticists and molecular biologists began to make use of GIS to try to understand how geographical and environmental features influence the structure of genetic data. The molecular biology and GIS may facilitate novel and complementary methods of dealing with some of the issues related to evolutionary processes. Power of the latter by exploiting the spatial dimension of the information they provide, proposing an alternative perspective that may lead to improved understanding of genomic functions. The visualization (exploratory spatial analysis) and the representation (cartography or thematic mapping) of spatially distributed genetic data are likely to highlight patterns of diversity and thus offer additional concrete

support for interpretation. Furthermore, spatial analysis may allow the discovery of relationships between genome regions and properties of the environmental surroundings for the examined populations of plants in general and durum wheat landraces in particular.

Analyses of genetic diversity of plants often consist in evaluating geographical patterns of diversity (biodiversity maps) generated from biological variables such as vegetation (McKendry & Machlis, 1991), or in habitat modeling (Jones et al., 1997). This is in fact the notion of biodiversity that evolved with the integration of genetic data and diversity to complement species diversity, ecosystem diversity and cultural diversity, which is determining how people interact with nature. This new dimension of biodiversity possibly reinforced the role of GIS, and especially the one of spatial analysis in the sense it multiplied in a phenomenal way the number of organisms' informative elements to be tested in relation to geographic and environmental information.

Arthur Mourant was the first to have the idea of making geographic maps of gene frequencies and to use them extensively (Cavalli- Sforza et al., 1994). Mourant (1954) led original works on blood groups and their hereditary clinical, social, and geographic patterns, he published in "*The Distribution of the Human Blood Groups*" which long was regarded as a revolutionary work. The study shown in "*The History and Geography of Human Genes*" of Cavalli-Sforza et al. (1950) proposed through mapping the worldwide geographic distribution of the genes an explanation of the understanding of how humans left Africa and populated the rest of the world, and also to the detecting of antique migrations, as for example the migration of Neolithic farmers from the Middle East towards Europe. The authors represented spatially the proportion of a given allele found in a population between several indigenous populations. They used 110 traits such as blood types, proteins and DNA markers. The spatial presentation was done by presenting the frequency of the alleles on maps according to the locations where the studied populations were sampled, and the points of equal gene frequencies were connected by "isogenic" curves. Two analyses were possible: mapping alleles is practical to understand evolutionary history of an allele; the correlation of allele frequencies with environmental parameters can be determining to discover specific genetic adaptations.

Smoothing or interpolation was then used as a spatial analysis. Interpolating surfaces was used in other genetic works to define specific genetic diversity: Bucci and Vendramin, (2000) to delineate genetically homogeneous regions and predict haplotype frequencies; and Hoffmann et al. (2003) define *Arabidopsis thaliana* areas of similar diversity across Europe based on nucleotide diversity. Hamann et al. (2000) exploited ordinary Kriging to predict performance of seed sources at un-sampled locations. They suggested exploring the composition of the environment constituting the dispersal zones (using temperatures and precipitations) to test if the genetic differentiation would fit the ecological one. This perception, directly related to what was exploited by Skøt et al. (2002) in their investigation of the interaction between environmental characteristics of a forage grass (*Lolium perenne*) and its molecular information. To increase the efficiency of breeding according to a given interesting property, marker-assisted selection was studied as a potential tool with the aim to understand the ability of *Lolium perenne* to survive and grow at low temperature, to acclimate to cold, to tolerate wind, snow cover, and ice encasement. Six AFLPs markers were identified to be involved in the resistance to cold. In addition, GIS was used in this study to display plants locations and to retrieve corresponding environmental variable values available on separate data layers. AFLPs markers were also used to show association with salt tolerance in wild barley (Pakniyat et al., 1997) applying the same methodology.

One way to consider gene-environment interaction (apart the classical GE interactions used in breeding) is to study the influence of the environment on the genome and try to understand how geographical and environmental features affect genetic structure. The landscape genetics is then created by David Galbraith (19??) of the Royal Botanical Gardens as the placement of genetic diversity into a spatial framework. This concept was adopted by several institutions studying genetic diversity by using GIS to analyze the geographical distribution of different genetic markers. One important work making precise definition of landscape genetics is "*combining landscape ecology and population genetics*" by Manel et al. (2003).

Landscape genetics is likely to facilitate our understanding of how geographical and environmental features structure genetic variation at both the population and individual levels, and has implications for ecology, evolution and conservation biology. This made possible the integration between spatial statistics, GIS and molecular markers. Since then several works used this approach (Hirao and Kudo, 2004; Watts et al., 2004; Spear et al., 2005). Spear et al. (2005) attributed much importance to GIS tools and concluded that GIS analyses should be added to the field of landscape genetics to examine the extent to which landscape features influence genetic structure. All previous works constitute full and direct recognition of GIS tools and methods' role in the framework of the analysis of genetic information in a spatial context. The importance here conferred to the management and the analysis of geographical information makes GIS a compulsory component of landscape genetics, together with molecular genetics and ecology. It is very difficult to list all the works combining GIS and genetics because most of these studies are published under the form of project reports, the most important is that the involvement of GIS in molecular genetic studies is increasing Joost et al. (2005). Kidd and Liu 2008 defined a 'geophylogeny' as a data structure within which phylogenetic and geographical data and models are explicitly linked. The developed 'geophylobuilder 1.0', an extension under ArcGIS to create geophylogeny from a tree and the associated geographical information.

Proches 2006 presented a principal component derived maps to generate latitudinal and longitudinal barriers in biogeography. Faleiro et al., 2008 used molecular markers and GIS to study native plant species. The GIS was used to plot accessions and made it possible to identify areas with great diversity. Cercueil et al., 2007 introduced new visual tool for investigating spatial variation of allele frequencies. They developed software called GENBMAP with the framework of the Wombling methods. The method is generally able to locate genetic boundaries or clines precisely. Manel et al., 2007 proved a method based on assignment tests applied in a moving window over an extensively sampled study area. For each individual, a spatially explicit probability surface is constructed, showing the estimated probability of finding its multilocus genotype across the landscape, and identifying putative migrants. Population boundaries are localized by estimating the mean slope of these probability surfaces over all individuals to identify areas with genetic discontinuities. At the university of Alberta, Canada they developed a macro under Arcview 3.2 using Avenue to run the spatial allele frequencies using the dominant marker data. Engler from Lausanne University developed a new tool to simulate the future distribution of species in the context of global warming. He used Arcmap GIS using ArcObjects. Linear relationship between genetic and geographic distance in a worldwide sample of human populations was found. A close relationship was shown to exist between the correlation of geographic distance and genetic differentiation (as measured by *F*ST) and the geographic pattern of heterozygosity across populations (Ramachandranet al., 2005). The spatial prediction of species distribution is an important tool for the conservation and management of the biodiversity. It uses a wide variety of statistical approaches together with geographic information systems (GIS) (revision in Austin, 2002).The gradient across different peninsular regions lead to postulate that it has a natural origin (Cánovas et al., 2004; De la Rúa et al., 2004). The northwards expansion of *A. m.mellifera* from Iberia to NW Europe, after the last glaciation period, was subsequently followed by the spreading out of *A. m. intermisa* from northern Africa to the Iberian Peninsula. According to this hypothesis the gradient should be found across the whole Peninsula and its nature should be explained in relation to natural factors as are climatic parameters and physical barriers (Canovas et al., 2008) - The predictive power of Generalized Linear Models (GLM) versus Canonical Correspondence Analysis (CCA) models of plant distribution in the Spring Mountains of Nevada, USA, are compared. Results show that GLM models give better predictions than CCA models because a species specific subset of explanatory variables can be selected in GLM, while in CCA, all species are modeled using the same set of composite environmental variables (axes). Although both techniques can be readily ported to a Geographical Information System (GIS), CCA models are more readily implemented for many species at once. Wagner et al. (2005) used variogram approach to analyze spatial genetic structure of populations using microsatellite data. This permitted to estimate the population genetic diversity and provide the spatial genetic structure accounting for autocorrelation. McVean (2009) provided a framework for

interpreting PCA axis in term of underlying geographical isolation and admixture. He also demonstrated a link between PCA and Wright's F_{st}. Kato and Yokoyama (1992) studied the Geographical variation in heading characters among wheat landraces. They found that almost 50% of the variation of a trait is explained by the geographical difference in origin using 158 wheat landraces. The difference among localities indicated that wheat landraces had been selected for early heading as an adaptation strategy to water stress and/or high temperature in early summer. Iwaki et al. (2001) studied the geographical variation of growth habit of 749 landraces from various parts of the world, with special reference to their adaptation and eco-geographical differentiation they found out that geographical variation of growth habit is closely related to the degree of winter coldness.

3. Material and Methods

3. 1. Durum wheat collection

We utilized ninety eight (98) durum landraces from Morocco and ninety (90) from Syria, representing the two countries' durum collection and representing the main Mediterranean environments: continental, temperate, and high altitude areas (Figure 6 and 7). The collections were executed by the genetic resources unit (GRU) of ICARDA in 1985 and 1987 for Morocco and Syria, respectively. During the collection missions, topographic data (Latitude, longitude and altitude) were recorded for each location. Physical address and the closest village were also registered (Supplementary table).

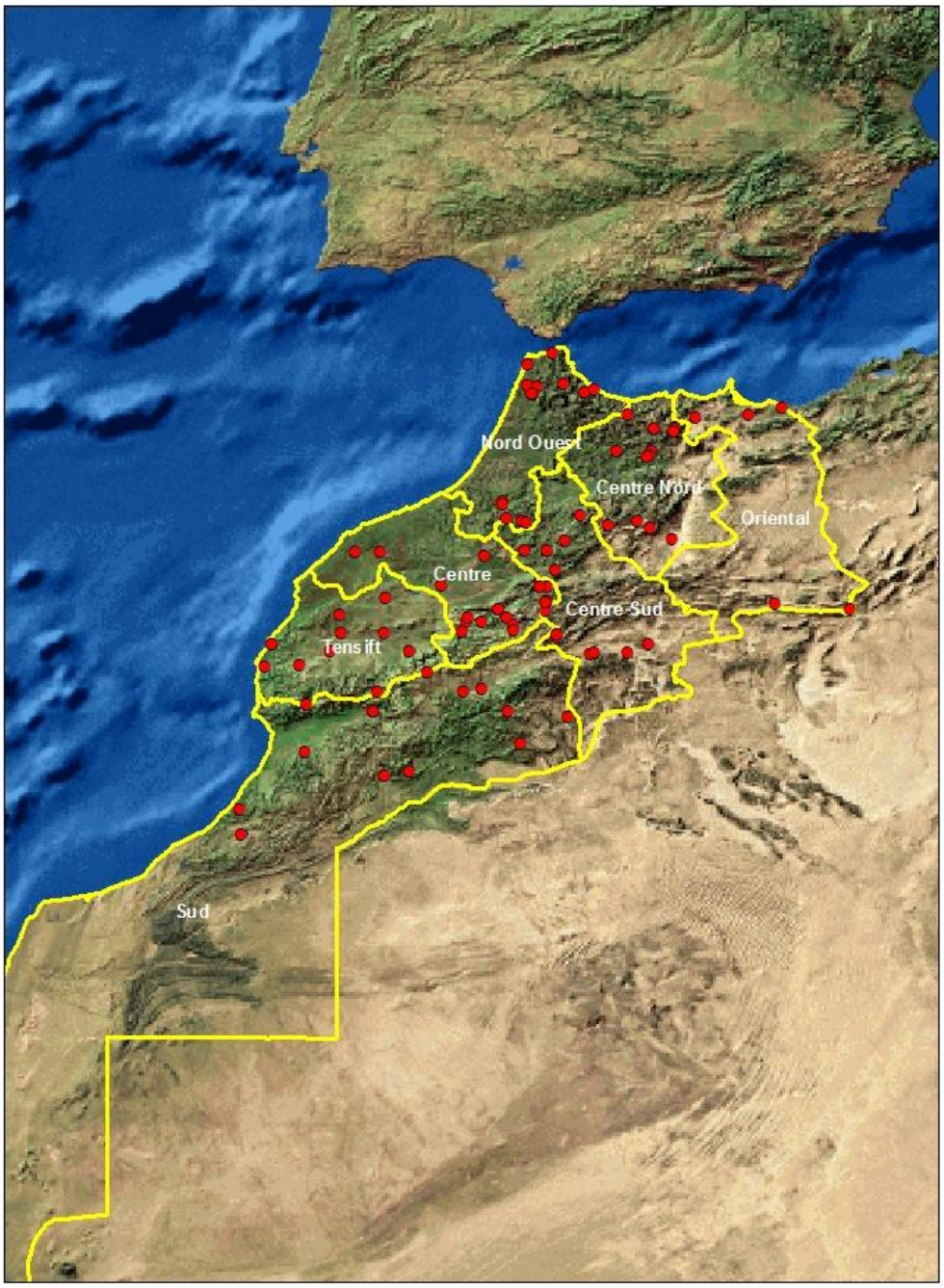


Figure 6: Distribution of Moroccan Durum landraces

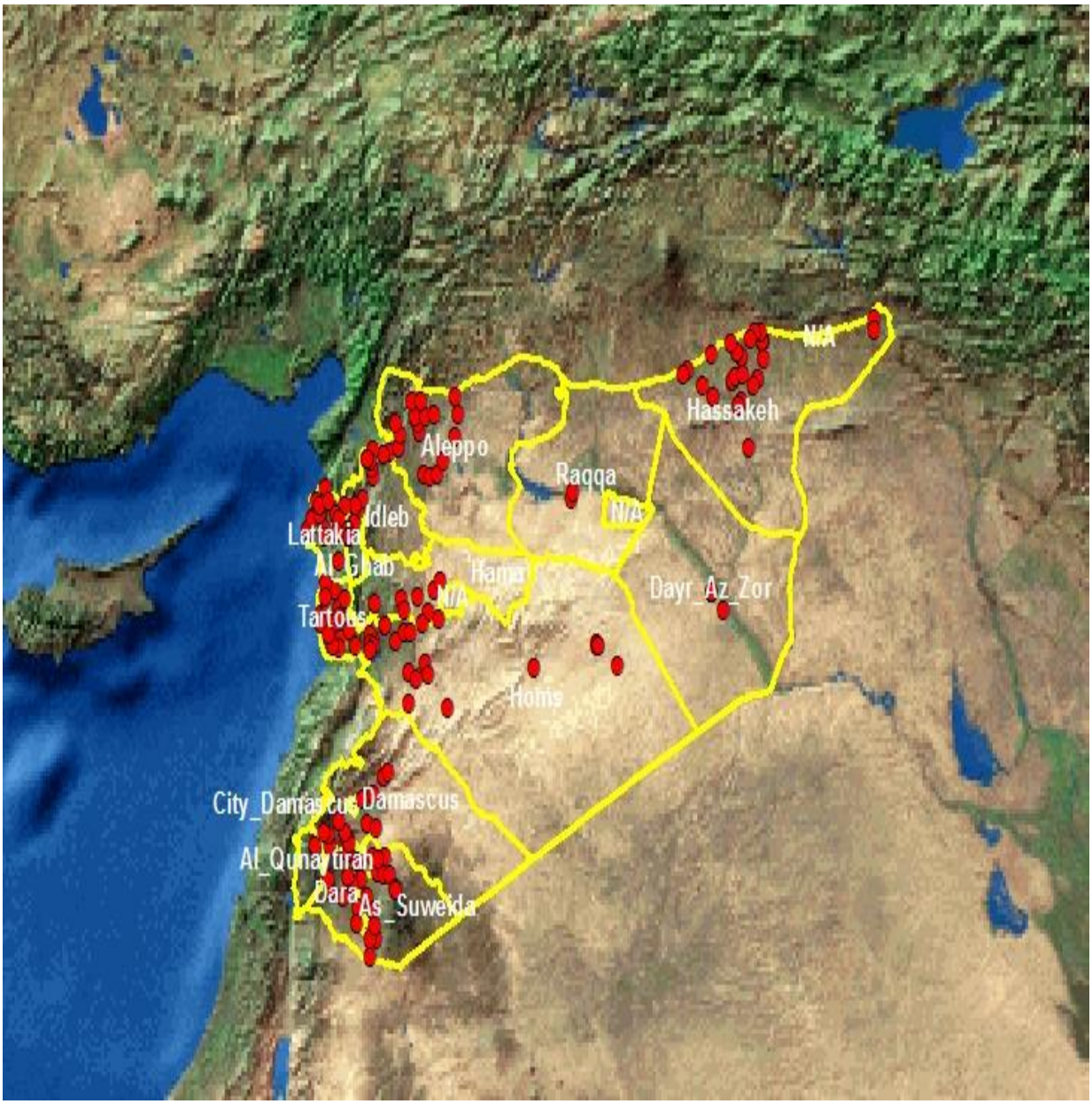


Figure 7: Distribution of Syrian durum landraces

3. 2. Field evaluation

3. 2. 1. Phenotypic traits

The two collections were evaluated for agronomy, quality, yield and yield components. The phenotypic data recorded by the GRU during the collection year was also used in the analysis. For the experimental design and data analyses; we used the augmented design AD (Federer 1956) with 5 checks (Omrabi5, Haurani, Korifla, Waha, and Gidara2). The AD is based on 10 blocks. Each trial had a total of 240 plots (N = (10 "Blocks" × 5 "Checks") + 188 "Landraces") arranged as a grid layout of 20 rows by 12 columns. Each block contained 24 plots, including all five check lines, and comprised a pair of adjacent rows in the layout.

The following observations on inner two rows basis were taken:
* Days to heading (days): number of days from emergence to the day when half of the spikes have appeared in 50% of the plants (**DH**).
* Days to maturity (days): number of days from emergence to the day when the peduncle was completely discolored in 90% of the plants (**DM**).
* Grain filling development period (days), **GFD** = **DM** - **DH**.
* Plant height (cm): plant height was measured from the ground level to the top of the spikes excluding awns (**PH**).
* Number of tillers: from inside rows, mean number of fertile tillers per one meter were counted and converted to square meter (**SPM2**).
* Peduncle length (cm): length from the last stem node to the base of the spike (**PL**).
* Spike length (cm): length was measured from the base to the top of the spike excluding the awns (**SL**).
* Number of grains per main spike (**KSPK**).
* Thousand grain weight (g): 200 grains were taken randomly from the harvested grain and converted to the weight of 1000 grains (**TKW**).
* Grain yield ($Kgha^{-1}$): the whole plot was harvested by hand and threshed, then cleaned and the grains were weighed (**GY**).
* Sedimentation (**SDS**). Sedimentation test is a method to estimate the strength of wheat gluten; it is based on the hydration capacity of flour in a low acidity media.

Gluten strength; sedimentation test (SDS ml**),** for measuring the gluten strength sedimentation test (ml) was done according to the method of Pena et al. (1990). Few grams ground by UDY cyclone grinder. One gram of the ground sample was shacked in the presence of lactic acid and sodium dodecyl sulfate. The height of the suspension after a standard shaking procedure and standing period is directly measured. The sediment height > 50 (ml) evaluated as very strong and the < 20 (ml) evaluated as very weak.
* Sedimentation index (**SDSI**): SDSI = SDS / Protein%
* Sedimentation n (**SDSn**) : SDSN = (SDS x Protein %) / 100 (Nachit et al. 1992). Firmness is the force required to cut cooked pasta. Good quality pasta and couscous should have the correct firmness or chewiness after cooking or steaming, respectively. The SDS index is used as surrogate for firmness test.
* Protein content (**PC**). Generally, high protein content is associated with good pasta, burghul, and couscous making values. The protein content was conveniently determined in all cereals by Near-Infra-Red (NIR) of the reflectance spectrometry, due to its rapidness and accurateness.
* Vitreousness (**VIT**). A high value for vitreousness is related to high semolina extraction. The vitreousness is expressed as percentage of vitreousness and it is determined visually. The vitreous kernel has to be 100% free of yellow berry sections.
* Yellow pigment (**YP**). The color of durum wheat is more or less yellow or amber; and is caused by the presence of carotenoid pigments, mainly xanthophylls and lutein. Yellow pigment was estimated according to AACC (1995): by extraction of pigments from the ground durum grains using water-saturated n-butanol, for overnight and the transmittance measured by direct spectrophotometer at 440nm.

Also flour color can be estimated visually in the semolina, or instrumentally by reflectance spectroscopy (NIR).

3. 2. 2. Physiological traits

For physiological traits, Spectral reflectance measurements were taken using a portable field spectro-radiometer (FieldSpec UV/VNIR, Analytical Spectral Devices, Boulder, CO). The spectro-radiometer was capable of measuring radiance from 350 to 1050 nm wavelengths with a sampling interval of 1.4 nm of the spectrum. Thus, 512 continuous data points were obtained with each reading. Measurements were taken during the middle of the day on cloudless days. The optical sensor was placed approximately 50 cm above the plant canopy in nadir position. The incident spectrum was taken from the light reflected from a white reference panel, and reflectance was calculated from the ratio of reflected light from the crop canopy against the total radiance reflectance from the white surface. According to the criteria developed at ICARDA durum breeding program (Motawaj 2007) four spectral measurements at two stages: Zadok 45 and 70 were taken randomly from four different places per genotype, and the mean of the four readings was used to calculate the spectral reflectance indices as following:

1. Chlorophyll content: **CHL** = R_{670} / R_{800}.
2. Water index: **WI** = R675/ R_{680}.
3. Carotene content: **CAROTENE** = R_{675}-R_{680}.
4. Chlorophyll Absorption Ratio Index: **CARI** = R703/R_{657}.
5. Soil Adjusted Vegetation Index (It is a modification of the index NDVI in order to compensate for the effect of soil): **SAVI**=[(R_{770}-R_{660})/R_{770}+R_{660}+L)](1+L).
6. Red-edge Vegetation Stress Index: **RVSI**= ((R_{718}+R_{748})/2)-R_{733}.
7. Ratio Nitrogen Vegetation Index: **RNVI**= ($R7_{62}$/R_{550}).
8. Relation of Carotene/Chlorophyll (Structural Independent Pigment Index): **SIPI**= (R_{800}-R_{435})/(R_{435}+R_{800}).
9. Relation of Carotene/Chlorophyll (Normalized Pigment Chlorophyll Index): **NPCI**=(R_{680}-R_{430})/($_{R430}$+R_{680}).
10. Chlorophyll Degradation (Normalized Phaeophytinization Index): **NPQI**=(R_{415}-R_{435})/(R_{415}+R_{435}).
11. Biomass (Simple Ratio): **SR**=R_{770}/R_{680}.
13. Photochemical Reflectance Index: **PRI**= (R_{531}-R_{570})/(R_{570}+R_{531}).
14. Ratio of **WI/NDVI**.
15. Yield of Photochemical Energy Conversion: **YPEC**= (F5-F1)/F5
16. Normalized difference vegetation index : **NDVI** = (NIR-VIS) / (NIR+VIS)

Where: VIS = the spectral reflectance in the visible wavelengths (680 nm) and NIR = the spectral reflectance in the near infra-red wavelengths (770 nm).
R is the spectral reflectance at X wavelengths.

17. **F0**= Minimal fluorescence.
18. **F1** = Fluorescence at first time.
19. **F2** = Fluorescence at 2nd time.
20. **F3** = Fluorescence at third time.
21. **F4** = Fluorescence at fourth time.
22. **F5** = Fluorescence at fifth time.
23. Maximal fluorescence: **Fm.**
24. Variable fluorescence: **Fv.**
25. Time for maximal fluorescence: **Tfm.**
26. Leaf water potential related to drought res.: **LWP**=Fm/F0.
27. Photochemical efficiency ratio: **Fv/Fm**.
28. Non Photochemical Quenching : **NPQ**=(Fm-F5)/F5.
29. Photochemical Quenching: **QP**= (F5-F1)/(F5-F0).

30. Non Photochemical Quenching: **QN**= (Fm-F5)/(Fm-F0).
31. Photochemical quenching: **Que**=F0/Fv = (Fm-Fv)/(Fm-F0).

3. 2. 3. Growing environment

The planting was conducted at Tel Hadya which was the main research station of ICARDA. Tel Hadya has a Mediterranean continental climate with average annual precipitation of 335 mm. It is at 35 Km south west of Aleppo city/Syria and located at 36°01' N latitude; 36°56' E longitude, and at 284 m above the sea level. The soil at Tel Hadya is fine to very fine clay. This station is characterized by the following climatic conditions: wet and cold in winter and warm and dry summer, a typical Mediterranean climate. Climatic data of 2004, 2005, 2006 and 2007 are given in graph (Figure 8):

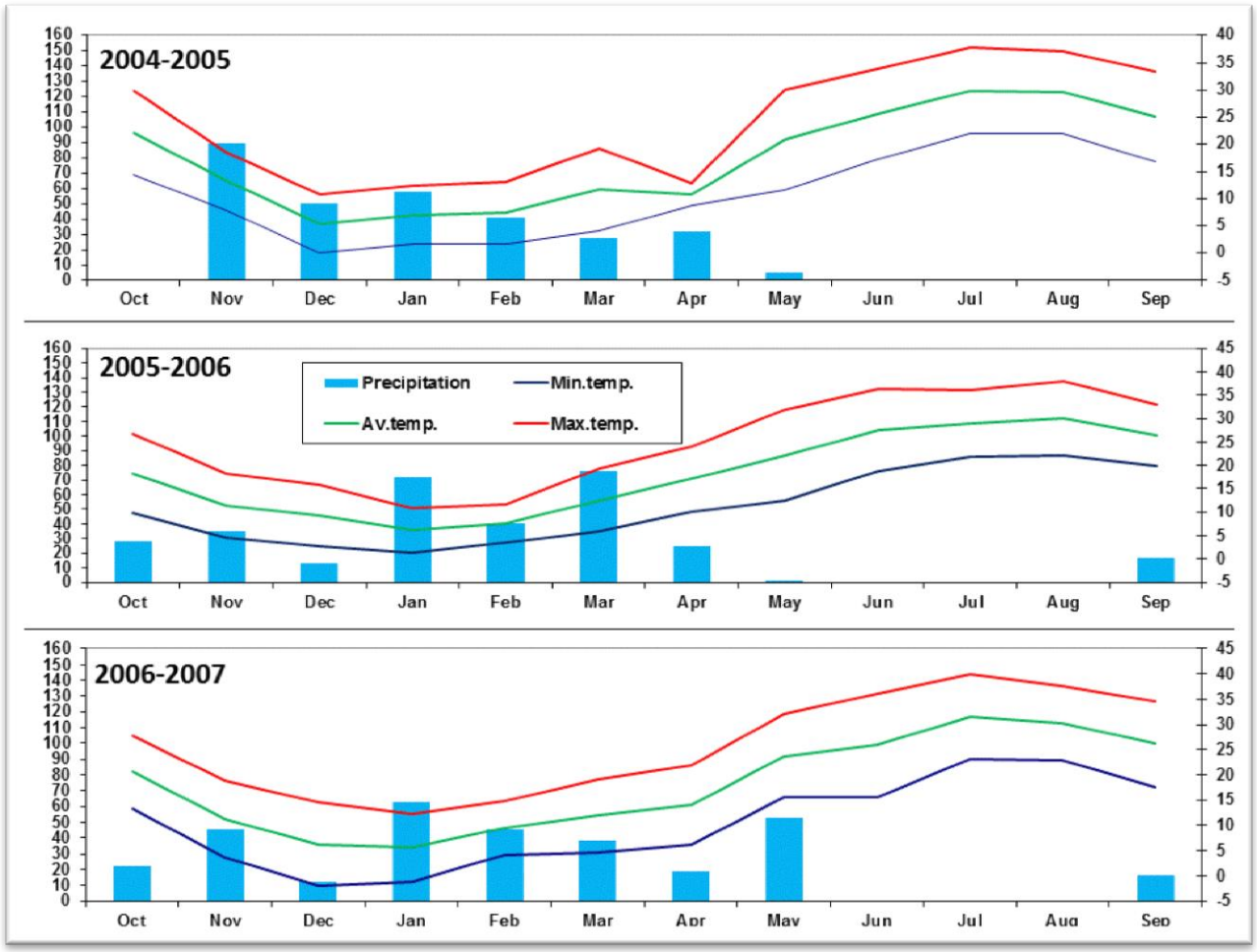


Figure 8: Climatic profiles for Tel Hadya experimental station during the three years of evaluation

We planted our landraces in rainfed which is representative for the Mediterranean continental dryland. This rainfed environment is used at the durum breeding program at ICARDA to screen for adaptation,

drought and the other associated biotic and abiotic stresses prevalent in the Mediterranean region. The date of sowing is usually mid- November and of harvesting is around mid-June.

Evaluation was made during four (4) years (2004, 2005, 2006 and 2007); plus 1985 or 1987 which were the years of the preliminary evaluation after collection at ICARDA Germplasm Resources Unit (Table 1). Three traits were recorded in five (5) environments or years (PC, VIT and TKW) and three (3) traits recorded at only one environment (SL, PL and SPM^2). Grain yield (GY) was measured in four (4) different years.

Table 1: Measured morphological traits

	Col	2004	2005	2006	2007
ASH		x	x	x	x
DH	x	x	x		
DM	x		x		
GFD	x		x		
GY		x	x	x	x
KSPK	x				x
PC	x	x	x	x	x
VIT	x	x	x	x	x
PH	x		x	x	x
SDS		x	x	x	x
SDSn		x	x	x	x
SDSI		x	x	x	x
TKW	x	x	x	x	x
YP		x	x	x	x
SPM2					x
PL					x
SL			x		

For physiology, ten (10) traits were recorded during three years 2006, 2007 and 2008 at the Zadoc scale 70. Fourteen (14) traits were scored at the Zadoc's scale 45 during only the 2008 season. The fluorescence was measured only during 2006 and at Zadoc's scale 70 (Table 2).

Table 2: Measured physiological traits

Zadoc scale	**45**		**70**		
Year	**2008**		**2006**	**2007**	**2008**

Area			**x**		
CARI	**x**			**x**	**x**
CHL	**x**		**x**	**x**	**x**
CAROTENE	**x**		**x**	**x**	**x**
F0			**x**		
F1			**x**		
F2			**x**		
F2			**x**		
F3			**x**		
F4			**x**		
F5			**x**		
Fm			**x**		
Fv/Fm			**x**		
Fv			**x**		
LWP			**x**		
NDVI	**x**		**x**	**x**	**x**
NPCI	**x**		**x**	**x**	**x**
NPQ			**x**		
NPQI	**x**		**x**	**x**	**x**
PRI	**x**		**x**	**x**	**x**
QN			**x**		
QP			**x**		
Que			**x**		
RNVI	**x**			**x**	**x**
RVSI	**x**			**x**	**x**
SAVI	**x**		**x**	**x**	**x**
SIPI	**x**			**x**	**x**
SR	**x**		**x**	**x**	**x**
Tfm			**x**		
WI/NDVI	**x**		**x**	**x**	**x**
WI	**x**		**x**	**x**	**x**
YPEC			**x**		

1

2

The field collected data was adjusted for field heterogeneity using the block adjustment method as described in Petersen (1985); this consisted on a mixed model where checks lines, landraces and blocks effect are considered as fixed effect.

3. 3. Genotyping

The DNA was extracted following the protocol developed at ICARDA durum wheat MAS laboratory. Briefly, 3-5 gm of leaf tissue per sample (each sample was collected from each landrace seedling 8 weeks after sowing) were ground in liquid nitrogen and incubated at 60 °C for 30 min with 5 volume (ml) of extraction buffer to 4 tissue volume, (100 mM Tris-HCl, 500 M NaCl, 50 mM EDTA, 1.25% SDS) in 15 ml polypropylene tubes. After cell disruption and incubation with hot isolation buffer, proteins were removed by chloroform: iso-amyl alcohol (24:1, v:v). Samples were incubated for 30 min by shaking and then centrifuged at 2800 rpm for 15 min. The aqueous layer was transferred to a new tube and 0.3% (v:v) of a 10 µg/ml of stock solution of RNAse A was added. Samples were incubated for 30 min at room temperature. One volume of cold ethanol (at -20°C) was added to DNA. After 30 min incubation -20°C, precipitated DNA was hooked out and placed in a 2 ml reaction tube containing 1 ml of 70% ethanol. After washing twice with 70% ethanol, the washing solution was removed and the DNA pellet was dried thoroughly and dissolved in 1% TE buffer. The DNA samples were diluted and stored at -20 °C. A DNA/RNA calculator was used to measure DNA concentration and purity (Nachit et al., 2001; Elouafi and Nachit, Motawaj, 2007 (PhD Thesis).

The two collections were genotyped by 53 Gatersleben wheat microsatellites (gwm), obtained from Röder et al.,(1995, 1998) from a conventional genomic library, distributed along the 14 chromosomes of the durum genome (Table 3). For that we used a DNA extraction protocol used at the durum wheat MAS laboratory as explained later in this chapter. We utilized ABI377 and each of the gels contains ninety six (96) landraces and two (2) checks genotypes. Samples were electrophoresed in an automatic DNA sequencer (ABI 377, Applied Biosystems). The ABI 377 is equipped with *GenScan* 3.0 software (Applied Biosystems) for data collection and fragment-size (bp) calculation to two decimals. Electropherograms obtained by *GenScan* 3.0 from the gel images were scored for allele size. Alleles were attributed according to the fragments size in base pair (bp).

Table 3: List of used SSRs and chromosomes localization

ID	LOCUS	CHROMOSOME	ID	LOCUS	CHROMOSOME
1	**GWM2**	2AS,3AS	27	**GWM335**	5B
2	**GWM6**	4BL,5A	28	**GWM44**	4A
3	**GWM33**	1AS,1BL	29	**GWM357**	1A
4	**GWM60**	7AS	30	**GWM368**	4B
5	**GWM63**	7A	31	**GWM369**	3A,4B
6	**GWM99**	1A	32	**GWM376**	3B
7	**GWM107**	3B,4B,6B	33	**GWM408**	5B
8	**GWM114**	3B	34	**GWM410**	2B,5A
9	**GWM129**	2B,5AS	35	**GWM413**	1A,1B
10	**GWM160**	4AL	36	**GWM448**	2A
11	**GWM165**	4A,4BS	37	**GWM471**	7A

12	**GWM169**	6AL	38	**GWM480**	3A
13	**GWM210**	2A,2B	39	**GWM493**	3B
14	**GWM219**	6B	40	**GWM494**	1B,3A,4A,6A
15	**GWM234**	5A,5BS	41	**GWM518**	6B
16	**GWM257**	2B	42	**GWM526**	2A,2B
17	**GWM260**	7AS	43	**GWM537**	5B,7B
18	**GWM264**	1A,1B,3B,7B	44	**GWM601**	4A
19	**GWM268**	1B	45	**GWM610**	4A
20	**GWM282**	7A	46	**GWM611**	7B
21	**GWM285**	3B	47	**GWM614**	2A,2B,4A
22	**GWM293**	5A,7B	48	**GWM617**	5A,6A
23	**GWM297**	7BS	49	**GWM639**	5A,5B
24	**GWM311**	2A,6B	50	**GWM644**	1B,3B,6B,7B
25	**GWM319**	2B	51	**GWM666**	1A,3A,5A,7A
26	**GWM344**	7A,7B	52		

3. 4. The GIS interface

ArcGIS is the latest version of Environmental Research Systems Institute's (ESRI) suite of GIS products. ArcGIS is designed as a scalable system for geographic data creation, management, and analysis. ESRI products have a large user base. The ESRI website states that there are over 1 million ESRI software users worldwide, and that 50,000 university students receive instruction utilizing ESRI products every year (ESRI 2002).

In previous versions, ESRI's desktop GIS, ArcView, and its enterprise level GIS, ArcInfo, were very different in terms of the primary geographic data model and user interface. In ArcGIS, however, all of the products use the same data model, GUI interface, and development environment (Limp 2001). In addition to ArcView and ArcInfo, there is a medium sized version of ArcGIS called ArcEditor. Each variety of ArcGIS consists of several individual applications that provide a set of functionality. ArcMap is a primary component of all three versions and is the interface for data display and analysis. An analysis tool developed for ArcMap can be used in all versions of ArcGIS.

The term Geographic Information System (GIS) is applied to systems that perform the computational treatment of geographic data and that store the geometry and the attributes of data that are geo-referenced, that is, situated on the earth surface and represented in a cartographic projection. The durum wheat landraces are one of these geo-referenced data, as they are collected at a precise location and having within their passports the spatial coordinates latitude, longitude and altitude.

Making GIS useful to people requires user interface. To make a successful interface, the designer should understand how the users think and work. The users will not use the algorithms, data structure or functions. Instead, users choose easy to use and friendly options for calculation. This is making a graphic user interface (GUI) an illusion because it hides the underlying architecture of the technology prominent in the programmer’s view and repackages it as something understandable and usable by analysts and decision makers. Non-programmers can directly manipulate visual representations of their data to retrieve

it from a database. These graphical representations facilitate browsing for needed information without having to use formal query languages or specify the location of the data within the database (Donelson 1978; Herot 1980; Friedell, Barnett, and Kramlich 1982; Friedell, 1984; McDonald 1984). Wu et al. (1989) introduce a visual query language for GIS. Their system provides a graphical way the user can browse a GIS database. The mouse is used to select layers and processing is determined by pointing to query commands within the interface. These graphical depictions of layers stored in a GIS are useful alternatives to directory listings of file names. In addition to querying, GUI systems are useful for laying out the logical structure of the database. In such systems, database designers manipulate graphical representations of entities, - relationships, and attributes to create and integrate conceptual models of database views (Wong and Kuo 1982; Reiner et al., 1984; King and Melville 1984; Goldman et al., 1985; Bryce and Hull 1986; Abiteboul and Hull 1986). The interface is then one important component of a GIS application (Figure 9). The other two components are: 1) Data input and integration: should be very well organized to facilitate the use, edit, and the update. 2) The modules: they are responsible to formulate the directives and functions asked by users through the GUI. These components relate in a hierarchical way. The interface defines how the system is operated and controlled. In an intermediate level a GIS must have spatial data processing mechanisms (input, edition, analysis, visualization, and output).

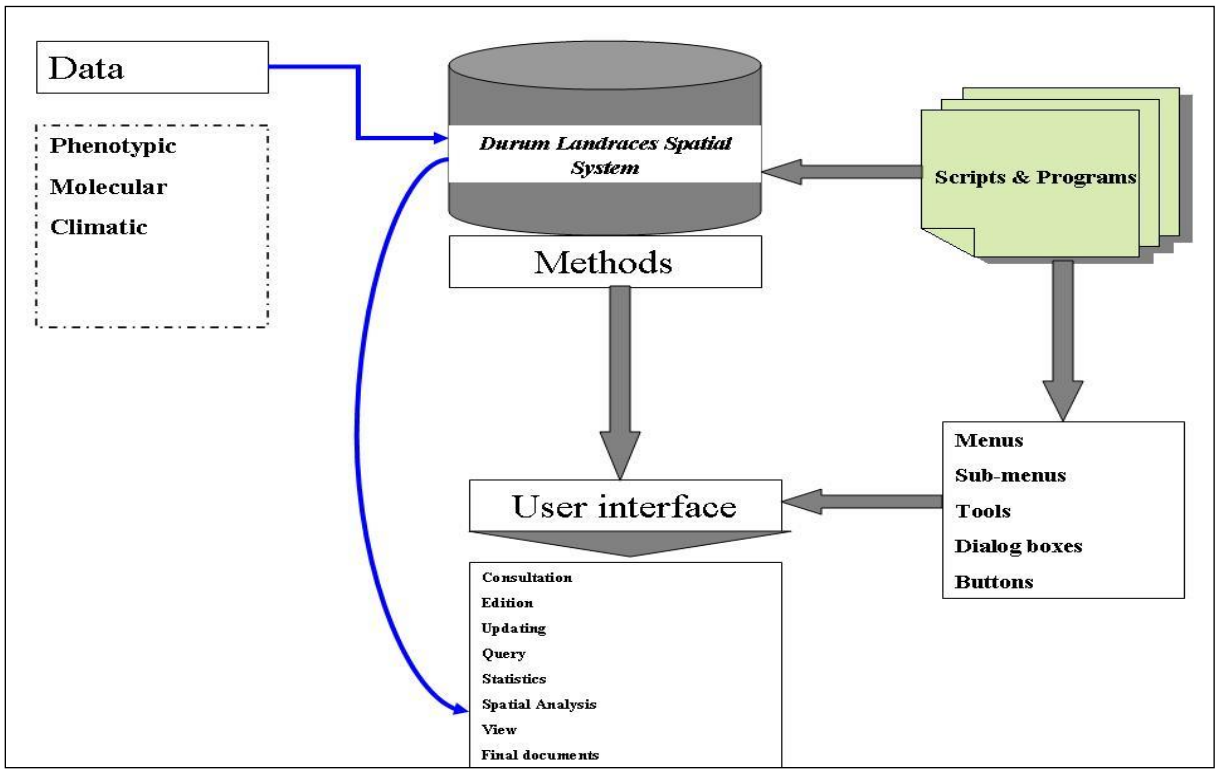


Figure 9: Conception of GIS application

Our GUI aims to give to the durum wheat breeder an analytical tool integrating the four mains GIS functionalities he needs to evaluate durum wheat landraces:

Spatial data manipulation: Spatial operations representing in general classic GIS capabilities. They aim at the maintenance and transformation of spatial data concerns the ability to input, manipulate, and

transform data once it has been created such as spatial queries & measurement, buffering and map layer overlay.

Data and Spatial data analysis: The emphasis of spatial analysis is to measure properties and relationships, taking into account the spatial localization of the phenomenon under study in a direct way. That is, the central idea is to incorporate space into the analysis to be made. It is in general a descriptive and exploratory task. Spatial autocorrelation is a key technique to understand the spatial entity of a trait of interest or an allele frequency.

Spatial statistical analysis: Spatial statistical methods permit rapid analysis and subsequent mapping of statistical quantities. A variety of interesting applications are used to illustrate how the integration of spatial statistics and the display capabilities of GIS enhance understanding of data and interpretation of the maps.

Spatial modeling or prediction: Spatial patterns of traits or genetic factors together with environmental variables can be of a good use to breeders in order to point out the germplasm of interest.

3. 4. 1. The ArcMap9.2 and VBA

ArcMap 9.2 was chosen to implement this application because of its capacity of handling numerical and alpha-numerical data; it has its own data base management relational system and of the use of Microsoft Visual Basic for Applications (VBA) as a programming language. VBA is a development environment that can be embedded into applications (Microsoft, 2002). VBA contains a set of programming tools based on the Microsoft Visual Basic development system and is designed to enable developers to build custom solutions using the full power of Microsoft Visual Basic. When using applications that host VBA, e.g. Word, Excel, Access, CorelDraw, ArcMap, automation and extension of the application functionality can be done. An example of this is creating tools with new or simplified functions. Software that includes VBA is called customizable applications, which mean applications that can be suited to fit specific business requirements. With VBA, customers can buy software and tailor it to meet a specific requirement, rather than building solutions from scratch. There are different ways of programming in VBA. Some examples of this are creating a toolbar, creating a macro, or using Visual Basic forms inside the VBA environment. VBA is mainly like VB, but the macros can easily be added to a toolbar within an existing program after they are created.

ArcMap gives the opportunity then to customize specific applications for users through ArcObject. VBA is a development environment that is provided with ArcGIS with which you can access ArcObjects. VBA macros allow the user to add further capabilities to ArcMap that are not available in the original interface or develop new analytical modules not present on it. VBA is built around *objects* (e.g. forms and controls), which have different *properties*, and *methods*. The methods are used to perform actions with the objects. A property is something that characterizes the object, e.g. its name. VB is so-called event-oriented programming, which means that something is executed when the user for example clicks a Command button or chooses from a so-called Combobox. These buttons and boxes are called controls and are connected to the code.

There are different levels of a customizing an application using VBA under ArcGIS:

Customize an interface: it doesn't need any programming. Project – based macros: You write the code behind a button or tool saved in a particular document.

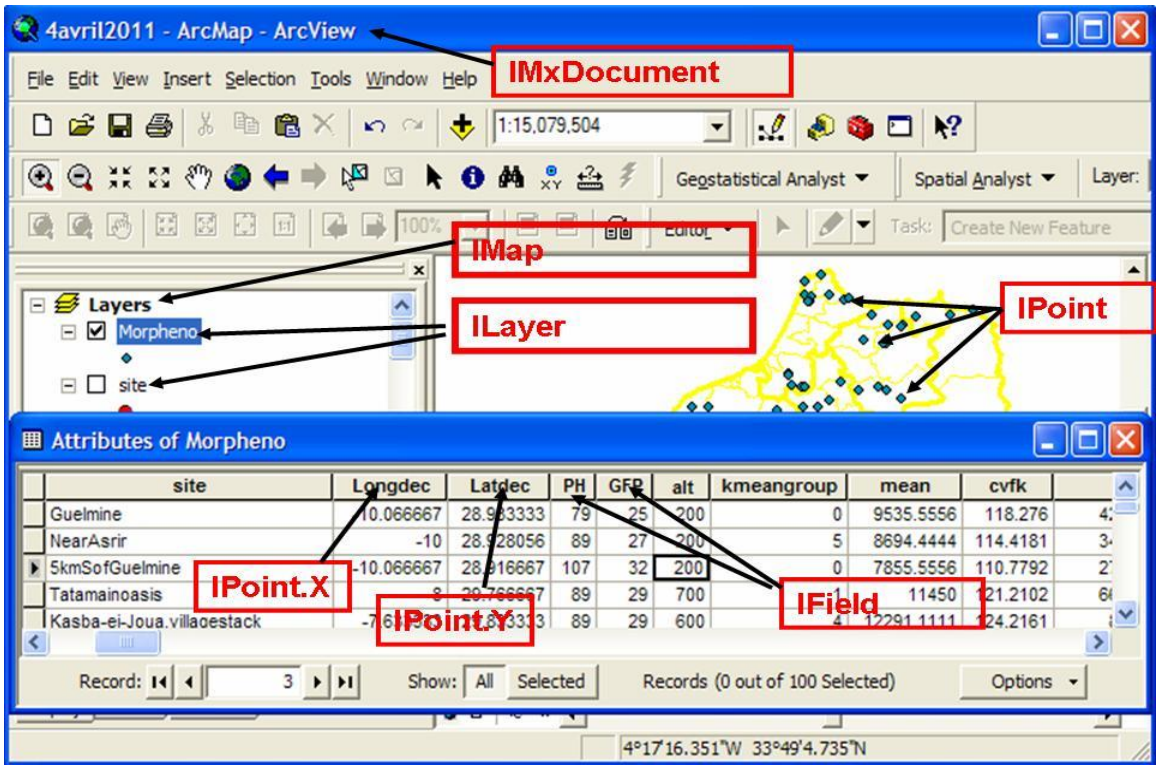


Figure 10: Example of interface of classes within ArcMap

The basic of anything we interact with within an ArcGIS component is an ArcObject: Maps, layers, geometries (points, lines, or polygons), tables, fields, raster etc. It comes from classes designed by ESRI. Each class has a logical grouping of properties and methods called interface. The "Map" class has an interface named IMap and through that interface you can get/set name of the map and you can add a layer having as well an interface named ILayer (Figure 10). Here is an example of a VBA code to read a point layer from an existing map on ArcMap:

```
Dim pMxDoc As IMxDocument, pEnumfeat As IEnumFeature
Dim pGeom As IGeometry
Dim pMap As IMap
Declare a pMap as a Map classe
Set pMxDoc = ThisDocument
This the open ArcMap document
Set pMap = pMxDoc.FocusMap
Dim pFeatureLayer As IFeatureLayer
Dim pILayer As ILayer
Declare a pLayer as a Layer classe
Dim pFeatureClass As IFeatureClass
'Set a UID for GeoFeatureLayers
Dim pId As New UID
pId = "{E156D7E5-22AF-11D3-9F99-00C04F6BC78E}" 'IGeoFeatureLayer
'Create an enumerator of GeoFeatureLayers
Dim pEnumLayer As IEnumLayer
Set pEnumLayer = pMap.Layers(pId)
'Load the input form
Load frmStability
'Populate the layer combo box with point layers existing in the map
```

```
pEnumLayer.Reset
Set pILayer = pEnumLayer.Next
Dim pointExists As Boolean
pointExists = False
Do While Not pILayer Is Nothing
Set pFeatureLayer = pILayer
Set pFeatureClass = pFeatureLayer.FeatureClass
If pFeatureClass.ShapeType = esriGeometryPoint Then
frmStability.cmbLayer.AddItem pILayer.Name
```
Populate Combobox with points layers

3. 4. 2. Environmental data in our GIS interface

The environmental maps of the two countries were extracted from Worldclim (Hijmans, R.J. et al., 2005) climatic database available for downloading free of charge from www.worldclim.org. The database is representative of current climate and it is an interpolation of observed data from 1950 to 2000. These maps were associated with the landraces coordinates to identify the climatic data for each landrace's spatial location.

The data layers were generated through interpolation of average monthly climate data from weather stations on a 30 arc-second resolution grid (often referred to as "1 km^2" resolution). Variables included are monthly total precipitation, monthly mean, minimum and maximum temperature, and nineteen (19) derived bioclimatic variables.

The WorldClim interpolated climate layers were made using:

1- Major climate databases compiled by the Global Historical Climatology Network (GHCN), the FAO, the WMO, the International Center for Tropical Agriculture (CIAT), R-HYdronet.

2- The SRTM elevation database (aggregated to 30 arc-seconds, "1 km").

3- The ANUSPLIN software which is a program for interpolating noisy multi-variate data using thin plate smoothing splines. We used latitude, longitude, and elevation as independent variables. For stations with multiple years' records, averages for the 1960-90 period were calculated. After removing stations with errors, the used database consisted of precipitation records from 47,554 locations, mean temperature from 24,542 locations, and minimum and maximum temperature for 14,835 locations.

A set of 'Bioclimatic variables' were derived from the monthly data. Bioclimatic variables (Table 4) are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modeling (e.g., BIOCLIM, GARP). The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).

Table 4: Bioclimatic variables extracted from Worldclim

Variable	Description
V1	Annual Mean Temperature
V2	Mean Monthly Temperature Range
V3	Isothermality (2/7) (* 100)
V4	Temperature Seasonality (STD * 100)

V5	Max Temperature of Warmest Month
V6	Min Temperature of Coldest Month
V7	Temperature Annual Range (5-6)
V8	Mean Temperature of Wettest Quarter
V9	Mean Temperature of Driest Quarter
V10	Mean Temperature of Warmest Quarter
V11	Mean Temperature of Coldest Quarter
V12	Annual Precipitation
V13	Precipitation of Wettest Month
V14	Precipitation of Driest Month
V15	Precipitation Seasonality (CV)
V16	Precipitation of Wettest Quarter
V17	Precipitation of Driest Quarter
V18	Precipitation of Warmest Quarter
V19	Precipitation of Coldest Quarter

3. 4. 3. Shape files in the GIS GUI

The spatial and phenotypic information about landraces was saved under a point format (Tables 5 and 6). Landraces are considered to be events and not linear or polygonal (zonal). The shape file contains first an identifier of the landrace, second the spatial information such as: nearest village, coordinates, altitude, collection date, and origin. Third, the file contains as well the phenotypic and physiological data. When a trait is measured during two different years, it symbolized by the trait code plus the year, example: GY04 and TKW07 are used for grain yield during 2004 and thousand kernel weight during 2007. The traits measured during the collecting year were coded as the trait code: PH. The identifier was unique in order to link easily between different shape files. We used the crop number taken from the ICARDA collection database as identifier.

Table 5: Geographic information-shape file (Captured from ArcGIS)

Shape	GIS_	PROVINCE	CROP_NO	IG	SITE	LONG_DEC	LAT_DEC	LON	LAT	ALT	COL_DATE
Point	3	Tiznit	20038	96337	5 km S of Guelmine	10.0667	28.9167	W10 04	N28 55	200	1985/05/03
Point	4	Tiznit	20039	96338	Tata main oasis	8	29.7667	W08 00	N29 46	700	1985/05/05
Point	5	Tiznit	20041	96340	Kasba-ej-Joua, village stack	7.63333	29.8333	W07 38	N29 50	600	1985/05/05
Point	6	Ouarzazate	20042	96341	Tanskit	6.20056	30.6925	W006 12 02	N30 41 33	850	1985/05/07
Point	7	Er Rachidia	20043	96342	Just W of Tinejdad	5.015	31.515	W005 00 54	N31 30 54	900	1985/05/08
Point	8	Beni Mellal	20045	96344	Mellah	6.81417	31.98	W006 48 51	N31 58 48	800	1985/05/08
Point	9	Er Rachidia	20046	96345	Fezna	4.46667	31.5333	W04 28	N31 32	740	1985/05/08
Point	10	Er Rachidia	20047	96346	3 km S of Aoufouss; outside main o	4.16667	31.65	W04 10	N31 39	750	1985/05/09
Point	11	El Jadida	20048	96347	5 km E of Boulaouane	8.05	32.9833	W08 03	N32 59	150	1985/05/12
Point	12	El Jadida	20050	96349	15 km W of Sidi Bennour	8.41667	32.9833	W08 25	N32 59	100	1985/05/12
Point	13	Agadir	20052	96351	Akermould	9.61667	31.6667	W09 37	N31 40	1	1985/05/13
Point	14	Agadir	20053	96352	Ain-el-Hajer, near Akermoud	9.61667	31.6667	W09 37	N31 40	1	1985/05/13
Point	15	Agadir	20054	96353	Ain-el-Hajer, near Akermoud	9.61667	31.6667	W09 37	N31 40	1	1985/05/13
Point	16	Tiznit	20055	96354	2 km E of Smimou	9.12583	30.7847	W09 07 33	N30 47 05	300	1985/05/13
Point	17	Tetouan	20056	96355	Tnine Sidi el Yamani	5.93333	35.3833	W05 56	N35 23	400	1985/05/13
Point	18	Tetouan	20057	96356	Tnine Sidi el Yamani	5.79139	35.3667	W005 47 29	N35 22	300	1985/05/13
Point	19	Agadir	20058	96357	Ounara	9.71833	31.3464	W009 43 06	N31 20 47	150	1985/05/13
Point	20	Marrakech	20059	96358	15 km N of Chichaoua	8.78139	31.5511	W008 46 53	N31 33 04	200	1985/05/14
Point	21	Agadir	20060	96359	30 km S of Chemaia	8.61667	31.8167	W08 37	N31 49	200	1985/05/14
Point	22	Agadir	20061	96360	Chemaia	8.63333	32.0833	W08 38	N32 05	300	1985/05/14
Point	23	Marrakech	20062	96361	10 km W of Ben Guerir	7.98333	32.3167	W07 59	N32 19	400	1985/05/14
Point	24	Marrakech	20063	96362	50 km S of Borouj	7.16667	32.5	W 07 10	N32 30	300	1985/05/14
Point	25	Khouribga	20064	96363	20 km N of Oued Zem	6.55	32.9167	W06 33	N32 55	750	1985/05/14
Point	26	Khenifra	20065	96364	El-Kbab	5.51667	32.7333	W05 31	N32 44	900	1985/05/15
Point	27	Bouarfa	20066	96365	Mengoub	2.35	32.25	W02 21	N32 15	900	1985/05/16
Point	28	Bouarfa	20067	96366	Figuig oasis	1.25	32.1667	W01 15	N32 10	800	1985/05/16

Table 6: Traits information shape file (Captured from ArcGIS)

CROP_NO	GFP	SDE	KPS	TKW	DH04RF	GY04RF	PRO04RF	SDS04RF	SDSI04RF	SDSN04RF
20038	32	1	39	39	138	2450	13.5	34	2.5	4.6
20039	29	5	47	33	145	2200	13.6	26	1.9	3.5
20041	29	5	40	33	148	2223	12.7	24	1.9	3.1
20042	32	5	49	41	148	2500	12.7	20	1.6	2.5
20043	28	5	44	37	142	3423	12.7	20	1.6	2.5
20045	24	5	42	39	142	3253	12.6	22	1.7	2.8
20046	31	5	42	38	138	3073	12.5	26	2.1	3.3
20047	22	5	41	34	142	2163	14	30	2.1	4.2
20048	22	7	43	33	142	2060	13.9	26	1.9	3.6
20050	21	7	41	39	150	2153	14	22	1.6	3.1
20052	23	5	33	37	144	2060	14.1	20	1.4	2.8
20053	20	7	44	31	145	2143	13.9	24	1.7	3.3
20054	25	5	39	38	152	2183	13.4	18	1.3	2.4
20055	27	5	39	32	147	2317	14	24	1.7	3.4
20056	23	7	52	32	147	2177	13.9	26	1.9	3.6
20057	22	7	48	30	147	2193	13.2	24	1.8	3.2
20058	25	7	42	29	142	2613	13.6	22	1.6	3
20059	18	5	40	38	140	2293	12.4	20	1.6	2.5
20060	21	7	38	30	142	2703	12.8	26	2	3.3
20061	22	7	49	28	142	2413	12.5	28	2.2	3.5
20062	28	5	58	38	141	2800	13.1	18	1.4	2.4
20063	27	7	43	33	140	2253	13.3	26	1.9	3.5
20064	32	5	52	40	138	3310	12.4	24	1.9	3
20065	25	7	37	43	142	2357	13.4	26	1.9	3.5
20066	24	5	52	35	142	2217	13.5	28	2.1	3.8
20067	24	5	41	33	142	2167	13.8	28	2	3.9
20068	23	5	47	40	141	2173	14	20	1.4	2.8
20069	28	7	43	37	155	2167	13.1	30	2.3	3.9

For the molecular file, the shape is constructed with the alleles of a locus. Durum is a tetraploid crop and has one allele on each genome (A and B). Each locus is stored in two columns. The file is having also the unique identifier and the same used for the phenotypic shape file. Missing genotypes are coded as zero (0). The file contains also the spatial coordinates. For example M1-1 is the allele on the genome A of markers M1 (Table 7).

Table 7: Marker information shape file (Captured from ArcGIS)

Sample_Sou	FID	X	Y	M1-1	M1-2	M2-1	M2-2	M3-1	M3-2
	0	10.066667	28.933333	114	114	204	204	117	117
ICDW-20037	1	10	28.928056	114	114	194	194	121	121
ICDW-20038	2	10.066667	28.916667	124	124	185	185	122	122
ICDW-20039	3	8	29.766667	114	114	206	206	117	117
ICDW-20041	4	7.633333	29.833333	112	112	206	206	121	121
ICDW-20042	5	6.200556	30.6925	114	114	204	204	121	121
ICDW-20043	6	5.015	31.515	114	114	0	0	122	122
ICDW-20045	7	6.814167	31.98	114	114	0	0	121	121
ICDW-20046	8	4.466667	31.533333	114	114	196	196	121	121
ICDW-20047	9	4.166667	31.65	114	114	204	204	117	117
ICDW-20048	10	8.05	32.983333	114	114	204	204	117	117
ICDW-20049	11	8.416667	32.983333	114	114	204	204	117	117
ICDW-20050	12	9.616667	31.666667	112	112	190	190	121	121
ICDW-20052	13	9.616667	31.666667	114	114	194	194	121	121
ICDW-20053	14	9.616667	31.666667	114	114	204	204	117	117
ICDW-20054	15	9.125833	30.784722	112	112	185	185	120	120
ICDW-20055	16	5.933333	35.383333	114	114	204	204	117	117
ICDW-20056	17	5.791389	35.366667	114	114	205	205	117	117
ICDW-20057	18	9.718333	31.346389	114	114	205	205	117	117
ICDW-20058	19	8.781389	31.551111	114	114	204	204	117	117
ICDW-20059	20	8.616667	31.816667	115	115	196	196	123	123
ICDW-20060	21	8.633333	32.083333	114	114	204	204	117	117
ICDW-20061	22	7.983333	32.316667	114	114	204	204	117	117
ICDW-20062	23	7.166667	32.5	114	114	196	196	121	174
ICDW-20063	24	6.55	32.916667	114	114	205	205	117	117
ICDW-20064	25	5.516667	32.733333	114	114	0	0	121	121
ICDW-20065	26	2.35	32.25	114	114	0	0	0	0
ICDW-20066	27	1.25	32.166667	114	114	204	204	117	117
ICDW-20067	28	2.238333	35.066389	114	114	204	204	117	117
ICDW-20068	29	2.733333	34.95	114	114	196	196	175	175
ICDW-20069	30	8.155556	30.701667	114	114	209	209	122	122
ICDW-20070	31	8.155556	30.701667	114	114	204	204	117	117

3. 4. 4. Methods developed within the GUI

The methods developed at our interface are shown in the Table 8. We divided the modules into two parts: Trait analysis is analyzing phenotypic data (agronomic and physiological data). This category gives relatedness between traits or individuals. It gives also the possibility to dissect the spatial pattern of traits. When we have multiple environments evaluation, some GE analyses are possible such as ANOVA or stability. Marker analysis on the other hand breaks down the genetic and spatial genetic structure of landraces. It also gives the possibility of running the marker-trait association. Other statistical analysis

that are not developed under the GIS interface such as mixed model were run using Genstat 12 (Payne et al. 2009)

Table 8: Menus developed within the Durum GIS interface

Trait analysis	Marker analysis
Statistics	**Individuals**
Descriptive Statistics	Moran's I
Pearson correlation	PCA
Spearman rank correlation	Spatial PCA
Regression	**Populations**
PCA	Populations Statistics
Path analysis	Genetic Distance
Multiple regression	PCA
K-mean clustering	One Locus PCA
Spatial Statistics	Spatial PCA
Moran's I	Population Centroid
Spatial PCA	**Marker-Trait association**
GxE	T-Test
ANOVA	PCA+Chi test
Stability	Multiple regression
Ranking Genotype	
Non Parametric stability	

3. 5. Statistical methods

3. 5. 1. Correlation

Correlation is a measure of relation between two variables. The correlation coefficients ranged from -1 for a perfect negative correlation to 1 for perfect positive correlation. A correlation coefficient of 0 means a lack of relation. The most widely used correlation is the Pearson (product moment) correlation which is defined as the covariance between two variables divided by the product of their standard deviations. The other correlation we used is the Spearman's rank correlation which is the Pearson correlation between the ranked variables. Correlation can be used to study or compare two different traits, same trait measured during two different. Also, correlation can be used as method for association between trait and an allele frequency (marker).

3. 5. 2. Regression

The simple regression is the statistical method attempting to determine the strength of the relationship between one dependent variable and other changing variables. Multiple regression is an extension of the simple regression problem to include more than one explanatory variable. In general, we will be considering linear regression with *m* independent variables, and our regression model will look like:

$$a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_m X_m + \epsilon = Y.$$

Our model states that our observations Y_i can be explained by a constant term a_0 plus a linear combination of the variables, with each variable having its own ``slope". Thus, a_2 includes how fast Y changes when X_2 changes, holding all other X_i fixed. In matrix notation, we can write the problem as:

$$\begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,m} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

This is nothing more than our general least squares problem, to which the solution is given as:

$$x = [\![A^T A]\!]^{-1} A^T b$$

The values of the regression coefficients tell us little, since they depend on the units chosen. We overcome this problem by normalizing the coefficients. These coefficients give an idea about, in a multivariate context, the contribution of a trait (or an allele frequency) on another composite trait such as GY.

3. 5. 3. ANOVA

The analysis of variance (ANOVA) is widely used to study GE data (Skroppa, 1984). An ANOVA allows the partitioning of total phenotypic variation into components due to genotype, environment, GE interaction and error. The relative sizes of these variance components can then be used to quantify the magnitude of the GE (Cooper and DeLacy, 1994). ANOVA is a basis for any study of GE, but it does not allow a final interpretation (Barnes et al., 1984).

3. 5. 4. Path analysis

Path analysis (PA) represents an early attempt at dealing with casual relationships between variables. It was developed by Wright in 1930's. Path analysis is an extension of the regression model, used to test the fit of the correlation matrix against two or more causal models which are being compared by the researcher. A path coefficient is a standardized regression coefficient (beta) showing the direct effect of an independent variable on a dependent variable in the path model. Thus when the model has two or more causal variables, path coefficients are partial regression coefficients, which measure the extent of effect of one variable on another in the path model controlling for other prior variables, using standardized data or a correlation matrix as input. Some assumptions need to be done to run path analysis; all the relationships are linear and the casual effect is one-way. The basic model for PA is the correlation matrix which is decomposed to direct effect and indirect effect. Direct effect is the path coefficient from one variable to another. Indirect effect is sequence of paths through one or more variables. Note that the sum of direct and indirect effects is the total casual part of the correlation between the two variables.

3. 5. 5. Stability

Phenotypic stability can be a good criterion for breeding or genotype selection. Several statistical methods have been developed to evaluate stability. The variance of a genotype evaluated across environment has been used as a measure of stability and a genotype with a low variance is considered stable. The mean of the estimated variance components of GE for all pairs of genotypes that include a specific genotype is the stability measure for that genotype (Plaisted, 1959, 1960). This approach involved the deletion of a genotype from the entire set of data and the GE interaction for the variance for the subset is the stability index for the deleted genotype. Francis and Kannenberg (1978) on the other hand used the coefficient of variation (CV) of each genotype as a measure of stability. A high yielding genotype with a low CV was considered stable. Other stability indices include Wricke's (1962) ecovalence, Shukla's (1972) stability variance, Perkins and Jinks' (1968) regression coefficient, Finlay and Wilkinson's (1963) and Eberhart and Russel's (1966) coefficients. In Finlay and Wilkinson (1963) model, the observed yields of the varieties were regressed on an environmental index defined as the difference between the marginal mean yield of the environments and the overall mean. The regression coefficient (bi) for each genotype was considered a measure of stability. A b-value approximating to 1.0 indicated average stability, genotypes with b = 1.0 and above average yield were considered as having general adaptation, while a genotype with b = 1 and below average yield was associated with poor adaptation to all environments. In this model, stability was defined by the regression coefficient, while adaptability was defined by the relative mean yield of the variety. In addition to the regression coefficient, Eberhart and Russell (1966) estimated the mean square of deviation from the regression as another stability parameter.

3. 5. 6. PCA

Principal Component Analysis (PCA) is an exploratory tool designed by Karl Pearson in 1901 to identify unknown trends in a multidimensional data set. Principal components analysis is the procedure to transform a number of possibly correlated variables to smaller number of uncorrelated variables called principal components. The first principal component should account of as much of the variability in the data as possible. Let's note p observations for n entries by $X = (X_1, X_2, X_3, \ldots, X_p)^t$:

$X_1 = (x_1^1, x_1^2, \ldots, x_1^n)$; $X_2 = (x_2^1, x_2^2, \ldots, x_2^n)$; …..; $X_p = (x_p^1, x_p^2, \ldots, x_p^n)$.

$$COV(X) = \frac{1}{n} \times XX^T$$

COV(X) is a positive symmetric matrix, so there is a vector $V \in R^n$, such that:

COV(X)* V = λ*V is an eigenvector of A and the corresponding scalar λ > 0 is the eigenvalue associated with V. We selected only eigenvectors V_j (j=1, 2,…, k<n) with large enough eigenvalue λ_j.

We project then the data points X^i to the hyper plane defined by selected eigenvectors V_j: $x_j^i = V_j^t * X^i$

Amount of variance explained by an eigenvalue is $(\boldsymbol{\lambda i / \sum_1^n \lambda i})$.

Applying PCA to a data table X correspond to the analysis of triplet (X, Q, D); where Q is a (p x p) scalar matrix (can be identity), D is an (n x n) scalar matrix and X is the (n x p) centred (PCA on covariance matrix) or standardized (PCA on correlation matrix) matrix.

Running PCA analysis consists on finding a vector u1 (first principal axis) so that:

$\mathbf{Q}(u_1) = \| XQu_1\|^2_D = u_1^t QX^t DXQu_1$

Under the constraint that $\| u1 \|^2_Q = u_1^t Qu^1 = 1$.

The solution vector u_j is obtained the right-hand eigenvectors of $QX^t DXQ$. The eigenvalue λ_j is the maximum of $\mathbf{Q}(u_1)$.

3. 5. 7. K-mean cluster

The k-means algorithm (MacQueen, 1967), in comparison with other partitional clustering algorithms (Fuzzy c-means clustering), is fast, doesn't require any specific preparation of the different data sets and is particularly easy to use. Its main weakness consists of the fact that it has to be told the number of clusters (*k*) to be found. Initially, it is necessary to define *k a priori* temporary centers (one for each cluster) which are located at random in the multidimensional scatter of points. All points belonging to the different data sets are associated with their nearest centre and this constitutes an early grouping together. Then each one of the *k* centers is calculated as the centroid of the points it «owns» and a new association is established with the nearest points of the data sets, and so on. The *k* centroids change their location step by step until they don't move any more.

The algorithm aims at minimizing an objective squared error function.

$J = \sum \sum \text{abs} (x_i^{(j)} - c_j)^2$

Where *k* = number of clusters and *n* = number of individuals

And abs $(x_i^{(j)} - c_j)^2$ is a measure of the distance between a data point $x_i^{(j)}$and the cluster centre *cj*.

k-mean can be used to classify landraces using traits or a matrix of allele frequencies resulting from markers characterization.

3. 5. 8. Descriptive locus statistics

The variation in alleles is critical to the survival of a species and allows organisms to adapt to changing environments. This variation is revealed by genetic diversity. The more variation, the better the chance that at least some of the individuals will have an allelic variant that is suited for a new environment. A large gene pool indicates a large genetic diversity, which is associated with a robust population able to survive intense selection. Meanwhile, low genetic diversity can cause reduced fitness and increased chances of extinction. Allele frequency, or the frequency at which alleles are found at any locus of interest, is used to estimate the frequency of a given genetic profile. Every diploid cell has two alleles, one inherited from each parent. If an individual has two different alleles at a specific locus, the individual is heterozygous at that locus; if the two alleles are the same, the individual is homozygous. Allele frequency is used to characterize the genetic diversity, or richness of the gene pool, in a population.

The measure of the amount of heterozygosity across loci is used as a general indicator of the amount of genetic variability in a population. Two measures of heterozygosity are defined:

Expected Heterozygosity (H_e) or genetic diversity (GD) is the probability that two alleles drawn at random are different alleles. It estimates the fraction of all individuals who would be heterozygous for any randomly chosen locus and is calculated as:

$$H_e = GD = 1 - \sum_{i=1}^{k} p_i^2$$

Where p_i is the frequency of the i^{th} allele and k is the total number of alleles. The expected heterozygosity over m loci (H_E) is

$$H_e = 1 - \frac{1}{m} \times \sum_{1}^{m} \sum_{1}^{k} p_i^2$$

Observed heterozygosity (H_o) of a population is measured by determining the proportion of loci that are heterozygote and the number of individuals that are heterozygote for each particular locus. For a single locus with two alleles, (H_O) is the number of heterozygotes at this locus divided by the total number of surveyed individuals. Over a series of several loci, H_O is the sum of H_o heterozygotes calculated for each locus divided by the number of considered loci.

$$Ho = \frac{Number\ of\ heterozygos\ at\ a\ locus}{Total\ number\ of\ individuals}$$

F-statistics are measures of genetic structure developed in the 1920s by Sewall Wright (University of Chicago), one of the primary founders of population genetics, related to statistical analysis of variance. For a locus, F is the ratio of the difference between expected and observed heterozygosity to the expected heterozygosity. F has values between 0 for no genetic drift and 1 fixation of alternative alleles:

$$F = \frac{H_e - H_O}{H_e}$$

For a population:

$$F = \frac{H_E - H_O}{H_E}$$

3. 5. 9. PCA for Multi-Locus data

Principal component analysis for population genetic data can be led on individuals or populations. The basic data or matrix for the PCA is matrix of allele frequencies. For individuals based PCA, consider n(i,j) is the number of copies of the j^{th} allele found in individual i at locus k. The matrix n has one row for each individual and one column for each allele j of the locus k. μ(j) and σ(j) are respectively the mean and standard deviation of j^{th} column of the matrix N. the normalized matrix M of n is then:

$$M(i,j) = \frac{N(i,j) - \mu(j)}{\sigma(j)}$$

For population based PCA, consider G(i,j) is the frequency of the j^{th} allele found in population i at locus k. We standardize by the mean and then the matrix M(i,j) is defined as:

M(i,j) = G(i, j) − μ(j) where μ(j) is the mean of the j^{th} column of the matrix G.

After defining the M matrix, we compute the C covariance matrix among individuals or populations as:

$$X = \frac{1}{n} \times MM^T$$

The last step is to calculate eigenvectors of X. each of the eigenvector will have n (number of individuals or populations) as length.

3. 5. 10. PCA to correct for stratification in association studies

There are three steps to use PCA analysis for association studies (Price et al., 2006). First, we run PCA on genotype data (see PCA for multi-locus data) to get small numbers of PCA axes which are continuous axes reflecting the genetic variation of data. Second, we adjust the candidate loci for association and the phenotype using the significant axes of variation. To make the adjustment to an axis, let's consider g_{ij} the genotype of individual j at locus i, a_j is the coordinate of individual j at axe a. So

G_{ji} adj = $g_{ji} - \gamma_i a_j$, where γ_i is the regression coefficient for predicting genotype across individuals j. We adjust with the same method the other axes of variation and the phenotype. The third part is to compute X^2 statistics between the adjusted phenotype and genotype; this is equal to (n-k-1) times the squared correlation between the two vectors with n is sample size and k is the number of used axes.

3. 6. Spatial statistics

3. 6. 1. Connectivity networks

Triangulation is a method for tessellation of domain. In fact triangulation is a common method for surface representation and for building a TIN. Triangulation produces a continuous surface. Important problems in triangulation algorithm are independency from starting point or dispersion of them. Further result must be repeatable and predictable. Therefore, the triangulation is made, the aim is to maximize the minimum angles and establish circle condition.

Delaunay triangulation: DT is the most widely used triangulation in scientific computing. A technique for creating a mesh of contiguous, no overlapping triangles from a dataset of points. Each triangle's circumscribing circle contains no points from the dataset in its interior.

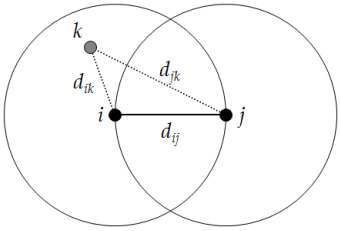
Distance-based connections: This procedure connects all points separated by a specified distance range. Any pair of points whose distance is between the specific minimal and maximal values (inclusive) will be connected; all pairs whose distance is outside of this range will remain unconnected.

Nearest neighbors: This procedure finds the nearest neighbor to every point. The nearest neighbor for a point is simply the point that is closest to it. A nearest-neighbor connections matrix does not have to be symmetric, because the nearest neighbor of one point is not necessarily the neighbor of the other point. Furthermore, a nearest-neighbor network does not have to completely span the points; usually it will not. One also has the option of specifying the number of neighbors to connect; the standard default is one, which is the traditional nearest-neighbor network, but one can choose to connect the closest two neighbors, or closest three, etc.

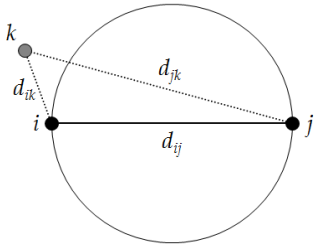
*Minimum spanning tree:*A minimum spanning tree is a connections matrix in which all of the points are connected in a single network without any reticulate (closed) loops and in which the sum of the distances along each connection is minimal. The procedure works by starting with a single point in the "connected" group and placing the remaining points in an "unconnected" group. The nearest neighbor connections are a subset of the minimum spanning tree.

Relative neighborhood network: The relative neighborhood network is a connection scheme in which two points are connected if the intersection between the two circles (or spheres in three dimensions) centered on the two points with radii equal to the distance between the points does not contain any additional points. Mathematically, another way to think of this is that points *i* and *j* are connected if the distance between them, *dij*, is less than the maximum of *dik* and *djk* for all other points *k*.

Connect *i* and *j* if *dij* < Maximum(*dik* , *djk*) for all *k*



Gabriel graph: In a Gabriel Graph, a connection scheme proposed by Gabriel and Sokal (1969), two points are connected when the circle (or sphere in three dimensions) associated with the diameter that has the two points as endpoints does not have another point within its circumference (volume).



Mathematically, two points i and j are connected if the square of the distance between them, d_{ij}^2, is less than the sum of the squared distance between each of these points and any other point k.

Connect *i* and *j* if $d_{ij}^2 < d_{ik}^2 + d_{jk}^2$ for all k.

The relative neighborhood network is a subset of the Gabriel graph. See Gabriel and Sokal (1969) and Matula and Sokal (1980) for more information on Gabriel Networks and their properties.

3. 6. 2. Spatial autocorrelation and Moran's I index

Autocorrelation is the statistical method to study and measure the dependence of the same variable over time, while spatial autocorrelation is to measure the degree of dependence of a variable in a geographic space. Moran's I is a measure of spatial autocorrelation developed by Patrick A.P. Moran. Moran introduced in 1950 the first measure of spatial autocorrelation in order to study stochastic phenomena, which are distributed in space in two or more dimensions. Moran's index has been subsequently used in almost all studies employing spatial autocorrelation. Moran's I is used to estimate the strength of this correlation between observations as a function of the distance separating them. Like a correlation coefficient the values of Moran's I range from +1 meaning strong positive spatial autocorrelation, to 0 meaning a random pattern to -1 indicating strong negative spatial autocorrelation. A positive (negative) spatial autocorrelation corresponds to a global (local) spatial structure (Thioulouse et al., 1995).

$$I = \frac{N\sum_i\sum_j W_{i,j}(X_i - \bar{X})(X_j - \bar{X})}{(\sum_i\sum_j W_{i,j})\sum_i(X_i - \bar{X})^2}$$

Where:

N is the number of studied locations
$\mathbf{X_i}$ is the variable value at a particular location
$\mathbf{X_j}$ is the variable value at another location
X is the mean of the variable
$\mathbf{W_{ij}}$ is a weight applied to the comparison between location *i* and location *j*

The expected value of Moran's I under hypothesis of no spatial autocorrelation is

$$I_0 = \frac{-1}{(N-1)}$$

With a matrix notation, Moran's I (x) index of a geo-referenced variable x of dimension n is:

$$I(x) = \frac{x^T W x}{W} \times \frac{n}{x^T x}$$

Where W is a weight matrix (w_{ij}, i=1 to n, j=1 to n) applied to comparison between location I and location j. If location i is adjacent to location j, the weight receives, for example, 1 and 0 if not. W can be a distance-based, and the weight will be the inverse of the distance between two locations. w is the sum of all terms in the matrix W. Another way of constructing **W**, is to use the spatial connectivity networks (Legendre and Legendre1998) as shown in part "3-5-2-1" of this chapter. Spatial auto-correlation measures the degree of clustering of data in the studied space and allows checking if the data is dispersed or clustered. When testing for SA a p-value is calculated to test for Null hypothesis, the NULL hypothesis is that the spatial distribution of data is Random. When the p-value is statistically significant, one can reject the null hypothesis. When the p-value is not statistically significant, one cannot reject the null hypothesis. It is possible that the spatial distribution of feature values is the result of random spatial processes.

3. 6. 3. High/Low Clustering (Getis-Ord General G)-hot spots (Getis-Ord Local G)

If one computes SA and finds out the the data is clustered. One question is essential: how the data is clustered? Methods developed by Getis and Ord (1992; 1996) not only provide hypothesis testing to determine whether clustering has occurred within data, but also provide information on the extent to which above and below average values cluster more strongly and identify local concentrations of clustering (Laffan, 2006; Mueller-Warrant et al., 2008). The Getis-Ord General G high/low clustering is calculated as:

$$G = \frac{\sum_{1=1}^{n}\sum_{i=1}^{n} w_{ij}x_i x_j}{\sum_{1=1}^{n}\sum_{i=1}^{n} x_i x_j} \ , j \neq i$$

Where x_i represents the value of feature *i*, x_j represents the value of feature *j*, and w_{ij} is the weight assigned to each pair of features x_i, x_j.

The z-score statics is computed as:

$$z_G = \frac{G - \mathrm{E}[G]}{\sqrt{\mathrm{V}[G]}}$$

Where

$$\mathrm{E}[G] = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \ \forall j \neq i$$
$$\mathrm{V}[G] = \mathrm{E}[G^2] - \mathrm{E}[G]^2$$

z-score is high positive (negative) and significant at 1, 5 and 10% means that high (low) values of the data are clustered spatially together and that there is a less than 1%, 5 and 10% likelihood that this high-clustered (low-clustered) pattern could be the result of random chance.

The Getis-Ord G_i^* test statistic is a local adaptation of global Getis-Ord General G and seeks to identify areas of hot and cold clustering based on local neighborhood values (Getis & Ord, 1996; Laffan, 2006). The G_i^* statistic is calculated as the summation of the differences between local sample values and the mean, and is observed as standard normal distribution *z*-score values:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

Where x_j is the attribute value for feature j, w_{ij} is the spatial weight between feature i and j, n is the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2}$$

The results of the Getis-OrdG_i^* testing may be best visualized in acartographic output format to easily identify local variation within the data. The resultant z-scores (G_i^*) and p-values allow to know where features with either high or low values cluster spatially. This score works by looking at each feature within the context of neighboring features. A feature with a high (low) value is interesting but may not be a statistically significant hot (low) spot. To be a statistically significant hot (low) spot, a feature will have a high (low) value and be surrounded by other features with high (low) values as well.

3. 6. 4. Local and global structure

For the analysis of spatial structure of a single variable *x*, total variance VAR(*x*) is partitioned between global variability GV(*x*) and local variance LV(*x*) according to Thioulouse et al. (1995):

VAR(*x*) = GV(*x*) + LV(*x*), with

$$VAR(x) = \sum_{i=1}^{n} p_i\ (x_i - \overline{x_p})^2$$

$$LV(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\ (x_i - x_j)^2$$

$$GV(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(x_i - \overline{x_p}\right) \times \left(x_j - \overline{x_p}\right)$$

Global variance can be seen as the covariance between x and the mean of its neighbors and that local variance can be seen as the covariance between each point and the mean of its neighbors.

3. 6. 5. Multispati

Examples from mapping the PCA scores or computing their spatial autocorrelation showed that the multivariate data contain spatial entity. Methods based on the principals of geostatistics combined with the multivariate analysis Wartenberg 1985, and Borcard and Legendre2002 permitted to identify the

spatial structure of multivariate data. This kind of analysis is easy to implement but requires that sample data should be well distributed over the studied space especially when we have natural barriers (Mountains, rivers). The use of the spatial connection networks is appropriate for the spatial multivariate analysis. Most of the methods running the spatial multivariate analysis are using the Moran' Í or Geary' C indexes.

Multivariate spatial analysis based on Moran's *I* (MULTISPATI) originates in a course in French (Chessel et al., 2004) and introduces the row-sum standardized weight matrix **W** in the analysis of a statistical triplet (**X**, **Q**, **D**). It is possible to extend the concept of lag vector to construct a lag matrix Lag(**X) = WX**. The two tables Lag(**X)** and **X** are fully matched, i.e. it contains the measurements of the same variables for the same sites. The principle of MULTISPATI consists of the analysis of this pair of tables by the co inertia analysis (Dolédec & Chessel 1994; Dray et al., 2003) of a pair of fully matched tables (Torre & Chessel 1994; Dray et al., 2003). MULTISPATI seeks for u_1 maximizing the quantity:

$\mathbf{Q}(u_1) = a_1{}^t D\ \mathrm{Lag}(a_1)$ with $a_1 = XQu_1$.

This analysis maximizes the scalar product between a linear combination of original variables ($\mathbf{a}_1 = \mathbf{XQu}_1$) and a linear combination of lagged variables ($\mathrm{Lag}(a_1) = \mathbf{WXQu}_1$). Then

$\mathbf{Q}(u_1) = I_D\ (a_1) \parallel a_1 \parallel^2_D$ (equa8)

This formulation shows that MULTISPATI finds coefficients ($\mathbf{u}_1$) to obtain a linear combination of variables ($\mathbf{a}_1 = \mathbf{XQu}_1$) which maximizes a compromise between the classical multivariate analysis ($\|\mathbf{a}_1\|_D^2$) and a generalized version of Moran's *I* ($I_D(\mathbf{a}_1)$). The only difference between the generalized I_D and the classical Moran's *I* is that the first one used a general matrix of weights **D** while the second considers only the usual case where **D**=1/*n*.

In practice, it is preferable to diagonalize the **Q**-symmetric matrix $\mathbf{H} = (1/2)(\mathbf{X}^t(\mathbf{W}^t\mathbf{D} + \mathbf{DW})\ \mathbf{XQ})$ instead of $\mathbf{X}^t\mathbf{DWXQ}$ which is not symmetric. The maxima of eq. 8 is equal and given by the first eigenvalue (λ_1) of **H**.

In the case of the normalized PCA, MULTISPATI is equivalent to Wartenberg's approach using a row-sum weighting scheme. In order to test the statistical significance of the spatial structure of the table **X**, a permutation procedure can be used. The statistic used is equal to *trace*(**X***t* **DWXQ**). The *p*-value is computed by comparing the observed value to those obtained by permutation of the rows of the table **X**. The MULTISPATI approach has been implemented in the R software as a function of the ade4 package (Chessel et al., 2004).

Trying to combine multiple analysis, spatial autocorrelation and GIS offers an opportunity to have a very practical system to analyze spatial multivariate data. Multivariate analysis dissect the structures of data, SAU help understanding the spatial pattern of these structures and GIS as a powerful tool to stock, analyze and visual spatial data.

3. 6. 6. Spatial principal components analysis sPCA

Let's have x a vector of allelic frequencies of n entities (individuals or populations). Moran's index of x will be computed using the Moran's I formula. Consider L the standardized matrix of W. w, the sum of all terms in W will be n which is the number of entities and then Moran's I formula will be:

$$I(x) = \frac{x^T\ L\ x}{x^T x}$$

From the paragraph (PCA), the solution of PCA problem is to find eigenvalues of $cov(x) = (1/n)\ XX^T$

These eigenvectors summarize the genetic variability of data but give no information about spatial patterns. As for the sPCA, finding a solution is equal to find eigenvectors of: $\frac{1}{2n} X^T (L + L^T) X$

where X is the matrix of p allelic frequencies for n entities. sPCA does not decompose the variance into decreasing additive components but separates the product of variance and spatial autocorrelation into positive, null and negative. The most important scores at sPCA analysis are first, the score with strongest variance and the highly positive spatial autocorrelation called the global score or global structure and second, the score with strong variance and the highly negative spatial autocorrelation called local score or local structure. As for PCA, a map of sPCA scores can help assess visually the spatial and genetic patterns of data (Cavalli-Sforza, 1966).

3. 6. 7. Geographic patterns using Monmonier's algorithm

Scientists used several analytical techniques to determine relationships between geography (or space) and a character of interest. As an example, spatial autocorrelation may reveal a spatial pattern but it is not explicit. SA or any other technique cannot establish a discontinuity of phenomena over space. Recent implementations of Monmonier's maximum-difference algorithm offer a particularly powerful example. This algorithm identifies boundaries from a distance matrix by visualizing data on a map. The algorithm begins with plotting sites to be used in the analysis using geographic coordinates into a map. Then a Voronoï tessellation are constructed, polygons for each site consisting of points on a plane nearer to the site's centroid than to any other centroid. From the tessellation the algorithm builds a Delaunay triangulation (Brassel & Reif, 1979), the fastest and 'most direct way to connect (triangulate) adjacent points on a map' (Manni et al., 2004). The distance (dissimilarity) matrix is mapped onto the triangulation such that each pairwise line between sites has an associated distance. Monmonier's algorithm then builds biogeographical boundaries beginning with the maximum pairwise distance and continuing until (1) the edge of the map is hit, (2) a loop is formed, or (3) a previously computed barrier is reached. Boundaries are drawn perpendicular to triangulation lines, and the growing boundary extends in the direction of the line with the largest pairwise distance (Manel et al., 2003; Manni et al., 2004). BARRIER 2.2 software (Manni & Guérard, 2004) was used to compute biogeographical boundaries by Monmonier's algorithm. Correlation distance matrices of agronomical, physiological, genetic and climatic were computed and used to study and determine barriers and discontinuities for durum wheat landraces.

3. 6. 8. Interpolating surfaces

Mapping techniques such as kriging can help understanding the spatial distribution of a trait, and allele frequency or yield stability. Landraces data are often collected at multiple points in space and time. These data are often correlated, and thus it will be important to take advantage of these dependencies to interpret them. Also, it is better to take in consideration all possible data that influences plant growth and development, physiology, phenology and yield components. Understanding the stochastic distribution of these data in time and space is therefore fundamental to solving problems of data interpretation in using these landraces in a crossing program. A lot of these analyses can be done through mapping. One can use mapping, in the framework of landraces evaluation in different ways. Table 9 gives examples of possible thematic mapping.

Table 9: Mapping examples in landraces diversity studies

Data	May it be used for mapping?

Trait	Map of the trait
	Map of the stability of the trait, GE effect
	Map of residuals from a regression model
	Map of the axis resulting from ordering analysis (PCA, Multispati)
	Map of clusters assignment
	Map of local spatial autocorrelation
Marker	Allele chart, map of allele frequency
	Population assignment of individuals
	Map of heterozygosity
	Map of axis resulting from ordering analysis (PCA, sPCA)

To map all events we used kriging which is a method for interpolation. Interpolation is the process of estimating a variable at an unmeasured location from observed values at surrounding locations. All interpolation algorithms (inverse distance squared, splines, radial basis functions, triangulation, etc.) estimate the value at a given location as a weighted sum of data values at surrounding locations. Almost all assign weights according to functions that give a decreasing weight with increasing separation distance. Kriging assigns weights according to a (moderately) data-driven weighting function, rather than an arbitrary function, but it is still just an interpolation algorithm and will give very similar results to others in many cases (Isaaks and Srivastava, 1989). In particular if the data locations are fairly dense and uniformly distributed throughout the study area, you will get fairly good estimates regardless of interpolation algorithm.

Semivariance, which is used by kriging, is a measure of the degree of spatial dependence between samples. The magnitude of the semivariance between points depends on the distance between the points. A smaller distance yields a smaller semivariance and a larger distance results in a larger semivariance. The plot of the semivariances as a function of distance from a point is referred to as a semivariogram. The semivariance increases as the distance increases until at a certain distance away from a point the semivariance will equal the variance around the average value, and will therefore no longer increase, causing a flat region to occur on the semivariogram called a sill.

Kriging is the estimation procedure used in geostatistics using known values and a semivariogram to determine unknown values. It was named after D. G. Krige from South Africa. The procedures involved in kriging incorporate measures of error and uncertainty when determine estimations. Based on the semivariogram used, optimal weights are assigned to unknown values in order to calculate unknown ones. Since the variogram changes with distance, the weights depend on the known sample distribution.

In ordinary kriging, which estimates the unknown value using a weighted linear combinations of the available sample.

$$\hat{v} = \sum_{j=1}^{n} w_j \times v \quad , \sum_{i=1}^{n} w_i = 1$$

The error of i^{th} estimate, r_i, is the difference of estimated value and true value at that same location:

$$r_i = \hat{v} - v_i$$

The average error of a set of k estimates is:

$$m_r = \frac{1}{k}\sum_{i=1}^{k} r_i = \frac{1}{k}\sum_{i=1}^{k} \hat{v}_i - v_i$$

The error variance is:

$$\hat{\sigma}_R^2 = \frac{1}{k}\sum_{i=1}^{k} (r_i - m_R)^2 = \frac{1}{k}\sum_{i=1}^{k} \left[\hat{v}_i - v_i - \frac{1}{k}\sum_{i=1}^{k} (\hat{v}_i - v_i) \right]^2$$

3. 7. Population genetic computations

3. 7. 1. Genetic structure and population genetics

We examined genetic structure of populations using two methods of population assignment using both Bayesian clustering method. We used STRUCTURE (Pritchard et al., 2000) and GENELAND version 1.0.5 (Guillot at al., 2005). GENELAND is a computer program developed under R but exists also as clickable user interface requiring no particular knowledge of R. GENELAND's main goal is to process individual multilocus genetic data to detect population structure, i.e. sub-populations at Hardy-Weinberg and linkage equilibrium. Although the concept of population refers here to genetic structure only, it is often realistic to assume that populations are spatially organized. Toward this aim, GENELAND is based on a spatially explicit model that can make use of both geographic and genetic information to estimate the number of populations in a dataset and delineate their spatial organization.

The general principal of Bayesian methods, which is the approach developed in STRUCTURE, is to consider data and parameters as random variables (Beaumont & Rannala, 2004). Bayesian statistical analysis is becoming very popular in quantitative genetics. Informally, Bayesian analysis is a natural extension of maximum likelihood. One reason that Bayesian methods have recently become very popular is that the very difficult issues of analytically obtaining the full posterior distribution for interesting problems has been complete circumvented by Markov Chain Monte Carlo (MCMC) methods. Bayesian statistics is concerned with generating the posterior distribution of the unknown parameters given both the data and some prior density for these parameters. As such, Bayesian statistics provides a much more complete picture of the uncertainty in the estimation of the unknown parameters, especially after the confounding effects of nuisance parameters are removed. These random variables have specific distributions, called *a priori* distribution. The critical feature of any Bayesian analysis is the choice of a prior. The key here is that when the data have sufficient signal, even a bad prior will still not greatly influence the posterior.

This is the main difference between non spatial Bayesian clustering methods, such as STRUCTURE, where this a priori distribution is uniform through the studied space. In GENELAND software, the a priori distribution is randomly modeled across space, using Poisson-Voronoi tessellation model. This model corresponds to the spatial patterns that can be expected when differentiation occurs by limited gene flow induced by the presence of physical barriers such as road, rivers, mountain ranges, human activity.

The parameters are inferred using MCMC iterations. The MCMC simulations are used to estimate the posterior probability that the data fit the hypothesis of K populations, $P(X|K)$. we tested values of K ranging from 2 to 6 with 3 independent runs per test, we used no admixture model with correlated allele frequencies (Falush et al., 2003), a 100,000 step burn-in followed by 10^6 steps of data collection. The

admixture model also calculates the fractional probability (*Q*) of individuals belonging to each population. In GENELAND, we tested the number of populations K ranging from 2 to 6 as well. We used correlated allele frequencies model. The burn-in was 100,000 iterations followed by 106 additional iterations, from which every 100th observation was sampled. The posterior probability of population membership for pixels was computed for the inferred k number of populations and the number of pixels was set to 2500 pixels along both the x and y axes. Finally, the posterior probability (probability $_{(i)}$ for the population i) of population membership was computed for pixels and the inferred population membership of individuals to model spatially the populations. F, individual F_{is} and pairwise F_{st} statistics (Weir and Cockerham 1984) relative to inferred populations obtained also from GENELAND. The spatial studied domain was divided into a grid of 2500 pixels and GENELAND calculates the posterior probability of every pixel to belong to a cluster (population).

3. 7. 2. Genetic distances

Genetic diversity:

$$GD = 1 - \sum_{i=1}^{n} F_i^2$$

ASD – Average Square Distance – (Goldstein, 1995):

$$ASD = \sum_{k=1}^{r} \sum_{i,j} \frac{(i-j)^2 x_{ij} y_{ij}}{r}$$

C_P (Prevosti et al., 1975):

$$C_P = \sum_{j=1}^{r} \sum_{i=1}^{m_j} \frac{\left| x_{ij} - y_{ij} \right|}{2r}$$

D_a(Nei et al., 1983):

$$Da = 1 - \frac{1}{r} \sum_{j=1}^{r} \sum_{i=1}^{m_j} \frac{2 \frac{(x_{ij} - y_{ij})^2}{(x_{ij} + y_{ij})}}{r}$$

D_C – Chord Distance – (Cavalli-Sforza, 1967):

$$D_C = \left(\frac{2}{\pi r}\right) \sum_{j=1}^{r} \sqrt{2\left(1 - \sum_{i=1}^{m_j} \sqrt{x_{ij} y_{ij}}\right.}$$

D_L (Latter, 1972):

$$D_L = -\ln(1 - F_{ST})$$

D_M – Nei minimum genetic distance – (Nei, 1973):

$$D_M = \frac{J_X + J_Y}{2} - J_{XY}$$

D_R – ROGERS distance – (Rogers, 1972):

$$D_R = \frac{1}{r}\sum_{j=1}^{r}\sqrt{\frac{\sum_{i=1}^{m_j}(x_{ij} - y_{ij})^2}{2}}$$

D_s – Nei standard genetic distance – (Nei, 1972):

$$D_S = -\ln\left(\frac{J_{XY}}{J_X J_Y}\right)$$

D_{sw} (Schriver, 1995):

$$D_{SW} = W_{XY} - \frac{(W_X + W_Y)}{2}$$

F_{ST} – LATTER'S FST distance – (Latter, 1972):

$$F_{ST} = \frac{\frac{(J_X + J_Y)}{2} - J_{XY}}{1 - J_{XY}}$$

X^2 (Sanghvi, 1953):

$$X^2 = \sum_{j=1}^{r}\sum_{i=1}^{m_j}\frac{2\frac{(x_{ij} - y_{ij})^2}{(x_{ij} + y_{ij})}}{r}$$

Where:

x_{ij} and y_{ij} are the frequencies of allele i at locus j for population x and y.

m_j is the number of alleles at locus j.

r is the number of evaluated loci.

$$J_X = \sum_{j=1}^{r} \sum_{i=1}^{m_j} \frac{x_{ij}^2}{r} \text{ (average of the heterozygosities for population } X\text{)}$$

$$J_Y = \sum_{j=1}^{r} \sum_{i=1}^{m_j} \frac{y_{ij}^2}{r} \text{ (average of the heterozygosities for population } Y\text{)}$$

$$J_{XY} = \sum_{j=1}^{r} \sum_{i=1}^{m_j} \frac{x_{ij} y_{ij}}{r}$$

1

$$W_X = \sum_{k=1}^{r} \sum_{i \neq j} \frac{|i-j| x_{ik} x_{jk}}{r}$$

$$W_Y = \sum_{k=1}^{r} \sum_{i \neq j} \frac{|i-j| y_{ik} y_{jk}}{r}$$

$$W_{XY} = \sum_{k=1}^{r} \sum_{i \neq j} \frac{|i-j| x_{ik} y_{jk}}{r}$$

2
3
4
5
6
7
8
9
10
11
12
13

4. Results

4.1. Phenotypic results

4. 1. 1. Agronomy

Grain yield (GY) was less diverse ranging from 2172 KG/Ha to 2529 KG/Ha. The most diverse traits for the Moroccan durum landraces were SPM2, KSPK, PL, SL and PH. For quality traits, PC was the most diverse and ranged from 19 to 36% (Table10).

Table 10: Descriptive statistics of measured traits

Trait	Min	Max	Mean	SD	Var
Grain Yield	2172.93	2529.68	2367.35	66.85	4469.40
Number of spike per square meter	100.00	440.00	218.88	50.34	2534.17
Number of kernels per main spike	8.00	32.00	17.79	5.61	31.48
Thousand kernel weight	30.29	47.07	39.18	4.543	20.641
Peduncle Length	1.00	8.00	2.27	1.56	2.44
Spike Length	4.00	9.00	6.84	1.20	1.43
Plant Height	86.91	114.34	105.54	5.96	35.54
Days to Heading	138.27	150.20	144.54	2.81	7.90
Days to Maturity	167.54	178.92	173.67	3.19	10.16
Grain Filling Duration	27.82	33.82	30.67	1.27	1.62
ASH content	2.97	3.04	3.02	0.01	0.0002
Protein Content	19.52	36.73	25.76	2.20	4.86
Sedimentation	2.94	5.17	3.83	0.31	0.10
Sedimentation N	1.33	2.62	1.76	0.16	0.03
Sedimentation Index	3.90	6.02	5.16	0.46	0.22
Yellow pigment	3.90	6.02	5.16	0.463	0.215
Vitreousness	93.36	94.42	94.08	0.19	0.04

Cluster analysis differentiated two main groups at the distance of 3.5 (Figure 11). GY was tightly linked to KSPK and PL within the first group where we could also find SPM2, GFD, ASH, YP and the three sedimentation traits. The second group contained SL, TKW, PC, VIT but also PH, DH and DM. Ash content in mature kernels could provide information on the integrated photosynthetic and retranslocation processes during grain filling (Araus & Nachit., 1998). In such a way, leaf and kernel ash content have been correlated with yield in wheat (Araus et al., 1998; Merah et al., 1999, 2001; Monneveux et al., 2004) grown under different water regimes.

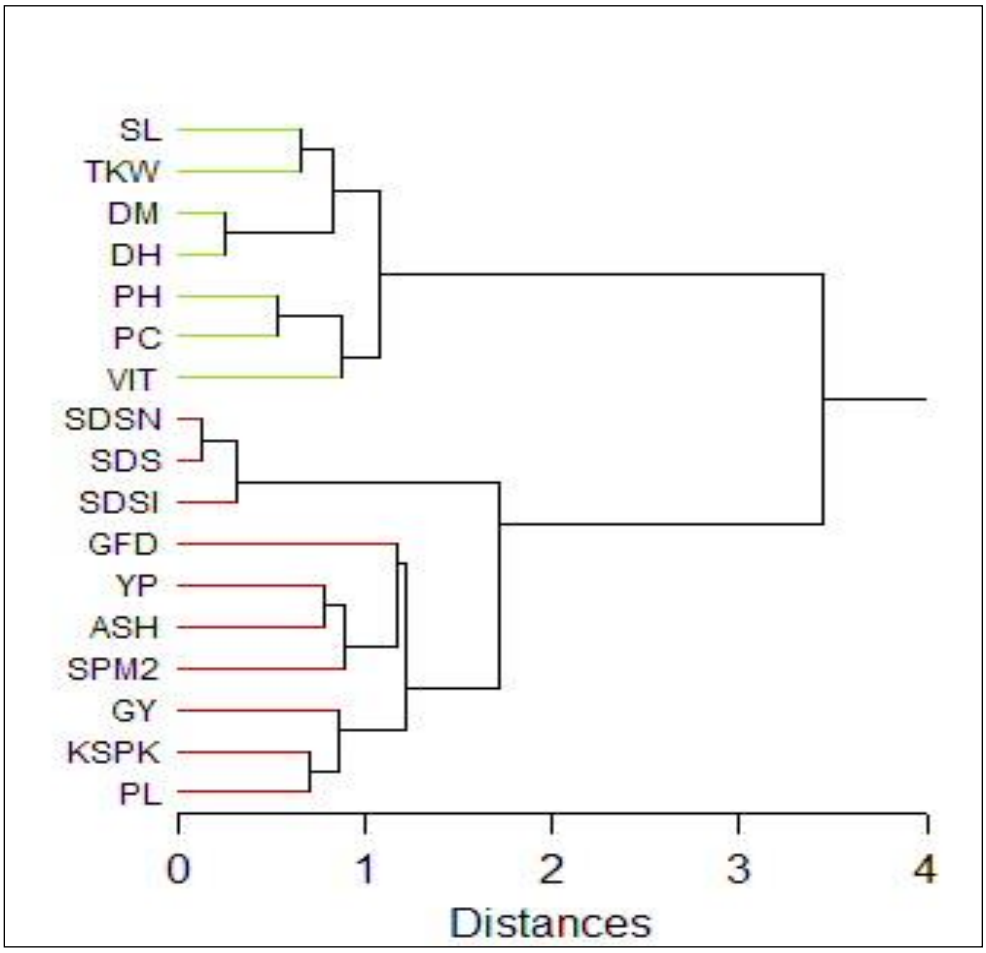


Figure 11: Cluster tree of GY, quality and agronomic traits

Modeling GY with the three yield components (TKW, SPM2 and KSPK) using multiple regressions had an R squared of 90%. The three resulting coefficients were highly significant and had a value of 0.64, 0.19 and 0.17 for TKW, SPM2 and KSPK respectively. The grain yield of durum wheat landraces reached then 90% of yield potential which is the product of kernel number by the number of kernels per square meter.

The multiple regression between GY and all agronomic traits gave only 4 significant effects ASH, DH, SL and TKW and explained 63% of the GY. Latitude and longitude were significant but having coefficients having the same absolute value with opposite signs (Table 11).

Table 11: Yield relationship with other traits (Agronomic, quality, yield, coordinates) using multiple regression

	Coefficient	Std Error	Std Coef	Tolerance	t	P
CONSTANT	7511.924	1349.77	0	.	5.565	0
ASH	-1233.646	474.154	-0.232	0.774	-2.602	0.011
DH	-7.325	2.481	-0.308	0.563	-2.952	0.004
SL	-30.276	5.516	-0.541	0.63	-5.489	0
TKW	5.101	1.711	0.347	0.453	2.981	0.004
LAT	-9.002	4.091	-0.231	0.557	-2.2	0.03
LONG	9.758	3.689	0.279	0.549	2.645	0.01

The residuals from modeling GY using the three yield components (TKW, SPM2 and KSPK) ranged between -67 kg/ha to 165 kg/ha. Seven landraces had low residuals from this model and were 1 from Tetouan in the North of the country, 2 from Taza, 1 from Fes, 2 from Beni Mellal inn the Atlas Mountains and one from the South in Tiznit. Modeling GY with the three yield components had large residuals in the high altitude of Rif mountain chain, extreme South of Morocco in the region of Tiznit and the Eastern part of the country (Figure 12).

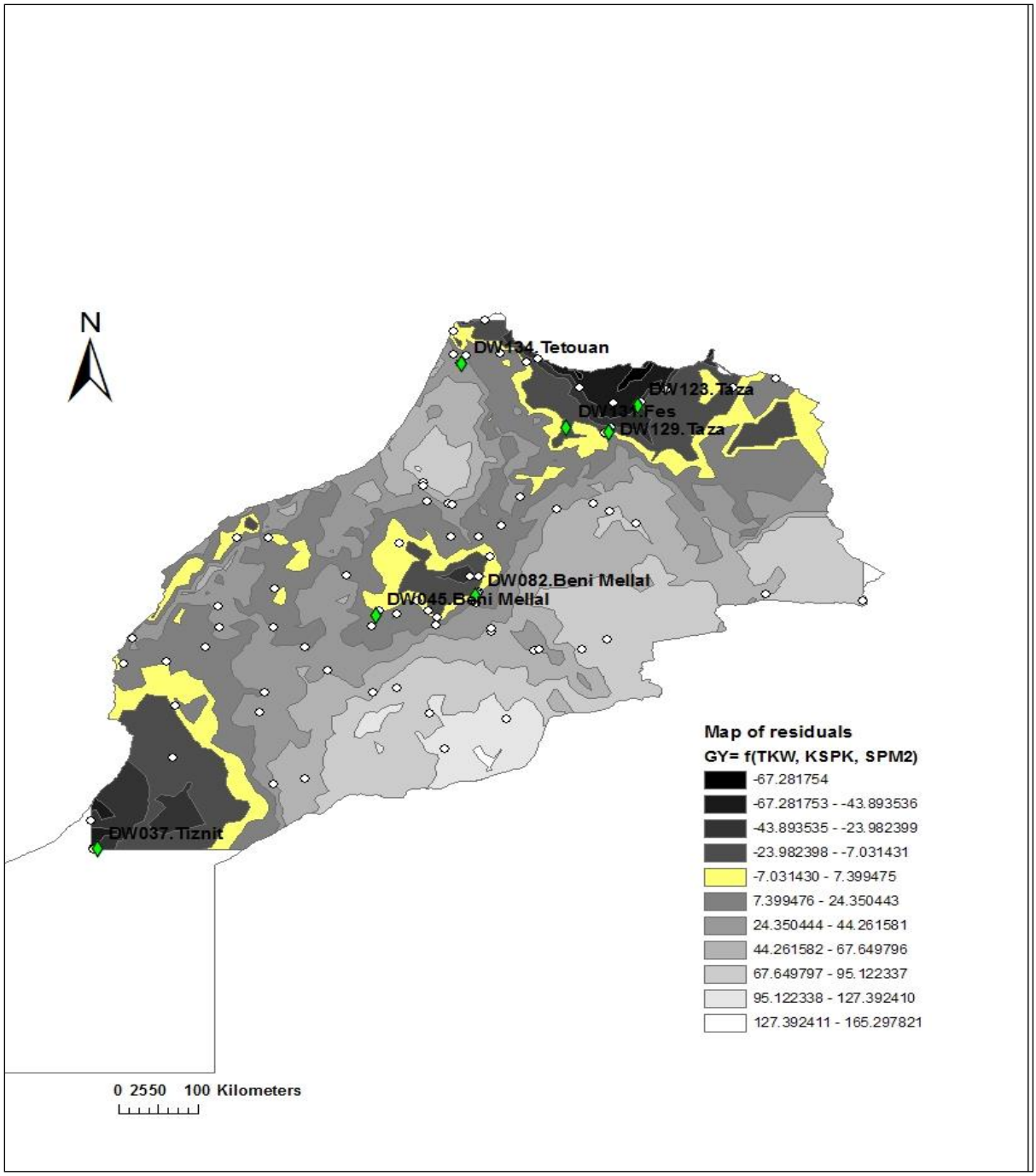


Figure 12: Map interpolated residuals from multiple regression of yield on its components (GY = f (TKW, KSPK, SPM2)

Correlation between GY and TKW was negative while correlation with SPM2 and KSPK were positive. The direct effects of the three yield components on GY were positive and highest effect was of KSPK. The gain on direct effect on yield by TKW was reduced by the negative indirect effect of TKW via KSPK and SPM2. On the other hand, the gain by direct effects of SPM2 and KSPK was reduced by the indirect effect of the two traits via TKW (Table 12). Gains in KM2, however, do not translate directly in yield potential gain due to partial compensation by decreased KW (Slafer et al., 1996). The lower KW observed with increased KM2 is not only due to a lower amount of assimilates per grain but is the result of an increased number of grains with a lower weight potential. It has been shown that competition for limited resources during the spike growth period, including light and nitrogen, is the major cause of KM2 potential loss (Slafer et al., 1996).

Table 12: Path coefficients (direct and indirect effects) of yield components to grain yield

	TKW	SPM2	KSPK
TKW	**0.156**	-0.071	-0.085
SPM2	-0.059	**0.128**	0.016
KSPK	-0.162	0.038	**0.294**

For correlations, six traits showed significant negative correlations with GY (ASH, PC, PH, DH, DM, and SL) and four traits showed positive correlations (GFD, PL, KSPK and SDSi). The highest positive correlation was founds between PH and PC and DM and DH and TKW. The highest negative correlations found between YP and SL, and DM and KSPK (Table 13). TKW was positively correlated with PC, DH, DM and SL and negatively associated with ASH, YP, SPM2, SL and KSPK. KSPK was linked positively to PL and negatively to PC, PH, DH, DM, TKW, VIT and SL.

Table 13: Pearson correlation between grain yield and agronomic, phonologic and quality traits

	GY	ASH	PC	PH	DH	DM	GFD	SDS	SDSn	SDSi	YP	TKW	VIT	SL	SPM2	PL	KSPK
GY	1.00	-0.25	-0.22	-0.35	-0.37	-0.25	0.12	0.08	0.04	0.13	0.12	-0.06	-0.04	-0.42	0.09	0.22	0.23
ASH	-0.25	1.00	0.59	0.45	0.20	0.05	-0.27	0.04	0.20	-0.14	0.41	-0.22	0.42	-0.23	0.21	-0.14	-0.06
PC	-0.22	0.59	1.00	0.65	0.50	0.41	-0.22	-0.07	0.15	-0.29	0.08	0.36	0.45	0.17	-0.11	-0.26	-0.39
PH	-0.35	0.45	0.65	1.00	0.65	0.55	-0.13	0.26	0.39	0.11	0.11	0.38	0.44	0.22	-0.12	-0.26	-0.47
DH	-0.37	0.20	0.50	0.65	1.00	0.88	0.01	-0.18	-0.11	-0.26	0.13	0.55	0.21	0.40	-0.36	-0.38	-0.66
DM	-0.25	0.05	0.41	0.55	0.88	1.00	0.30	-0.10	-0.05	-0.14	0.00	0.67	0.21	0.42	-0.36	-0.41	-0.73
GFD	0.12	-0.27	-0.22	-0.13	0.01	0.30	1.00	0.16	0.10	0.23	0.01	0.02	-0.24	0.04	-0.05	-0.08	-0.06
SDS	0.08	0.04	-0.07	0.26	-0.18	-0.10	0.16	1.00	0.97	0.97	0.07	-0.11	0.32	-0.18	0.17	0.02	0.11
SDSn	0.04	0.20	0.15	0.39	-0.11	-0.05	0.10	0.97	1.00	0.88	0.11	-0.07	0.43	-0.18	0.18	-0.02	0.06
SDSi	0.13	-0.14	-0.29	0.11	-0.26	-0.14	0.23	0.97	0.88	1.00	0.02	-0.15	0.19	-0.19	0.16	0.08	0.16
YP	0.12	0.41	0.08	0.11	0.13	0.00	0.01	0.07	0.11	0.02	1.00	-0.40	0.24	-0.63	0.15	0.09	0.08
TKW	-0.06	-0.22	0.36	0.38	0.55	0.67	0.02	-0.11	-0.07	-0.15	-0.40	1.00	0.26	0.58	-0.46	-0.23	-0.55
VIT	-0.04	0.42	0.45	0.44	0.21	0.21	-0.24	0.32	0.43	0.19	0.24	0.26	1.00	-0.17	0.07	-0.02	-0.25
SL	-0.42	-0.23	0.17	0.22	0.40	0.42	0.04	-0.18	-0.18	-0.19	-0.63	0.58	-0.17	1.00	-0.40	-0.20	-0.34
SPM2	0.09	0.21	-0.11	-0.12	-0.36	-0.36	-0.05	0.17	0.18	0.16	0.15	-0.46	0.07	-0.40	1.00	0.08	0.13
PL	0.22	-0.14	-0.26	-0.26	-0.38	-0.41	-0.08	0.02	-0.02	0.08	0.09	-0.23	-0.02	-0.20	0.08	1.00	0.42
KSPK	0.23	-0.06	-0.39	-0.47	-0.66	-0.73	-0.06	0.11	0.06	0.16	0.08	-0.55	-0.25	-0.34	0.13	0.42	1.00

Positive direct effect DE was found on GY for quality traits (TKW, PC, SDS). This DE was reduced by the indirect effect IE of these traits via SDS and SL for TKW; via SDSn, SDSi and ASH for PC; and SDSn and SDSi for SDS. On the other hand, SL and VIT had a negative DE on GY (Table 14).

Table 14: Path coefficients (direct in diagonal and indirect effects in column) of agronomic, phonologic and quality traits to grain yield

	ASH	**PC**	**PH**	**DH**	**DM**	**GFD**	**SDS**	**SDSn**	**SDSi**	**YP**	**TKW**	**VIT**	**SL**	**SPM2**	**PL**	**KSPK**
ASH	**-0.17**	-0.10	-0.08	-0.04	-0.01	0.05	-0.01	-0.03	0.02	-0.07	0.04	-0.07	0.04	-0.04	0.02	0.01
PC	0.11	**0.19**	0.12	0.09	0.08	-0.04	-0.01	0.03	-0.05	0.02	0.07	0.08	0.03	-0.02	-0.05	-0.07
PH	-0.09	-0.12	**-0.19**	-0.12	-0.10	0.03	-0.05	-0.07	-0.02	-0.02	-0.07	-0.08	-0.04	0.02	0.05	0.09
DH	-0.07	-0.17	-0.23	**-0.35**	-0.31	0.00	0.06	0.04	0.09	-0.05	-0.19	-0.07	-0.14	0.13	0.13	0.23
DM	0.00	0.02	0.03	0.05	**0.05**	0.02	-0.01	0.00	-0.01	0.00	0.04	0.01	0.02	-0.02	-0.02	-0.04
GFD	-0.02	-0.01	-0.01	0.00	0.02	**0.06**	0.01	0.01	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00
SDS	0.15	-0.27	1.09	-0.76	-0.40	0.68	**4.13**	4.01	4.00	0.28	-0.46	1.33	-0.76	0.71	0.10	0.46
SDSn	-0.43	-0.33	-0.85	0.24	0.10	-0.22	-2.12	**-2.18**	-1.92	-0.24	0.15	-0.93	0.40	-0.38	0.05	-0.12
SDSi	0.28	0.58	-0.22	0.51	0.28	-0.46	-1.93	-1.76	**-1.99**	-0.05	0.30	-0.37	0.38	-0.32	-0.15	-0.32
YP	0.05	0.01	0.01	0.02	0.00	0.00	0.01	0.01	0.00	**0.12**	-0.05	0.03	-0.08	0.02	0.01	0.01
TKW	-0.11	0.19	0.20	0.29	0.35	0.01	-0.06	-0.04	-0.08	-0.21	**0.52**	0.14	0.30	-0.24	-0.12	-0.29
VIT	-0.08	-0.08	-0.08	-0.04	-0.04	0.04	-0.06	-0.08	-0.03	-0.04	-0.05	**-0.18**	0.03	-0.01	0.00	0.05
SL	0.13	-0.10	-0.13	-0.23	-0.24	-0.03	0.11	0.11	0.11	0.37	-0.33	0.10	**-0.58**	0.23	0.11	0.20
SPM2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.01**	0.00	0.00
PL	-0.01	-0.02	-0.02	-0.03	-0.04	-0.01	0.00	0.00	0.01	0.01	-0.02	0.00	-0.02	0.01	**0.09**	0.04
KSPK	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**-0.01**

Relationship between two traits may change across space. One method of assessing this change is to map residuals from a linear regression model. The three figures (Figure 13, 14, 15) show the spatial distribution of residuals from linear regression of GY on TKW, SPM2 and KSPK. When the residuals are very small, the regression model between GY and the other trait is strong and the two traits are highly correlated in the corresponding landrace's collection site. Large residuals means that in the corresponding sites, the correlation between GY and the other traits is weak and that the linear regression model can't explain the variation of GY.

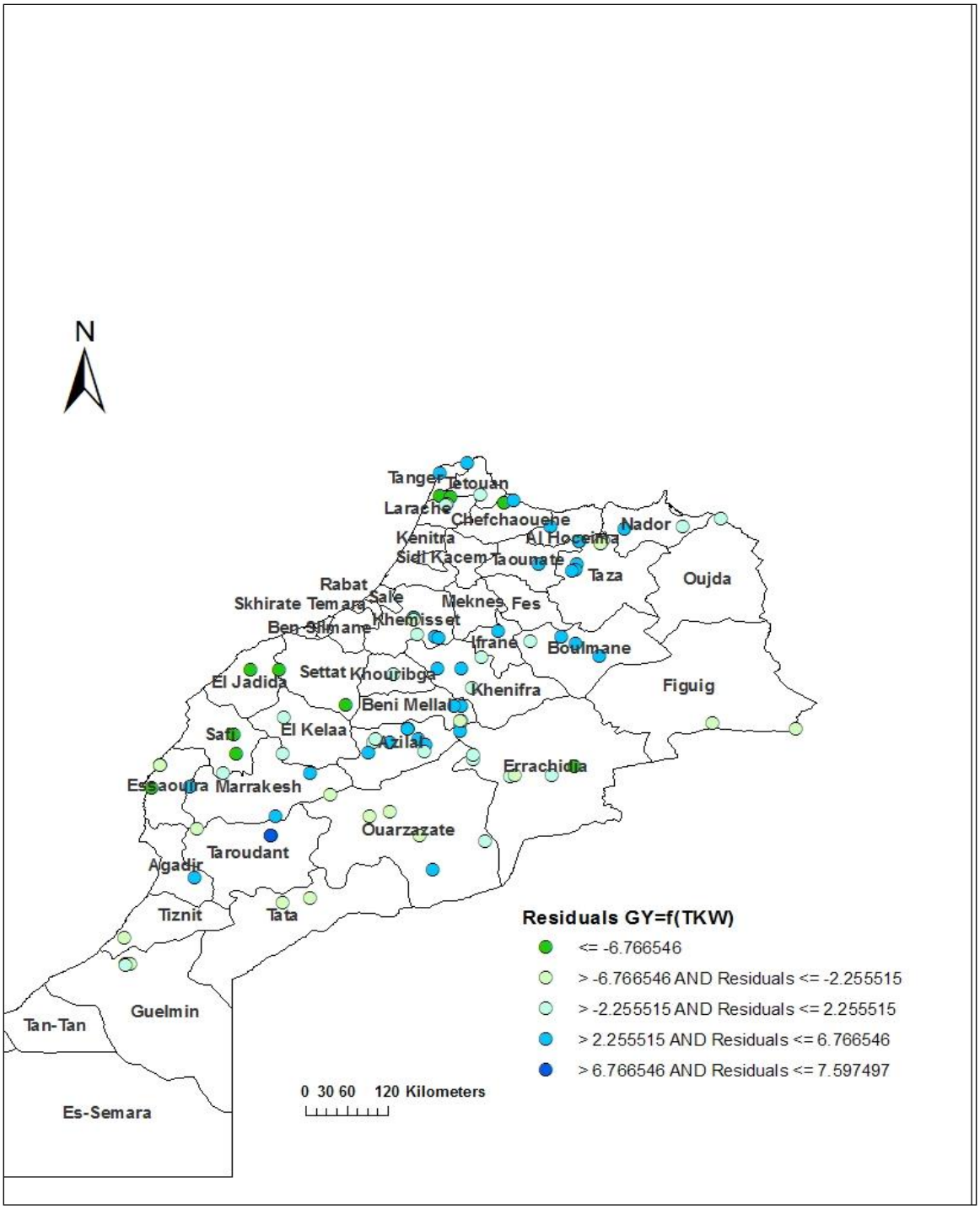


Figure 13: Map of residuals from regression of yield on TKW

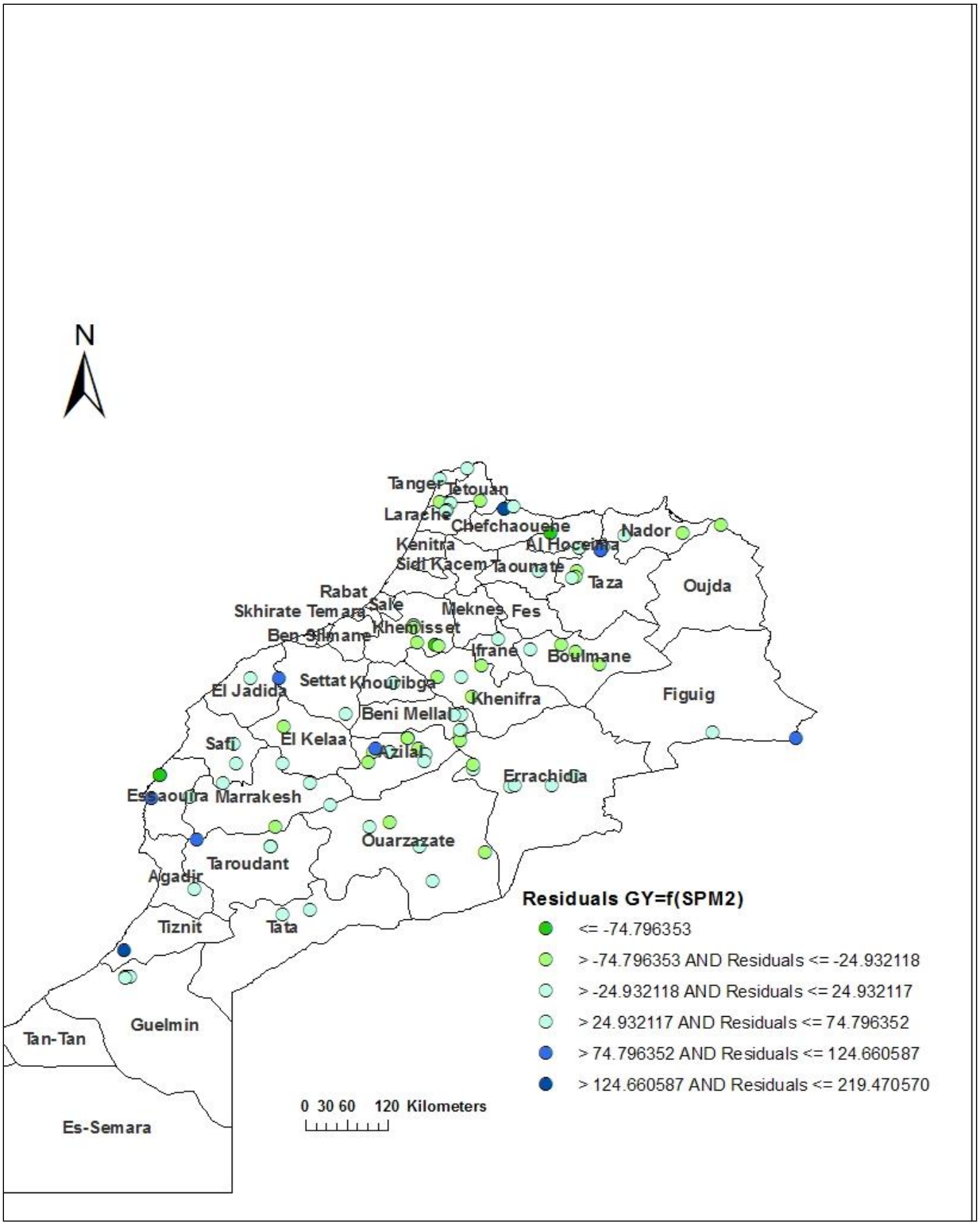


Figure 14: Map of residuals from regression of yield on SPM2

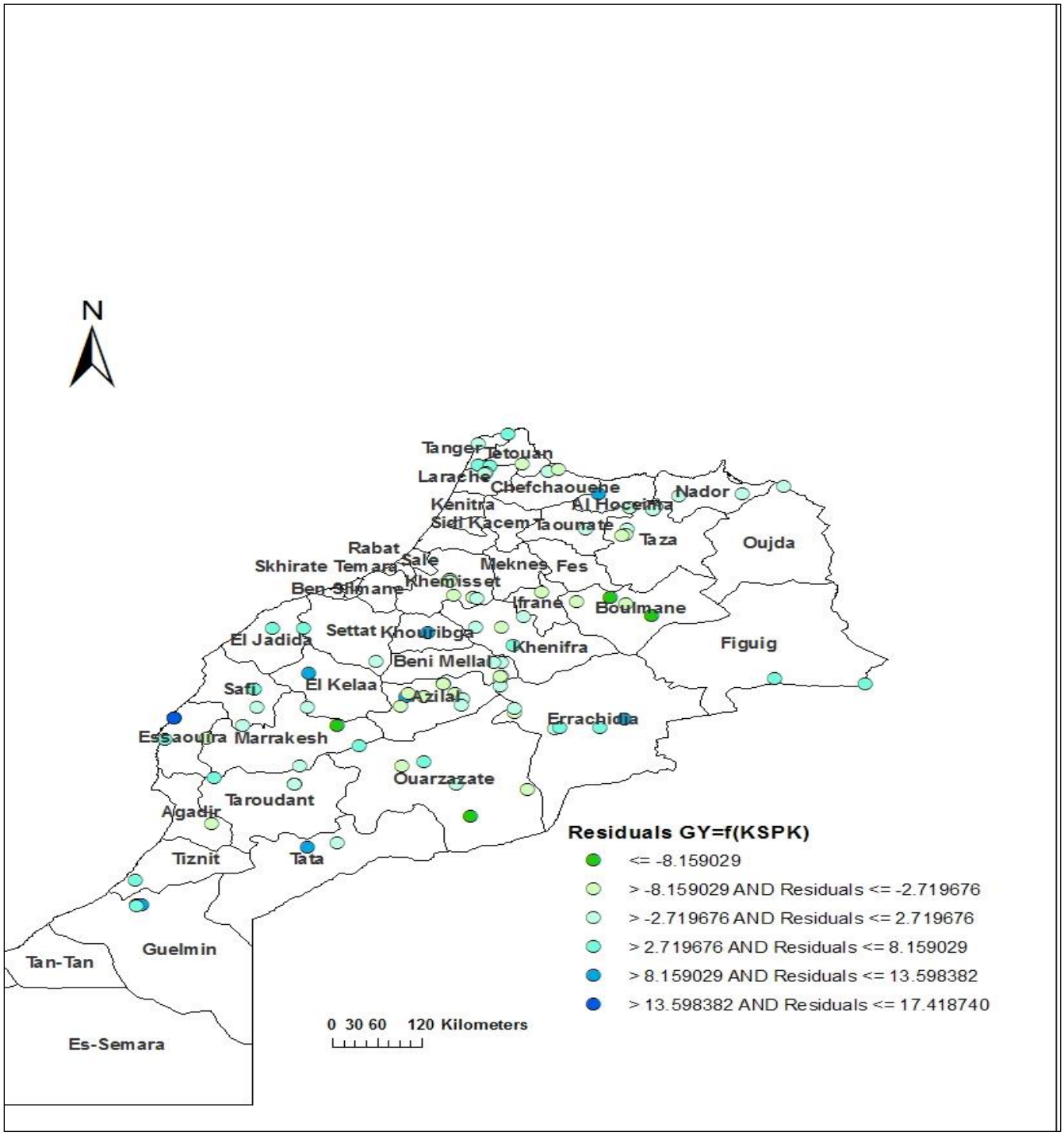


Figure 15: Map of residuals from regression of yield on KSPK

4. 1. 2.Physiology

Large variability was shown by the physiological traits for the Moroccan durum landraces. Most of the traits exhibited large range (Min, Max). Nevertheless, variability was small for traits with small scale (Carotene70, WI70, SAVI70…). Table 15 shows different range for measured physiological traits.

Table 15: Descriptive statistics of measured physiological traits

Trait	Min	Max	Mean	SD	Var
CARI70	0.965	0.970	0.969	0.001	0.000
CAROTENE70	-0.001	-0.001	-0.001	0.000	0.000
CHL70	0.084	0.248	0.153	0.028	0.001
NDVI70	0.803	0.846	0.833	0.009	0.000
WI70	1.118	1.146	1.131	0.005	0.000
WINDVI70	1.376	1.424	1.385	0.009	0.000
NPCI70	0.1174	0.1987	0.1526	0.0189	0.0004
PRI70	0.0027	0.0248	0.0167	0.0050	0.0000
SAVI70	0.2826	0.2956	0.2918	0.0028	0.0000
SIPI70	0.8501	0.9213	0.8890	0.0154	0.0002
SR70	17.9145	19.2565	18.4270	0.3214	0.1033
RNVI70	2.8459	3.5687	3.2191	0.1485	0.0221
RVSI70	4.9393	5.6794	5.2106	0.1589	0.0252
F070	561.0000	870.0000	647.2000	58.4045	3411.0909
F170	638.0000	1008.0000	742.3100	62.0306	3847.7918
F270	717.0000	1162.0000	843.7300	74.6229	5568.5829
F370	1011.0000	1680.0000	1204.2600	119.2294	14215.6489
F470	1471.0000	2288.0000	1762.9400	171.8088	29518.2590
F570	1858.0000	3214.0000	2701.7500	387.8195	150403.9874
FM70	2640.0000	4083.0000	3579.1400	397.6319	158111.1317
FVFM70	0.7070	0.8540	0.8164	0.0297	0.0009
FV70	1883.0000	3450.0000	2931.9400	409.1796	167427.9358
LWP70	3.4141	6.8574	5.5768	0.8103	0.6566
AREA70	40500.0000	214000.0000	90097.0000	18751.6004	351622516.1616
CARI45	0.6572	1.0126	0.9708	0.0474	0.0022
CAROTENE45	-0.0025	-0.0001	-0.0011	0.0005	0.0000
CHL45	0.0235	0.5081	0.1453	0.0685	0.0047
NDVI45	0.3638	0.9752	0.8758	0.0888	0.0079
NPCI45	-0.1630	0.6175	-0.0067	0.1357	0.0184
NPQ70	0.1531	0.5113	0.3330	0.0789	0.0062
NPQI45	-0.1035	-0.0298	-0.0612	0.0139	0.0002
PRI45	-0.1772	0.0803	0.0447	0.0380	0.0014

QN70	0.1727	0.4354	0.3036	0.0579	0.0033
QP70	0.9297	0.9702	0.9530	0.0083	0.0001
QUE70	0.1707	0.4142	0.2264	0.0464	0.0022
RNVI45	1.2600	4.9628	3.1491	0.6471	0.4188
RVSI45	2.1661	12.1344	4.6437	1.6217	2.6301
SAVI45	-0.0587	0.0021	-0.0243	0.0109	0.0001
SIPI45	0.7381	0.9692	0.8845	0.0407	0.0017
SR45	2.1438	79.7261	19.7169	10.5879	112.1029
TFM70	233.0000	588.0000	359.9200	55.0283	3028.1147
WINDVI45	1.1709	2.7676	1.2857	0.2064	0.0426
WI45	0.9801	1.2783	1.1092	0.0380	0.0014
YPEC70	0.5888	0.7827	0.7196	0.0458	0.0021

Cluster tree using physiological traits and GY had two groups containing both two sub-groups. We could state that the fluorescence traits were in one group. The physiological traits were in general grouped according to the Zadoc scale 45 or 70. The GY was highly linked to CHL, WI/NDVI and NPCI of both stages (Zadoc scale 45 and 70). The quenching traits were also affected to the GY group (Figure 16). Most of traits at Zadoc 70 were linked to GY.

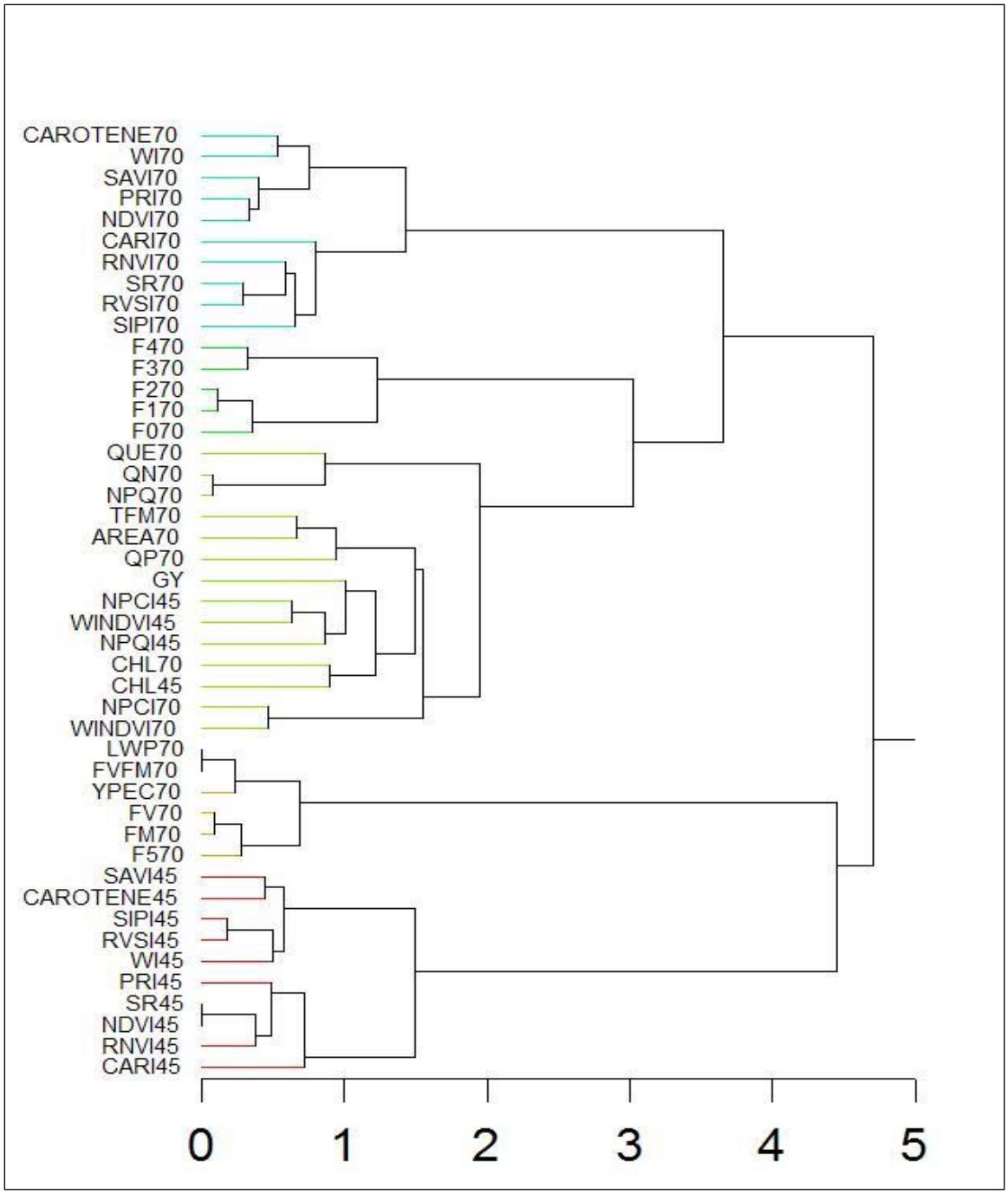


Figure 16: Cluster tree of physiological traits with GY

Only 4 physiological traits had significant coefficients as independent variables (F070, LWP70, PRI45, YPEC70) in a multiple regression for the GY (Table 16). The model resulting had an R square of 44.5%. The physiological traits couldn't explain more than 50% of GY while the explanation by agronomic traits reached 63%.

Table 16: Yield relationship with physiological traits using multiple regression

	Coefficient	Std Error	Std Coef	Tolerance	t	P
CONSTANT	1881.768	443.993	0	.	4.238	0
F070	0.49	0.169	0.428	0.395	2.896	0.005
LWP70	131.129	33.287	1.589	0.053	3.939	0
PRI45	-1050.849	507.183	-0.597	0.104	-2.072	0.041
YPEC70	-1686.297	510.509	-1.156	0.07	-3.303	0.001

4. 1. 3. Agronomy & physiology

Combining all the traits (agronomic and physiologic) in the cluster analysis differentiated two groups: the first one contains the GY, which was again tightly linked not only to KSPK and PL but also to QP, NPCI and WI/NDVI at Zadoc scale70. The second group differentiated the physiological traits into two small groups: one for Zadoc scale 45 and the other for Zadoc scale 70. All the agronomic traits belonging to the second group were clustered together. Most of the physiological traits linked to GY were measured at Zadoc scale 70.

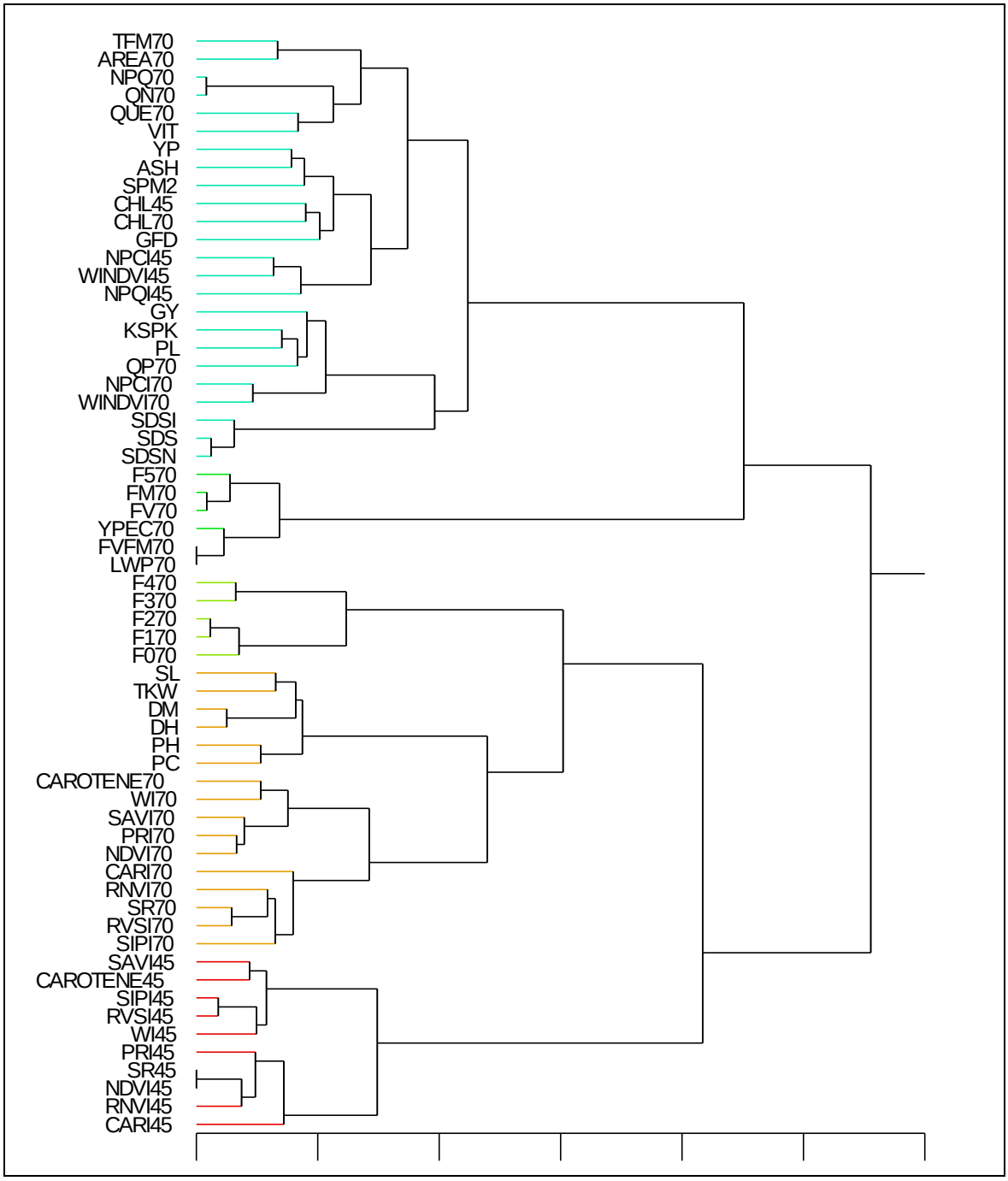


Figure 17: Cluster tree of all measured traits

Modeling GY using multiple regression and using all traits as dependent variables explained 74.9%. Only 14 traits had a significant coefficients (Table 17). Only 10% more could be explained by adding the physiology to the regression model. Breeders can explain most of the GY variability using only the agronomy which is less expensive compared to physiology.

Table 17: Yield relationship with all measured traits using multiple regression

	Coefficient	Std Error	Std Coef	Tolerance	t	P
CONSTANT	13762.757	4148.19	0	.	3.318	0.001
DH	-6.39	2.608	-0.269	0.45	-2.45	0.016
TKW	7.965	2.025	0.541	0.286	3.933	0
VIT	-104.569	43.832	-0.292	0.36	-2.386	0.019
SL	-40.818	6.886	-0.73	0.357	-5.928	0
PL	7.567	3.706	0.177	0.721	2.042	0.044
F170	-3.958	1.302	-3.673	0.004	-3.04	0.003
F470	0.316	0.125	0.811	0.052	2.518	0.014
F570	0.58	0.163	3.362	0.006	3.553	0.001
LWP70	-196.293	71.45	-2.379	0.007	-2.747	0.007
NDVI45	-416.503	161.805	-0.553	0.117	-2.574	0.012
NPQ70	1507.76	347.341	1.779	0.032	4.341	0
NPQI45	713.294	416.5	0.148	0.721	1.713	0.091
WINDVI45	-138.299	69.113	-0.427	0.119	-2.001	0.049
SDS	377.958	135.176	12.472	0	2.796	0.006
SDSN	-1460.988	512.314	-6.731	0.001	-2.852	0.006
SDSI	-2515.868	926.265	-6.11	0.001	-2.716	0.008

Residuals from two multiple regressions analysis, the first one regressing GY on agronomic traits and the second one regressing GY on physiological traits were correlated at 0.594 with a p-value of 0 (Figure 18).

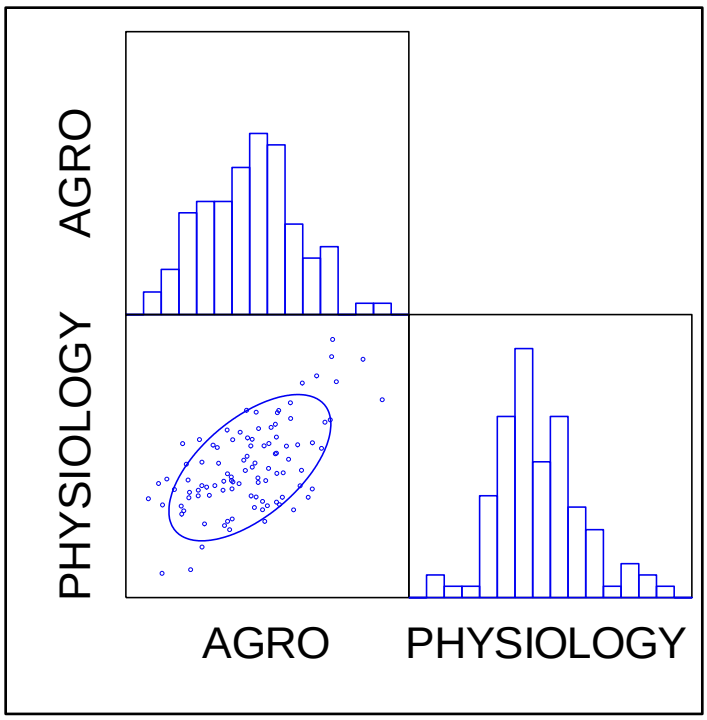


Figure 18: Correlation between residuals resulting from multiple regression of GY on agronomic traits (AGRO) and on physiological traits (PHYSIOLOGY)

As for linear regression, model found by multiple regression may change across space. Also, mapping residuals from this model can help evaluating this spatial change (Figure 19).

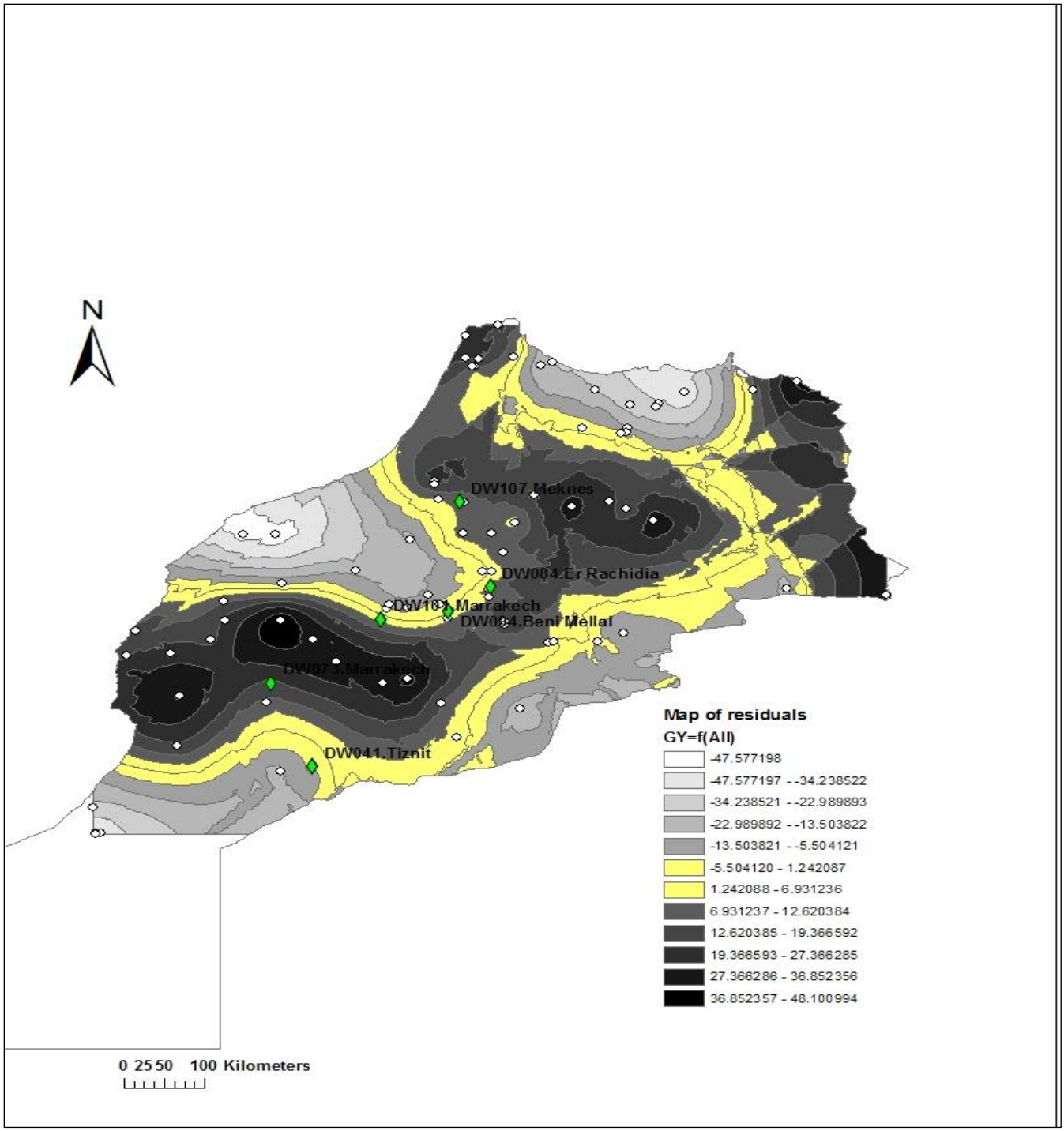


Figure 19: Map of residuals of multiple regression of GY on All traits (Yellow regions with the lowest residuals)

Clustering Moroccan landraces using both agronomic and physiological traits is groupping landraces according to their origins (locations of collection) or to the agro-ecological regions of adaptation (Figure 20).

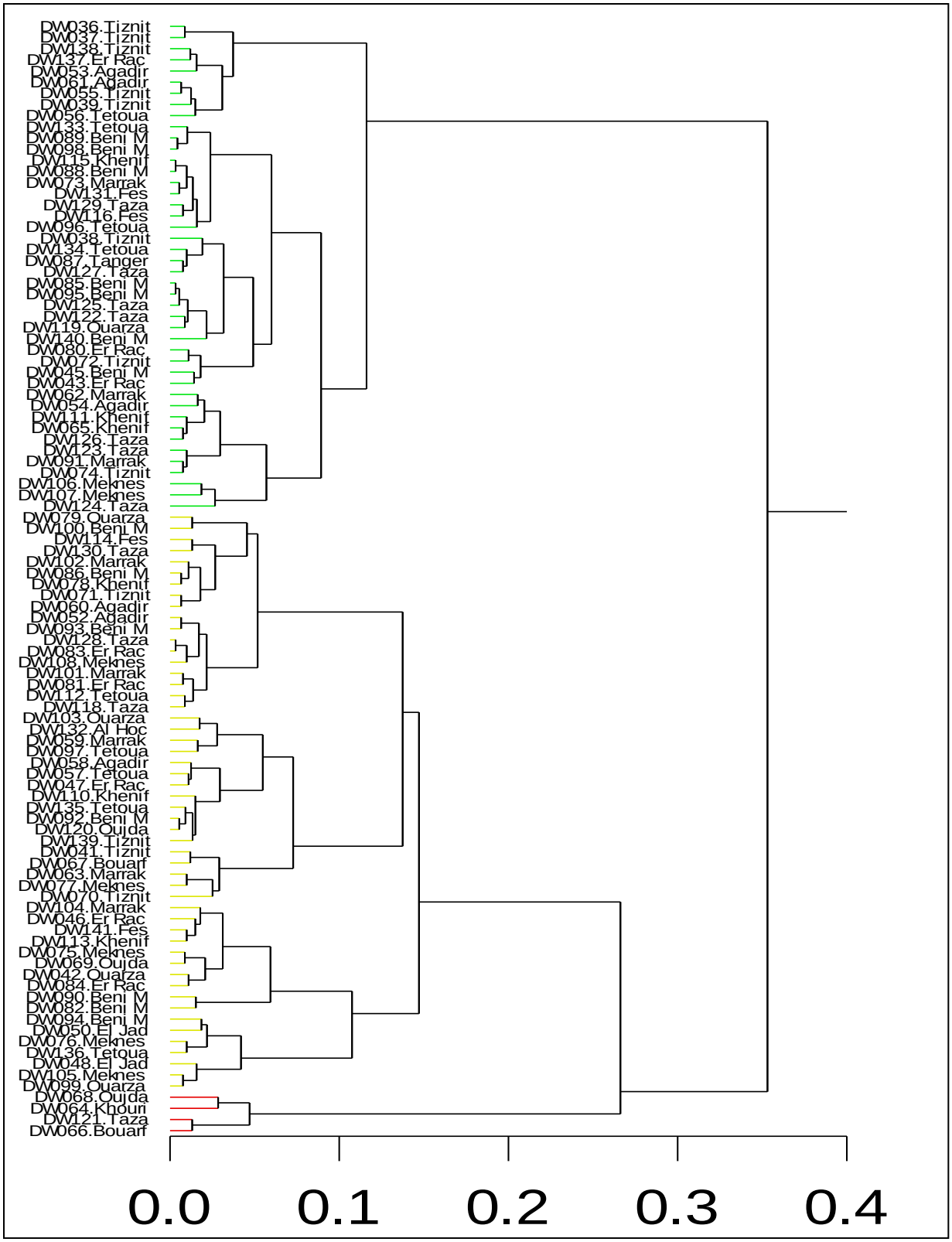


Figure 20: Cluster tree of Moroccan durum landraces using agronomic and physiological traits

4. 1. 4. GxE analysis

The GxE analysis showed that for the agronomic traits only TKW and DM had low G.E components (30% and 27% respect.). The maximum was found for VIT (98%), GY (94%) and ASH (91%). The variance components were equally divided into two parts G and G.E for two traits (DH and GFD) (Table 18). For physiology, the GxE component was very high for all traits and ranged from 100% (CHL70) to 50% for (SIPI70). Nine out of the thirteen studied traits had a GxE component of more than 90% (Table 18).

Table 18: Variance components of genotype and genotype by environment of studied traits

	G	GxE	Total
ASH	8.3	91.7	100
DH	49.4	50.6	100
DM	72.9	27.1	100
GFD	48.5	51.5	100
GY	5.5	94.5	100
PC	24.2	75.8	100
VIT	1.6	98.4	100
PH	35.3	64.7	100
SDS	30.5	69.5	100
SDSn	24.9	75.1	100
SDSI	35.2	64.8	100
TKW	69.1	30.9	100
YP	35.4	64.6	100
SIPI70	50	50	100
NPCI70	10.1	89.9	100
SAVI70	5.2	94.8	100
RNVI70	23.3	76.7	100
RVSI70	6.3	93.7	100
CARI70	4.6	95.4	100
CAROTENE70	2.8	97.2	100
CHL70	0	100	100
NDVI70	6.7	93.3	100
PRI70	12.2	87.8	100
SR70	2	98	100
WI70	9.7	90.3	100
WI/NDVI70	2.5	97.5	100

The variance components showed that the GxE effect was high and positive for landraces originated from the Eastern (Oujda, Figuig, Errachidia and Ouarzezate) part of Morocco and from the 'Moyen Atlas'

region of Meknes, Fes, Ifrane and Khenifra. The map of the GxE effect by landrace shows a spatial pattern across Morocco from the East (Positive effect) to the West (Negative effect). The extreme values of effects recorded as a minimum at landraces from Eljadida, Settat and Alhoceima; and as a maximum values at landraces from Errachidia, Boulmane and Figuig. The null value of the GxE effect by landrace was remarked at Marrakech, Essaouira, Agadir, Taza, Taouenate and Beni Mellal (Figure 21).

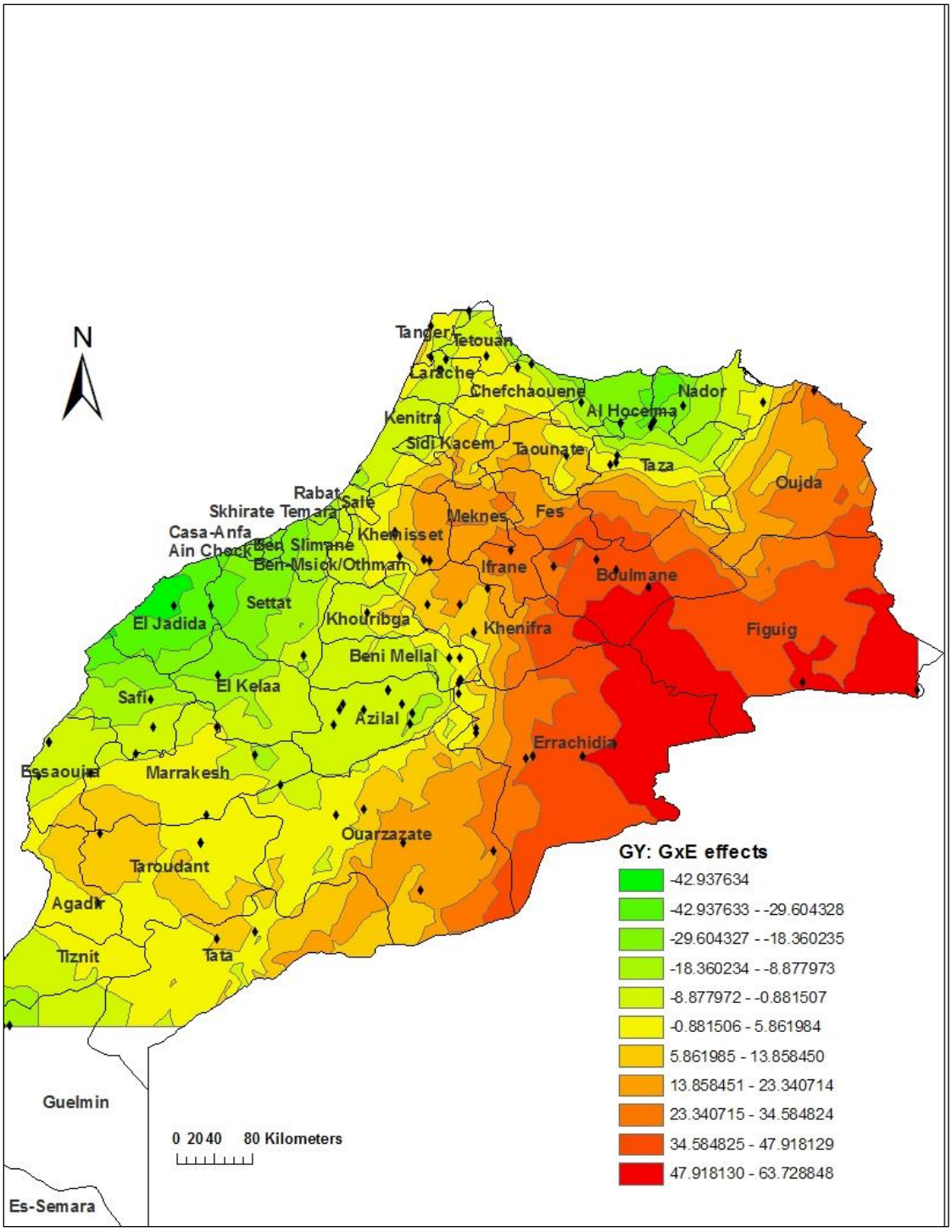


Figure 21: Map of mean sensitivity of a landrace to environment

For GY, the most stable Moroccan landraces were mainly from Errachidia, Ouarzazete, Taza, Tiznit. Using the non-parametric stability, the results were the same according to the mean rank over all

environments and the variance of rank. Landraces DW043 from Errachidia, DW103 from Ouarzazete and DW046 from Errachidia are the most stable for GY. Using the coefficient of variation of (Francis and Kanenberg), a landrace is stable when its coefficient is high and the most stable landraces were the ones originated from high altitudes of Atlas and Rif mountainous chains.

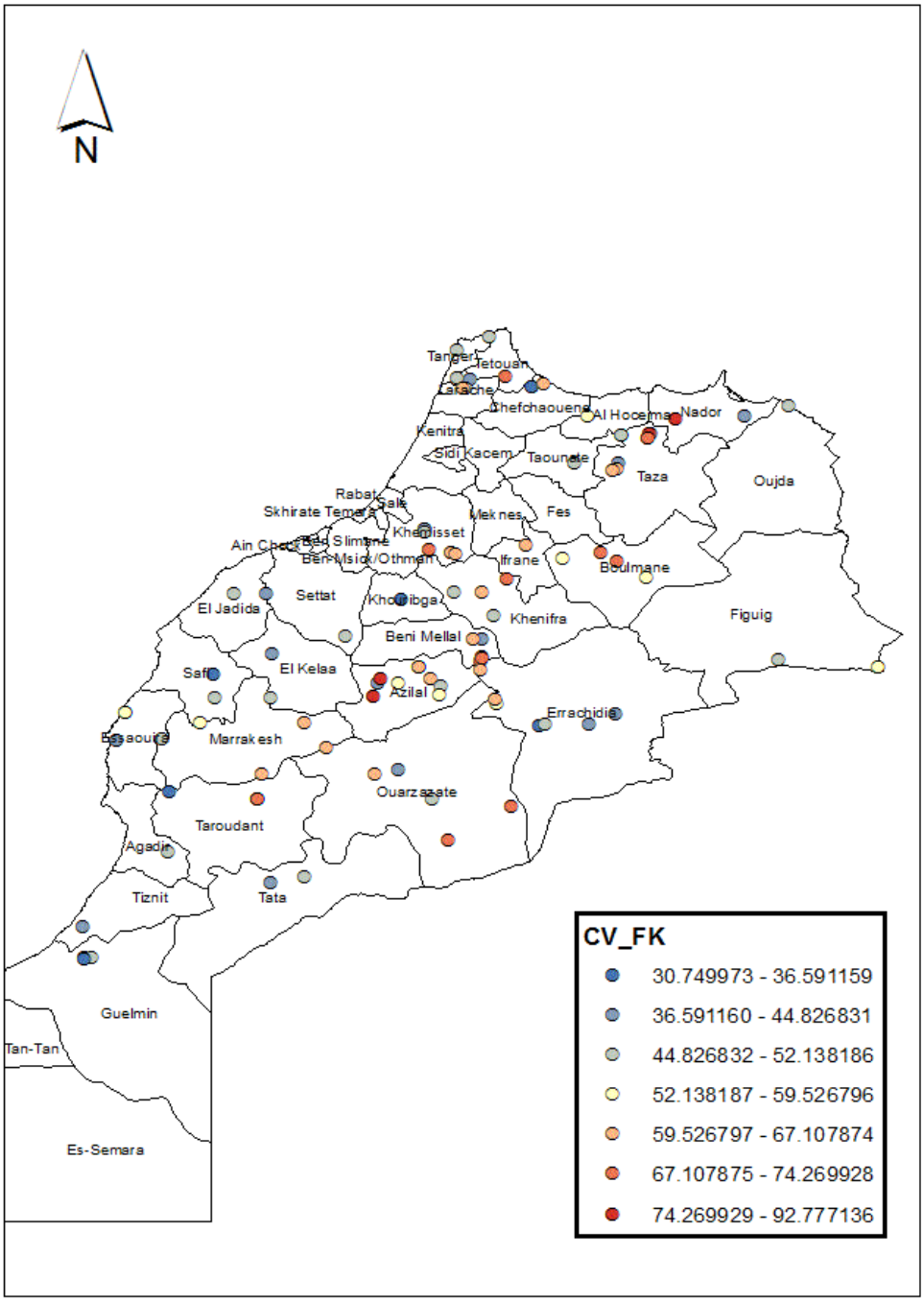


Figure 22: Map of coefficient of variation (Francis and Kanenberg) of Moroccan durum landraces

4. 1. 5. Spatial analysis of phenotypic data

4. 1. 5. 1. Spatial networks

The graphs (Figure 23) shows different spatial connectivity networks for the Moroccan wheat landraces. The fact that two landraces are spatially connected depends not only on the distance separating them but also on the method used for constructing the spatial network. Spatial analysis (spatial autocorrelation, Spatial PCA, spatial multiple analysis) depends a lot on the choice of the spatial connectivity networks. In this study, the six spatial networks were used to study the spatial phenomena of durum landraces. In most of the cases results were similar.

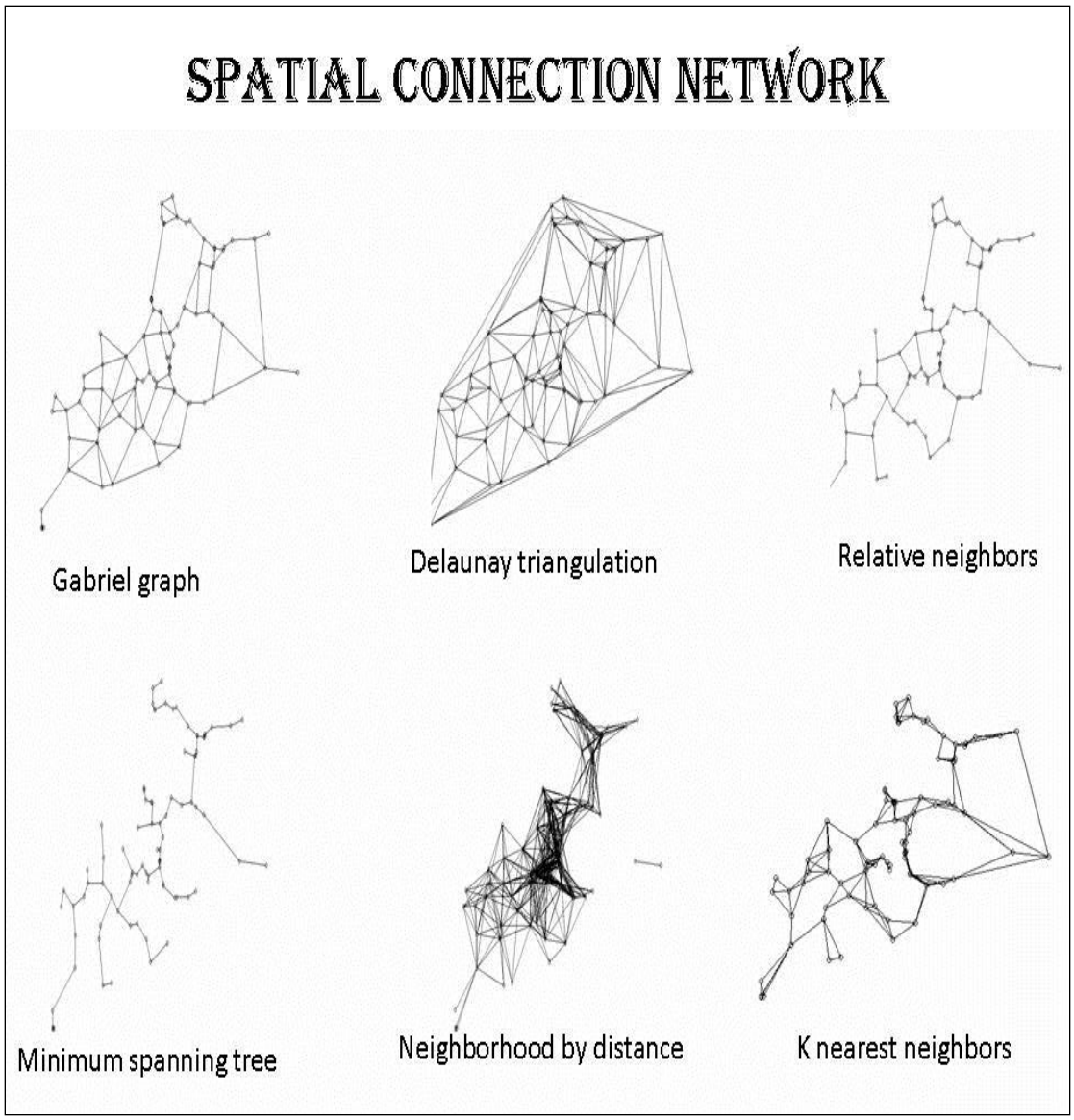


Figure 23: Examples of different spatial connectivity network of Moroccan durum landraces

4. 1. 5. 2. Spatial autocorrelation

We computed the spatial autocorrelation of traits/Variables using three different spatial connectivity networks; nearest neighbors with 3 and 5 neighbors, Gabriel graph and minimum spanning tree. The results were similar for all of them. Only 8 of the 18 studied agronomic traits had a significant SAU. The maximum found for DM (0.375) and minimum for PC (0.129). The G.E by landraces effect had also a significant positive SAU. These results showed that those variables had a global pattern over Morocco and that the agronomic traits cannot distinguish between neighboring landraces. This result is applicable for the G.E by landrace, which match with the results showed by the map generated by the G.E effect in the 'GxE section' (Table 19).

Four of the five variables generated by the stability of yield analysis had a significant positive SAU (V-FK with 0.21, CV-FK with 0.14, B-FW with 0.21 and R2 with 0.18). Using the rank, only the rank of GY04 and GY07. The average and the variance of rank over the four years didn't show any significant SAU. As for the agronomy, the SAU was positive or null for all the 44 physiological traits but only significant for 14 and 13 of them were of the Zadoc scale 70. The values ranged from 0.365 for SAVI70 to 0.126 for F270. Moran'I values from PCA-AGRO were no significant.

Table 19: Significant spatial autocorrelation of mean traits

Trait	I	Prob(I)
DM	0.375	0.000
DH	0.327	0.000
KSPK	0.271	0.000
TKW	0.250	0.000
GFD	0.197	0.003
VIT	0.153	0.018
GxE by LA	0.152	0.014
PH	0.140	0.029
PC	0.129	0.045
SAVI70	0.365	0.000
WI70	0.365	0.000
RVSI70	0.299	0.000
SR70	0.277	0.000
NDVI70	0.261	0.000
PRI70	0.251	0.000
NPCI70	0.179	0.006
RNVI70	0.169	0.010
CAROTENE70	0.148	0.021
QN70	0.146	0.024
RVSI45	0.145	0.021
WI/NDVI70	0.143	0.022
NPQ70	0.134	0.038
F270	0.126	0.043

The SA for grain yield changed significantly from year to year and ranged from 0.39 to 0.03 (where it was not significant) (Table 20). The same result was found for PH and PC. For traits showed reasonable constant SA across evaluation years: DH (3 years), DM (2 years) and KSPK (2 years) and TKW (5 years).

Table 20: Significant spatial autocorrelation of measured traits in different years

	I	Prob(I)
DH	0.298	0.000
DH04	0.325	0.000
DH05	0.402	0.000
DM	0.410	0.000
DM05	0.279	0.000
GY04	0.393	0.000
GY05	0.033	0.531
GY06	0.176	0.007
GY07	0.246	0.000
KSPK	0.178	0.007
KSPK07	0.271	0.000
PC	0.116	0.072
PC04	0.269	0.000
PC05	0.240	0.000
PC06	0.053	0.368
PC07	0.088	0.160
PH	0.144	0.025
PH05	0.077	0.203
PH06	0.247	0.000
PH07	0.446	0.000
TKW	0.185	0.005
TKW04	0.311	0.000
TKW05	0.196	0.003
TKW06	0.229	0.001
TKW07	0.207	0.002

Multispati-PCA applied on agronomic traits found 17 Eigen values from which 7 were positive. The multivariate spatial autocorrelation test using Monte Carlo and 10^6 iterations showed almost a null non-significant value. So no spatial pattern revealed by the agronomic traits. Also SAU of the kept axes by Multispati-PCA had no significant Moran'I values which match with the non-significance of the SAU of PCA-AGRO.

4. 1. 5. 3. Spatial clustering

When SA is not significant, we cannot reject absolutely the Null hypothesis. Also, when SA has a significant p-value (Table 20), we are sure that the spatial distribution of a trait is clustered and there is very little chance that this cluster can be a result of random process. General G analysis of Getis-Ord gives an idea about which part of our data is clustered; low values for low clustering or high value for high cluster. This analysis confirms some of the results of SA. Most of the traits with non-significant SA have a random clustering according the z-score of general G (GY, ASH, SDS, SDSN, SNSi and VIT). Other traits with non-significant SA showed high or low clustering but the pattern is not significant (PC, SPM2 and YP). On the other hands, some traits showed a very significant high (PH, DH, DM, TKW and SL) and low (KSPK) clustering (Table 21).

Table 21: High and low clusters using Getis-Ord General G

Trait	z-score	p-value	Cluster
GY	-1.196	0.232	Random
ASH	-0.786	0.432	Random
PC	1.664	0.096	High
PH	2.750	0.006	High
DH	4.241	0.000	High
DM	4.540	0.000	High
GFD	0.618	0.536	Random
SDS	-0.765	0.444	Random
SDSN	-0.583	0.560	Random
SDSI	-0.834	0.405	Random
YP	-1.672	0.095	Low
TKW	4.527	0.000	High
VIT	-0.010	0.992	Random
SL	2.796	0.005	High
SPM2	-1.873	0.061	Low
PL	-1.670	0.095	Low
KSPK	-2.932	0.003	Low

Now that we know which traits are clustered and how, we can push the study further and check where these trait 'values cluster across Morocco. For this, we computer the Local G* of of Getis-Ord to find hot (regions with high values) and cold spots (regions with high values). Only region with of z-score superior or inferior to 2.5 standard deviation are significant at 1%. This means that in those regions (locations of durum landraces collection), there is less that % likelihood that the observed pattern is by chance.

For PH (Figure 24), only one low value site was significant in the region of Errashidia. Other regions were detected to be regions of high PH especially in BeniMellal and Azilal.

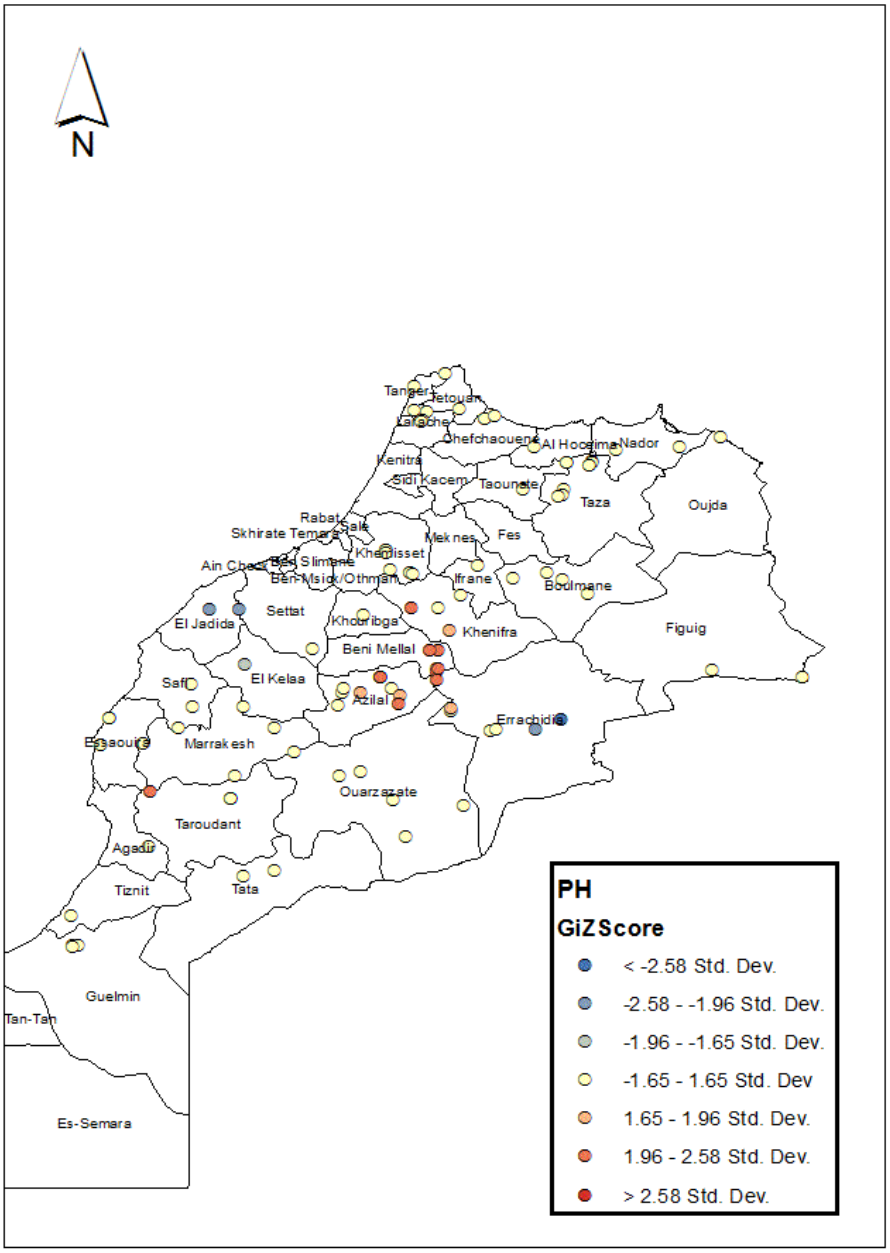


Figure 24: Spatial hotspots for Plant height

For DH (Figure 25), several regions were found to be statically significant hotspots. These regions were in the high altitude of Atlas chain in the regions of Khenifra, Azilal and BeniMellal. Landraces originated from Tiznit and Guelmin were significant regions of low DH.

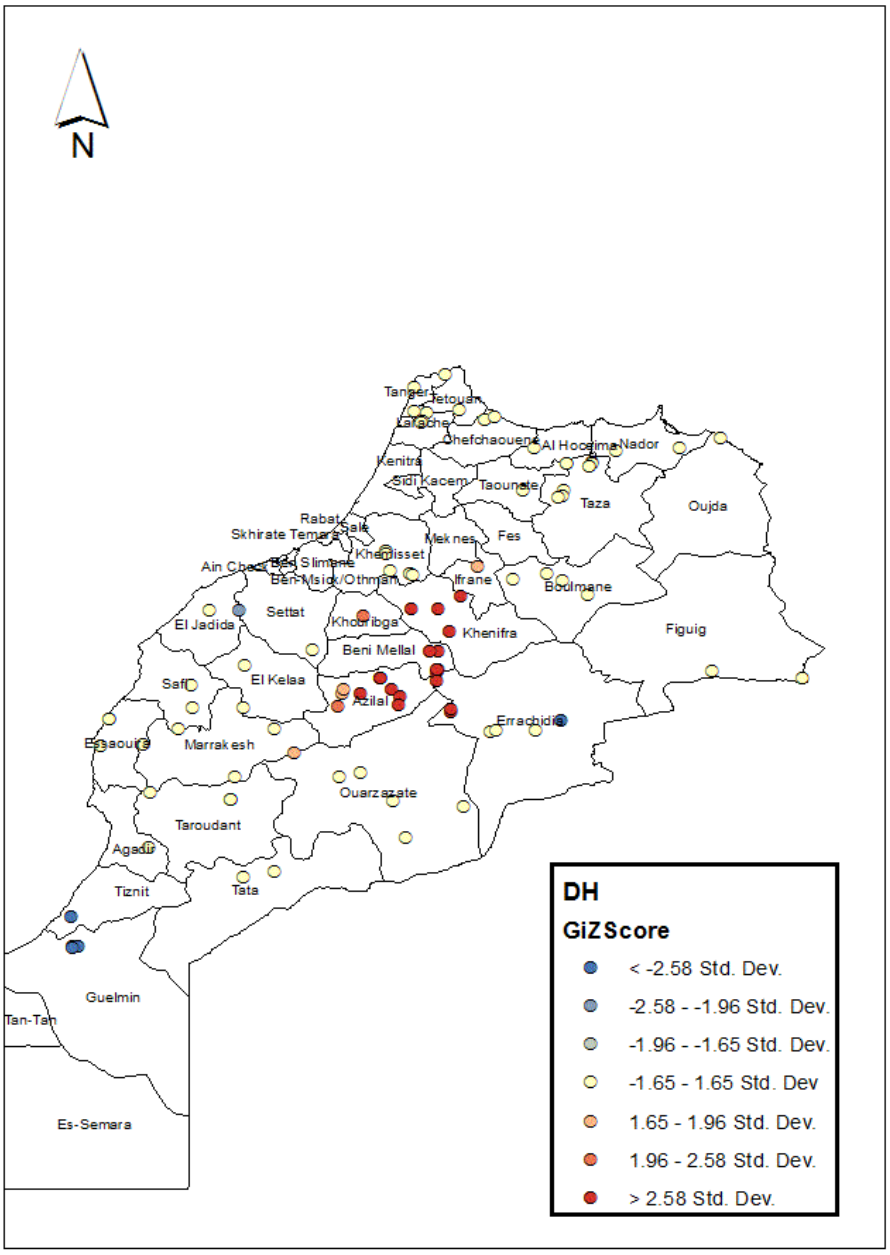


Figure 25: Spatial hotspots for days to heading

For DM (Figure 26), almost the same pattern of DH was found mainly for hotspots. Nevertheless, some non-significant hotspots for DH were very significant hotspots for DM. Those were declared in the regions of Ifrane and Khouribga. On the other hand, the region of low spot clusters was more extended and found the regions along the Atlantic Ocean: Essaouira, Safi, Tiznit and Guemin.

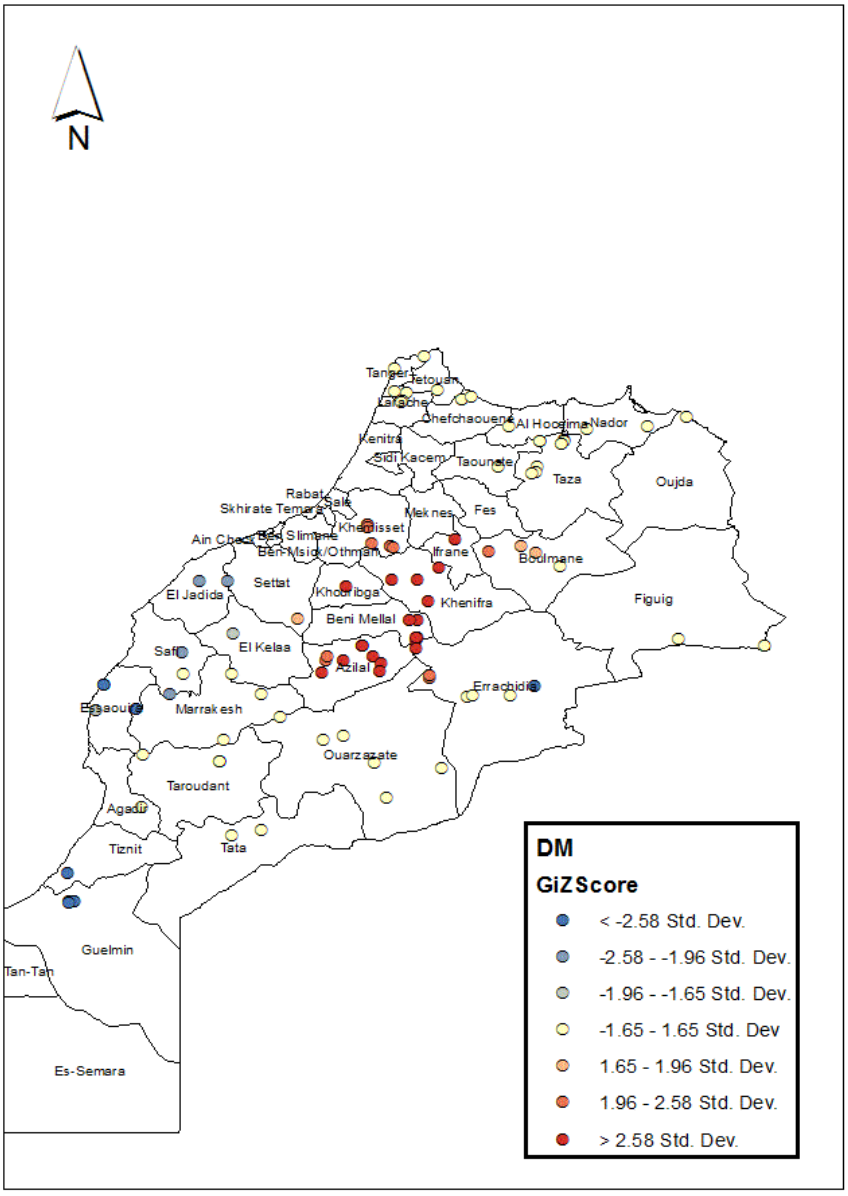


Figure 26: Spatial hotspots for days to maturity

TKW showed a large region expressing hotspots (Figure 27). This region is found in the high altitude of Rif and Atlas mountain chain. The low values areas were in the Atlantic regions of Aljadida, Safi and Essaouira.

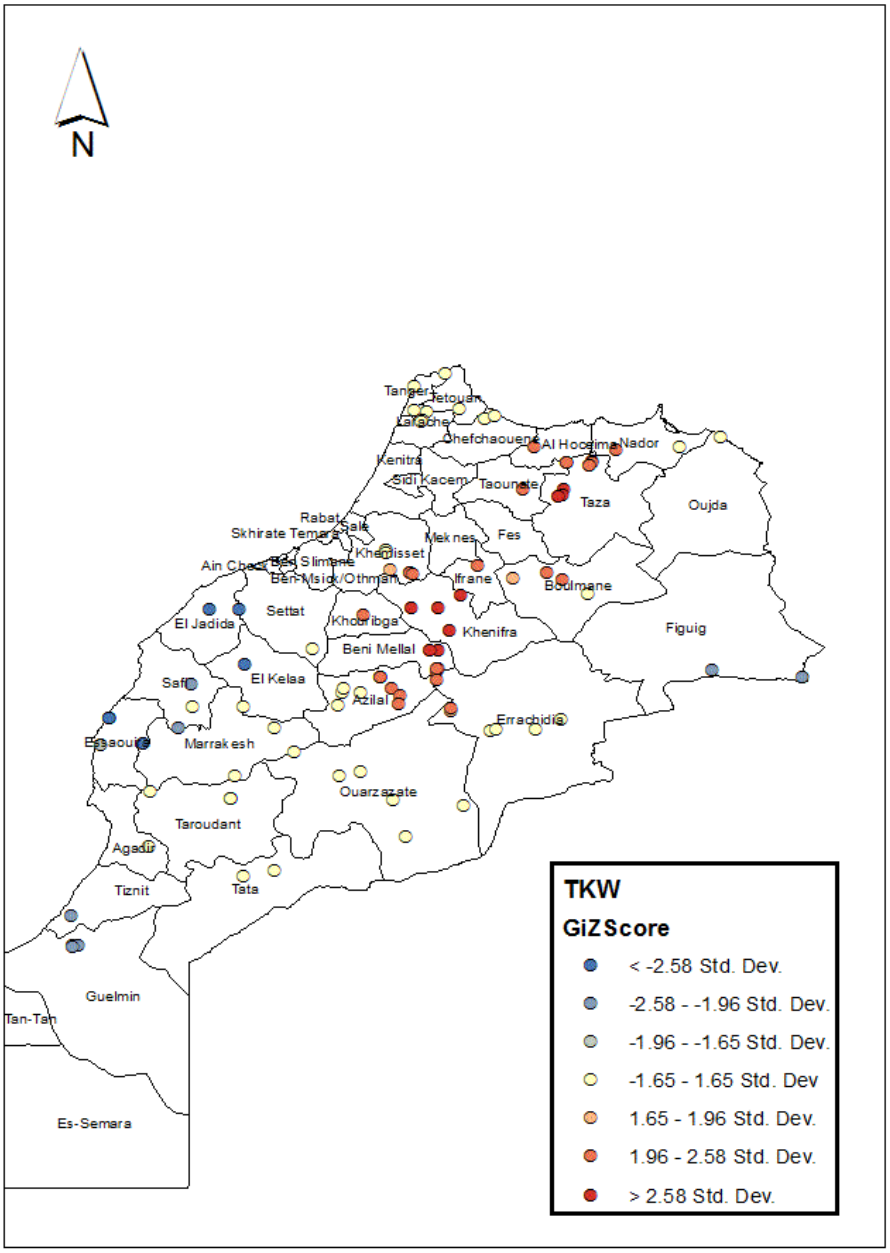


Figure 27: Spatial hotspots for thousand kernel weight

No significant hot (or low) spots were found for SL (Figure 28). While for KSPK (Figure 29), the spatial pattern of hot and low spots was almost the inverse of the one found for TKW. Landraces with high number of kernel per spike were concentrated in the Atlantic regions of Aljadida, Safi and Essaouira. Low values of KSPK were mainly focused in the high altitude of Rif and Atlas mountain chain. Low spots were not significant.

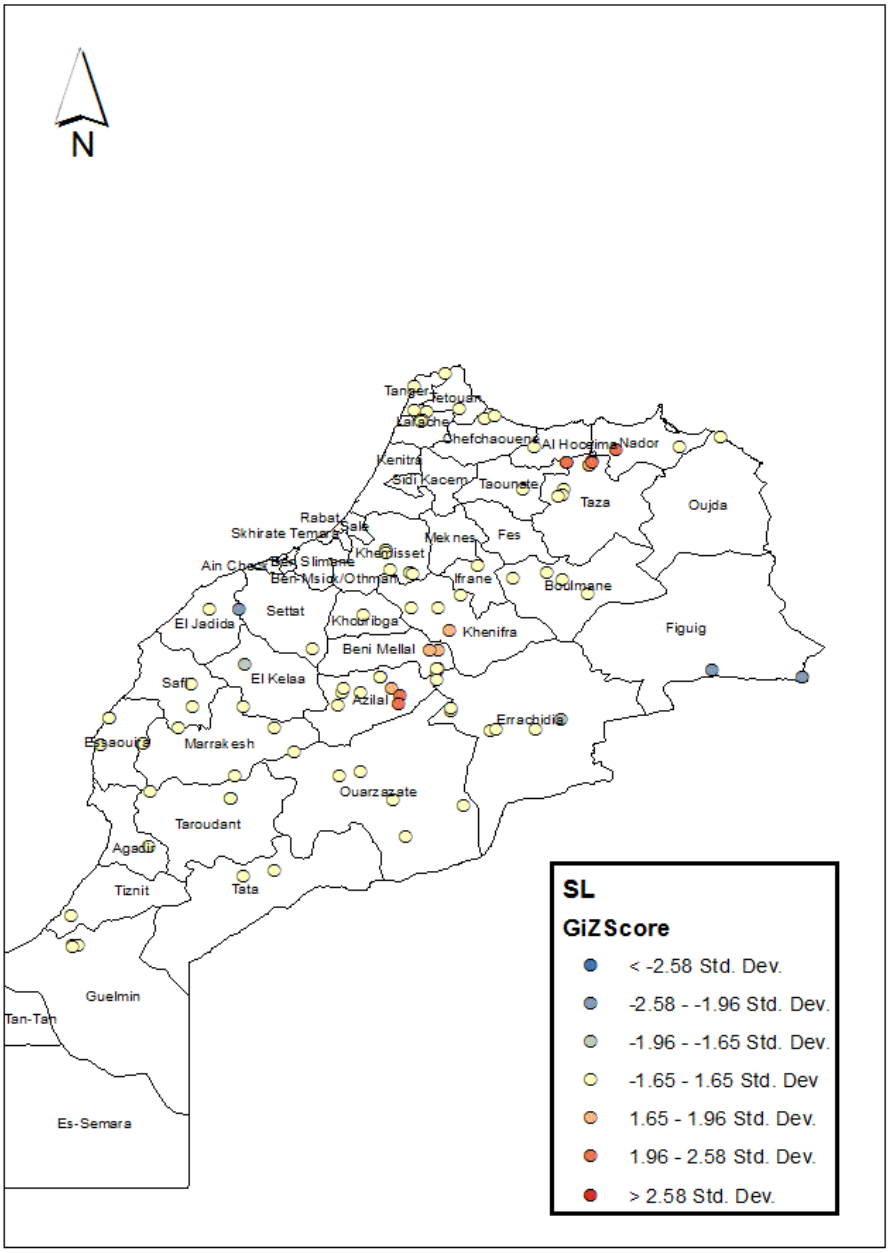


Figure 28: Spatial hotspots for spike length

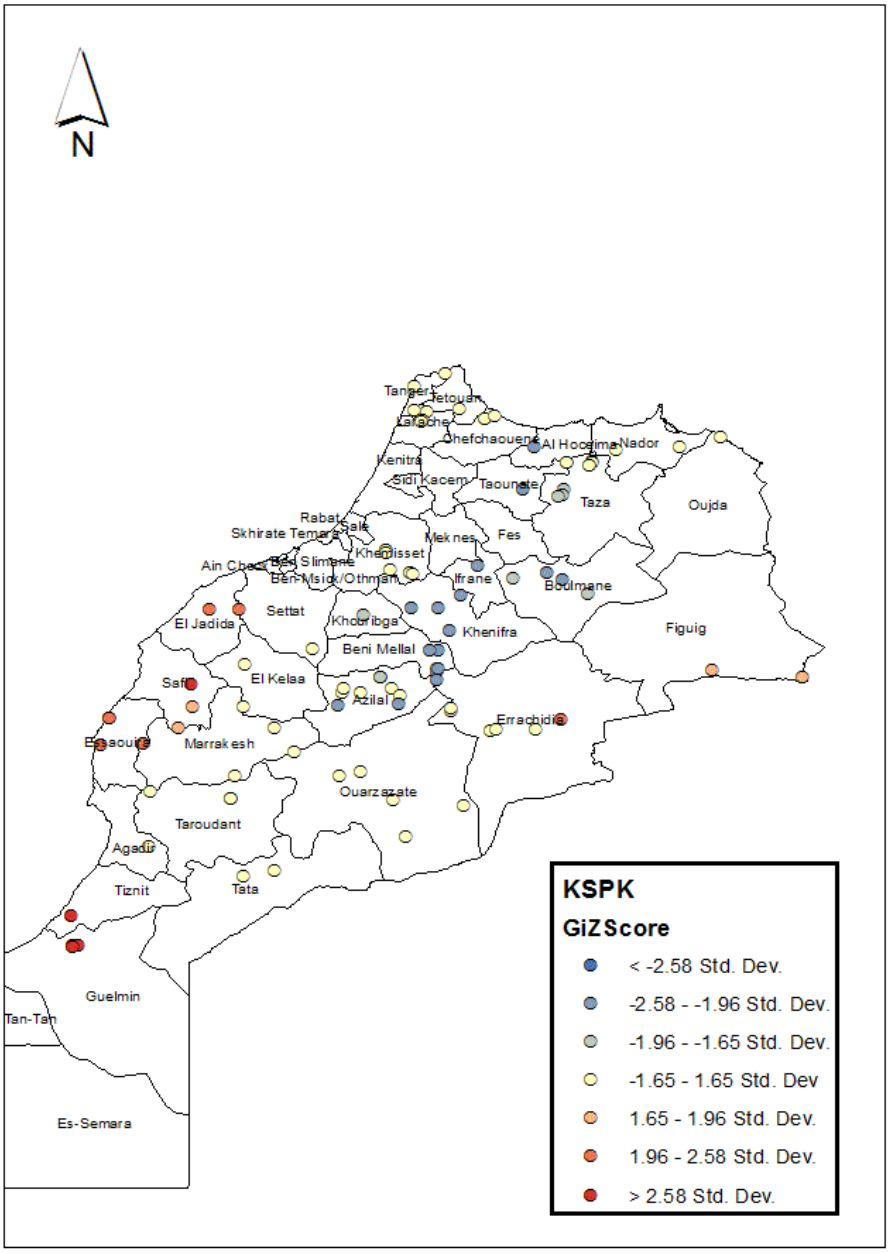


Figure 29: Spatial hotspots for number of kernel per spike

4. 1. 5. 4. Spatial modeling

When a trait is measured at different locations, one can use geo-statistical methods to map and predict this trait in areas where it was not measured. Kriging is the most appropriate for mapping. When the trait has no evidence of spatial pattern, the mapping (prediction) is not accurate. In our case, GY presents no significant spatial autocorrelation across Morocco. Using Variogram and Kriging techniques permitted to have a predicted map of GY across Morocco (Figure 30).

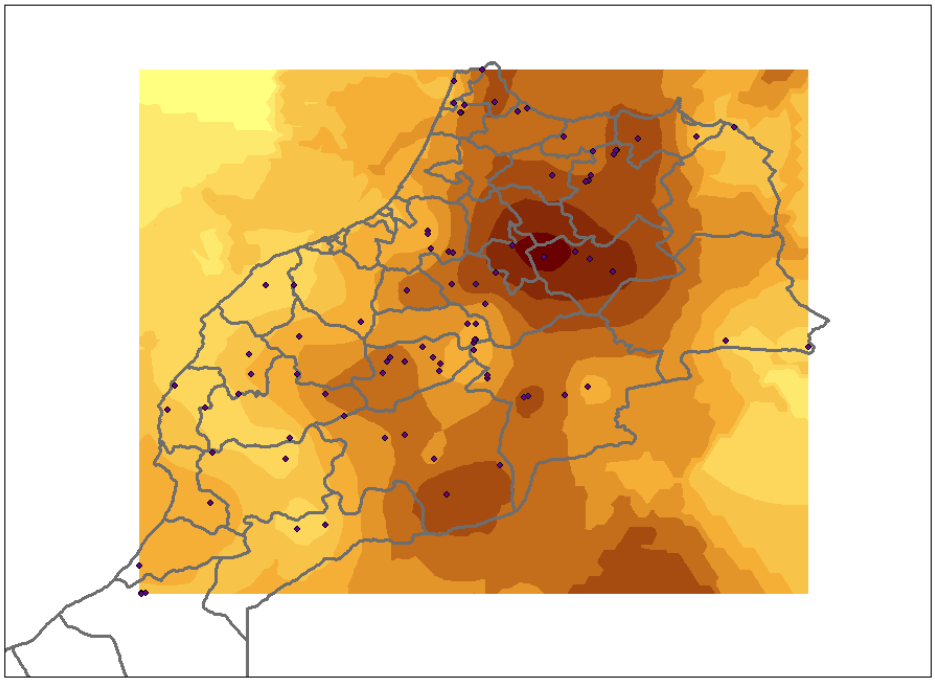


Figure 30: Map of interpolated GY (High "dark" to low yield "light")

Modeling Grain yield with traits having a strong spatial autocorrelation can help refining the spatial prediction. We first run a multiple regression (Table 22)on grain yield using four traits with positive and significant SA (TKW, GFD, PH and SL). Second, we computed the predicted map of the four traits used as independent variables in the multiple regression model.

Table 22: Coefficients of multiple regression of GY on TKW, GFD, PH and SL

Variable	Trait	Coefficients
Dependent	GY05RF	
Independent	TKW05RF	66.41
Independent	GFD05RF	3.67
Independent	PH05RF	-0.56
Independent	SL05RF	-87.74
Constant		329.66

The resulting maps are more accurate since the traits have positive SA and then a global pattern across Morocco. The last step is the use the raster calculator in ArcGIS 9.2 to compute a predicted map of GY using maps of TKW, GFD, PH and SL and coefficients from multiple regression (Figure 31).

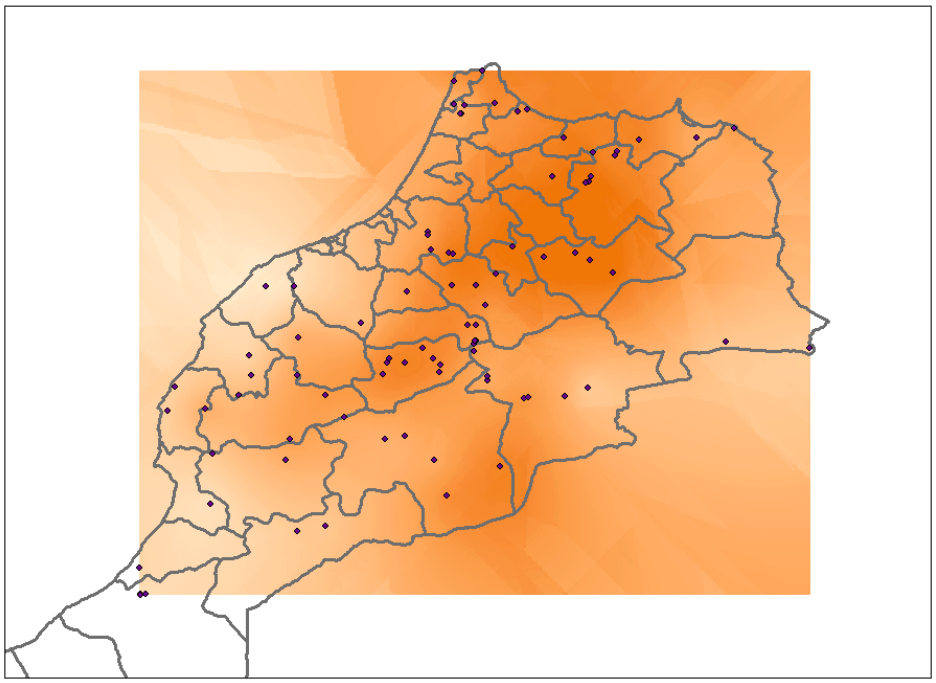


Figure 31: Predicted map of GY using multiple regression coefficients and raster calculation (High "dark" to low yield "light")

4. 1. 5. 5. Geographical barriers of traits

To study the geographic barriers of any spatial phenomena, one should construct Delaunay triangulation and Voronoï tessellation. The Voronoï tessellation represents a polygonal neighborhood for each population that is constituted of those points, on the plane, that are closer to such point than to any other one. This tessellation determines which populations are neighbors, adjacent. Two points are adjacent if the corresponding Voronoï polygons have a common edge (Green in Figure 32). The corresponding Delaunay triangulation is in Black in the same figure. The sample of our points corresponding to durum landraces are labeled with a number. The blue points are the virtual points used to obtain a closed tessellation enclosing all the points.

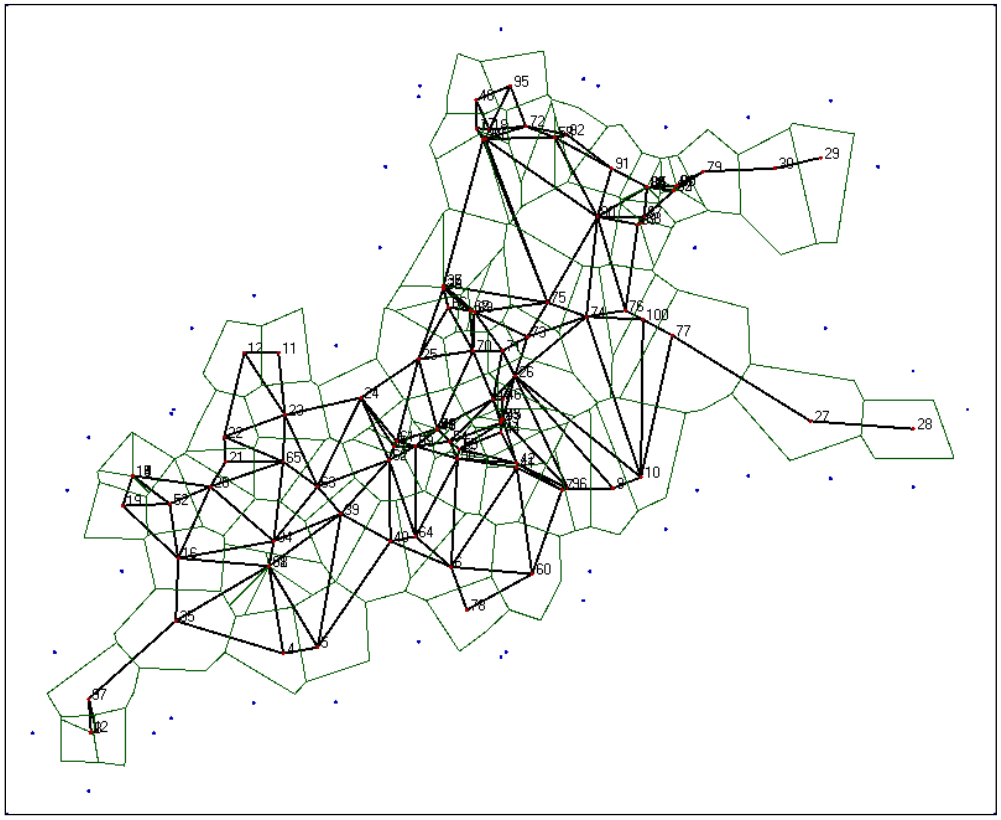


Figure 32: Voronoï tessellation and Delaunay triangulation for Moroccan durum landraces

Several barriers were found for DH (Figure 33). The first one is between the Atlas Mountains and the Atlantic Ocean, the second in the Southern part of the country starting from the Ocean side at Safi until the Desert in Tata. The barrier present around the Atlas chain is surrounding the hotspots found for DH earlier in this paragraph.

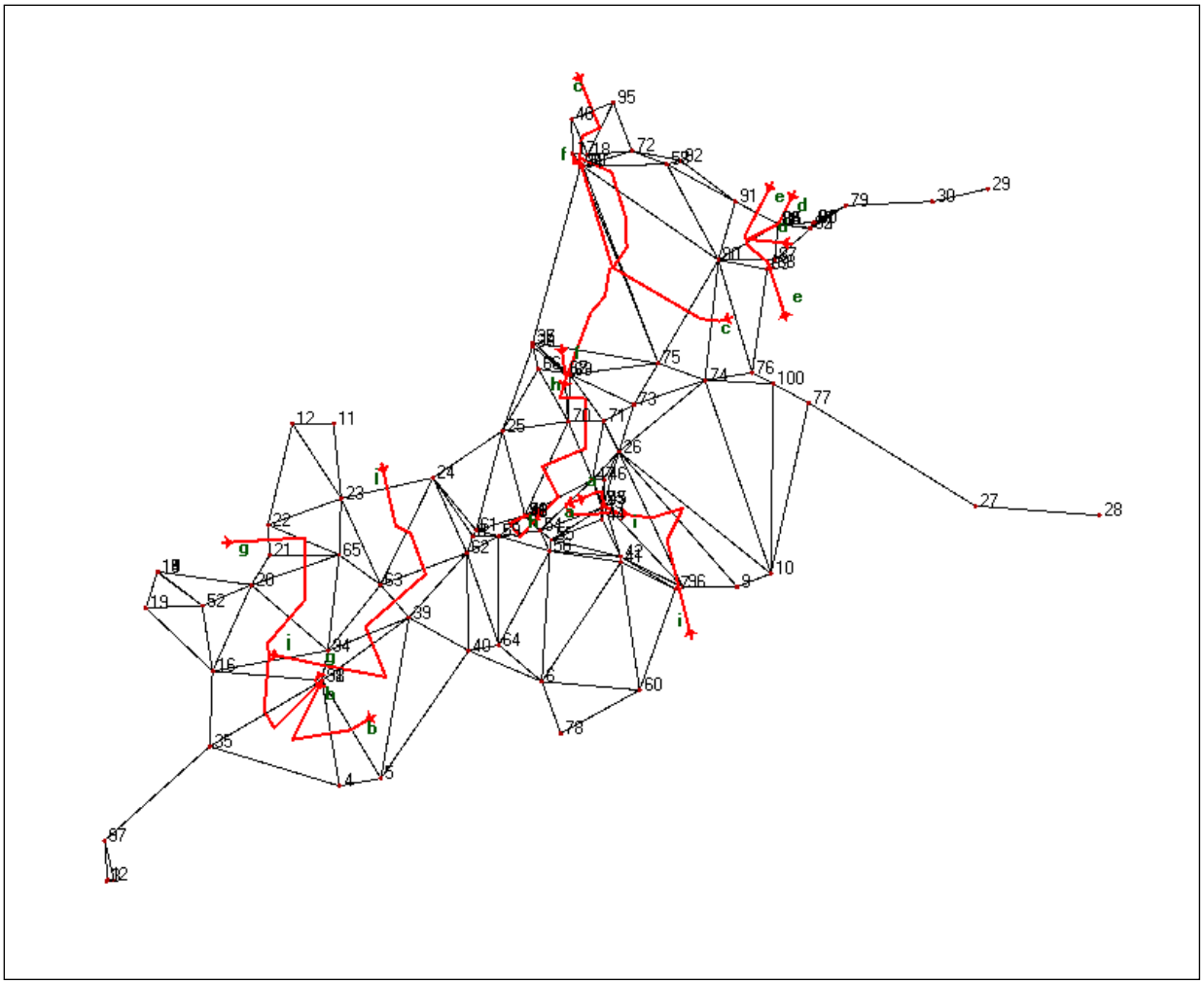


Figure 33: Days to heading barriers using Monmonier's Algorithm

For TKW (Figure 33), on large barrier was found in the South-East of the country. This barrier separates between hotspots in the North, Low spots in the South and the West. Another discontinuity found in the north (the Rif Mountains). The last one Overlaps as well with the regions of high values of TKW (Figure 34).

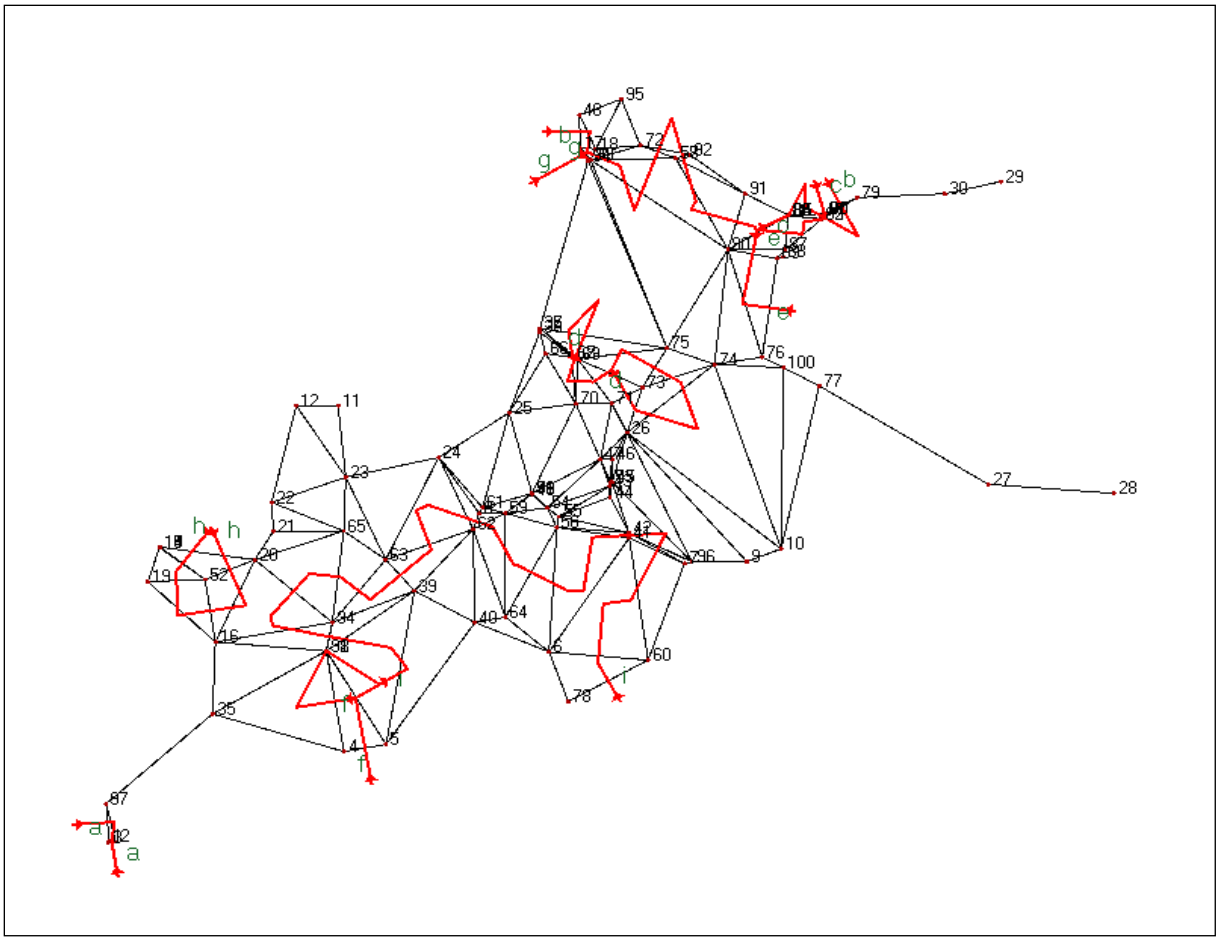


Figure 34: Thousand kernel weight barriers using Monmonier's Algorithm

4. 2. Phenotypic / Climate relationships

4. 2. 1. Correlation

All extracted long-term climatic variables showed a large diversity in locations where Moroccan durum landraces were collected specially during the wheat cycle in Morocco (November to May/June). All monthly minimal temperatures (Tmin) had a significant negative correlation with altitude and most of maximal temperatures (Tmax) and precipitations (prec) were correlated negatively and positively with the altitude respectively. Tmin were highly positively correlated from month to month. The same pattern, but less high, of correlations was found for Tmax from month to month (Table 23).

Table 23: Descriptive statistics of the extracted long-term climatic variable for Moroccan durum wheat landraces

Code	Variable	Mean	Median	Max	Min	Var
tmin1	Minimal temperature-January	1.91	1.65	9.60	-7.50	15.48
tmax1	Maximal temperature-January	14.29	15.05	19.40	6.90	8.43
prec1	Precipitation-January	56.52	52.00	153.00	3.00	1166.39
tmin2	Minimal temperature-February	3.03	3.00	10.10	-5.90	13.73
tmax2	Maximal temperature-February	15.73	16.45	21.60	8.60	8.84

prec2	Precipitation-February	55.93	51.00	144.00	3.00	1139.40
tmin3	Minimal temperature-March	5.41	5.60	12.00	-3.80	13.81
tmax3	Maximal temperature-March	18.21	18.50	25.00	11.20	8.93
prec3	Precipitation-March	59.74	58.50	153.00	5.00	1151.31
tmin4	Minimal temperature-April	7.70	7.70	13.30	-0.80	12.02
tmax4	Maximal temperature-April	20.98	20.75	29.60	14.80	9.76
prec4	Precipitation-April	48.03	54.00	89.00	2.00	498.23
tmin5	Minimal temperature-May	10.25	10.15	15.70	2.00	12.49
tmax5	Maximal temperature-May	24.13	23.60	34.20	18.20	12.13
prec5	Precipitation-May	27.57	30.50	59.00	2.00	181.50
tmin6	Minimal temperature-June	13.76	14.40	20.10	6.30	10.58
tmax6	Maximal temperature-June	28.53	28.30	39.60	21.30	13.62
prec6	Precipitation-June	10.03	10.00	29.00	1.00	34.94
tmin7	Minimal temperature-July	16.58	16.80	23.80	10.40	9.39
tmax7	Maximal temperature-July	32.97	32.80	43.80	22.50	17.02
prec7	Precipitation-July	3.00	2.00	13.00	0.00	7.31
tmin8	Minimal temperature-August	17.07	17.50	23.80	10.60	8.83
tmax8	Maximal temperature-August	32.75	32.40	42.40	22.60	14.38
prec8	Precipitation-August	4.66	4.00	14.00	0.00	10.23
tmin9	Minimal temperature-September	14.15	14.65	19.40	6.40	9.74
tmax9	Maximal temperature-September	28.64	28.50	35.90	22.90	7.80
prec9	Precipitation-September	16.24	15.00	34.00	5.00	52.53
tmin10	Minimal temperature-October	10.64	10.90	15.90	2.20	11.74
tmax10	Maximal temperature-October	23.42	23.25	29.70	16.40	8.74
prec10	Precipitation-October	39.43	40.00	79.00	10.00	239.58
tmin11	Minimal temperature-November	6.49	6.60	13.10	-2.80	14.32
tmax11	Maximal temperature-November	18.43	18.75	24.10	10.70	8.59
prec11	Precipitation-November	57.95	57.00	117.00	13.00	707.16
tmin12	Minimal temperature-December	3.40	3.10	10.70	-5.60	14.46
tmax12	Maximal temperature-December	14.89	15.80	19.60	7.00	9.57
prec12	Precipitation-December	67.52	66.00	155.00	7.00	1269.73
bio1	Annual Mean Temperature	15.97	16.40	21.74	8.97	7.67
bio2	Mean Monthly Temperature Range	13.55	14.06	17.32	6.37	7.42
bio3	Isothermality (V2/V7) (* 100)	43.55	43.99	49.68	38.05	7.00
bio4	Temperature Seasonality (STD * 100)	606.05	622.63	818.34	245.60	15178.32
bio5	Max Temperature of Warmest Month	33.10	32.80	43.80	22.90	16.31
bio6	Min Temperature of Coldest Month	1.90	1.65	9.60	-7.50	15.48
bio7	Temperature Annual Range (V5-V6)	31.20	33.40	41.10	13.30	39.38
bio8	Mean Temperature of Wettest Quarter	10.85	11.41	21.57	3.67	12.02
bio9	Mean Temperature of Driest Quarter	23.34	23.13	29.48	11.80	8.23
bio10	Mean Temperature of Warmest Quarter	23.80	23.45	32.20	18.75	8.45
bio11	Mean Temperature of Coldest Quarter	8.87	8.99	14.35	0.58	10.43
bio12	Annual Precipitation	446.62	476.00	874.00	63.00	44356.08
bio13	Precipitation of Wettest Month	70.56	66.50	155.00	15.00	1262.11
bio14	Precipitation of Driest Month	2.92	2.00	13.00	0.00	7.39
bio15	Precipitation Seasonality (CV)	66.15	66.46	88.35	40.21	131.78
bio16	Precipitation of Wettest Quarter	190.58	184.50	450.00	35.00	10004.87
bio17	Precipitation of Driest Quarter	17.62	15.00	49.00	3.00	120.02

bio18	Precipitation of Warmest Quarter	21.41	20.50	49.00	3.00	130.93
bio19	Precipitation of Coldest Quarter	179.97	167.00	443.00	13.00	10523.83

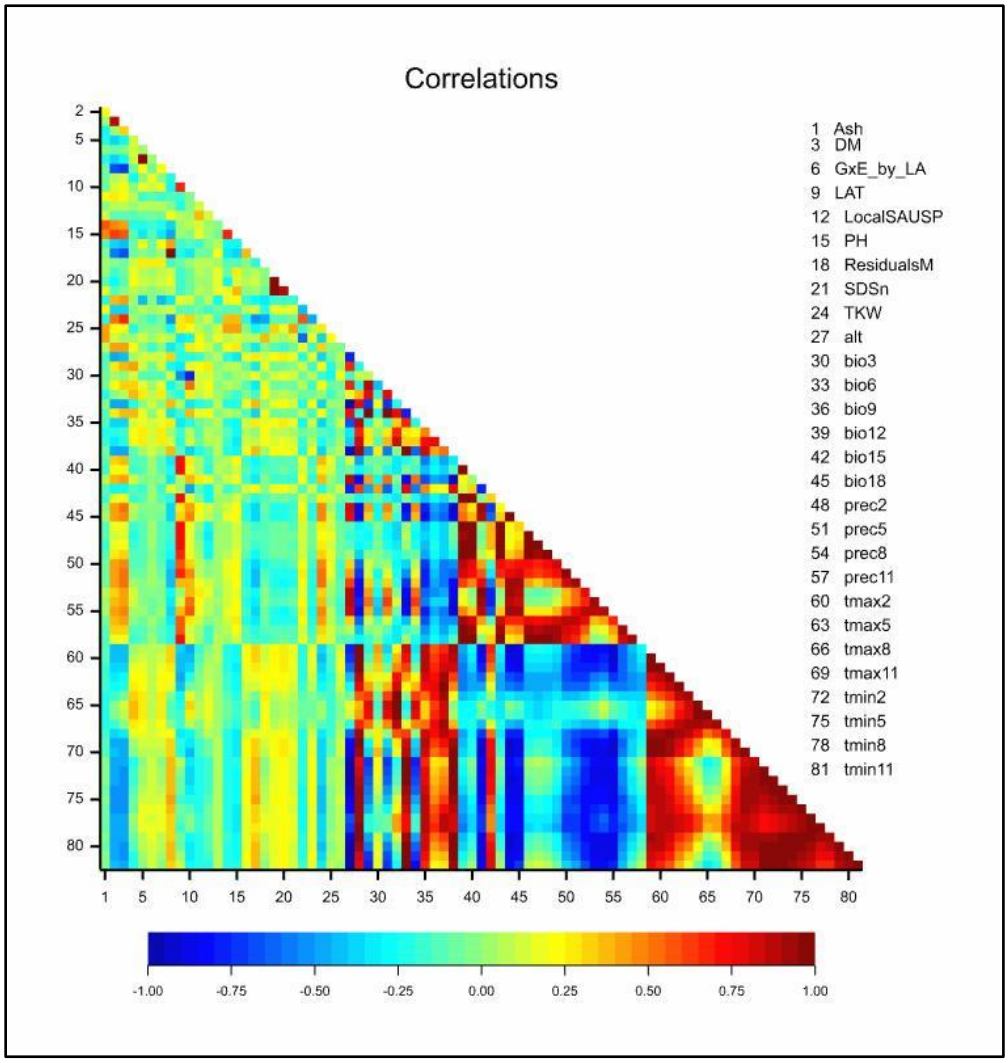


Figure 35: Correlations between phenotypic traits and long-term climate data of landrace's origin

As for the traits, TKW was positively correlated to precipitations in all months but negatively associated to Tmax and Tmin in most of the months during durum development cycle. The same associations were found for PH, PC, DH and DM but the correlations were of less magnitude. Opposite correlation patterns were dissected for KSPK and SPM2 (Figure 35).

4 .2. 2. GxE with genotypic covariates

The long-term climatic variables (during and around the durum wheat cycle in Morocco) and the three spatial coordinates (Lat, Long and Atl) were used as genotypic co-variables in a linear mixed model to compute how much a co-variable is reducing the genotypic variance component in a GxE model.

Table 24: Reduction of the genotypic effect using a long term environmental characteristic as a genotypic co-variable

Environmental variable	Variance components	DH	DM	GY	PC	PH	SDS	TKW	Y
LAT	**G**				98.5		84.7	95.8	82
	Cov				1.5		**15.3**	**4.2**	**17**
	TOTAL				100.0		100.0	100.0	100
LONG	**G**		99.4				60.0	88.9	84
	Cov		0.6				**40.0**	**11.1**	**16**
	TOTAL		100.0				100.0	100.0	100
Tmin1	**G**	98.9	99.4				89.4		
	Cov	1.1	0.6				**10.6**		
	TOTAL	100.0	100.0				100.0		
Tmin2	**G**	98.7	99.3				87.7		
	Cov	1.3	0.7				**12.3**		
	TOTAL	100.0	100.0				100.0		
Tmin3	**G**	98.5	99.2				88.8		
	Cov	1.5	0.8				**11.2**		
	TOTAL	100.0	100.0				100.0		
Tmin4	**G**	97.7	99.0				85.9		
	Cov	2.3	1.0				**14.1**		
	TOTAL	100.0	100.0				100.0		
Tmin5	**G**	97.7	99.0			99.4	81.2		
	Cov	2.3	1.0			0.6	**18.8**		
	TOTAL	100.0	100.0			100.0	100.0		
Tmin6	**G**	97.1	99.2			97.7	93.3		
	Cov	2.9	0.8			2.3	**6.7**		
	TOTAL	100.0	100.0			100.0	100.0		
Tmin10	**G**	98.3	99.2				79.5		
	Cov	1.7	0.8				**20.5**		
	TOTAL	100.0	100.0				100.0		
Tmin11	**G**	98.7	99.2				91.1		
	Cov	1.3	0.8				**8.9**		
	TOTAL	100.0	100.0				100.0		
Tmin12	**G**	98.8	99.3				91.2		

	Cov	1.2	0.7				**8.8**		
	TOTAL	100.0	100.0				100.0		
Tmax1	**G**	95.5	98.4				81.5		
	Cov	**4.5**	1.6				**18.5**		
	TOTAL	100.0	100.0				100.0		
Tmax2	**G**	94.3	98.7				78.4		
	Cov	**5.7**	1.3				**21.6**		
	TOTAL	100.0	100.0				100.0		
Tmax3	**G**	93.4	98.8			98.0	84.7		
	Cov	**6.6**	1.2			2.0	**15.3**		
	TOTAL	100.0	100.0			100.0	100.0		
Tmax4	**G**	96.0	98.9	96.7	97.8	97.7			
	Cov	4.0	1.1	3.3	2.2	2.3			
	TOTAL	100.0	100.0	100.0	100.0	100.0			
Tmax5	**G**	98.1	99.2	77.2	98.7	97.0			
	Cov	1.9	0.8	**22.8**	0.4	3.0			
	TOTAL	100.0	100.0	100.0	100.0	100.0			
Tmax6	**G**	98.3	98.7			98.1			
	Cov	1.7	1.3			1.9			
	TOTAL	100.0	100.0			100.0			
Tmax10	**G**	92.4	99.4			89.6			
	Cov	**7.6**	0.6			**10.4**			
	TOTAL	100.0	100.0			100.0			
Tmax11	**G**	95.6	98.7			99.6	81.1		
	Cov	4.4	1.3			0.4	**18.9**		
	TOTAL	100.0	100.0			100.0	100.0		
Tmax12	**G**	97.3	99.0			98.1	81.6		
	Cov	2.7	1.0			1.9	**18.4**		
	TOTAL	100.0	100.0			100.0	100.0		
Prec6	**G**		99.1	97.9		99.4	92.2		
	Cov		0.9	2.1		0.6	**7.8**		
	TOTAL		100.0	100.0		100.0	100.0		
Bio1	**G**	94.7	99.0			95.5	71.9		
	Cov	**5.3**	1.0			**4.5**	**28.1**		
	TOTAL	100.0	100.0			100.0	100.0		
Bio2	**G**	96.7	94.9	88.1	98.0			99.2	98
	Cov	3.3	**5.1**	**11.9**	2.0			0.8	1.
	TOTAL	100.0	100.0	100.0	100.0			100.0	100
Bio3	**G**	98.7	98.3				91.4	98.1	90
	Cov	1.3	1.7				**8.6**	1.9	**9.**
	TOTAL	100.0	100.0				100.0	100.0	100

Bio5	**G**	98.0	98.4		98.5	98.6			
	Cov	2.0	1.6		1.5	1.4			
	TOTAL	100.0	100.0		100.0	100.0			
Bio6	**G**	98.9	99.4				89.4		
	Cov	1.1	0.6				**10.6**		
	TOTAL	100.0	100.0				100.0		
Bio7	**G**	98.2	98.6						
	Cov	1.8	1.4						
	TOTAL	100.0	100.0						
Bio8	**G**	98.3	99.7	99.8		99.4	93.8		
	Cov	1.7	0.3	0.2		0.6	**6.2**		
	TOTAL	100.0	100.0	100.0		100.0	100.0		
Bio9	**G**	94.9	98.3	89.1	98.6	97.6		99.5	
	Cov	**5.1**	1.7	**10.9**	1.4	3.5		0.5	
	TOTAL	100.0	100.0	100.0	100.0	100.0		100.0	
Bio10	**G**	95.6	98.9	92.5	98.8	96.4			
	Cov	**4.4**	1.1	**7.5**	1.2	3.6			
	TOTAL	100.0	100.0	100.0	100.0	100.0			
Bio11	**G**	98.1	99.1				82.9		
	Cov	1.9	0.9				**17.1**		
	TOTAL	100.0	100.0				100.0		
Bio14	**G**		99.4			99.4	97.1		
	Cov		0.6			0.6	2.9		
	TOTAL		100.0			100.0	100.0		

For the GY, the climatic variables reducing the genotypic variance components were Tmax5 (23%), the mean monthly temperature range BIO2 (12%), mean temperature of the driest quarter BIO9 (11%) and mean temperature of the driest quarter BIO10 (7.5%). Tmax1, Tmax2 and Tmax3 were contributing to genotypic variance of DH at around 5%. Also, BIO1 and BIO10 were reducing the DH genotypic variance. For DM, only long-term mean monthly temperature range was able to explain of the GV. Only Tmax10 and annual mean temperature BIO1 were able to explain some of the GV of PH (10% and 4.5% respectively). Most of the long-term climatic variables explained a large component of the SDS genotypic variance. The proportion of the explained variance ranged from 28% for the annual mean temperature BIO1 to 6% for the mean temperature of the wettest quarter BIO8. Latitude and longitude explained 15% and 40% of the SDS genotypic variance respectively. For TKW, only the geographic coordinates were able to explain some of the genotypic variance (4% for latitude and 11% for longitude). The geographic coordinates were also found to explain some of the genotypic variability of YP. The isothermality BIO3 was also reducing the GV of YP by 9% (Table 24). This analysis showed that long-term climate profiles of the site of landrace's collection and the geographic coordinates were able to explain a part of the genotypic variability of several traits including yield. This also can help understanding the adaptation and phenotypic variability of Moroccan landraces, and the area of traits variability using climatic variables.

4. 2. 3. Spatial pattern of climate variables

Four spatial discontinuities were computed using the long-term climatic variables of Moroccan durum landraces. Three of them were very clear and divide the Morocco in three parts: the atlas chain is explaining these barriers and concluding one part east and the other West of the chain. The third barrier is between the Haouz and Souss regions in Southern Morocco (Figure 36).

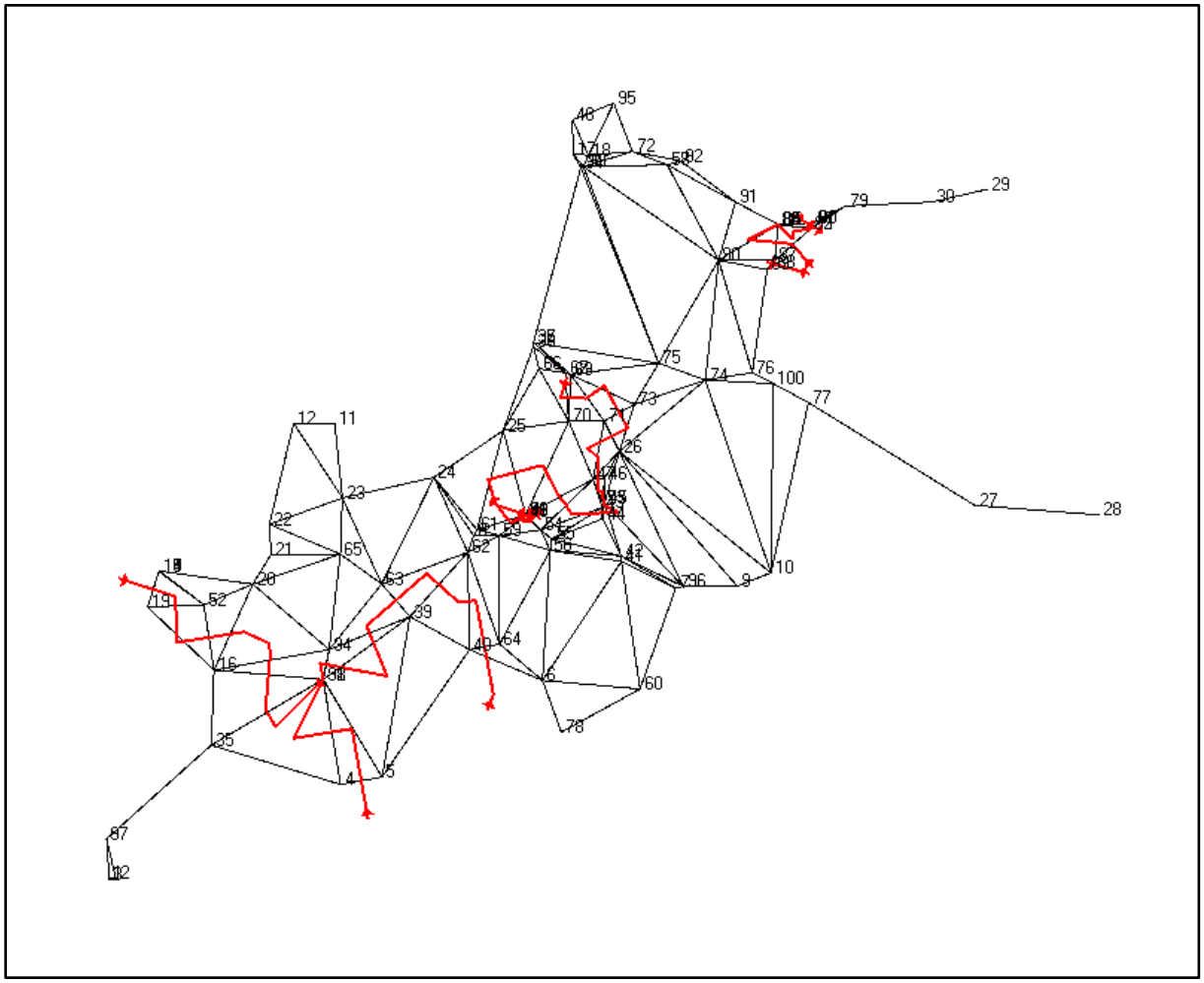


Figure 36: Climatic barriers for durum landrace's long term climate (using all climatic variables)

Using only minimal temperature (Figure 37), the studied space in Morocco was divided into four parts. Two of the barriers found using Tmin were identical to the one found for maximal temperature (Figure 38).

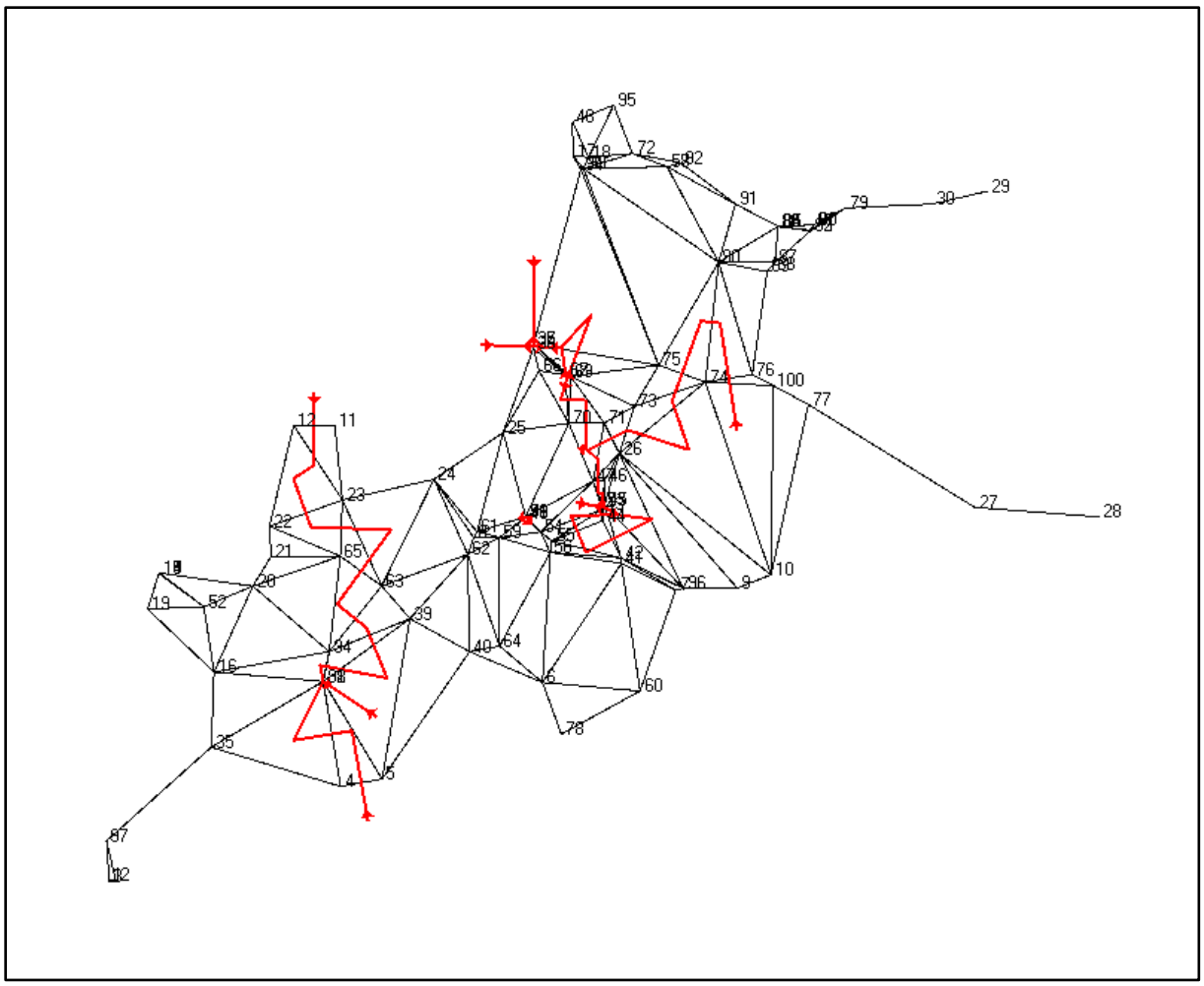


Figure 37: Climatic barriers for durum landrace's long term climate (using Tmin)

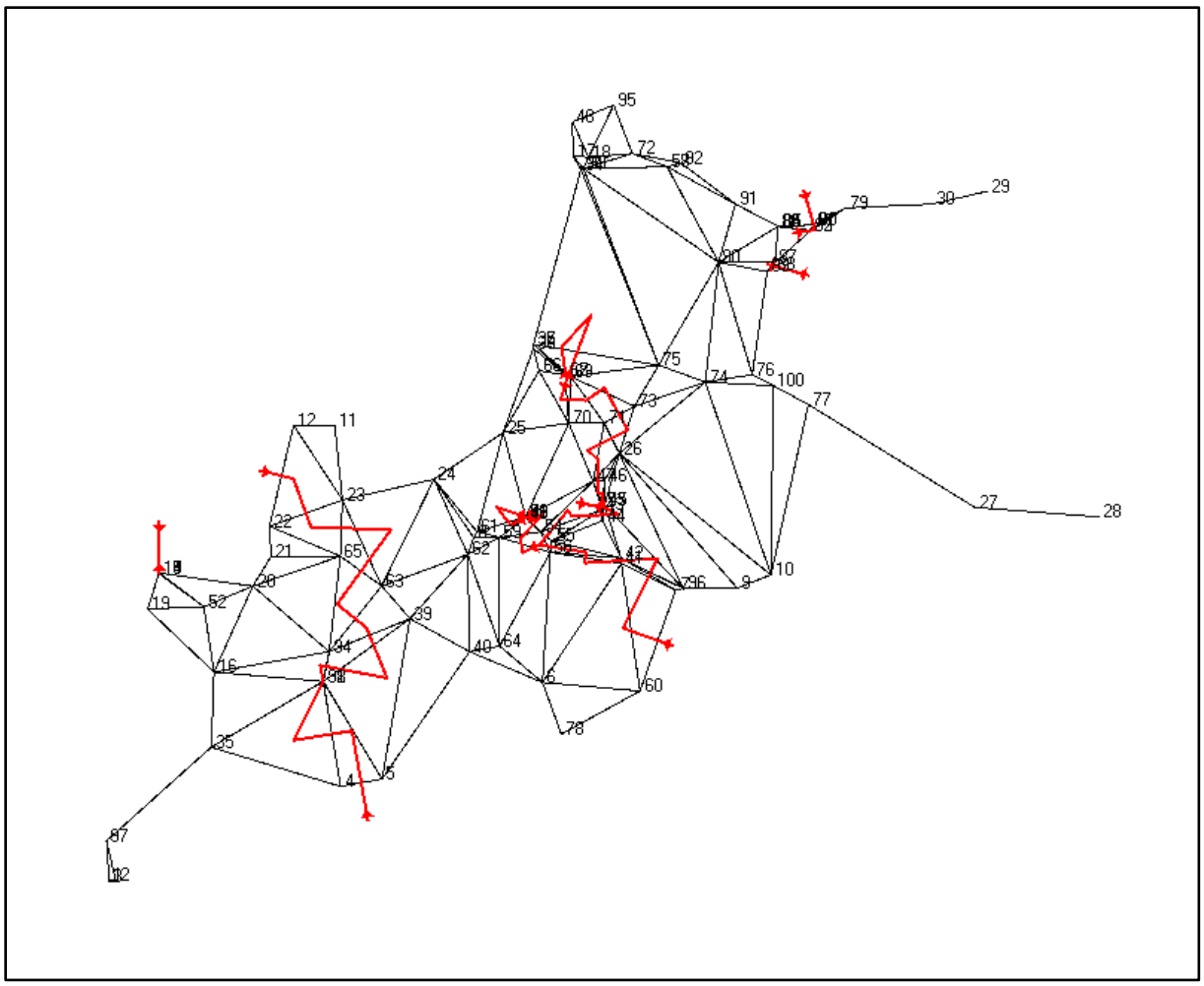


Figure 38: Climatic barriers for durum landrace's long term climate (using Tmax)

4. 3. Genotyping results

4. 3. 1. Locus description

The total number of amplified alleles at the 51 microsatellites loci found for the 188 Moroccan and Syrian landraces was 1208 alleles. The number of alleles per locus ranged from 5 for *gwm471* to 60 for *gwm368*. For the specific allele number of the Syrian and the Moroccan populations, the number of alleles per locus was higher in the Syrian landraces than in the Moroccan ones. The marker with the highest heterogeneous proportion was *gwm494*; and the lowest was *gwm257*. For the genetic diversity, the expected heterogeneity was significantly ($p < 0.001$) higher than the observed one, with a mean of difference of 0.56. We observed 741 alleles in the Moroccan population. For the range, the highest value was scored with 34 alleles at the locus *gwm610;* and the lowest with 2 alleles at *gwm165*. The average per locus was 14 alleles. Six of the studied markers were 100% homogenous (*gwm357*, *gwm169*, *gwm369*, *gwm471*, *gwm165*, *gwm257*) and 3 were higher than 80% heterogeneous (*gwm493*, *gwm494*, *gwm264*). Expected heterogeneity ranged from 0.99 for *gwm335* and *gwm644* to 0.11 for *gwm257;* and H_E was 0.763 (Table 25). H_e was significantly ($p < 0.001$) higher that H_o and with a mean difference of 0.5.

Table 25: Locus descriptive parameters for the dataset and for the Moroccan and the Syrian durum wheat landraces populations

Locus name	Chromosome	Dataset		Morocco					Syria				
		Na	ASR (bp)	Na	ASR (bp)	H0	He	F	Na	ASR (bp)	H0	He	F
gwm2	2AS,3AS	11	110-124	5	110-124	0.01	0.16	0.94	9	112-121	0	0.77	1
gwm6	4BL,5A	21	185-214	14	185-209	0.03	0.85	0.96	17	187-214	0.07	0.94	0.93
gwm33	1AS,1BL	30	105-190	20	115-190	0.11	0.87	0.87	20	105-177	0.09	0.86	0.9
gwm60	7AS	19	189-232	17	189-232	0.06	0.85	0.93	12	205-222	0.11	0.89	0.87
gwm63	7A	11	246-277	10	246-277	0.13	0.84	0.84	7	255-271	0.08	0.82	0.9
gwm99	1A	28	104-138	23	104-138	0.19	0.88	0.78	21	104-133	0.08	0.95	0.92
gwm107	3B,4B,6B	6	185-205	5	185-205	0.01	0.32	0.97	4	186-191	0	0.33	1
gwm114	3B	16	114-132	13	114-129	0.06	0.84	0.93	9	115-132	0.02	0.54	0.96
gwm129	2B,5AS	17	200-237	12	200-234	0.02	0.81	0.97	9	221-237	0	0.82	1
gwm160	4AL	23	169-209	16	169-207	0.29	0.82	0.65	22	172-209	0.2	0.87	0.77
gwm165	4A,4BS	10	182-193	2	190-192	0	0.25	1	9	182-193	0.02	0.84	0.97
gwm169	6AL	22	178-228	11	178-197	0	0.79	1	19	180-228	0.02	0.9	0.98
gwm210	2A,2B	10	164-191	7	165-191	0.27	0.37	0.29	7	164-187	0.06	0.36	0.85
gwm219	6B	28	153-190	19	153-190	0.06	0.9	0.93	19	154-182	0.11	0.74	0.85
gwm234	5A,5BS	14	99-211	12	99-202	0.03	0.79	0.96	6	100-211	0.24	0.29	0.15
gwm257	2B	9	191-200	3	193-196	0	0.12	1	9	191-200	0	0.78	1
gwm260	7AS	19	134-165	15	137-164	0.03	0.82	0.96	13	134-165	0.01	0.91	0.99
gwm264	1A,1B,3B,7B	32	102-231	27	102-212	0.86	0.89	0.04	11	161-231	0.09	0.75	0.88
gwm268	1B	18	169-248	5	176-246	0.01	0.16	0.94	15	169-248	0.07	0.88	0.92
gwm282	7A	19	110-194	12	110-124	0.04	0.82	0.95	16	112-194	0.02	0.9	0.98
gwm285	3B	18	212-237	17	216-228	0.03	0.83	0.96	3	212-237	0	0.42	1
gwm293	5A,7B	17	135-205	10	136-205	0.03	0.69	0.96	12	135-137	0.06	0.76	0.93
gwm297	7BS	39	149-178	18	149-178	0.13	0.88	0.85	34	150-175	0.2	0.91	0.78

gwm311	2A,6B	14	110-168	11	117-166	0.09	0.62	0.85	9	110-168	0.06	0.47	0.88
gwm319	2B	13	169-198	11	172-198	0.1	0.82	0.88	11	169-198	0.04	0.84	0.95
gwm344	7A,7B	45	99-124	31	99-122	0.52	0.9	0.42	26	111-124	0.9	0.8	0
gwm335	5B	38	136-256	12	151-256	0.04	1	0.96	29	136-244	0.13	0.98	0.86
gwm44	4A	15	173-274	9	204-274	0.01	0.67	0.98	10	173-271	0.02	0.46	0.95
gwm357	1A	33	101-146	18	101-146	0	0.95	1	28	101-136	0.04	0.99	0.96
gwm368	4B	60	103-131	31	107-125	0.12	0.95	0.87	45	103-131	0.18	0.94	0.81
gwm369	3A,4B	12	232-288	9	243-268	0	0.65	1	11	232-288	0	0.85	1
gwm376	3B	15	186-296	10	186-296	0.09	0.75	0.88	13	187-293	0.07	0.49	0.86
gwm408	5B	27	118-145	18	136-145	0.79	0.87	0.1	17	118-145	0.9	0.82	0
gwm410	2B,5A	11	136-190	9	136-185	0.01	0.8	0.99	11	148-190	0	0.78	1
gwm413	1A,1B	31	234-342	20	234-341	0.1	0.9	0.89	25	234-342	0.14	0.9	0.84
gwm448	2A	43	82-98	23	89-98	0.16	0.89	0.82	36	82-98	0.28	0.96	0.71
gwm471	7A	5	202-247	5	204-240	0	0.48	1	3	202-247	0	0.63	1
gwm480	3A	20	105-191	8	105-185	0.08	0.87	0.91	18	110-191	0.06	0.9	0.94
gwm493	3B	20	172-181	13	172-181	0.88	0.82	0	18	172-174	0.82	0.88	0.07
gwm494	1B,3A,4A,6A	45	130-176	26	138-176	0.87	0.89	0.03	37	130-176	0.88	0.92	0.04
gwm518	6B	19	173-208	14	174-206	0.16	0.83	0.8	16	173-208	0.02	0.89	0.97
gwm526	2A,2B	31	126-228	20	126-226	0.1	0.91	0.89	21	126-228	0.03	0.95	0.97
gwm537	5B,7B	10	129-158	8	132-149	0.41	0.67	0.39	7	129-158	0.33	0.44	0.24
gwm601	4A	25	202-238	17	207-229	0.09	0.88	0.9	23	202-238	0.29	0.91	0.68
gwm610	4A	43	121-140	34	123-140	0.44	0.96	0.54	32	121-128	0.73	0.93	0.21
gwm611	7B	12	149-186	9	149-181	0.03	0.79	0.96	10	153-186	0.01	0.88	0.99
gwm614	2A,2B,4A	36	122-216	24	134-216	0.35	0.82	0.57	27	122-178	0.21	0.83	0.74
gwm617	5A,6A	31	142-158	18	142-158	0.43	0.95	0.55	28	147-156	0.54	0.88	0.38
gwm639	5A,5B	21	111-190	20	111-174	0.64	0.83	0.22	10	111-190	0.66	0.74	0.11
gwm644	1B,3B,6B,7B	8	128-184	8	131-184	0.03	0.99	0.97	3	128-182	0	0.96	1
gwm666	1A,3A,5A,7A	16	110-162	12	110-162	0.37	0.67	0.45	9	135-148	0.56	0.67	0.18

1 *Na: the number of observed alleles by loci. ASR (bp): Allele size range in base pair. H_o: Observed heterozygosity. H_e: Expected heterozygosity*

4. 3. 2. Population structure

For the STRUCTURE analysis, it separates clearly between the Moroccan and the Syrian landraces. However, 16 landraces originated from Syria and mainly from the coastal areas were grouped with the Moroccan landrace population; and one landrace (Sherieh) was assigned equally to both populations. These results indicate that durum wheat in North Africa may have been introduced from the coastal areas of the Middle East where the Phoenicians had lived and later immigrated to the South-Western Mediterranean countries. This is supported by the strong relativeness found in this study between the coastal durum wheat landraces from Syria and the Moroccan ones. On the other hand, two landraces from Morocco (ICDW20038-Tiznit, ICDW20041-Tiznit) were grouped with the Syrian population (Figure 39a). In GENELAND analysis results by using 10^6 iterations at 85% runs, the number of populations detected was 2; and the remaining 15% runs detected 3 populations. Distinction between the two populations was very clear if we consider number of populations K=2 (Figure 39b).

In this chapter we will be discussing only the Moroccan landraces population. Syrian population is discussed in our paper Kehel et al. 2013 attached to this document.

GENELAND estimated six Moroccan sub-populations (Figure 40). Maps of posterior probabilities of the six subpopulations are shown in Figure 41 and were named ***P1M***, ***P2M***, ***P3M***, ***P4M***, ***P5M*** and ***P6M***. Eleven landraces were attributed at more than 90% to subpopulation1 (***P1M***), 7 of them are from Tensift and two from Doukkala regions (South Casablanca-Marrakech region), region that is influenced by the Atlantic ocean, but this subpopulation contains as well 3 other landraces from the North-Eastern region of Oujda and Figuig, which is also influenced by the Mediterranean Sea. The second subpopulation (***P2M***) was found at 96% and contains 2 landraces originated from the irrigated areas of the South-Eastern Atlas high plateaus and 3 landraces from the highlands of Boulman and Nador regions in the eastern plateaus of Middle-Atlas and Rif Mountains. The third subpopulation (***P3M***) consists of 8 landraces from southern warm areas of Morocco (Tata, Tiznit, Goulmine). As for the subpopulation (***P4M***), it has the largest number of landraces (46). The landraces of this cluster are originated mainly from the western mountainous cold areas of the Atlas Mountains and Rif chains. Further, for most of the 14 landraces of subpopulation 5 (***P5M***), they were originated from the southern Atlantic lowland region of Morocco (Taroudant, Agadir and Essaouira); and 3 landraces from the northern Atlantic lowland region (Larache). These latter ones were assigned as well to ***P4M*** at 40%. The sixth subpopulation (***P6M***) had 15 landraces from Moroccan pre-and anti-Atlas areas (Beni Mellal, Khenifra, Errachidia and Ouarzazate) representing the continental areas of the pre-and anti-Atlas plateaus of South-East Morocco.

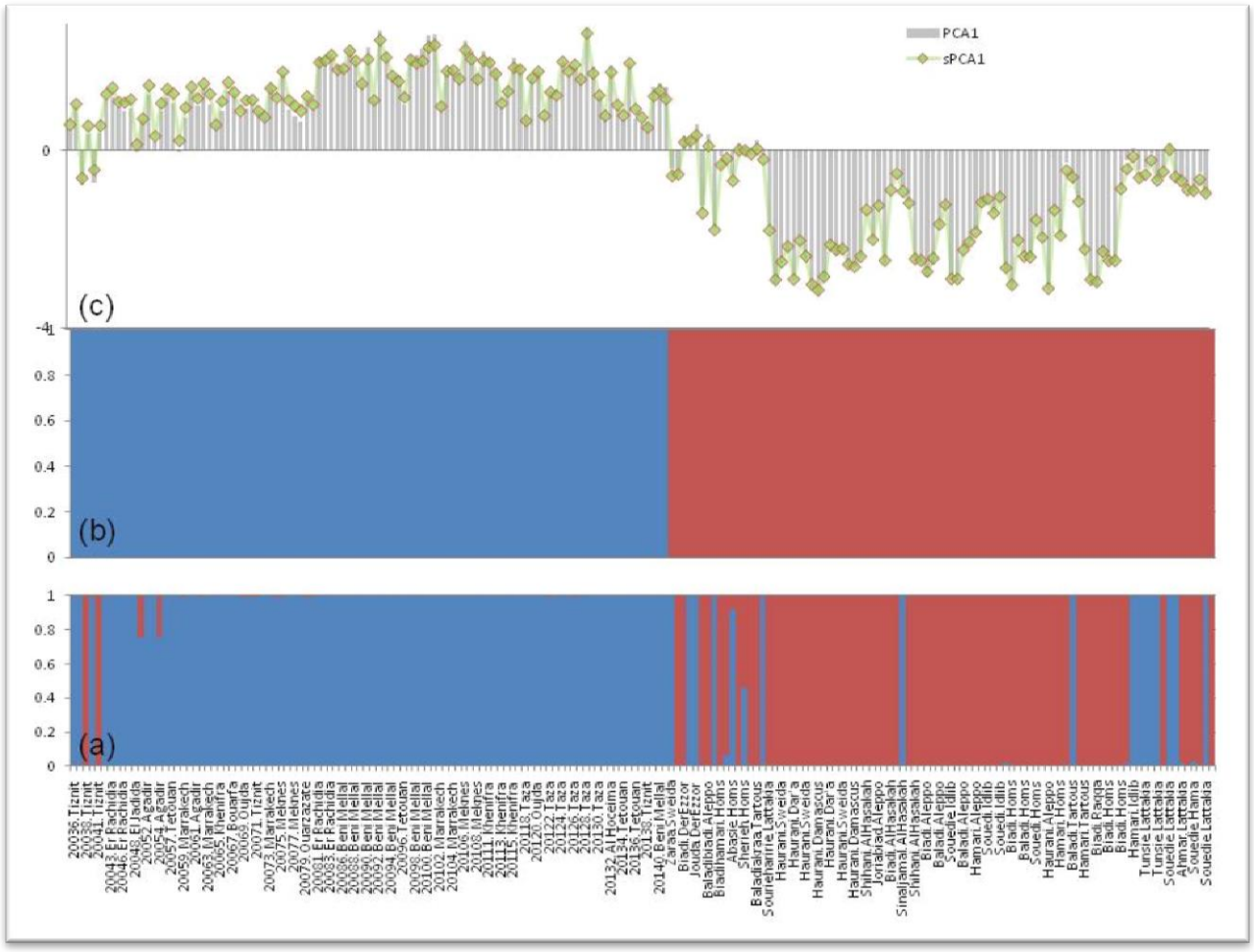


Figure 39: Genetic structure of the Moroccan and Syrian durum wheat landraces

a) STRUCTURE chart; b) GENELAND chart: c) first principal component (Grey bars) and first spatial principal component (Green points).

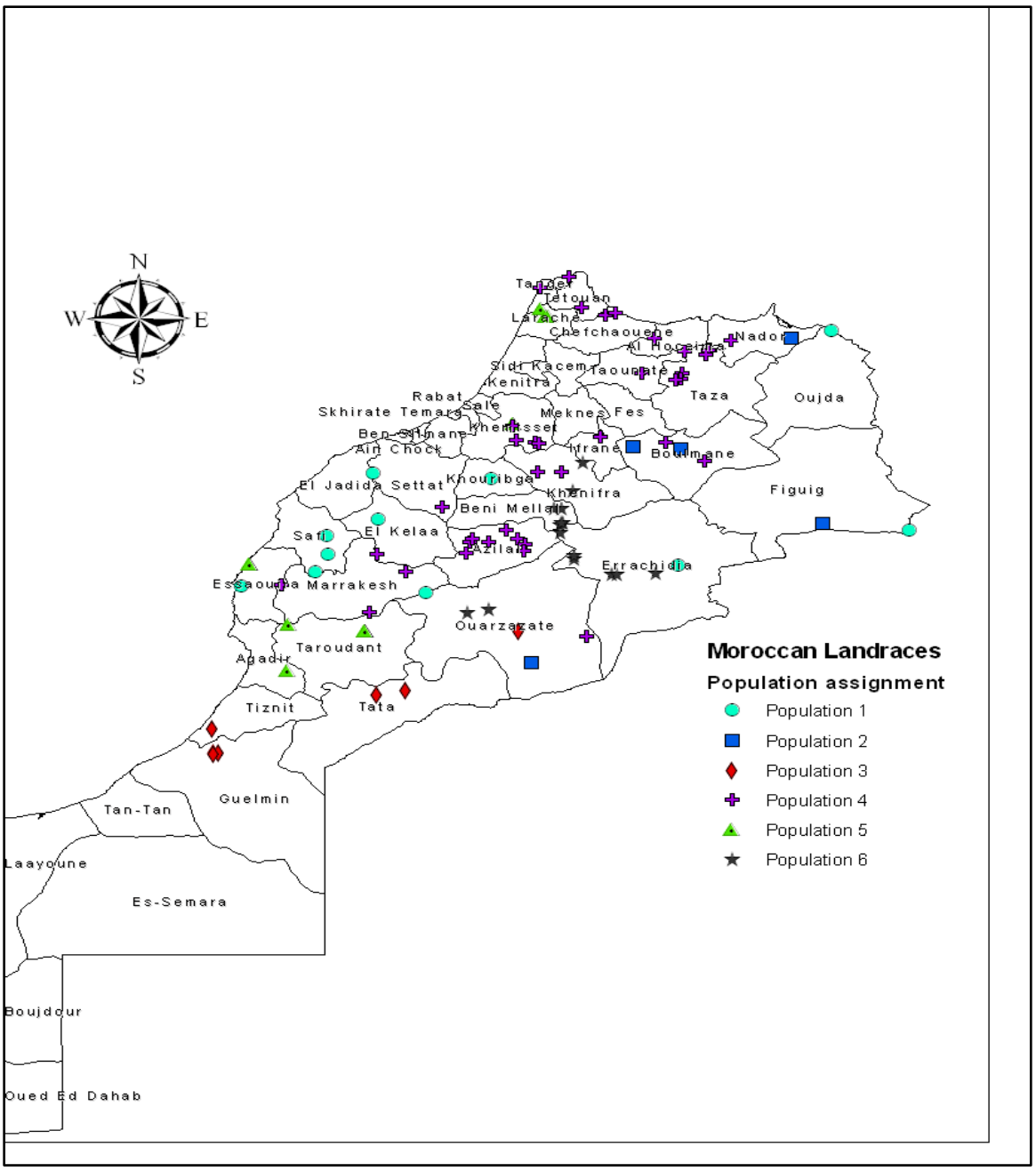


Figure 40: Moroccan landraces sub-populations

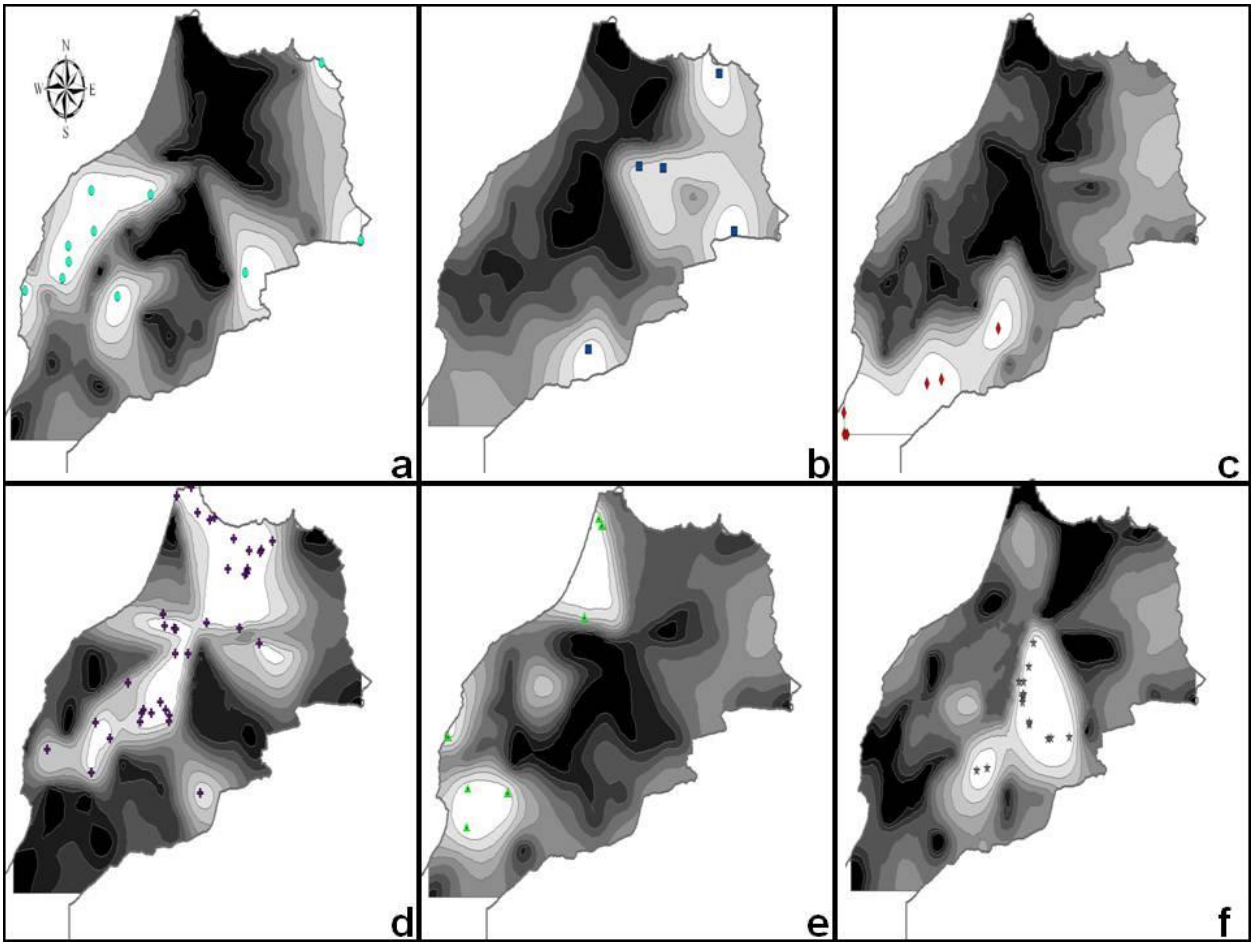


Figure 41: Moroccan landraces population probability assignments on the Morocco studied area

high probability: white; low probability: dark.
a: P1M, b: P2M, c: P3M, d:P4M, e: P5M and f: P6M

Heterogeneity, number of alleles, individuals and the geographical regions of collection of landraces composing each subpopulation are summarized in Table 26. **P2M**, **P3M** and **P5M**, found at the Eastern and southern parts of Morocco had the largest number of alleles per locus and large values for heterogeneity compared to **P4M** found in the highland areas.

Table 26: Moroccan durum populations information

	P1M	P2M	P3M	P4M	P5M	P6M
Geographical region	**Safi, Eljadida, Elkelaa, Essaouira, Marrakech, Khribga, Figuig, Oujda**	**Ouarzazat, Bouleman, Nador, Figuig**	**Tata, Tiznite, Goulmine**	**Azilal, Khmiset, Khnifra, Marrakech, Ifrane, Boulman, Taza, Tawnate, Housseima, Tetouan, Chaouen, Tanger**	**Taroudante, Agadir, Essaouira**	**Beni Mellal, Khenifra Errachidia, Ouarzazat**
Number of individuals	11	5	7	46	14	15
Total number of Alleles	205	140	217	471	405	302
Number of Loci with Ho=0	32	37	24	9	16	17
Number of Loci with Ho>0.5	9	4	9	7	9	7
Number of Loci with GD=0	1	2	0	0	0	2
Number of Loci with He>0.5	33	44	43	41	45	43
GD	0.564	0.727	0.716	0.656	0.775	0.685
H0	0.184	0.102	0.17	0.187	0.223	0.182

4. 3. 3. Spatial entity of molecular markers

The SAU, under the form of Moran's I, applied to allele frequencies showed that 30% of the alleles showed no SAU with a value of $I = I_0 = -0.01$. Eighty alleles had a significant positive SAU with a maximum value of 0.43 observed at the allele 217 of *gwm285* and a z-score of 8 (p-value =0).Allele 118 of gwm375 had a very significant SAU(I=0.2, z-score=4.8, p-value =0). Figure 42 shows the spatial distribution of the allele 217 of *gwm28.*

The high values of z-score for these two alleles indicated that the values of the alleles (0, 1 and 2) are clustered and there is less than 1% that this spatial cluster is a result of random process.

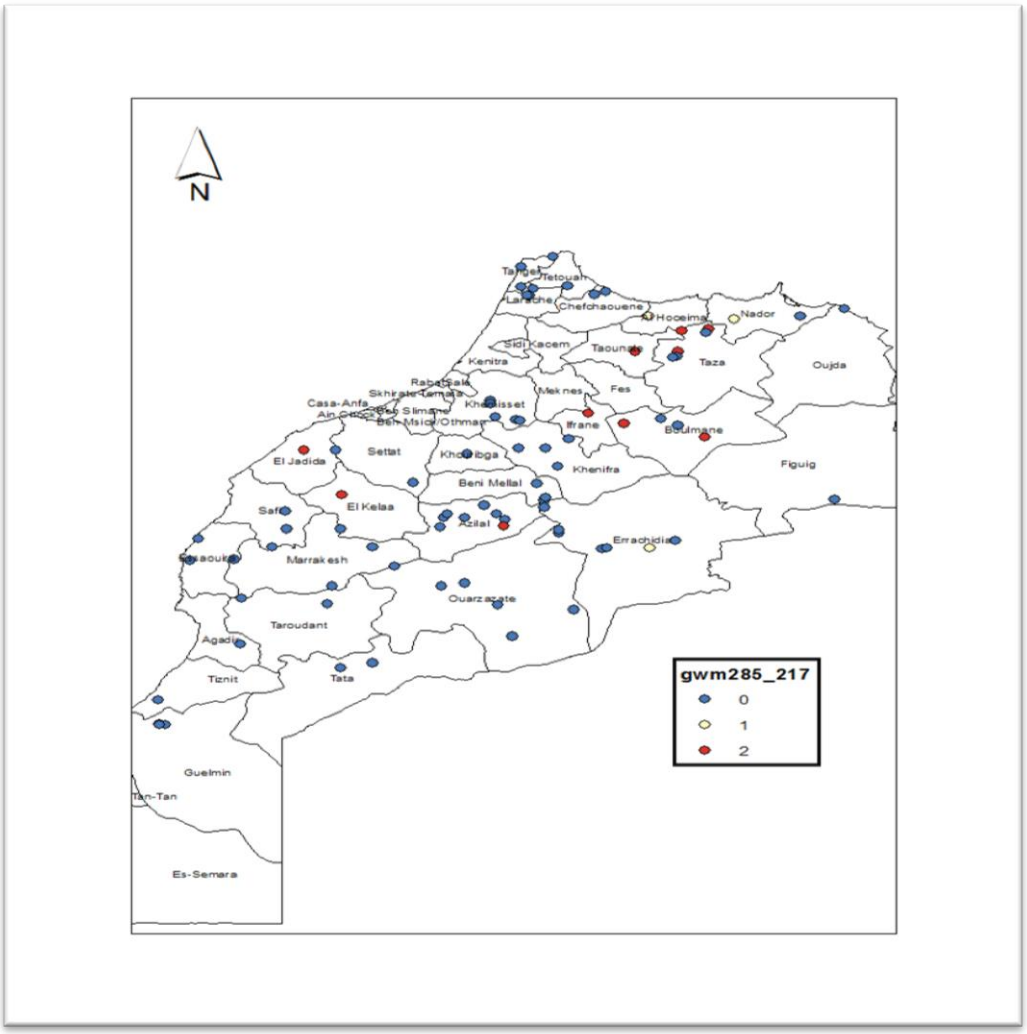


Figure 42: Spatial distribution of alleles 217 of GWM285

Using the Local z-score of Getis-Ord General G, we found that high values (2) for both alleles are high positive with a significant p-value at 1%. We can conclude that the high values of these two alleles (value=2) are clustered spatially (high clusters). Durum wheat landraces are homozygotes for these alleles and are collected in the same areas.

Using Getis-Ord Local G, significant hot spots for allele 217 of GWM 285 (selected region in Figure 43) were found to be located in the mountains area of RIF chain and Fes-Saiss region. We consider a hotspot where a landrace has interesting value (2 for the allele in our case) but also neighboring landraces have similar significant patterns. If a landrace has high value and neighboring landraces have different values, the location is not considered as hotspot.

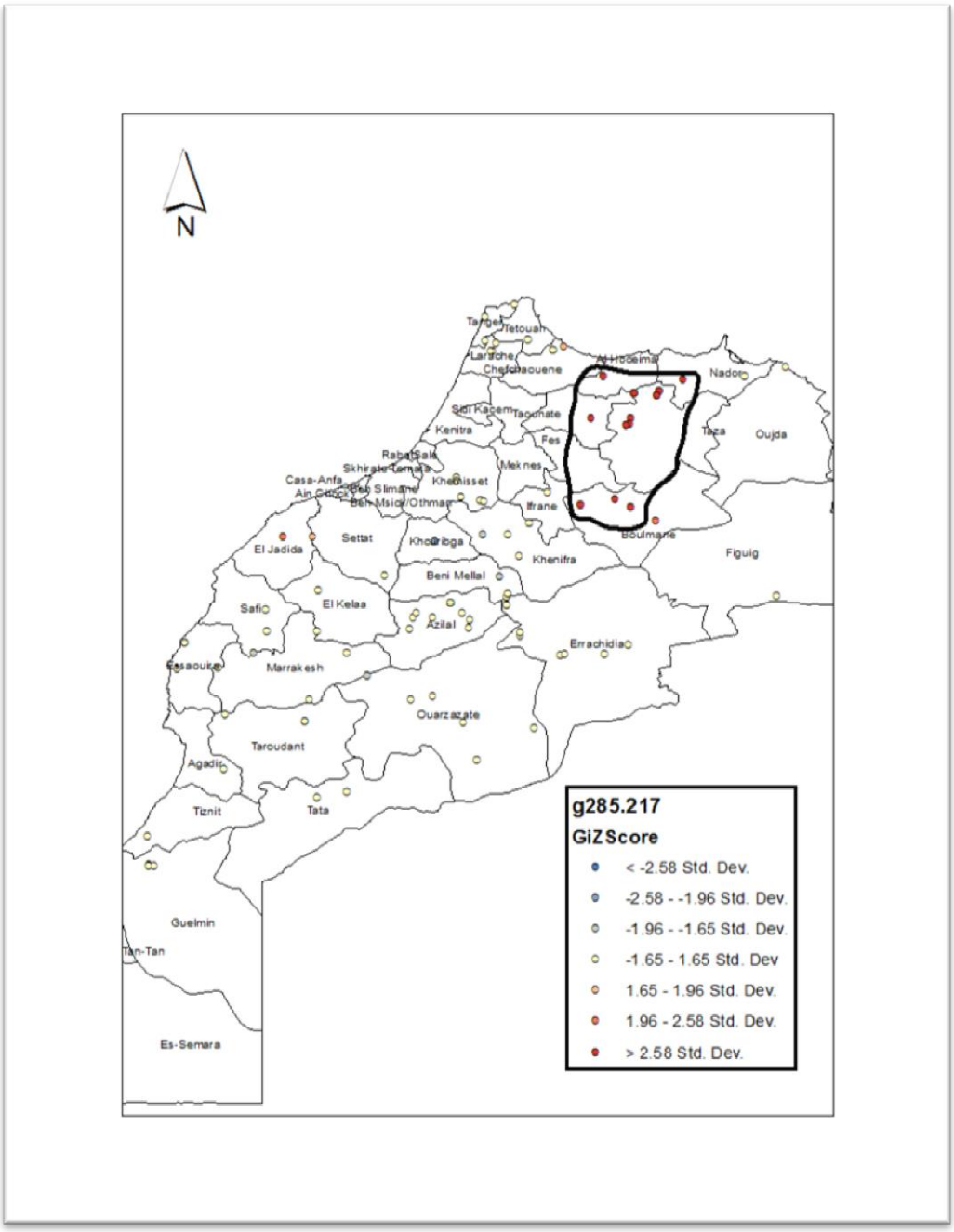


Figure 43: Hotspots for allele 217 of GWM285

For allele 18 of GWM357 (Figure 44), the hotspot area was found in the high altitude area of Atlas (Beni Mellal, Azilal and Errachidia).

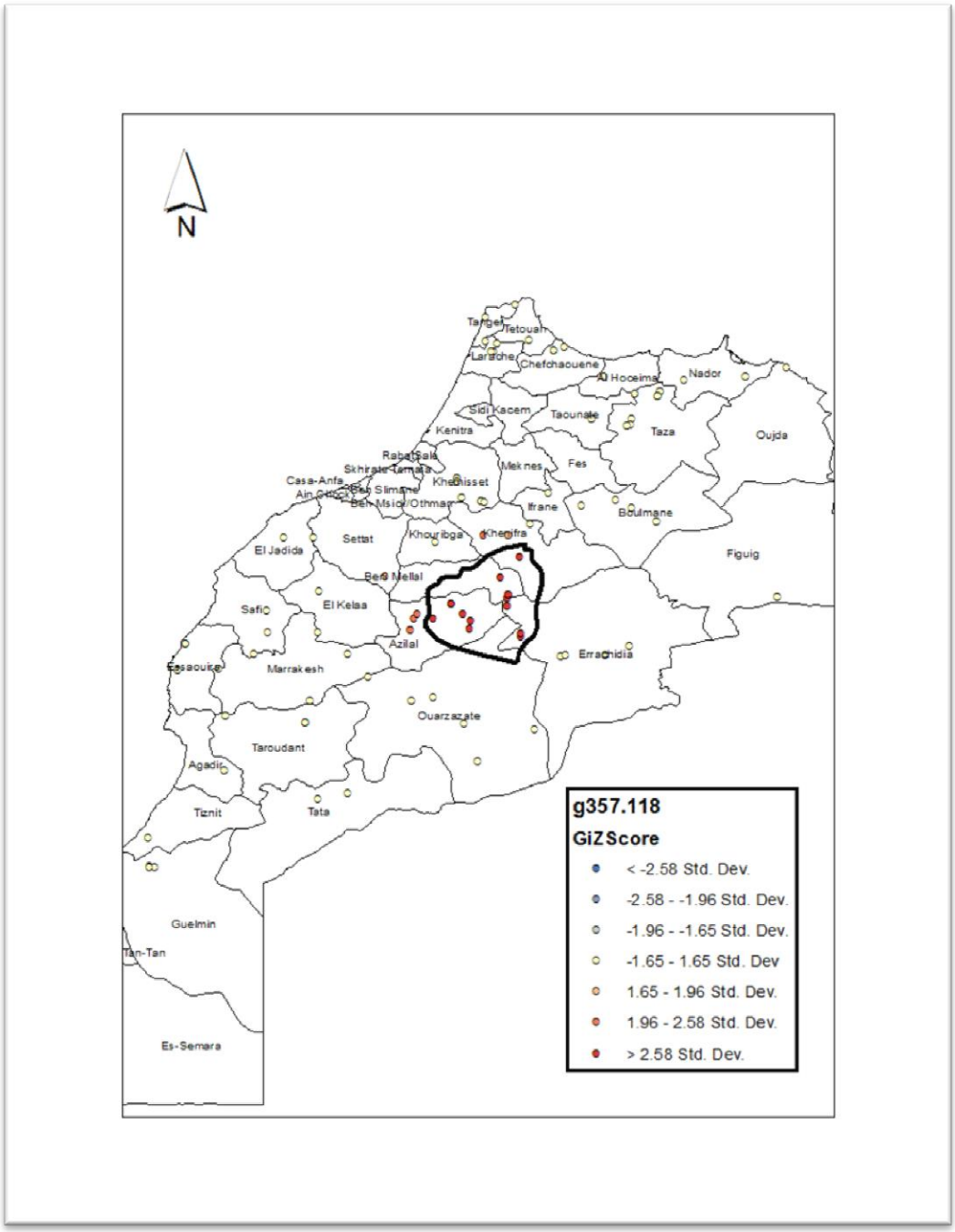


Figure 44: Hotspots for allele 118 of GWM375

4. 3. 4. Multivariate analysis

In the PCA analysis, the two first eigenvalues explained only 6% of the total variance. But plotting the first axis against the second axis and first axis against the third axis (Figures 45 and 46) gave clear evidence of the structure and of the difference between the two populations (Moroccan and Syrian) especially the first axis, which was positive for Moroccan landraces and negative for the Syrian ones (Figure 39c). Some exceptions were stated; and some of them were matched with those found in the STRUCTURE results (Figure 39c). Using the ANOVA t-test, the means of the first axis significantly differ between the two populations, Syrian and Moroccan, with a *p-value* of 0.

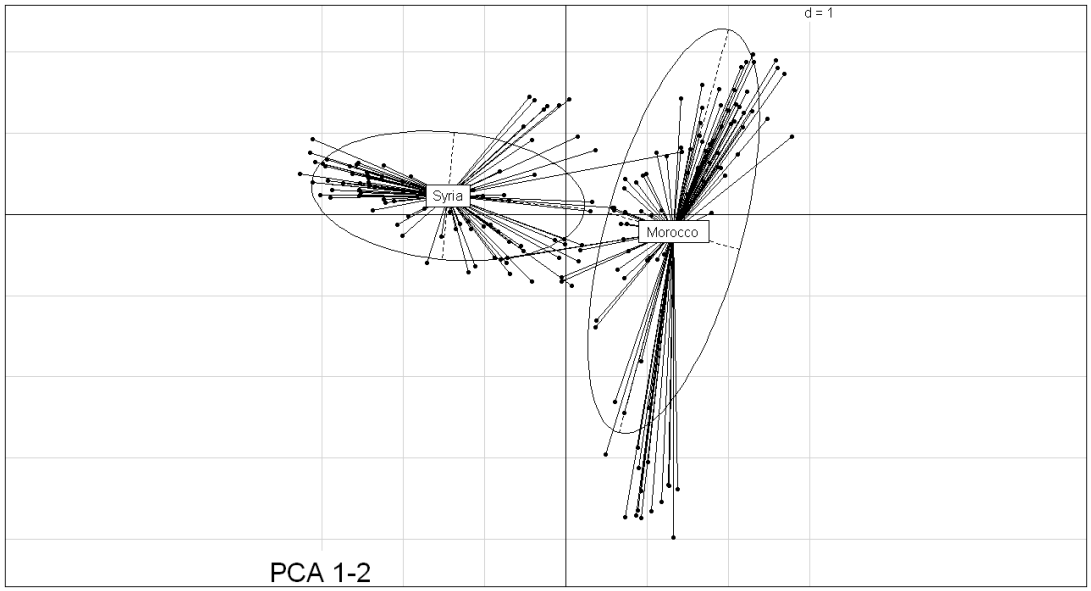


Figure 45: PCA plot of Moroccan and Syrian landraces (axis1 VS axis2)

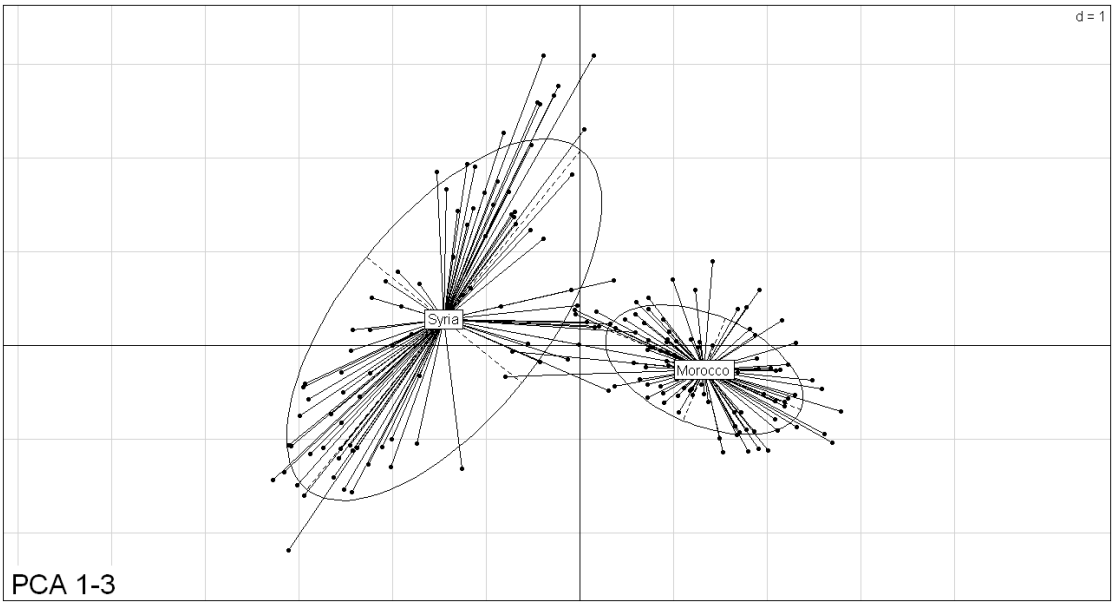


Figure 46: PCA plot of Moroccan and Syrian landraces (axis1 VS axis3)

In general, in PCA each principal component is associated with an eigenvalue that quantifies the amount of variance explained by the component. The bar-plot of the eigenvalues sorted in decreasing order is the basic tool used to choose which principal components to interpret: it describes how the total genetic variance is distributed across the principal axes.

In the spatial Principal component analysis sPCA, positive and negative Eigenvalues are computed and plotted (Figure 47). Positive Eigenvalues are associated with global genetic structure while negative Eigenvalues are associated with local genetic structure.

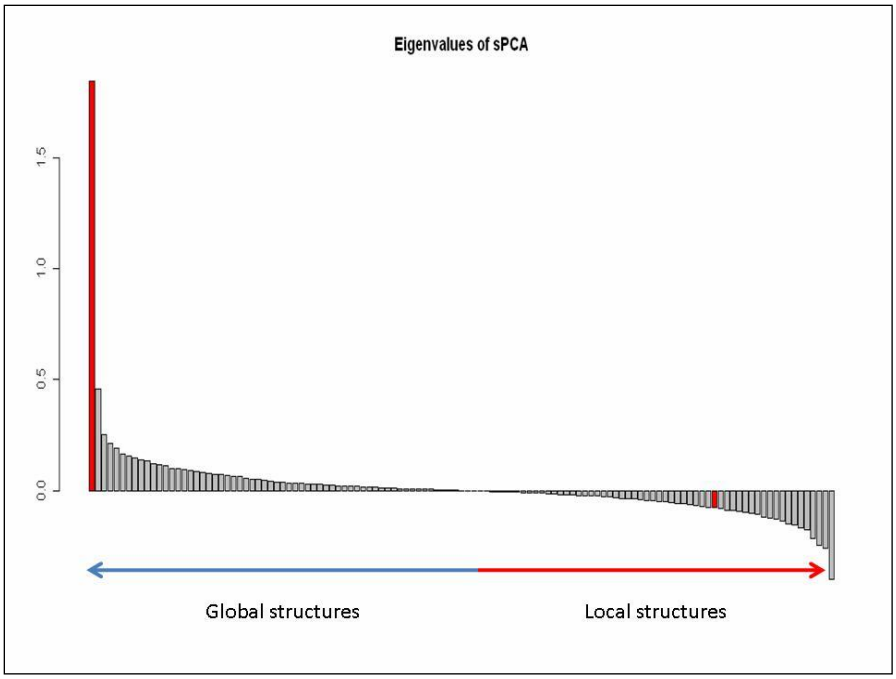


Figure 47: Histogram of sPCA eigenvalues

The sPCA analysis gave similar results as the PCA analysis with approximately the same exceptions (Figure 2c). We used for sPCA a minimum distance connection network, in order to not connect Syrian and Moroccan landraces as they were spatially not linked.

In this case, the sum of all Eigenvalues used has no sense (as compared with PCA). This sum can be very low if there is no genetic structure in the data or if; for example, we have similar global and local structures. A suitable method of selecting useful Eigenvalues to be interpreted is to assess score that account for large genetic variability and enough spatial structure (using spatial autocorrelation). This can be done using the plot (Figure 48) of variance against SA.

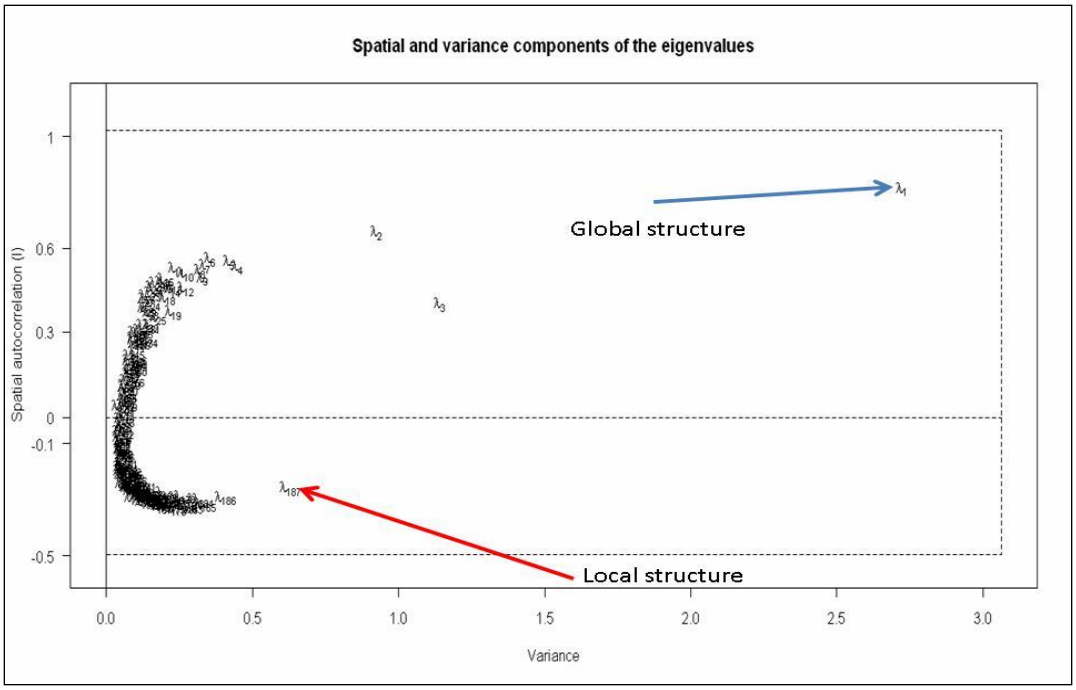


Figure 48: Plot of variance component of the sPCA Eigen values versus spatial autocorrelation

The Lag vector computes for a given landraces is the mean frequency of its neighbors. Plotting the first and last components against their Lag vectors showed positive and negative regression coefficients and demonstrated the significance of the global and local patterns of the two components respectively (Figures 49and 50).

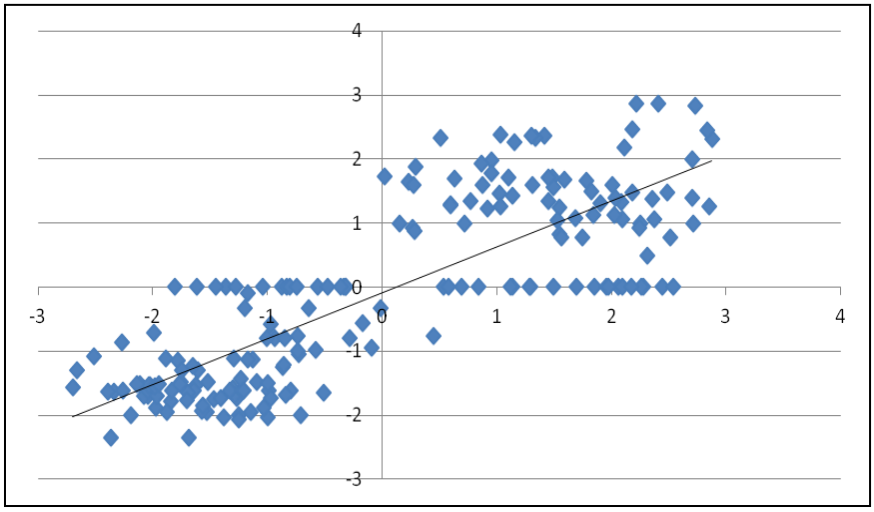


Figure 49: Plot of the first component (x-axis) and its lag vector (y-axis)

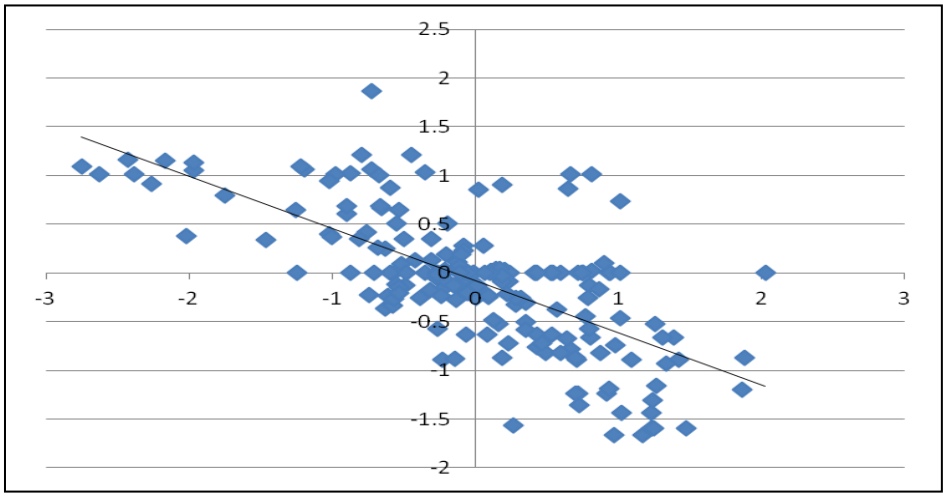


Figure 50: Plot of the last component (x-axis) and its lag vector (y-axis)

For the Moroccan durum landraces population, we used the matrix of allele frequencies without standardization for PCA analysis. The first three eigenvalues were 52.23, 11.75, and 4.78 and explaining respectively 30.9, 7.0, and 2.9 % of total genetic variability. The corresponding axes are symbolized **PC1**, **PC2** and **PC3** in this study.

For sPCA analysis, the spatial network, we used Gabriel graph. The first and the last eigenvalues (**λ1**; **λ97**) had the strongest variance and (positive for **λ1**; negative for **λ97**) spatial autocorrelation(Figure 48). The global and local tests presented by Jombart et al. (2008) showed a significant global test (*p = 0.02*) and non-significant local test (*p = 0.26*). Therefore, only the global structure is significant and only **λ1** is interpretable in the case of Moroccan landraces and the first sPCA axis **sPC1** is used for evaluation analysis. The SAU of **sPC1** was 0.47, for **PC1** was 0.28, for **PC2** was -0.02 and **PC3** was 0.27, which means a global structure is given by **PC1** and **PC3**. In addition to the positive value of SAU of **PC1**, mapping **PC1** over the studied areas of Morocco showed a very strong spatial pattern schematized by a positive component for landraces from the high altitude (*RIF* and *ATLAS* mountains) landraces had a negative one elsewhere. The same spatial structure was found for **sPC1** (Figure 51).

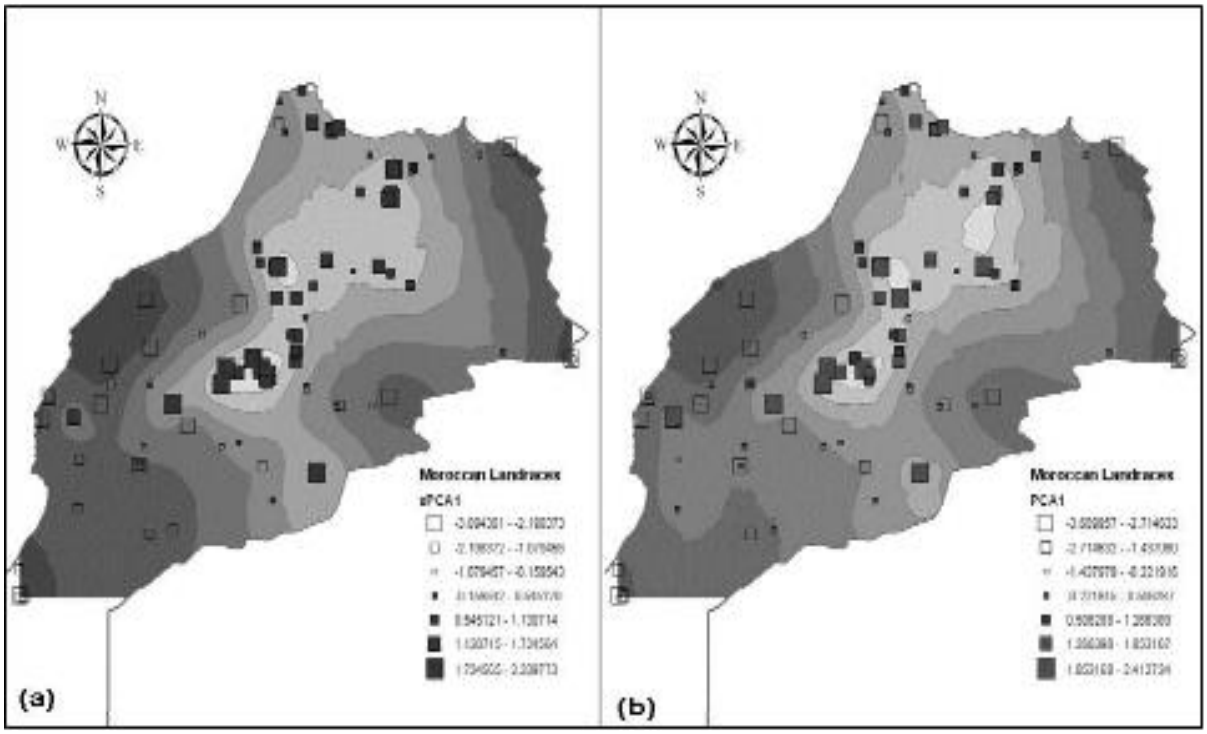


Figure 51: Maps of the first spatial principal (a) and principal (b) components for the Moroccan durum wheat landraces

Correlation between **PC1** and **sPC1** coordinates was very highly significant (*p< 0.001*) with a coefficient of 0.87 and R^2 of 75.7. A correlation with groups between the two axes and using subpopulations found by GENELAND, as factor was also highly significant (*p <0.001*) with a coefficient of 0.61 and R^2 of 37.2. The t-test (Table ??) showed that only **P6M** could not be differentiated by the four axes (**sPC1**, **PC1**, **PC2** and **PC3**). sPC1 an PC1 could discriminate between 4 out of the six Moroccan sub-populations (Table 27).

Table 27: T-test value for populations found for Moroccan durum landraces

	PC1	PC2	PC3	sPC1
P1M	< 0.001	0.002	0.034	< 0.001
P2M	0.221	0.751	0.763	0.321
P3M	< 0.001	0.233	0.275	< 0.001
P4M	< 0.001	0.043	0.002	< 0.001
P5M	0.03	0.094	0.89	0.002
P6M	0.514	0.006	< 0.001	0.186

Alleles 202 of GWM234, 137 of GWM 293, 114 of GWM282 and 114 GMW2 are the alleles contributing most to the global structure found by sPCA (Figure 52). This means that these alleles can reveal a big amount of the first spatial principal component. Those alleles can be used to discover global spatial genetic patterns found in durum wheat landraces as they have greatest weight in revealing genetic structure.

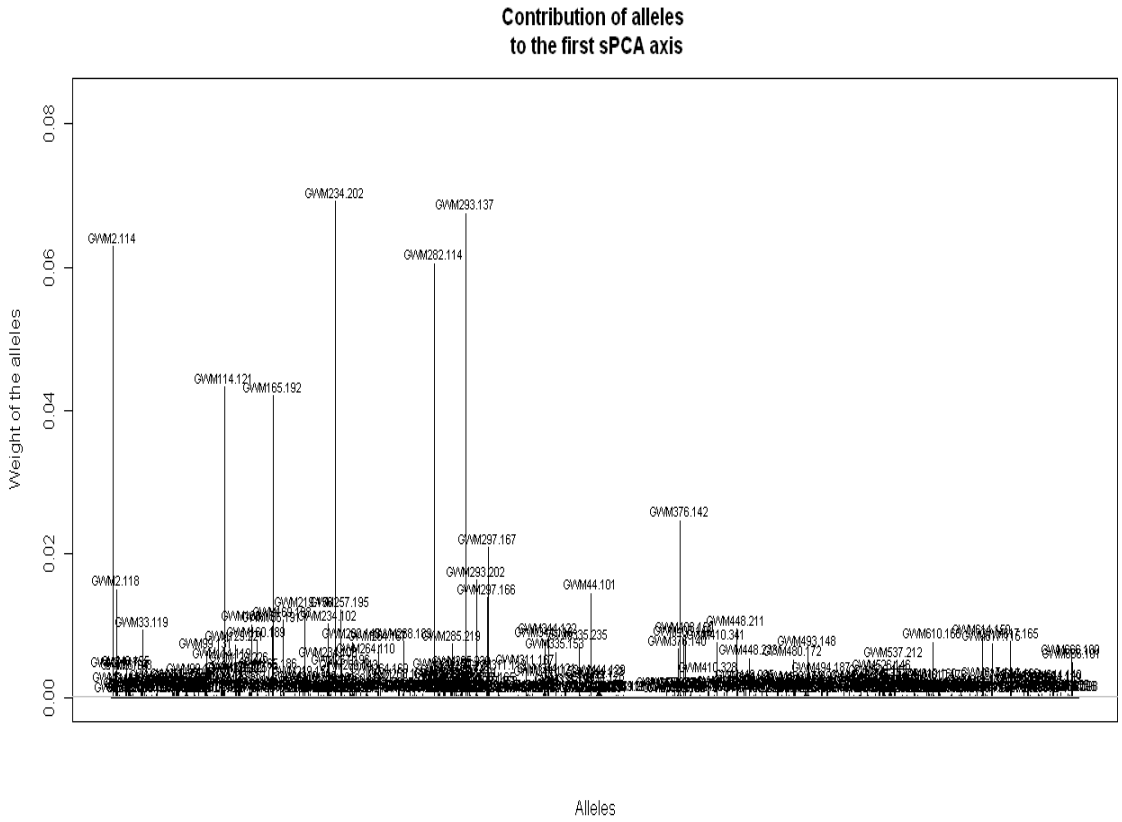


Figure 52: Contribution of alleles to the first sPCA component

On the other hand, most of the local structure discovered by the last spatial principal component was expressed by alleles 156 of GWM219, 118 of GWM 357, 119 of GWM33 and 121 of locus GWM33 (Figure 53). This showed that the alleles cited above can be used to discriminate between neighboring durum landraces.

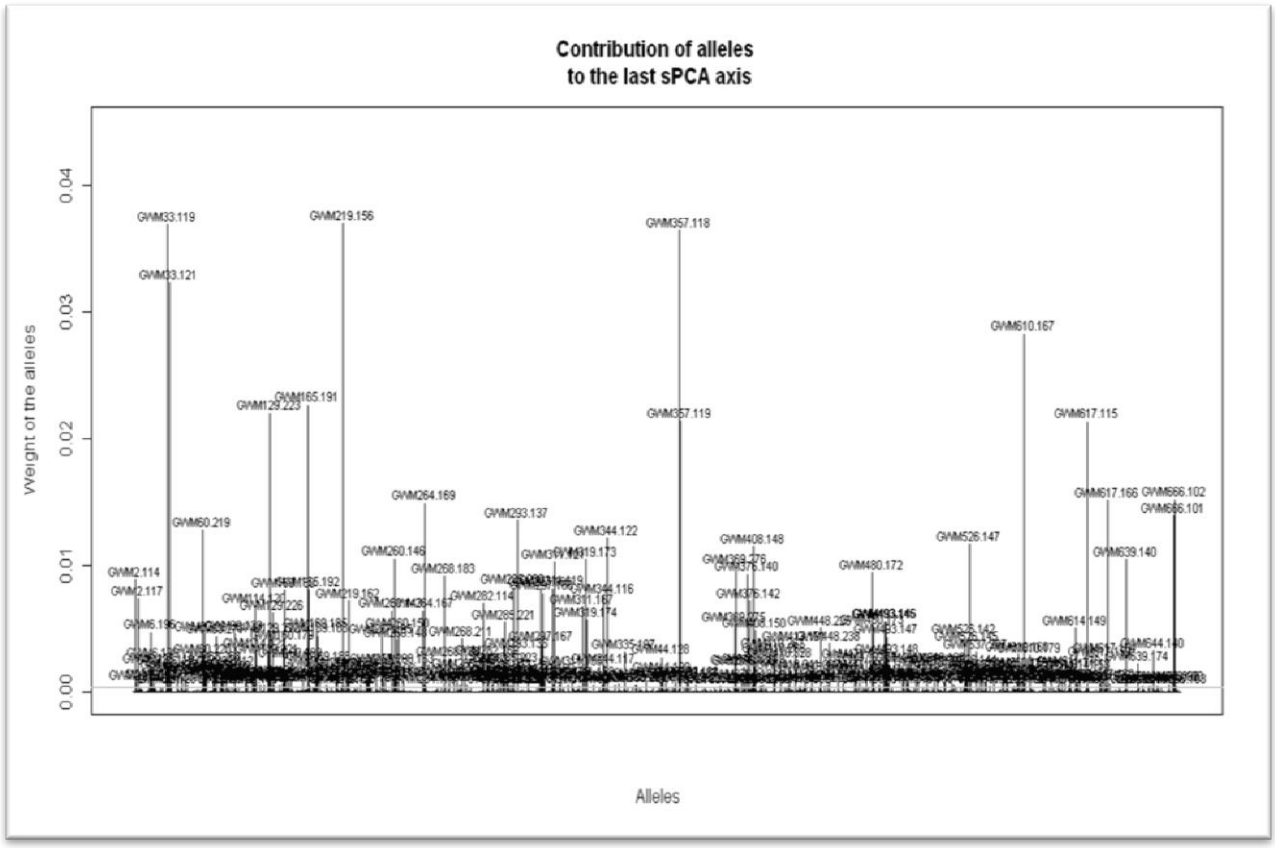


Figure 53: Contribution of alleles to the last sPCA component

Using the SSRs used earlier for Bayesian and multivariate structure to study genetic discontinuities among Moroccan durum wheat landraces, a significant barrier resulted from analysis using Monmonier's algorithm. This genetic barrier coincides mainly with the Moroccan altitude pattern and fellow the delineation of the two main mountainous chains in the country (Rif in the North and Atlas in the middle). Overlapping this barrier with the patterns found by the sPCA showed obviously that this genetic barrier distinguish between positive and negative values of the first spatial principal component (Figure 54). If we use the global structure of sPCA to study to genetic structure of the Moroccan durum landraces. We could identify easily two sub-populations: one having positive score and the other having negative score. The Monmonier barriers (or path) is delaminating these two sub-populations.

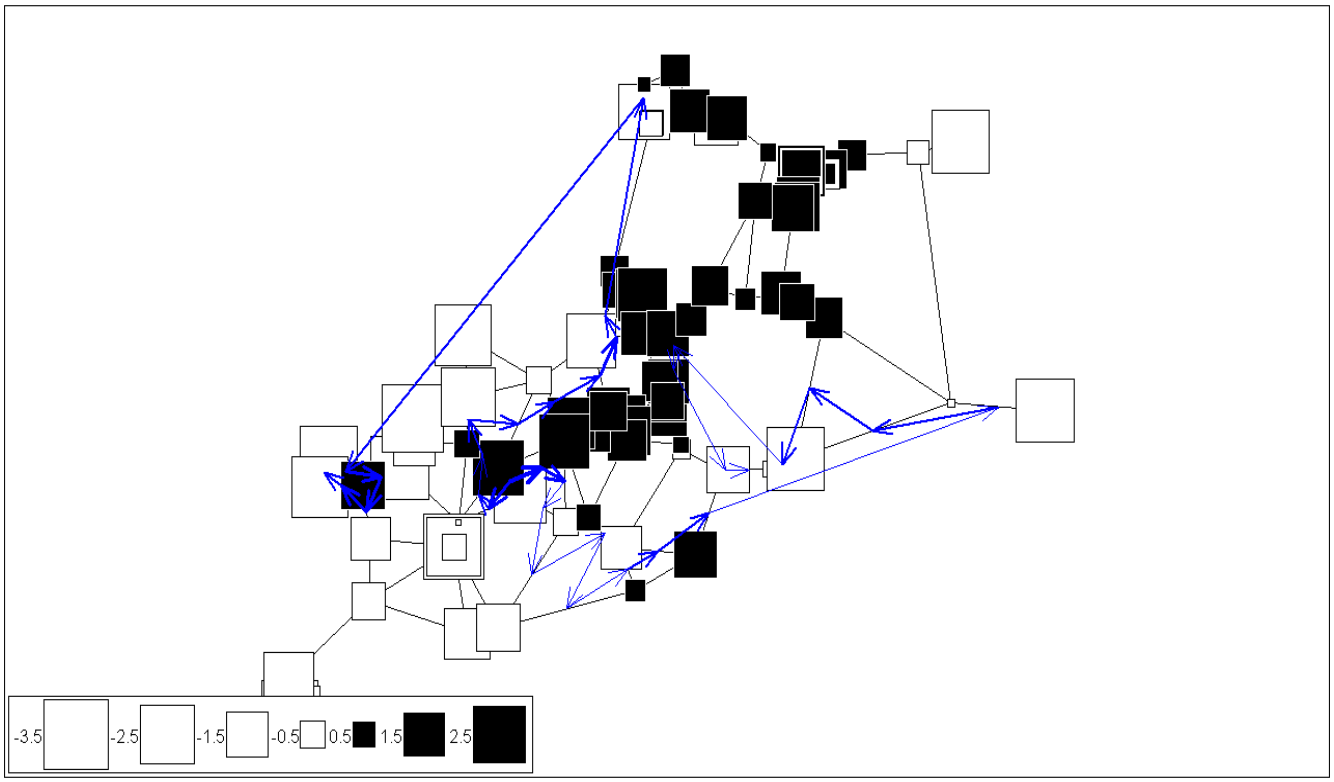


Figure 54: Moroccan durum landraces molecular barriers

The squares size is the first spatial principal component

In an attempt to compare between durum landraces genetic and geographic distances, we computed the spatial autocorrelation using the genetic similarity for each class of geographic distance (Figure 55). For this analysis we used 10 distance classes used for analysis. The distance classes were constructed using equal distances with unequal sample sizes. SA was computed for each distance class and the average was 0.71. The minimum was observed for class 1 with SA=0.66 and the maximum for class 9 with a SA=0.78. For all classes we could notice that our markers present a global structure within the Moroccan durum landraces. This result is in accordance with what was found using multivariate or Bayesian statistics. Also, the correlation between genetic and geographic distance was of 0.2647. The probability of observing a correlation greater or equal and less or equal to 0.2647 was 0. 0009 and 1 respectively using a permutation test with number of permutation equal to 10^6.

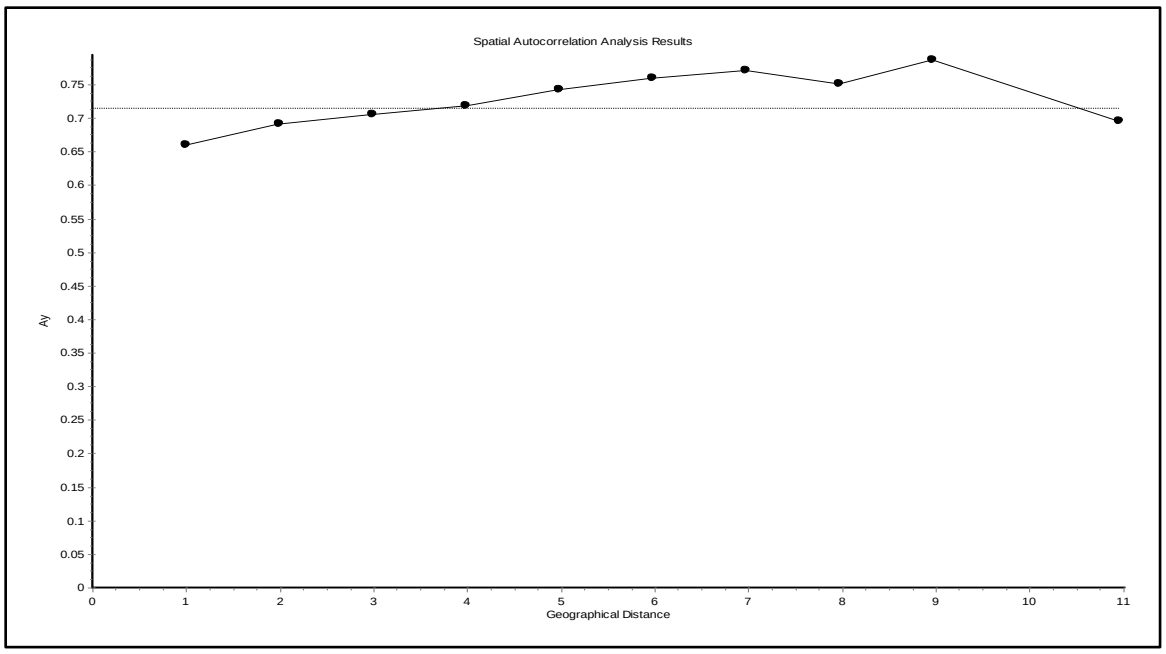


Figure 55: Spatial autocorrelation in relation with distance classes

4. 3. 5. Evaluation of populations

According to the phenotypic data, the 6 populations found for the Moroccan landraces were very diverse. The most diverse traits were KSPK ranging from 16 to 24 and SPM2 from 196 spikes per square meter to 249. For TKW, population4 had an average of 41.59 g while population3 had an average TKW of 34.21 g.

The mean GY, ASH content, YP and PC were not different between the populations. On the other hand, almost all the physiological traits showed diversity from population to another (Table 28).

Table 28: Agronomic and physiological traits of the Moroccan durum wheat landraces populations

	P1M	P2M	P3M	P4M	P5M	P6M
INDIVIDUALS	11	5	7	46	14	15
GY	2379.77	2382.88	2355.7	2364.89	2363.22	2380.75
ASH	3.03	3.02	3.02	3.02	3.02	3.02
PC	15.43	15.51	15.29	15.52	15.57	15.46
PH	100.78	104.16	102.46	107.16	106.57	105.22
DH	142.09	144.68	141.18	145.52	144.05	145.36
DM	170.76	174.2	169.76	175.24	172.53	174.17
GFD	30.49	30.62	31.35	30.61	30.32	31.22
SDS	25.58	25.77	27.67	25.81	25.74	24.92
SDSn	3.8	3.85	4.05	3.83	3.85	3.7
SDSi	1.75	1.75	1.92	1.77	1.74	1.71
YP	5.3	5.21	5.39	5.08	5.29	5.05
TKW	34.74	39.31	34.21	41.59	37.99	39.44
VIT	94.13	94.13	93.99	94.13	94.05	93.94
SL	6	6.8	6	7.02	6.93	7.2
SPM2	249.45	196	237.14	210.87	232.29	203.47
PL	2.73	2	4	2	2.36	2.07
KSPK	23.27	16	24.57	14.89	19.57	18
Area70	85090.91	86920	109357.1	87052.17	95235.71	94113.33
CARI45	0.956675	0.924831	0.979904	0.97382	0.976422	0.97682
CARI70	0.967726	0.968395	0.968094	0.968788	0.968887	0.969324
CAROTENE45	-0.00128	-0.00121	-0.00084	-0.00109	-0.00107	-0.00103
CAROTENE70	-0.00129	-0.00126	-0.00128	-0.00125	-0.00127	-0.00126
CHL45	0.173504	0.140872	0.13208	0.140012	0.13492	0.15304
CHL70	0.157956	0.151317	0.155666	0.14603	0.159783	0.163767
F070	646.6364	639	601.7143	669.6957	624.5714	623.1333
F170	739.1818	735.6	703.7143	764.0652	723.7857	715.3333
F270	839.5455	838.4	800.7143	867.8696	823.4286	811.2
F370	1193.455	1203.2	1178.714	1228.152	1193.357	1159.4
F470	1736	1768.4	1818.571	1772.739	1799.214	1704.6
F570	2651.091	2831.8	2955.143	2661.413	2845.071	2593.467
Fm70	3482.182	3704.8	3787.143	3553	3698.786	3548.133
Fv/Fm70	0.810455	0.8246	0.840571	0.80887	0.829286	0.8226

Fv70	2835.545	3065.8	3185.429	2883.304	3074.214	2925
LWP70	5.461384	5.812233	6.305396	5.348235	5.966078	5.691019
NDVI45	0.838484	0.786485	0.902385	0.884677	0.89059	0.880406
NDVI70	0.818641	0.833057	0.824749	0.836595	0.831477	0.836529
NPCI45	0.027418	0.126779	-0.04487	-0.02179	-0.03277	0.011621
NPCI70	0.175392	0.151468	0.16932	0.145782	0.156461	0.143835
NPQ70	0.323666	0.311632	0.283721	0.344091	0.303297	0.378992
NPQI45	-0.06614	-0.05226	-0.06404	-0.05971	-0.06542	-0.06015
PRI45	0.033143	0.008558	0.05417	0.048395	0.053405	0.040645
PRI70	0.01018	0.018058	0.011313	0.01879	0.015329	0.019
QN70	0.297931	0.287282	0.262749	0.313868	0.280667	0.332347
QP70	0.953987	0.955011	0.956804	0.951848	0.954967	0.952116
Que70	0.237049	0.213276	0.190026	0.237755	0.206939	0.216095
RNVI45	3.018173	2.690412	3.40331	3.211007	3.24962	2.993527
RNVI70	3.204673	3.143548	3.167306	3.238294	3.239056	3.194133
RVSI45	3.622197	4.15969	5.306875	4.883478	4.616621	4.673644
RVSI70	5.130384	5.17638	5.103979	5.282179	5.150821	5.170597
SAVI45	-0.03154	-0.02206	-0.02456	-0.02222	-0.02476	-0.02388
SAVI70	0.287289	0.292446	0.289101	0.293393	0.291245	0.292487
SIPI45	0.862551	0.873374	0.895691	0.889129	0.886794	0.884459
SIPI70	0.880637	0.884061	0.883419	0.893251	0.886774	0.889796
SR45	14.98961	15.18685	24.83558	21.17202	20.02593	18.24835
SR70	18.22158	18.34366	18.14793	18.54481	18.34371	18.44833
Tfm70	350.0909	334.8	355	359.913	380	365
WI/NDVI45	1.305039	1.50981	1.241112	1.282083	1.254132	1.262504
WI/NDVI70	1.398014	1.383562	1.391306	1.382095	1.385921	1.381671
WI45	1.080534	1.088401	1.118348	1.116134	1.113886	1.10966
WI70	1.125393	1.12958	1.128736	1.133169	1.128629	1.131372
YPEC70	0.714691	0.735613	0.760503	0.707082	0.742246	0.718512

The variance component of the interaction genotype and the environment G.E was diverse and explained until almost 100% of the total genetic variability for two populations 2 and 4. Two other populations (populations 1 and 3) showed also high GxE (>88%). Finally populations P5M and P6M revealed less GxE and then the landraces forming these two population are less influenced by the environment. TKW exhibited reasonable (maxi 50%) GxE across all populations but was minimum at population 5 and 1. GFD was highly affected by the environment at 4 populations but relatively less affected at 2 populations (pop 2 and 3). GxE explained more than 50% for PH at all the populations with a maximum of 84% at P4M and minimum of 48% at P6M. All of remaining traits evaluated for GxE showed very diverse GxE across populations (Table 29).

Table 29: GxE variance components of agronomic traits per population

		P1M	P2M	P3M	P4M	P5M	P6M
PC	**G**	5.1	1	4.9	26.3	38.8	46.9
	G.E	94.9	99	95.1	73.7	61.2	53.1
VIT	**G**	7.2	12.1	1.4	1.0	11.3	1.0
	G.E	92.8	87.9	98.6	99.0	88.7	99.0
TKW	**G**	65.2	54.5	49.8	56.8	69.5	54.0
	G.E	34.8	45.5	50.2	43.2	30.5	46.0
ASH	**G**	24.8	1.0	6.3	1.0	30.9	41.3
	G.E	75.2	99.0	93.7	99.0	69.1	58.7
GY	**G**	11.2	1.0	11.8	1.0	26.0	37.2
	G.E	88.8	99.0	88.2	99.0	74.0	62.8
SDS	**G**	64.5	80.9	60.6	13.6	30.6	25.8
	G.E	35.5	19.1	39.4	86.4	69.4	74.2
SDSI	**G**	71.5	79.7	62.5	12.9	43.5	38.1
	G.E	28.5	20.3	37.5	87.1	56.5	61.9
SDSn	**G**	48.2	75.5	55.3	15.6	22.0	20.0
	G.E	51.8	24.5	44.7	84.4	78.0	80.0
YP	**G**	33.2	41.4	57.7	34.3	57.2	28.9
	G.E	66.8	58.6	42.3	65.7	42.8	71.1
DH	**G**	68.3	35.5	17.4	26.0	49.4	65.3
	G.E	31.7	64.5	82.6	74.0	50.6	34.7
DM	**G**	63.7	89.7	1.0	51.2	55.7	77.2
	G.E	36.3	10.3	99.0	48.8	44.3	22.8
GFD	**G**	1.0	46.5	45.2	2.8	1.0	1.0
	G.E	99.0	53.5	54.8	97.2	99.0	99.0
PH	**G**	40.0	38.4	25.6	16.0	42.5	52.0
	G.E	60.0	61.6	74.4	84.0	57.5	48.0

The variability found in the posterior probability of a landrace to belong to one or the other population in Morocco was explained by the environmental factors Tmin, Tmax, Rainfall and Altitude. Genotypes were also used to check if the variability maybe explained by the genotypes themselves. The genotype factor was absent in the variance components of the first three populations. Altitude contributed at 100% to population1 and population3, and at 80% to population5. Genotype counted for 17% of the total variance of Pop5. The probability to belong to population2 is completely explained by Rainfall. The four environmental parameters Altitude, Tmax, Tmin and Rainfall contributed to probability4 with 26, 11, 31 and 26 % respectively; while the genotype had only a contribution of 6% of the total variance component. The total variance component of probability6 was divided into four environmental parts (43% for Altitude, 32% for Rainfall, 15% for Tmax and 10 for Tmin). The total variance components of **PC**1 and **sPC1** were divided between genotype, Rainfall and Tmin (Figure 56). This showed that most of the genetic variability and structure of the Moroccan durum wheat landraces can be explained by the long-term climatic factors and Altitude.

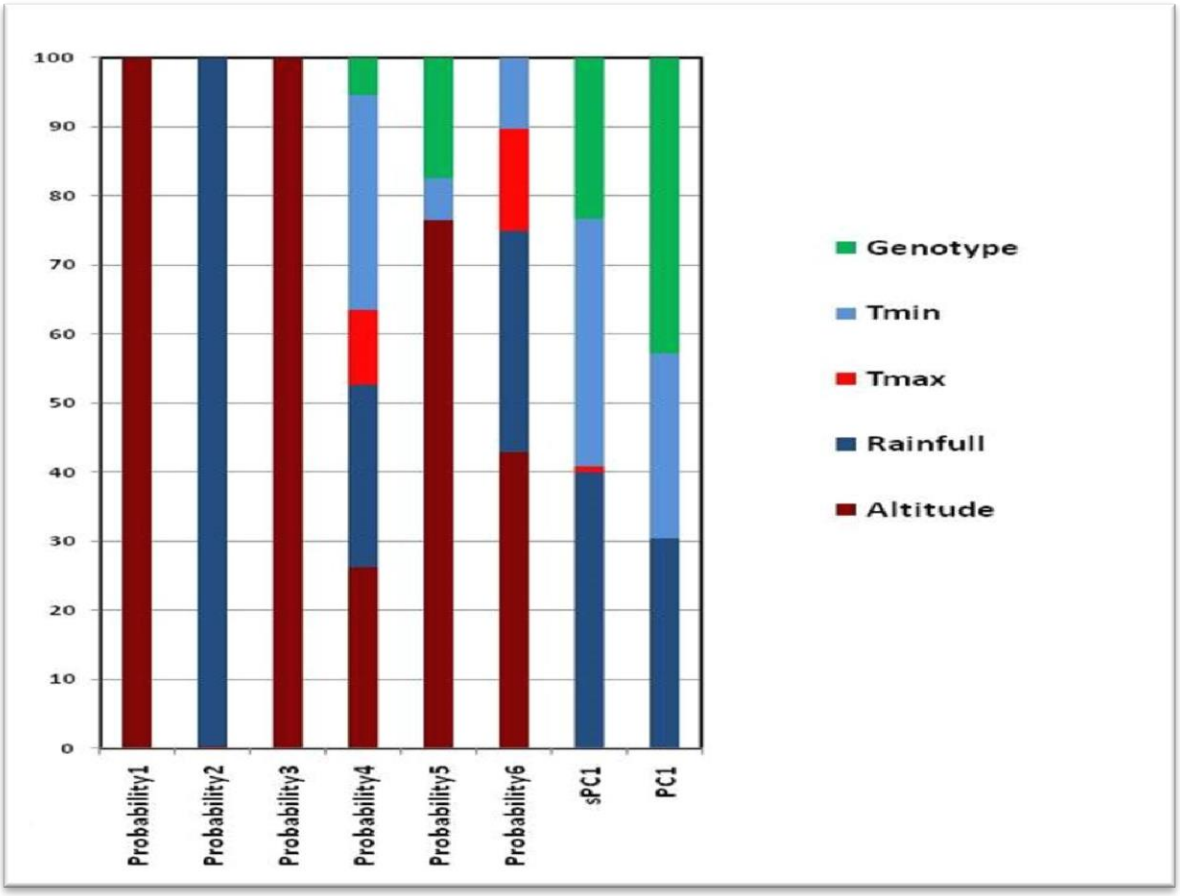


Figure 56: Population assignment probabilities, the first spatial and non-spatial principal explained by the genotype and some environmental factors for Moroccan durum landraces

4. 4. GIS user interface for durum landraces evaluation

The graphic user interface was developed using visual Basic for applications VBA under ArcGIS. It includes two main menus: trait analysis and marker analysis (Figure 57). The GUI uses directly data stored in shape files. Some outputs are stored in the shape file itself or presented in different outputs forms: window, text or Excel file.

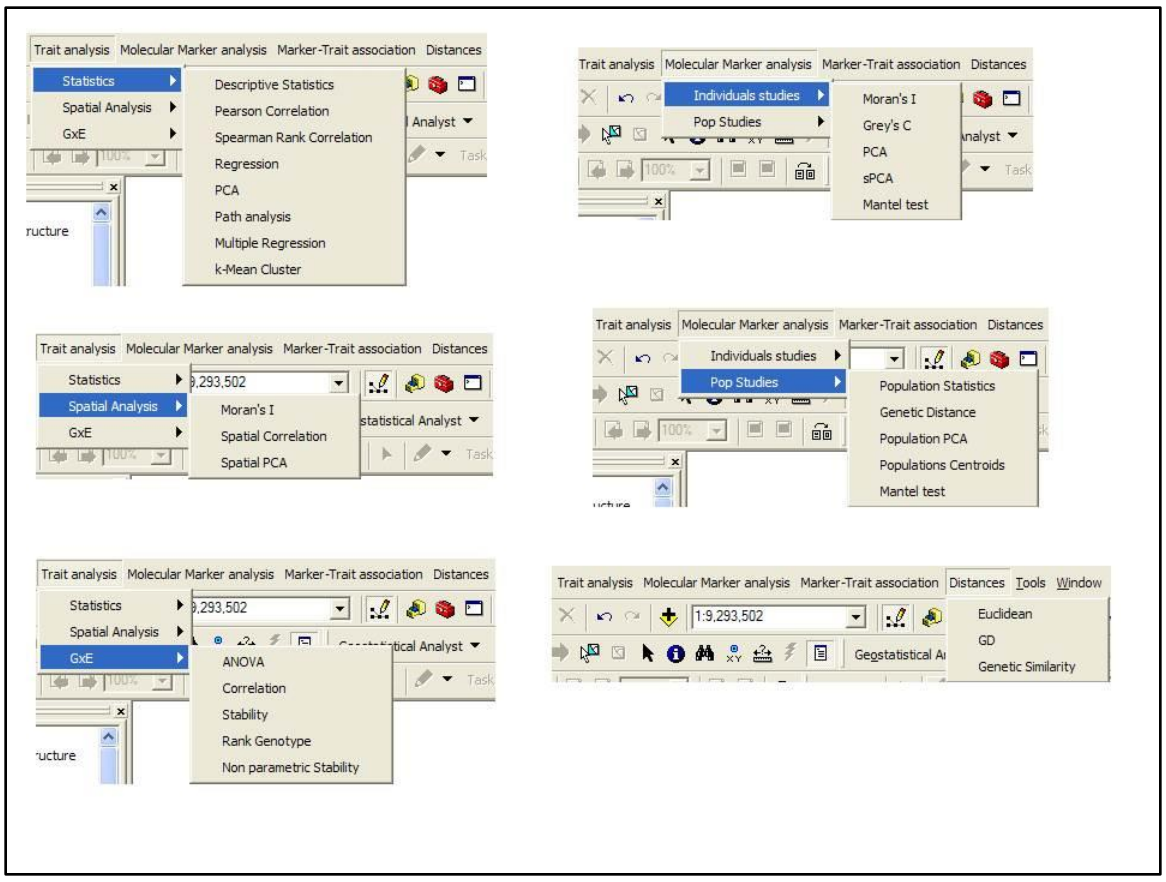


Figure 57: ArcGIS interface for durum wheat landraces evaluation

4. 4. 1. Trait analysis

This first module in the Trait analysis menu gives all descriptive statistics of a trait. The output is an ArcGIS window containing mean, average, standard deviation, variance and skewness (Figure 58).

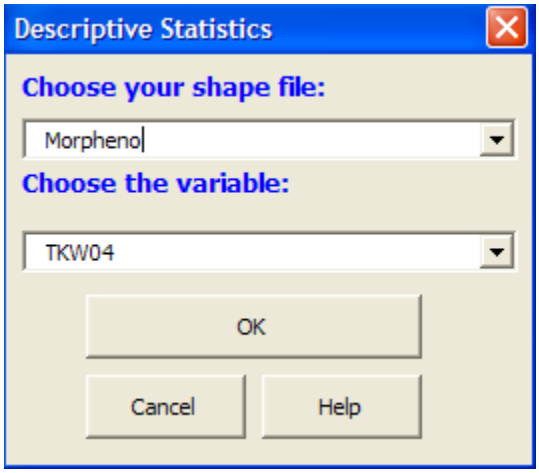


Figure 58: Descriptive statistics program's window

Two correlations are possible between two traits: Pearson (Figure 59) and Spearman (Figure 60) correlations. Outputs from the two programs are displayed in an ArcGIS notice window.

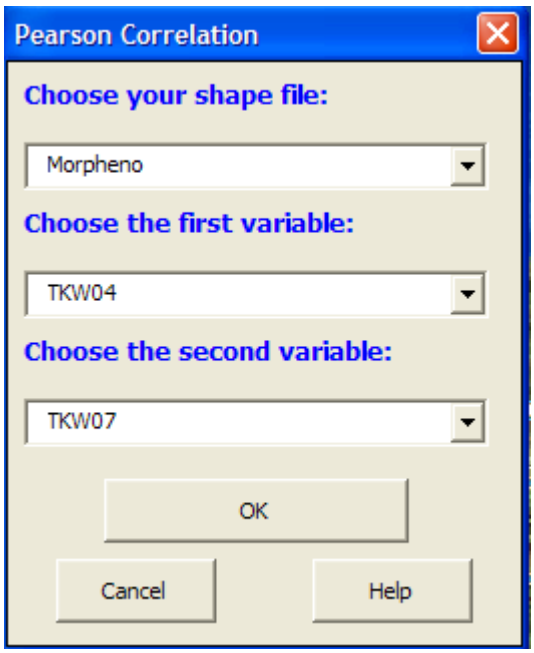


Figure 59: Pearson correlation program's window

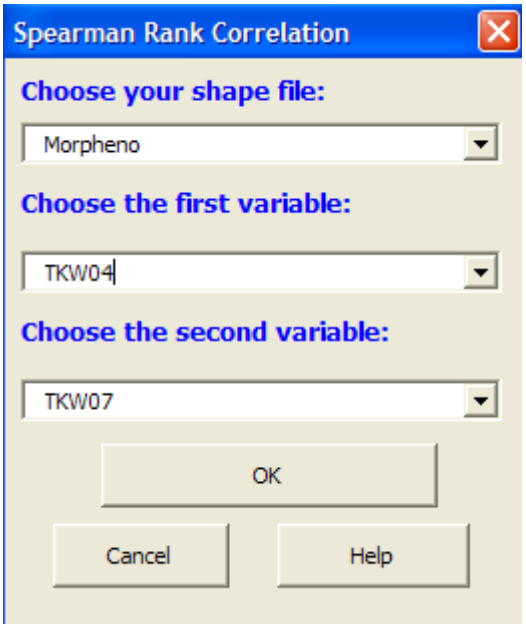


Figure 60: Spearman correlation program's window

The regression module was not developed during this study but was downloaded from (http://arcscripts.esri.com/details.asp?dbid=12405). This program allows the computation of simple linear regression (bivariate) between two numeric attributes (Figure 61). The program provides: complete set of statistics in including calculated t-values and p-values for slope and intercept and correlation coefficient (Table 30), Four new fields are added in the shape file table that contain the estimated best fit line [Fit], upper [HIGH95] and lower [LOW95] confidence intervals and calculated residuals [RESIDUALS] (Table 30)
, and Optional automatic generation residual diagnostic plot (Figure 62).

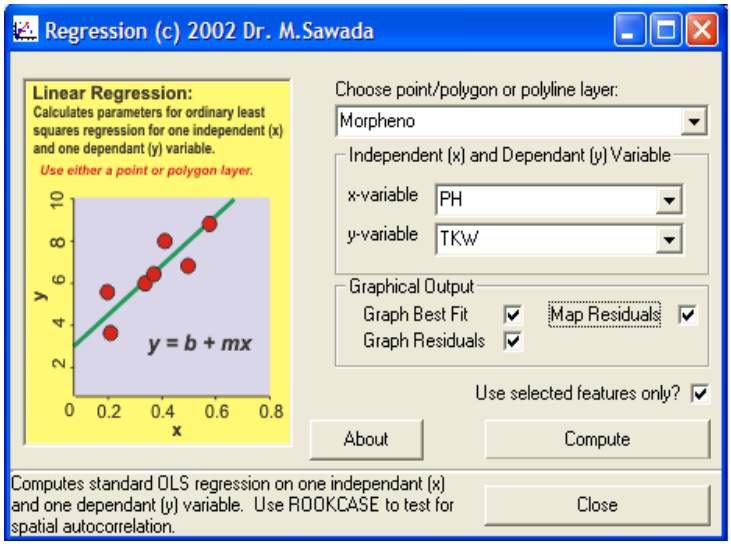


Figure 61: Linear regression program's window

Table 30: . Linear regression output (Table of regression parameters)

Attributes of ROOKOUT1

OID	Parametr	Value
0	SLOPE (BETA)	0.216035
1	Student's t-value of Beta	3.595377
2	Standard Error of Beta	0.060087
3	p-value Beta [Ho b=0]	0.000509
4	Y INTERCEPT	17.947725
5	Student's t-value of Intercept	3.03551
6	Standard Error of Intercept	5.912589
7	p-value Intercept [Ho b=0]	0.003076
8	CORRELATION (R)	0.341371
9	Student's t-value of R	3.595377
10	p-value R [Ho = 0; df = n-1]	0.000509
11	R-squared	0.116534
12	Standard Error of XY	4.985683
13	Number of Observations (n)	100
14	Mean X-Variable	98.05
15	Mean Y-Variable	39.13
16	Sum of x-values	9805
17	Sum of x-values squared (x^2)	968265

Record: 6 Show: All

Table 31: Linear regression output (Table of regression parameters)

Attributes of Morpheno

Residuals	LOW95	FIT	HIGH95	mean	cvfk
-4.014525	32.567419	35.014525	37.461631	9535.5556	118.276
-2.174879	35.728889	37.174879	38.620869	8694.4444	114.4181
-2.063517	39.626186	41.063517	42.500849	7855.5556	110.7792
-4.174879	35.728889	37.174879	38.620869	11450	121.2102
-4.174879	35.728889	37.174879	38.620869	12291.1111	124.2161
0.368554	39.356747	40.631446	41.906146	12750	123.1453
0.689263	34.489478	36.310737	38.131997	14324.4444	116.3187
0.096837	37.918176	38.903163	39.88815	14523.3333	118.9627
2.76944	32.890954	35.23056	37.570166	12668.8889	117.7249
-6.631446	39.356747	40.631446	41.906146	10743.3333	120.2343
-7.631446	39.356747	40.631446	41.906146	5797.7778	103.5246
-0.767305	38.730188	39.767305	40.804421	5201.1111	104.9416
-2.335234	38.351656	39.335234	40.318812	8253.3333	118.5564
-7.903163	37.918176	38.903163	39.88815	11858.8889	120.1197
-2.631446	39.356747	40.631446	41.906146	4872.2222	105.5169
-8.199375	39.061502	40.199375	41.337249	11466.6667	115.5318
-7.767305	38.730188	39.767305	40.804421	11181.1111	119.9056
-9.335234	38.351656	39.335234	40.318812	9586.6667	114.2177
-9.039021	36.89507	38.039021	39.182973	8215.5556	105.5726
-2.199375	39.061502	40.199375	41.337249	6481.1111	104.8584
-6.310737	34.489478	36.310737	38.131997	8534.4444	108.4604
-9.174879	35.728889	37.174879	38.620869	8915.5556	108.0413
2.985475	32.567419	35.014525	37.461631	8471.1111	110.5085
-6.335234	38.351656	39.335234	40.318812	9895.5556	117.0689
3.257192	35.1154	36.742808	38.370217	9197.7778	104.3515
1.504412	39.877582	41.495588	43.113594	8591.1111	113.2696
-3.039021	36.89507	38.039021	39.182973	9988.8889	118.1517
-5.471092	37.429971	38.471092	39.512213	15800	126.5299

Record: 0 Show: All Selected Records

The outputs from the regression programs stored in the shape file are potential variables for mapping and understanding the spatial explanation of relationship between two traits. The mapping of residuals gives an idea geographical spots where the regression model can clearly explain the relation or where the model is weak.

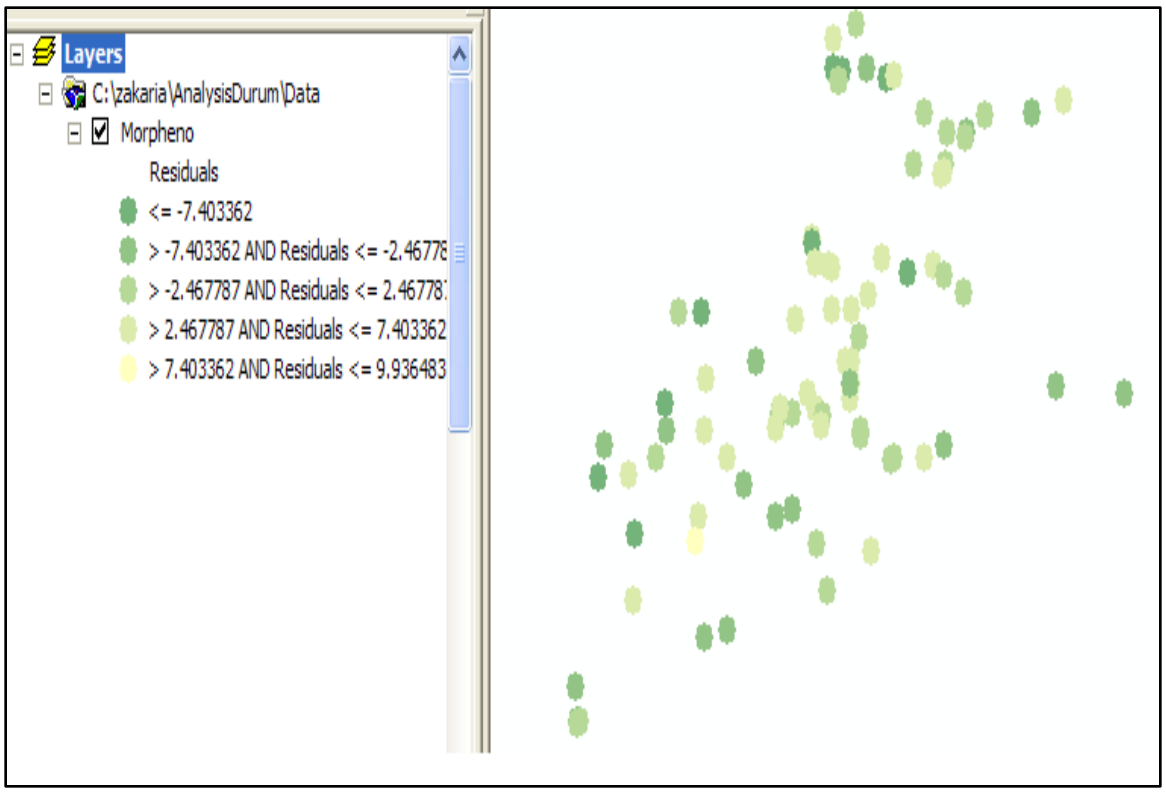


Figure 62: Linear regression output (map of residuals)

The multiple regression program models a dependent variable into a linear equation using a set of independent ones. Coefficients, means, standard deviations and standardized coefficients are given in an output under the form of Excel file. Residuals can also be computed and used later in mapping.

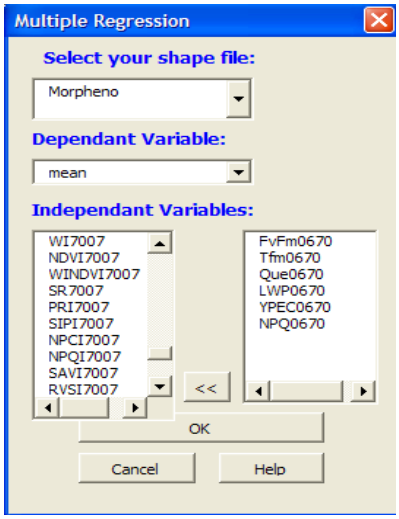


Figure 63: Multiple regression program's window

Table 32: Multiple regression output

	Parameters	Coefficients	Mean	Standard Deviation	Normalized Coeff
Dependent	**GY04RF**		2892.69697	681.2997682	
Independent	**DH04RF**	-50.81105352	147.7070707	5.934922405	-0.442624046
Independent	**PRO04RF**	-111.0847057	13.10909091	0.734778063	-0.119804246
Independent	**SDS04RF**	-136.592597	23.87878788	3.476953225	-0.697088261
Independent	**SDSI04RF**	1336.603772	1.834343434	0.26626763	0.522375516
Independent	**SDSN04RF**	-13.57803793	3.103030303	0.51703519	-0.010304309
Independent	**VIT04RF**	13.87672833	90.41414141	7.205962079	0.146771191
Independent	**YP04RF**	-82.16510999	4.875757576	0.835407942	-0.100750637
Independent	**TKW04RF**	28.47499842	46.46565657	6.465055576	0.270207706
Independent	**ASH04RF**	281.367679	2.834343434	0.094584755	0.039062237
	Constant	-4550.632696			

Path program (Figure 64) of a trait using a set of other traits computes correlation (Table 33) between all traits and divides this correlations to a matrix of direct and indirect effects (Table 34) of one trait via another one.

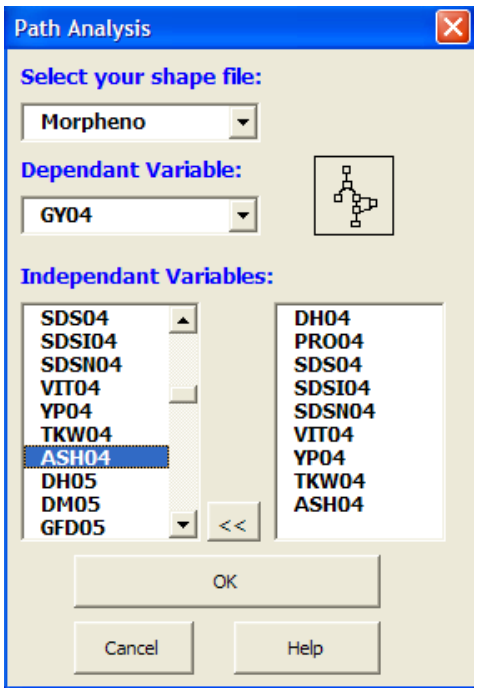


Figure 64: Path analysis program's windows

Table 33: Path analysis output (correlation matrix and variance)

	GY	ASH	PC	PH	DH	DM	GFD	SDS
GY	1.000	-0.250	-0.219	-0.346	-0.372	-0.251	0.124	0.084
ASH	-0.250	1.000	0.585	0.451	0.203	0.049	-0.272	0.035
PC	-0.219	0.585	1.000	0.652	0.497	0.408	-0.224	-0.066
PH	-0.346	0.451	0.652	1.000	0.647	0.551	-0.135	0.264
DH	-0.372	0.203	0.497	0.647	1.000	0.881	0.014	-0.184
DM	-0.251	0.049	0.408	0.551	0.881	1.000	0.304	-0.097
GFD	0.124	-0.272	-0.224	-0.135	0.014	0.304	1.000	0.165
SDS	0.084	0.035	-0.066	0.264	-0.184	-0.097	0.165	1.000
SDSn	0.044	0.196	0.151	0.388	-0.109	-0.048	0.099	0.972
SDSi	0.132	-0.143	-0.293	0.113	-0.258	-0.142	0.232	0.968
YP	0.119	0.413	0.082	0.105	0.135	0.001	0.010	0.068
TKW	-0.065	-0.219	0.363	0.381	0.550	0.669	0.020	-0.112
VIT	-0.037	0.424	0.447	0.442	0.211	0.209	-0.237	0.321
SL	-0.417	-0.226	0.170	0.223	0.397	0.420	0.043	-0.184
SPM2	0.094	0.207	-0.106	-0.125	-0.362	-0.357	-0.050	0.172
PL	0.223	-0.140	-0.256	-0.262	-0.379	-0.410	-0.075	0.024
KSPK	0.225	-0.057	-0.390	-0.467	-0.656	-0.726	-0.063	0.112
	GY	**ASH**	**PC**	**PH**	**DH**	**DM**	**GFD**	**SDS**
VAR	4469.395	0.000	0.109	35.539	7.897	10.163	1.617	4.866

Table 34: Path analysis output (path coefficients)

	ASH	PC	PH	DH	DM	GFD	SDS	SDSn	SD
ASH	**-0.174**	-0.102	-0.079	-0.035	-0.009	0.047	-0.006	-0.034	
PC	0.109	**0.187**	0.122	0.093	0.076	-0.042	-0.012	0.028	
PH	-0.085	-0.123	**-0.189**	-0.122	-0.104	0.025	-0.050	-0.073	
DH	-0.071	-0.175	-0.228	**-0.352**	-0.310	-0.005	0.065	0.038	
DM	0.003	0.021	0.029	0.046	**0.052**	0.016	-0.005	-0.003	
GFD	-0.015	-0.013	-0.008	0.001	0.017	**0.056**	0.009	0.006	
SDS	0.145	-0.273	1.092	-0.759	-0.400	0.681	**4.131**	4.014	
SDSn	-0.428	-0.328	-0.845	0.237	0.105	-0.215	-2.118	**-2.180**	
SDSi	0.285	0.584	-0.225	0.514	0.284	-0.463	-1.930	-1.759	
YP	0.051	0.010	0.013	0.017	0.000	0.001	0.008	0.013	
TKW	-0.115	0.190	0.200	0.288	0.351	0.011	-0.059	-0.036	
VIT	-0.076	-0.080	-0.079	-0.038	-0.037	0.042	-0.058	-0.076	
SL	0.131	-0.098	-0.129	-0.230	-0.243	-0.025	0.107	0.106	
SPM2	0.002	-0.001	-0.001	-0.004	-0.004	-0.001	0.002	0.002	
PL	-0.012	-0.022	-0.022	-0.032	-0.035	-0.006	0.002	-0.002	
KSPK	0.000	0.003	0.004	0.005	0.006	0.000	-0.001	0.000	

K-mean clustering (Figure 65) programs groups, using a set of traits, individuals or landraces into groups or clusters. The number of cluster needs to be defined by the user. A variable (Table 35) containing the cluster to which each landrace is affected is added to the shape file table. This output can also be used directly to map cluster affectations of landraces and detect visually if this cluster affectation presents any spatial pattern.

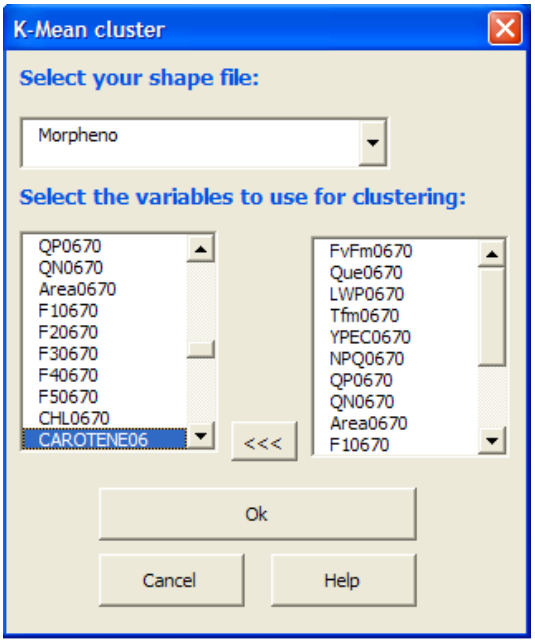


Figure 65: k-mean clustering program's window

Table 35: k-mean clustering output (field of groups added to shape file)

kmeangroup
0
5
0
1
4
5
1
1
1
0
3
2
0
5
1
5
0
0
0
3
1
1
0
0
5
0
0
0

Stability program (Figure 66) is using a shapefile containing a trait measured at different locations or/and during different years. It computes the stability of a landrace. The computed stability parameters (two in our case) can be mapped to dissect spatial pattern of landraces plasticity and further be used to compare such stability with geographic pattern of climate. The outputs from analysis (mean trait, CVFK and WR) are stored in the original shape file (Table 36).

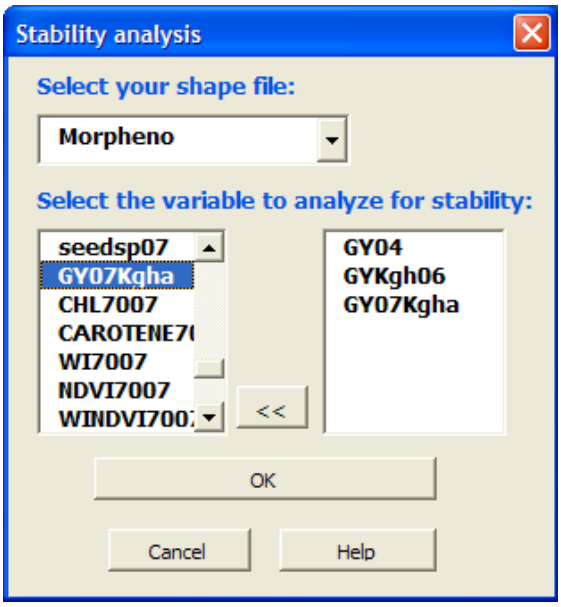


Figure 66: Stability analysis program's window

Table 36: Stability analysis output (Field for mean, CV and WR added to shape file)

Attributes of Morpheno

site	Longdec	Latdec	alt	mean	cvfk	wr	gha	awi
Guelmine	-10.066667	28.933333	200	9535.5556	118.276	429384937.6293	1	
NearAsrir	-10	28.928056	200	8694.4444	114.4181	341287470.3896	1	
5kmSofGuelmine	-10.066667	28.916667	200	7855.5556	110.7792	275200634.3991	1	
Tatamainoasis	-8	29.766667	700	11450	121.2102	662041537.3925	1	
Kasba-ej-Joua,villagestack	-7.633333	29.833333	600	12291.1111	124.2161	807747593.478	1	
Tanskit	-6.200556	30.6925	850	12750	123.1453	874178970.9576	1	
JustWofTinejdad	-5.015	31.515	900	14324.4444	116.3187	1001929057.2076	1	
Mellah	-6.814167	31.98	800	14523.3333	118.9627	1102976480.9881	1	
Fezna	-4.466667	31.533333	740	12668.8889	117.7249	906928732.6485	1	
3kmSofAoufouss;outsidemainoasis	-4.166667	31.65	750	10743.3333	120.2343	781710455.2759	1	
5kmEofBoulaouane	-8.05	32.983333	150	5797.7778	103.5246	414208834.0349	1	
15kmWofSidiBennour	-8.416667	32.983333	100	5201.1111	104.9416	450089056.8147	1	
Akermould	-9.616667	31.666667	1	8253.3333	118.5564	730297596.3323	1	
Ain-el-Hajer,nearAkermoud	-9.616667	31.666667	1	11858.8889	120.1197	1141926141.4788	1	
Ain-el-Hajer,nearAkermoud	-9.616667	31.666667	1	4872.2222	105.5169	647126120.5331	1	
2kmEofSmimou	-9.125833	30.784722	300	11466.6667	115.5318	1212606375.9258	1	

PCA program can be run on both landraces and traits. Outputs from both analyses can be plotted to study the relationships between traits or landraces (Figure). The program is giving the Eigen values tables and the principal coordinates.

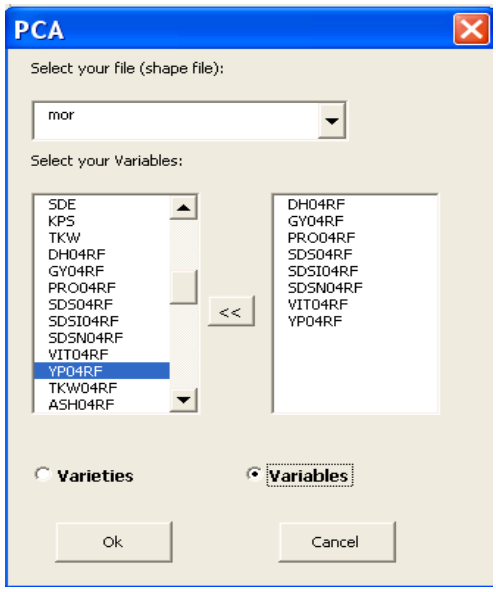


Figure 67: PCA analysis program's window

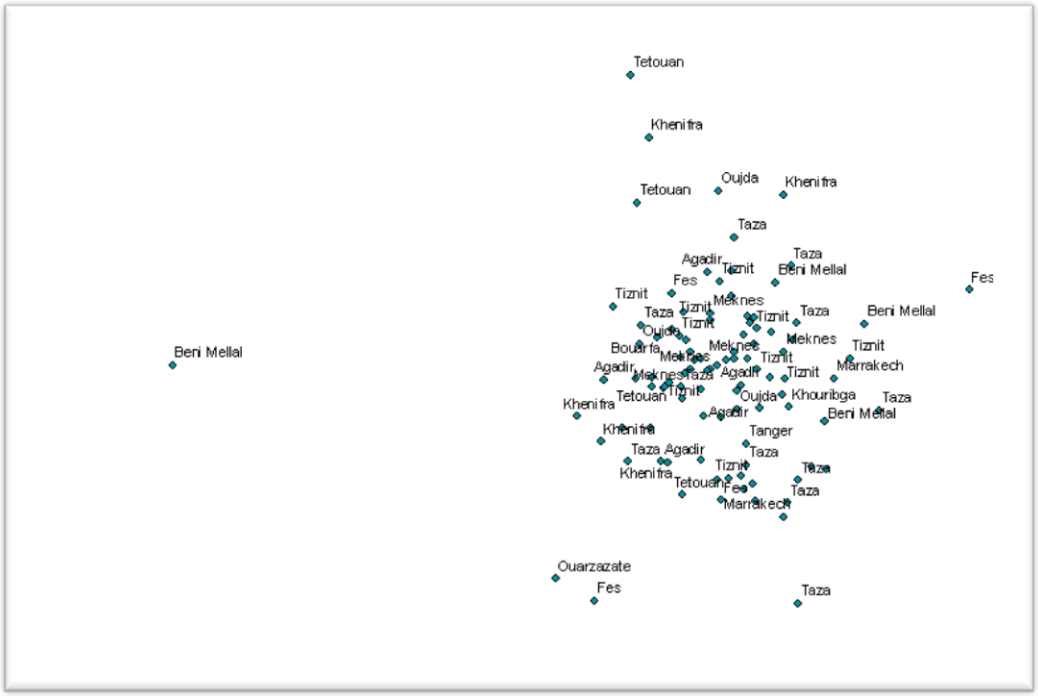


Figure 68: PCA analysis program's output

4. 4. 2. Marker analysis

Markers analysis is very useful tools for analyzing landraces diversity and structure. Different tools were developed in this context. The program of allele frequencies and PCA (Figure 69) on individuals is using a shape file with markers data; it is computing different parameters such as number of alleles and allele's frequencies to be used later in the principal components analysis PCA to study the genetic variability and structure of landraces.

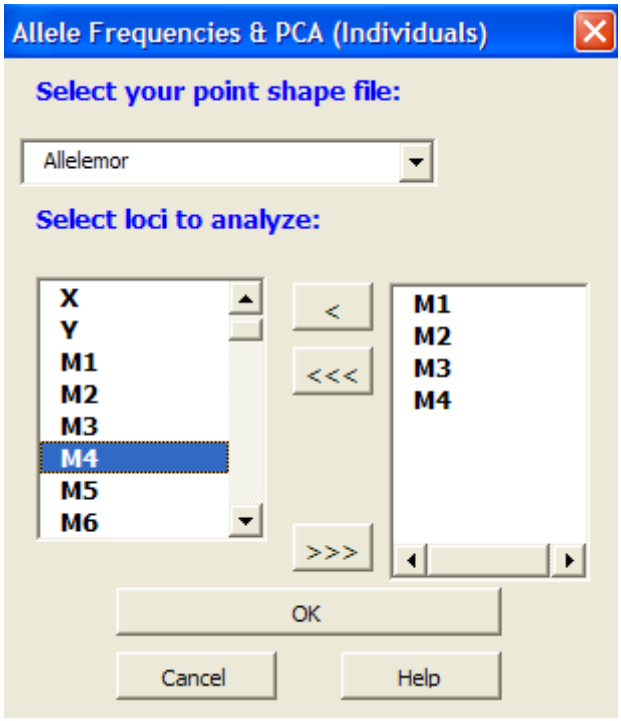


Figure 69: Allele frequencies and PCA program's window on individuals

Different distance methods were implemented under the distance program (Figure). Most of distances methods are using coordinate (Euclidean, Gamma, Pearson), only one method is using molecular markers (Genetic Nei's similarity). Genetic similarity can be used to further study spatial genetic structure. In the example bellow (Figure), one can compute genetic similarity between landraces and construct a map for each similarity with a particular landrace. Assuming that neighboring landraces tend to have similar genetic profile, if the map presents higher genetic similarities (Dark color) near the studied landrace (Highlighted landrace), the hypothesis is validated. If on the other hands, lower similarities are found near the studies landraces, the hypothesis is not valid or the landrace was not originated from the region or adapted to a given micro-climate.

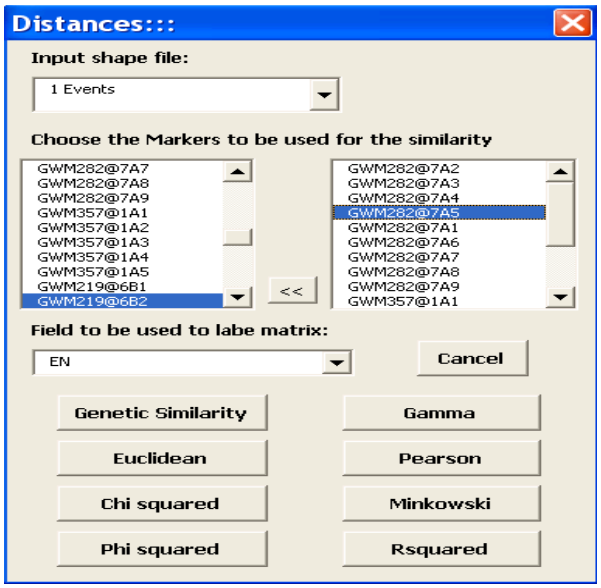
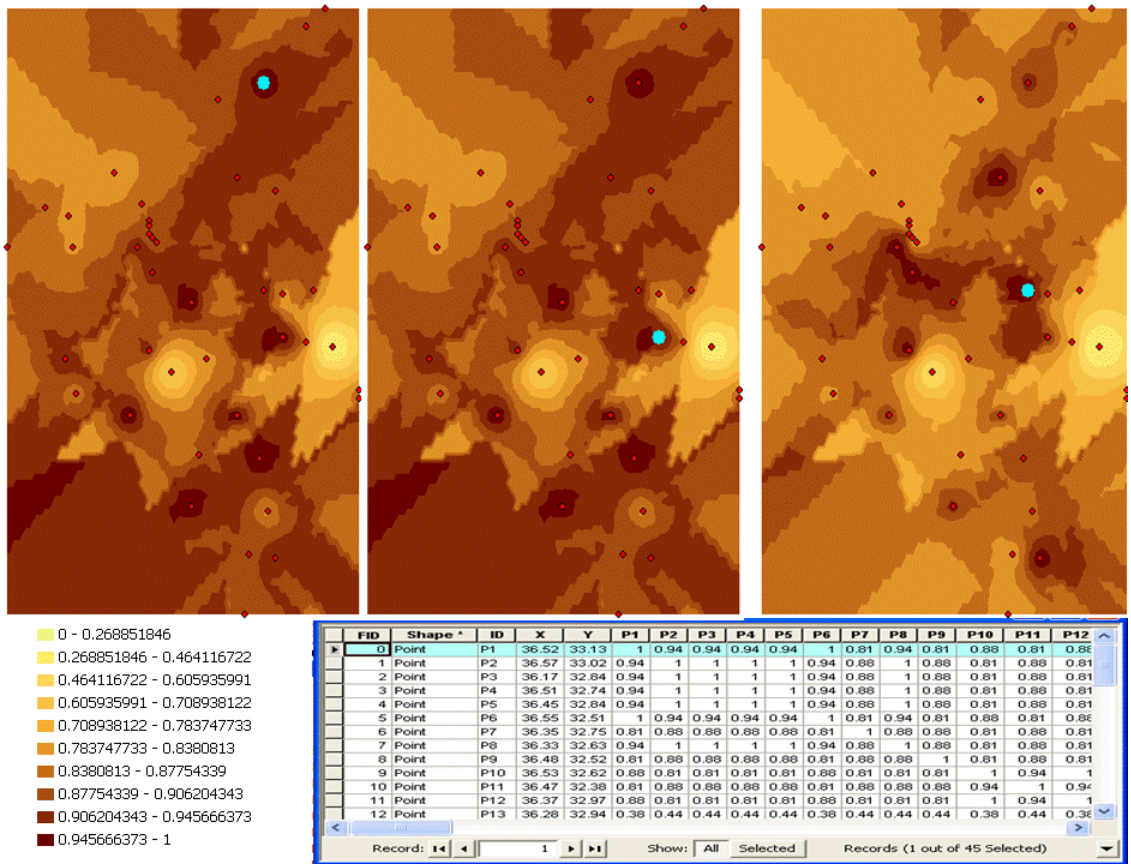


Figure 70: Distance calculation program's window

Table 37: Genetic similarity output

Haurani nawawi	1	0.874	0.874	0.84	0.874	0.857	0.782	0.866	0.84	0.874	0.832	0.866	0.79	0.908	0.849	0.857
Haurani 27		1	0.975	0.824	0.966	0.866	0.815	0.866	0.84	0.84	0.824	0.832	0.79	0.891	0.95	0.916
Normal haurani			1	0.824	0.966	0.866	0.815	0.866	0.84	0.84	0.824	0.832	0.79	0.891	0.941	0.916
Hamari ahmar				1	0.832	0.807	0.756	0.824	0.79	0.782	0.773	0.79	0.748	0.832	0.824	0.832
Akbash					1	0.882	0.832	0.874	0.857	0.857	0.84	0.849	0.798	0.908	0.958	0.933
Kishk						1	0.798	0.899	0.857	0.832	0.849	0.84	0.824	0.899	0.874	0.849
Baladia hamra (A)							1	0.79	0.824	0.798	0.807	0.807	0.731	0.824	0.832	0.815
Hedba 3								1	0.849	0.857	0.84	0.857	0.899	0.908	0.849	0.84
Oued zenati 368									1	0.874	0.84	0.874	0.79	0.857	0.84	0.824
Romanou 2										1	0.882	0.941	0.798	0.874	0.832	0.824
Mavragani-Iraklion											1	0.891	0.782	0.857	0.824	0.807
Moundros-2												1	0.798	0.866	0.84	0.815
Atsiki-3													1	0.832	0.773	0.765
Local Iraklion														1	0.899	0.891
Tripolino															1	0.924
Scorsonera																1



FID	Shape *	ID	X	Y	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
0	Point	P1	36.52	33.13	1	0.94	0.94	0.94	0.94	1	0.81	0.94	0.81	0.88	0.81	0.88
1	Point	P2	36.57	33.02	0.94	1	1	1	1	0.94	0.88	1	0.88	0.81	0.88	0.81
2	Point	P3	36.17	32.84	0.94	1	1	1	1	0.94	0.88	1	0.88	0.81	0.88	0.81
3	Point	P4	36.51	32.74	0.94	1	1	1	1	0.94	0.88	1	0.88	0.81	0.88	0.81
4	Point	P5	36.45	32.84	0.94	1	1	1	1	0.94	0.88	1	0.88	0.81	0.88	0.81
5	Point	P6	36.55	32.51	1	0.94	0.94	0.94	0.94	1	0.81	0.94	0.81	0.88	0.81	0.88
6	Point	P7	36.35	32.75	0.81	0.88	0.88	0.88	0.88	0.81	1	0.88	0.88	0.81	0.88	0.81
7	Point	P8	36.33	32.63	0.94	1	1	1	1	0.94	0.88	1	0.88	0.81	0.88	0.81
8	Point	P9	36.48	32.52	0.81	0.88	0.88	0.88	0.88	0.81	0.88	0.88	1	0.81	0.88	0.81
9	Point	P10	36.53	32.62	0.88	0.81	0.81	0.81	0.81	0.88	0.81	0.81	0.81	1	0.94	1
10	Point	P11	36.47	32.38	0.81	0.88	0.88	0.88	0.88	0.81	0.88	0.88	0.88	0.94	1	0.94
11	Point	P12	36.37	32.97	0.88	0.81	0.81	0.81	0.81	0.88	0.81	0.81	0.81	1	0.94	1
12	Point	P13	36.28	32.94	0.38	0.44	0.44	0.44	0.44	0.38	0.44	0.44	0.44	0.38	0.44	0.38

Record: 1 Show: All Selected Records (1 out of 45 Selected)

Figure 71: Genetic similarity maps

For populations, the module is giving several outputs: number of alleles per locus, number of alleles per locus per population, allele's frequencies, heterozygosities (observed and expected), number of effective alleles, PIC, F-stat and genetic distance between populations (4 methods). A file is constructed to run a PCA and spatial PCA analysis to study to genetic structure.

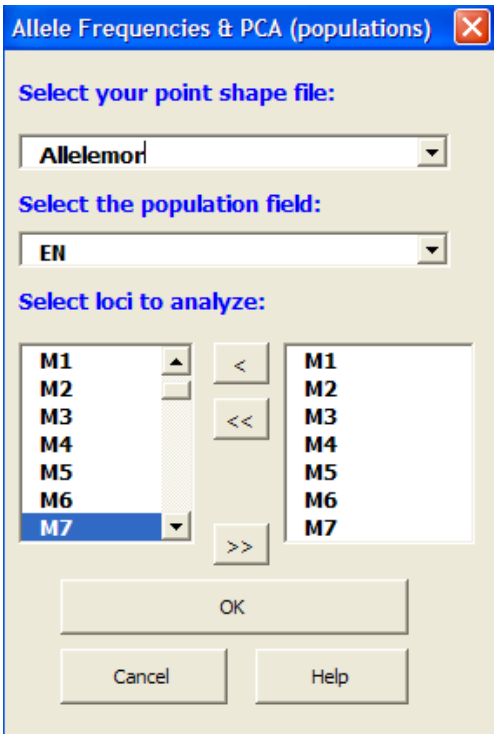


Figure 72: Allele frequencies and PCA program's window on populations

Table 38: Excel output file from PCA & AF analysis (overall view)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	114	114			118	118	218	218	270	270	130	130	187	187	121	121	222	222	180	180
1	114	114	197	206	123	175	217	217	271	271	109	109	187	187	121	121	224	224	181	205
1	117	117	193	193			220	220	270	270	129	129	188	188	120	120	223	223	180	180
1	117	117	194	194	121	121	219	219	267	267	129	129	187	187	120	120	223	223	180	180
1	118	118	194	194	121	121	219	219	268	268	129	129	187	187	121	121	223	223	179	179
1	117	117	194	194	121	121	219	219	267	267	109	133	187	187	121	121	223	223	180	180
1	118	118	195	195	121	121	218	218	270	270	129	129	187	187	121	121	223	223	181	181
1	118	118	196	196	121	121	219	219	268	268			188	188	121	121	223	223		
1	121	121	195	195	121	121	219	219	268	268	118	129	188	188	120	120	223	223	180	180
1	120	120			107	117	217	217			119	119	188	188			224	224		
1	119	119			121	121	221	221					187	187	121	121	224	224	191	209
1	117	117			121	121	219	219			128	128	187	187	121	121	223	223	172	195
1	119	119			121	173	219	219	268	268			187	187	121	121	222	222	182	207
1	118	118	197	197	121	121	219	219	270	270			187	187	121	121	223	223	180	180
2	114	114	196	196	121	121	219	219	270	270	132	132	187	187	119	119	222	222	181	205
2	121	121	206	206	119	119	218	218	270	270	131	131	187	187	121	121	224	224	177	189
2	117	117	193	206	121	154	218	218	270	270	128	128	187	187	121	121	224	224	179	179
2	115	115	209	209	155	155	221	221			115	115	188	188	120	120	224	224	178	178
2	115	115	206	206	105	155	218	218	271	271	133	133	187	187	120	120	223	223	178	178
2	118	118	207	207	155	155	216	216	268	268	131	131	187	187	121	121	222	222	179	179
2	118	118	195	195	158	158	213	213	268	268	133	133	187	187	121	121	222	222	178	178
2	118	118	196	209	175	175	216	216	271	271	115	131	187	187	115	121	224	224	177	177
2	118	118					216	216	271	271	112	112	187	187	121	121	224	224	178	178
2	117	117	195	195	155	155	218	218	268	268	117	117	187	187	121	121	224	224	178	178
3	112	112			122	122	219	219	270	270	115	115	187	187	120	120	223	223	190	190
3	114	114	205	205					271	271	131	131	187	187					181	181
3	114	114			118	118			271	271	132	132	187	187	118	118	226	226	179	179
3	114	114	197	197	119	119	216	216					187	187	127	127	228	228	178	198
3	114	114	208	208	118	163			256	269	130	130	188	188	119	119			180	180
3	114	114	199	199	119	119			270	270	115	115	187	187	121	121	237	237	178	178
3	114	114			122	122	217	217	271	271					132	132	223	223	179	179
3	114	114			121	121			268	268	110	110	187	187	121	121	222	222	181	181

Alleles per pop | Number of Allele Loci | PCA Matrix | Genetic distances | Allele frequencies | Data | POP Statistics

Table 39: Excel output file from PCA & AF analysis (Alleles and number of alleles per loci)

Marker	Nbre of Alleles	Alleles								
1	9	112	114	115	116	117	118	119	120	1
2	17	187	191	193	194	195	196	197	198	1
3	20	105	107	117	118	119	120	121	122	1
4	12	205	206	207	213	215	216	217	218	2
5	7	255	256	267	268	269	270	271		
6	21	104	105	106	107	108	109	110	112	1
7	4	186	187	188	191					
8	9	115	117	118	119	120	121	127	130	1
9	9	221	222	223	224	225	226	228	235	2
10	21	172	177	178	179	180	181	182	183	1

Table 40: Excel output file from PCA & AF analysis (number of allele per locus and per population)

Pop/Locus						
1	**pop1**	**pop2**	**pop3**	**pop4**	**pop5**	**pop6**
2	6	6	9	7	3	3
3	7	7	11	6	5	2
4	5	5	5	7	4	4
5	4	3	6	7	4	1
6	7	7	11	11	3	4
7	2	2	3	3	2	2
8	2	4	7	4	2	2
9	3	3	7	4	5	3
10	10	6	13	6	5	5

Table 41: Excel output file from PCA & AF analysis (locus information per population)

PoP id	Locus id	Allele id	Allele count	all freq	GD/He	Ho	F	r	Ne	PIC
1	**1**	**114**	4	0.142857	0.785714	0	1	0.44	4.666667	0.18258
1	**1**	**117**	8	0.285714						
1	**1**	**118**	8	0.285714						
1	**1**	**119**	4	0.142857						
1	**1**	**120**	2	0.071429						
1	**1**	**121**	2	0.071429						
1	**2**	**193**	2	0.071429	0.910714	0.071429	0.921569	0.439252	11.2	0.084024
1	**2**	**194**	6	0.214286						
1	**2**	**195**	4	0.142857						
1	**2**	**196**	2	0.071429						
1	**2**	**197**	3	0.107143						
1	**2**	**206**	1	0.035714						
1	**3**	**107**	1	0.035714	0.528061	0.214286	0.594203	0.205342	2.118919	0.461269
1	**3**	**117**	1	0.035714						
1	**3**	**118**	2	0.071429						
1	**3**	**121**	19	0.678571						
1	**3**	**123**	1	0.035714						
1	**3**	**173**	1	0.035714						
1	**3**	**175**	1	0.035714						
1	**4**	**217**	4	0.142857	0.622449	0	1	0.383648	2.648649	0.342514
1	**4**	**218**	4	0.142857						
1	**4**	**219**	16	0.571429						
1	**4**	**220**	2	0.071429						
1	**4**	**221**	2	0.071429						

Table 42: Excel output file from PCA & AF analysis (Genetic distances)

		pop2	pop3	pop4	pop5	pop6
Prevosti and al., 1975	**pop1**	0.479286	0.457143	0.544549	0.518571	0.56
	pop2		0.453333	0.58	0.58	0.565
	pop3			0.496491	0.485556	0.448333
	pop4				0.530526	0.474474
	pop5					0.56
LATTER'S FST distance - (Latter, 1972)	**pop1**	0.156781	0.118631	0.197995	0.166911	0.19637
	pop2		0.096488	0.182048	0.161506	0.174969
	pop3			0.122902	0.103613	0.0907

	pop4				0.156905	0.15952
	pop5					0.174224
Nei minimum genetic distance - (Nei, 1973)	**pop1**	0.123633	0.098376	0.162611	0.139133	0.162847
	pop2		0.077633	0.145722	0.133	0.1405
	pop3			0.100985	0.087438	0.07266
	pop4				0.127341	0.124551
	pop5					0.146
DL-Latter, 1972 (-ln (1-Fst))	**pop1**	0.170528	0.126279	0.22064	0.182615	0.218616
	pop2		0.101466	0.200952	0.176148	0.192334
	pop3			0.131136	0.109384	0.09508
	pop4				0.170676	0.173782
	pop5					0.191432

Table 43: Excel output file from PCA & AF analysis (Allele frequencies, PCA input)

Loci	Allele	1	2	3	4	5	6	MEAN	SDEV
1	**112**	0	0	0.055556	0	0	0	0.000617	0.022436
1	**114**	0.142857	0.1	0.444444	0.052632	0.2	0.4	0.014888	0.25565
1	**115**	0	0.2	0.055556	0	0	0	0.00284	0.08335
1	**116**	0	0	0.111111	0	0	0	0.001235	0.044871
1	**117**	0.285714	0.2	0.055556	0.184211	0	0	0.008061	0.156709
1	**118**	0.285714	0.4	0.166667	0.578947	0.2	0.4	0.02257	0.345435
1	**119**	0.142857	0	0.111111	0.078947	0.2	0	0.005921	0.110216
1	**120**	0.071429	0	0	0.052632	0.2	0.2	0.005823	0.116885
1	**121**	0.071429	0.1	0	0.026316	0.2	0	0.004419	0.093449
2	**187**	0	0	0.055556	0.013158	0.2	0	0.002986	0.083375
2	**191**	0	0	0	0	0	0.1	0.001111	0.040384
2	**193**	0.071429	0.05	0.055556	0	0	0	0.001966	0.040857
2	**194**	0.214286	0	0	0	0	0	0.002381	0.086537
2	**195**	0.142857	0.2	0.111111	0	0	0	0.005044	0.106714
2	**196**	0.071429	0.15	0	0.052632	0	0.2	0.005267	0.104518
2	**197**	0.107143	0	0.055556	0.184211	0	0	0.003855	0.087478
2	**198**	0	0	0.111111	0.302632	0	0.3	0.00793	0.174637
2	**199**	0	0	0.055556	0.210526	0	0	0.002956	0.087452
2	**200**	0	0	0	0.078947	0	0	0.000877	0.031882
2	**201**	0	0	0	0.026316	0.1	0	0.001404	0.041533
2	**205**	0	0	0.055556	0	0	0	0.000617	0.022436
2	**206**	0.035714	0.25	0.055556	0	0	0	0.003792	0.10357

The alleles frequencies output can be used in the second stage of the program analysis to run PCA. This leads to a table of Eigen values and the variance they explain (Table) and also the projected coordinates of landraces to principal axes (Table).

Table 44: Excel output file from PCA & AF analysis (Eigen values)

id	Eigenvalue	Variance
1	52.23560427	30.87136
2	11.75121629	6.944996
3	4.788937215	2.830273
4	4.042309834	2.389014
5	3.735338935	2.207594
6	3.495714066	2.065975
7	3.060215428	1.808594
8	2.91591139	1.72331
9	2.580944307	1.525344
10	2.55673493	1.511036
11	2.383630367	1.408731
12	2.31632765	1.368955

Table 45: Excel output file from PCA & AF analysis (PC scores)

axis	1	2	3	4	5
IND1	0.077	0.193	0.091	0.016	0.011
IND2	0.089	0.122	-0.009	0.052	0.084
IND3	0.041	0.031	-0.106	-0.063	0.041
IND4	0.074	0.088	-0.012	-0.045	-0.017
IND5	0.059	0.050	-0.177	-0.058	0.011
IND6	0.078	0.084	-0.084	-0.047	0.069
IND7	0.100	0.039	-0.216	-0.121	-0.041
IND8	0.102	0.032	-0.226	-0.152	-0.026
IND9	0.095	0.024	-0.170	-0.150	0.145
IND10	0.085	0.216	0.065	0.077	-0.025
IND11	0.094	0.211	0.062	0.043	-0.058
IND12	0.054	0.031	-0.143	0.035	0.082
IND13	0.095	0.014	-0.110	0.033	0.187
IND14	0.088	0.194	0.061	0.048	-0.018
IND15	0.044	0.007	-0.081	0.030	0.083

IND16	0.075	0.172	0.071	-0.041	0.064
IND17	0.083	0.205	0.083	0.007	-0.042
IND18	0.092	0.227	0.057	0.041	-0.049
IND19	0.060	0.039	-0.159	-0.020	-0.049
IND20	0.078	0.230	0.100	0.037	0.022

Conclusions

Landraces are valuable genetic resources in the hand of breeders to develop new varieties adapted to climate change, with high productivity and good quality traits. To be used in a breeding program, a landrace should be studies deeply for valuable trait or allele and also asset its adaptation. Since landraces are collected in specific geographic locations, it is a big advantage of using spatial statistics in the process of adaptation's study. Also, landraces are characterized by different traits, molecular markers and environmental variables. This makes identifying adaptation of landraces a multi-disciplinary science and also the amount of spatial and non-spatial data generated from these studies make geographic information systems necessary to store, analyze and present data generated. In this study, we evaluated Moroccan durum wheat landraces for agronomy, physiology and genetic diversity. Several conclusions could be thrived from this evaluation.

Traits showed a large diversity in Morocco specially quality traits and yield components; and physiological traits. Yield was tightly associated with Ash content and grain filling period and number of kernels and yield reached 90% of the potential yield for durum Moroccan landraces. Yield was highly linked to Chlorophyll content, ratio of Water index to normalized difference vegetation index and normalized pigment chlorophyll index of both stages (Zadoc scale 45 and 70). We could explain in this study much of the yield variation using agronomic traits. Most of traits exhibited a large genotype by environment effect except for the phenology and quality traits. We studied the long climate profiles of collection sites of Moroccan landraces. We identified in this study the climatic variables that are explaining the genotypic variation of several traits. This could be useful to identify areas of variability of a trait of interest for the Moroccan durum wheat collection.

Some traits exhibited a significant SAU across the country: dates to heading, maturity, number of kernel per spike, kernel weight, grain filling duration, the effect of environment on a landrace as a stability parameter, and plant height. Multivariate analysis using space showed no significant spatial patter for phenotypic traits. Phenology and quality traits showed a very significant high clusters across Morocco which means that landraces with high values for these traits were collected in the same geographic regions. We could also identify in this study areas of Morocco where values for specific trait are significantly low or high. Several traits presented as well clear geographic discontinuities over Morocco; and these barriers were mainly driven by environmental variables.

We studies genetic structure and diversity using two common methods (Bayesian and Eigen) and showed that similar spatial genetic patterns were found using the two approaches for the Moroccan population. The axis of the Eigen analysis differentiated clearly between clusters revealed by the Bayesian method. The Eigen analysis is easy to implement in any software, has no assumption on data, and can help in understanding diversity and structure of a given population. The resulting axes are continuous and can be used to correct phenotype trait and genotypic data for association studies (Price et al., 2006). This study showed clearly the geographic distribution of landraces in Morocco and Syria and confirmed that in general, landraces tended to group according not to their geographical origin (Moraguees et al., 2006), but also to their agro-ecological adaptation. The use of spatial genetic structure helped largely to understand the mechanisms of adaptation of durum wheat landraces; and that environment (topography, landscape) has a considerable effect on population structure (Coulon et al., 2006). We also analyzed genetic discontinuities through barriers using Monmonier's algorithm and results showed similar spatial pattern found by the other two methods. Also this genetic barrier was driven mainly by the Altitude pattern for the Moroccan country. Moroccan durum landraces showed a clear spatial pattern differentiating between

landraces originated from the mountains and oasis and landraces from lowlands. The use of explicit methods such spatial PCA, Monmonier, SAU or the spatial Bayesian genetic structure method ameliorate the precision of assessing pattern of spatially distributed phenomena like landraces diversity.

These analyses techniques, aided by marker-trait association, are a powerful tool in the hand of the breeders for deciding on the choice of the parental material in a crossing program (Castillo et al., 2010, Zarkti et al., 2010). The amplified alleles found in this study were more than twice than the durum wheat elite collection population (Maccaferri et al., 2005). This may be explained that our populations consisted of diverse landraces; whereas the mentioned previous work was mainly of improved genotypes. The genetic diversity found in the Moroccan landraces was higher than the diversity found by Moraguees and colleagues for a group of Mediterranean durum landraces using low and high molecular weight loci. Moroccan durum wheat landraces hold large genetic variability and considerable number of alleles with the probability of having some of these alleles associated with stress tolerance, yield, and/or grain quality (Nachit et al., 2004, Pagnotta et al., 2004). The spatial autocorrelation (SAU) applied to an individual allele did not express, for most of the alleles, the global spatial structure we have in our data (Smouse & Peakall 1999). Some of these alleles presenting global structure in the Moroccan durum landraces showed significant regions of clustered homozygote and heterozygote landraces.

Six subpopulations were detected for Moroccan landraces collection using spatial and non-spatial models of Bayesian genetic structure. Most of the probability of belonging to one or the other subpopulations was almost fully explained by the long-term climate profile. The global genetic structure was significantly higher ($p=0.02$) than the local one using spatial principal components analysis, consequently neighboring landraces tend to have a similar genetic profile. The identified subpopulations found were very diverse especially for quality traits, phenology and yield components. Interaction with the environments was very different from subpopulation to another for most of the traits. For GY, GxE was low for subpopulations located in dry areas and this is due most probably to the fact that the testing environment (Tel Hadya-Syria) was dry and stresses in the four years of planting. The subpopulations were also very diverse in number of alleles and more alleles could be found in landraces collected in high altitude and hot regions of Morocco.

At the end of this study, we developed a graphic user's interface for ArcGIS 9.2 using VBA for evaluating phenotypically and genetically landraces. The interface is a useful tool to analyze and study phenotypic and genotypic diversity of durum wheat landraces. This GUI permits different methods of analyzing traits for multivariate analysis or GxE, population genetics statistics and some spatial statistics such as spatial autocorrelation, spatial PCA. Several outputs from analysis are stored in the spatial files and can be used further for mapping, predicting and editing thematic maps.

This study supports the use of geographic information systems together with existing phenotypic data and genetic markers to assess quickly and efficiently large number of genetic resources entries held by gene-banks in particular in the context of climate change. The use of climate in dissecting variations found in durum landraces can help projecting genetic diversity using modeling and future climatic scenarios under changing climate.

References

Abiteboul, S. & Hull, R. (1986) "IFO: A Formal Semantic Database Model", Technical Report TR-84-304. University of Southern California, Computer Science Department.

Agorastos, AG. & Goulas, CK. (2005) Line selection for exploiting durum wheat (T. turgidum L. var. durum) local landraces in modern variety development program. **Euphytica**, 146: 117-124.

Agrama, H.A., Eizenga, G.C. & Yan, W. (2007) Association mapping of yield and its components in rice cultivars. **Mol. Breed.**, 19:341–356.

Ahmadizadeh, M. (2010) Genetic diversity of durum wheat landraces for antioxidant activity and some physiological traits under drought stress. **M.Sc. Thesis on Plant Breeding**. Islamic Azad University Ardabil Branch, IRAN.

Ahrens, J.F. & Loomis, W.E. (1963) Floral induction and development in winter wheat. **Crop Sci.**, 3: 463-466.

Allison, J.C.S. & Daynard, T.B. (1976) Effect of photoperiod on development and number of spikelets of a temperate and some low-latitude wheats. **Ann. Appl. Biol**., 83: 93-102.

Araus, JL., Amaro, T., Casadesús, J., Asbati, A. & Nachit, MM. (1998) Relationships between ash content, carbon isotope discrimination and yield in durum wheat. **Australian Journal of Plant Physiology**, 25:835–842.

Araus, JL., Amaro, T., Zuhair, Y. & Nachit, MM. (1997) Effect of leaf structure and water status on carbon isotope discrimination field-grown durum wheat. **Plant Cell and Environment**, 20:1484–1494.

Araus, JL., Casadesus, J., Asbati, A. & Nachit, MM. (2002) Basis of the relationship between ash content in the flag leaf and carbon isotope discriminitation in kernels of durum wheat. **Photosynthetica**, 39:591–596.

Araus, JL., Casadesús, J. & Bort, J. (2001) Recent tools for the screening of physiological traits determining yield. In: Reynolds MP, Ortiz-Monasterio JI, McNab A, editors**. Application of physiology in wheat breeding**. Mexico, DF: CIMMYT; pp. 59–77.

Araus JL., Ferrio JP., Buxó R., Voltas J. (2007) The historical perspective of dryland agriculture: lessons learned from 10,000 years of wheat cultivation. J Exp Bot, 58(2):131-45.

Araus JL. (1996) Integrative physiological criteria associated with yield potential. In: Reynolds MP, Rajaram S, McNab A, editors. Increasing yield potential in wheat: breaking the barriers.Mexico, DF: CIMMYT, pp. 150–167.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. **Ecol. Model.,** 157, 101–118.

Austin, R.B., Bingham, J., Blackwell, R.D., Evans, L.T., Ford, M.A., Morgan, C.I. & Taylor, M. (1980) Genetic improvements in winter wheat yields since 1900 and associated physiological changes. **J. Agric. Sci.**, 94: 675-689.

Autrique, E., Nachit, MM., Monneveux, P., Tanksley, SD., Sorrells, ME. (1996) Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. **Crop Sci.**, 36:735–742

Autrique, E., Nachit, M.M., Monneveux, P., Tanksley, S.D. and Sorrells, M.E. (1996) Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. **Crop Sci.**, 36: 735-742.

Avise, J.C. (1994) Molecular Markers, Natural History, and Evolution. **Chapman & Hall**, New York.

Beaumont, M.A. and Rannala, B. (2004) The Bayesian revolution in genetics. **Nature Reviews Genetics**, 5: 251- 261.

Belay, G., Tesemma, T., Becker, HC. and Merker, A. (1993) Variation and interrelationships of agronomic traits in Ethiopian tetraploid wheat landraces. **Euphytica**, 71: 181 - 188.

Bertorelle, G. and Barbujani, G. (1995) Analysis of DNA diversity by spatial autocorrelation. **Genetics**, 140, 811- 819.

Bertranpetit, J., Cavalli-Sforza, LL. (1991) A genetic reconstruction of the history of the population of the Iberian Peninsula. **Annals of Human Genetics**, 55 : 51-67.

Biesantz, A., Limberg, P. and Kyzeridis, N. (1990) Evaluation of Greek and Turkish durum wheat landraces. p. 45–55. In J.P. Srivastava and A.B. Damania (ed.) Wheat genetic resources: Meeting diverse needs, **John Wiley & Sons**, Chichester, UK.

Blanco, A., De Giovanni, C., Laddomada, B., Sciancalepore, A., Simeone, R., Devos, K.M. & Gale, M.D. (1996) Quantitative trait loci influencing grain protein content in tetraploid wheats. **Plant Breeding**, 115: 310-316.

Boggini, G., Dal Belin Peruffo, A., Mellini, F. and Pogna, N.E. (1987) Storage, Protein composition, morphophysiological, and quality characters of 24 old durum wheat varieties from Sicily. **Rachis**, 6: 30-35.

Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. **Ecological Modelling**, 153 : 51-68.

Bozzini, A. (1988) Origin, distribution and production of durum wheat in the world. In: Durum wheat: Chemistry and Technology, G. Fabriani and C. Lintas (ed.). **AACC, St. Paul, MN.**, p: 1-16.

Bryce, D., and Hull R. (1986) "SNAP: A Graphics-Based Schema Manager", Proceedings of the Second IEEE International Conference on Data Engineering, pp. 151-164.

Bucci, G. and Vendramin, G., (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. **Molecular Ecology**, 9 (7):923-934.

Buckler, E.S., and Thornsberry, J.M. (2002). Plant molecular diversity and applications to genomics. **Curr. Opin. Plant Biol.**, 5, 107–111.

Caloz, R. and Collet, C. (1997) Geographic information systems (GIS) and remote sensing in aquatic botany: methodological aspects. **Aquatic Botany**, 58 (3/4):209-228.

Camus-Kulandaivelu, L., J. B. Veyrieras, D. Madur, V. Combes, M. Fourmann et al. (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. **Genetics**, 172 2449–2463.

Cánovas, F., De la Rúa P., Serrano J., Galián J. (2008) Geographic patterns of mitochondrial DNA variation in Apis mellifera iberiensis (Hymenoptera: Apidae). **J. Zool. Syst. Evol. Res.**, 46, 24–30.

Cardon, LR., Palmer, LJ. (2003) Population stratification and spurious allelic association. **Lancet**, 361:598 604.

Cavalli-Sforza, LL. (1966) Population structure and human evolution. **Proceedings of the Royal Society of London Series B**, 164 : 362-379.

Cavalli-Sforza, L., Menozzi, P., Piazza, A. (1994) The history and geography of human genes, **Princeton University Press**, Princeton, New Jersey.

Cavalli-Sforza, L.L., and Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. **Amer. J. Hum. Genet.**, 19: 233-257.

Cercueil, A., Francois, O. and Manel, S. (2007) The genetical bandwidth mapping: A spatial and graphical representation of population genetic structure based on the wombling method. **Theoretical Population Biology**, 71: 332-341.

Chao, S., Zhang, W., Dubcovsky, J. and Sorrells, M. (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (Triticum aestivum L.) germplasm representing different market classes. **Crop Sci.**, 47:1018–1030.

Chapman, C.G.D.(1986) The role of genetic resources in wheat breeding. **Plant Genet. Res. News.**, 65: 2-5.

Chessel, D., Dufour, AB. and Thioulouse, J. (2004). The ade4 package-I- one-table methods. **R News**, 4 : 5-10.

Ciaffi, M., Benedettelli, S., Giorgi, B., Porceddu, E. and Lafiandra, D. (1991) Seed storage proteins of Triticum turgidum ssp. dicoccoides and their effect on the technological quality in durum wheat. **Plant Breeding**, 107: 309-319

Cooper, M. & DeLacy, I.H. (1994) Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theor. Appl. Genet.**, 88: 561-572.

Coulon, A., Guillot, G., Cosson, J.F., Angibault, J., Au- lagnier, S. and Cargnelutti, B. (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. **Molecular Ecology**, 15, 1669- 1679.

Crow, J. F. (1986) Basic concepts in population, quantitative, and evolutionary genetics. **W. H. Freeman**, New York.

De la Rúa, P., Hernández-García, R., Pedersen B.V., Galián J. and Serrano, J. (2004) Molecular diversity of honeybee Apis mellifera iberica L (Hymenoptera: Apidae) from western Andalusia. **Archivos de Zootecnia**, 53, 195–203.

DeWan, A., Liu, M., Hartman, S., Zhang, S. S.-M., Liu, D. T. L., Zhao, C., Tam, P. O. S., Chan, W. M., Lam, D. S. C., Snyder, M., Barnstable, C., Pang, C. P., and Hoh, J. (2006) HTRA1 promoter polymorphism in wet age-related macular degeneration. **Science**, 314, 989-992.

Doledec, S., Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species environment relationships. **Freshwater Biology**, 31 : 277-294.

Donelson, W.C. (1978) Spatial Management of Information. **Proceedings of ACM SIGGRAPH 1978**, pp. 203-209.

Donmez, E., Sears, RG., Shroyer, JP. and Paulsen, GM. (2000) Evaluation of Winter Durum Wheat for Kansas. **Kansas State University Agricultural Experiment Station and Cooperative Extension Service**. Publication No. 00-172-S.

D'Ovidio, R., Tanzarella, O.A and Porceddu, E. (1990) Rapid and efficient detection of genetic polymorphism in wheat through amplification by polymerase chain reaction. **Plant Molecular biology**, 15: 169-171.

Dray, S., Chessel, D. and Thioulouse, J. (2003) Co-inertia analysis and the linking of ecological tables. **Ecology**, 84 : 3078-3089.

Dupanloup, I., Schneider, S. and Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. **Molecular Ecology**, 11, 2571-2581.

Duwayri, M., Migdadi, H., Sadder, M., Kafawin, O., Ajlouni, M., Amri, A. and Nachit, M. (2007) Use of SSR markers for characterizing cultivated durum wheat and its naturally occurring hybrids with wild wheat. **Jordan Journal of Agricultural Sciences**, 3.

Eberhart, S.A. & Russell, W.A. (1966) Stability parameters for comparing varieties. **Crop Sci.**, 6: 36-40.

Eberhart, S.A. & Russell, W.A. (1969) Yield and stability for a 10-line diallel of single-cross and double-cross maize hybrids. **Crop Sci.**, 9: 357-361.

Ehrlich, P. R., and Raven, P. H. (1969) Differentiation of populations. **Science**, 165:1228-1232.

Elouafi, I. and Nachit, M.M. (2004) A Genetic Linkage Map of the Durum x Triticum dicoccoides Backcross Population Based on SSRs and AFLP Markers and QTL Analysis for Milling Traits. **Theor. Appl. Genet.**, 108(3), 401–413.

Epperson, BK. (2003) Covariances among join-count spatial autocorrelation measures. **Theor Popul Biol.**, 64(1):81-7.

Escudero, A., Romao, R., de la Cruz, M. and Maestre, FT. (2005) Spatial pattern and neighbor effects on Helianthemum squamatum seedlings in a semiarid Mediterranean gypsum community. **J Veg Sci.**, 16:383–390.

Esquinas-Alcazar, J.T. (1987) Plant Genetic Resources: A Base for Food Security. **Review Ceres**, 118(29,4): 39-45.

Evans, L.T., Wardlaw, I.F. & Fischer, R.A. (1975) Wheat. In L.T. Evans, ed. Crop physiology, p. 101-149. Cambridge, UK, **Cambridge University Press**.

Falush, D., Stephens, M. and Pritchard, J.K. (2003) In- ference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. **Genetics**, 164, 1567-1587.

Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., et al. (2000). **Genome Res.,** 10: 220–227.

Federer, W. T. (1956) Augmented (or hoonuiaku) designs**. Hawaiian Planters' Record**, 55, 191-208.

FELDMAN, M. and SEARS, E. R. (1981) The wild gene resources of wheat. **Sci. Am.**, 244: 102-112.

FieldSpec UV/VNIR, Analytical Spectral Devices, Boulder, CO

Finlay, K.W. & Wilkinson, G.N. (1963) The analysis of adaptation in a plant-breeding programme. **Aust. J. Agric. Res.**, 14: 742-754.

Fischer, R.A. (1985) Number of kernels in wheat crops and the influence of solar radiation and temperature. **J. Agric. Sci.**, 105: 447-461.

Flint-Garcia, S.A., Thornsberry, J.M. and Buckler, E.S. (2003) Structure of linkage disequilibrium in plants. **Ann. Rev. Plant Biol.**, 54:357–374.

Francois, O., Ancelet, S. and Guillot, G. (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. **Genetics**, 174:805-816.

Frankel, O.H. (1977) Natural variation and its conservation. In A. Muhammed & R.C. von Botstel, eds. Genetic diversity of plants, p. 21-24. New York, NY, USA**, Plenum Press**.

Frankham, R., Briscoe, DA. and Ballou, JD. (2002) Introduction to conservation genetics. **Cambridge University Press**, New York, New York, USA.

Friedell, M. (1984) Automatic Synthesis of Graphical Object Descriptions. **ACM Computer Graphics**, 18(3), pp. 53-62.

Friedell, M., Barnett, J. and Kramlich, D. (1982) Context-Sensitive Graphic Presentations of Information. **ACM Computer Graphics**, 16(3). pp. 181-188.

Futuyma, D. (1998) **Evolutionary Biology**. Third edition. Sunderland, MA, Sinauer Associates.

Gabriel, K. R. and Sokal, R. R. (1969) A new statistical approach to geographic variation analysis. **Systematic Zoology (Society of Systematic Biologists**, 18 (3): 259–270.

Getis, A. and Ord, JK. (1996) Local spatial statistics: an overview. In: Longley P, Batty M (eds) Spatial analysis: modelling in a GIS environment pp. 261–277. **GeoInformation International**, Cambridge.

Getis, A. and Ord, J. K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. **Geographical Analysis**, 24, Nº 3.

Godt, M. J. W. and J. L. Hamrick. 1998. Allozyme diversity in the endangered pitcher plant Sarracenia rubra ssp. alabamensis (Sarraceniaceae) and its close relative S. rubra ssp. rubra. Am. J. Bot. 85:802-810.

Goldman, K. et al. (1985) ISIS: Interface for a Semantic Information System. **Proceedings of ACM SIGMOD International Conference on the Management of Data**. pp. 328-342.

Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. **Genetics**, 139: 463-471.

Goodchild, F., Parks, B.O. and Steyaert, L.T. (1993) Environmental Modeling with GIS. **Oxford University Press, New York**.

Goodchild, M.F., (1992) Geographical information science. **International Journal of Geographical Information Systems**, 6 (1):31-45.

Goodchild, M.F. and Haining, R.P. (2004) GIS and spatial analysis : Converging perspectives. **Papers in Regional Science**, 83:363-385.

Guillot, G., Estoup, A., Mortier, F. and Cosson, J.F. (2005) A spatial statistical model for landscape genetics. Genetics, 170, 1261-1280.

Guillot, G., Estoup, A., Mortier, F. and Cosson, J.F. (2005) A spatial statistical model for landscape genetics. **Genetics**, 170, 1261-1280.

Gupta, P.K., Mir, RR., Mohan, A. and Kumar, J. (2008) Wheat Genomics: Present Status and Future Prospects. **International Journal of Plant Genomics**, vol. 2008, Article ID 896451, 36 pages.

Gupta, SK., Charpe, A., Koul, S., Prabhu, KV. and Haq, QMR. (2005) Development and validation of molecular markers linked to an Aegilops umbellulata-derived leaf rust resistance gene, Lr9, for marker-assisted selection in bread wheat. **Genome**, 48:823–830.

Haldane, JBS. (1948) The theory of a cline. **J. Genet.**, 48: 277-284.

Hamann, A., Koshy, M., Namkoong, G. and Ying, C., (2000) Genotype x environment interactions in Alnus rubra: developing seed zones and seed-transfer guidelines with spatial statistics and GIS. **Forest Ecology and Management**, 136 (1-3):107-119.

Hamrick, J. L. and Godt, MJW. (1996) Effects of life history traits on genetic diversity in plant species. **Phil. Trans. Roy. Soc. London Biol. Sci.**, 351:1291-1298.

Hanft, J.M. & Wych, R.D. 1982. Visual indicators of physiological maturity of hard red spring wheat. **Crop Sci.**, 22: 584-587.

Hanski, IA. and Simberlo, D. (1997) Metapopulation biology : Ecology, Genetics and Evolution, **Academic Press**, chap. The metapopulation approach, its history, conceptual domain, and application to conservation, pp. 5-26.

Harlan, JR. (1992) Origin and processes of domestication. In Chapman G. P. (ed).Grass evolution and domestication. **Cambridge University Press**. Cambridge.

Haun, J.R. (1973) Visual quantification of wheat development. **Agron. J.**, 65: 116-117.

Herot, C.F. (1980) Spatial Management of Data. **ACM Transactions on Database Systems**, 5(4), pp. 493-514.

Hijmans, R.J., Cameron, SE., Parra, JL., Jones, PG. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology**, 25: 1965-1978.

Hirao, A.S., Kudo, G. (2004) Landscape genetics of alpine-snow bed plants: comparisons along geographic and snowmelt gradients. **Heredity**, 93 (3):290-298.

Hoffmann, M.H., Glass, A., Tomiuk, J., Schmuths, H., Fritsch, R.M. and Bachmann, K. (2003) Analysis of molecular data of Arabidopsis thaliana (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). **Molecular Ecology**, 12:1007-1019.

Isaaks, E. H. and Srivastava, RM. (1989) An Introduction to Applied Geostatistics. **Oxford Univ. Press**, New York, Oxford.

Iwaki, K., Haruna, S., Niwa, T. and Kato, K. (2001) Adaptation and ecological differentiation in wheat with special reference to geographical variation of growth habit and Vrn genotype. **Plant Breeding**, 120: 107–114.

Jombart, T., Devillard, S., Dufour, A.B. and Pontier, D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. **Heredity**, 101, 92- 103.

Jones, P.G., Beebe, S.E., Tohme, J. (1997) The use of geographical information systems in biodiversity exploration and conservation**. Biodiversity and Conservation**, 6:947-958.

Jones, R.J., Roessler, J. & Quattar, S. (1985) Thermal environment during endosperm cell division in maize: effects on number of endosperm cells and starch granules. **Crop Sci.**, 25: 830-834.

Joost, S. & the Econogene Consortium (2005) Combining biotechnologies and GIScience to contribute to sheep and goat genetic resources conservation. International Workshop on the role of biotechnology for the characterization and conservation of crop, forestry, animal and fishery genetic resources, **FAO**, Torino, pp. 109-116.

Karlsson et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. **Nature Genetics**.

Kato, K. & Yokoyama, W. (1992) Geographical variation in heading characters among wheat landraces, Triticum aestivum L. and its implication for their adaptability. **Theor. Applied Genet.**, 84: 259-265.

Kehel, Z. , Garcia-Ferrer, A. and Nachit, M. (2013) Using Bayesian and Eigen approaches to study spatial genetic structure of Moroccan and Syrian durum wheat landraces. **American Journal of Molecular Biology**, 3, 17-31.

Kerth, G, Petit E (2005) Colonization and dispersal in a social species, the Bechstein's bat (Myotis bechsteinii). **Molecular Ecology**, 14: 3943–3950

Kidd, D.M. & Liu, X. (2008) Geophylobuilder 1.0: an ArcGIS extension for creating geophylogenies. **Molecular Ecology Resources**, 8: 88-91.

Kidd, D.M. & Ritchie, M.G. (2000) Inferring the patterns and causes of geographic variation in Ephippiger ephippiger (Orthoptera, Tettigoniidae) using geographical information systems (GIS). **Biological Journal of the Linnean Society**, 71:269-295.

King, R. & Melville, S. (1984) The Semantics-Knowledgeable Interface. **Proceedings of the Conference on Very Large Databases**. pp. 30-37.

Kirby, E.J.M. & Appleyard, M. (1984) Cereal development guide. **Stoneleigh, Kenilworth, UK, NAC Arable Unit**. 95 pp.

Knowles, B. (1992) Bat hibernacula on Lake Superior's North Shore, Minnesota. **Canadian Field-Naturalist**. 106(2):252-254.

Kraakman, A.T.W., Niks, R.E., Van den Berg, P.M.M.M., Stam, P. & Van Eeuwijk, F.A. (2004) Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. **Genetics**, 168, 435–446.

Krumm, M., Moazami, V. & Martin, P. (1990) Influence of potassium nutrition on concentrations of water soluble carbohyd-rates, potassium, calcium, and magnesium and the osmotic potential in sap extracted from wheat (Triticum aestivum) ears during pre-anthesis development. **Plant Soil**, 124: 281-285.

Laffan, SW. (2006) Assessing regional scale weed distributions, with an Australian example using Nassella trichotoma. **Weed Research**, vol. 46, pp. 194 – 206.

Large, E.C. (1954) Growth stages in cereals. Illustration of the "Feekes" Scale. **Plant Pathol.**, 3: 128-129.

Latter, B. D. (1972) Selection infinite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. **Genetics**, 70: 475–490.

Legendre, P. and Legendre, L. (1998) Numerical ecology**. Elsevier Science B. V.**, Amsterdam.

Limp, W. F. (2001) Spatial Database Break Computing Barriers Enterprisewide. **Center for advance spatial Technologies**, University of Arkansas, Fayetteville.

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2001) Geographic Information Systems and Science. **Wiley**, Chichester.

Loveless, M. D., and Hamrick, JL. (1984) Ecological determinants of genetic structure of plant populations. **Ann. Rev. Ecol. Syst.**, 15:65-95.

MacArthur, R. H. and Wilson, E. O. (1967) The Theory of Island Biogeography. **Princeton, N.J.: Princeton University Press**.

Maccaferri, M., Sanguineti, MC., Nachit, M. et al. (2008) Quantitative trait loci for grain yield and adaptation of durum wheat (Triticum durum Desf.) across a wide range of water availability. **Genetics**, 178: 489-511.

MacKendry, J.E., Machlis, G.E. (1991) The role of geography in extending biodiversity gap analysis. **Applied Geography**, 11:135-152.

MacKey, J. (2005) Wheat, its concept, evolution and taxonomy. In: Royo C, Nachit M, Di Fonzo N, Araus JL, Pfeiffer WH, Slafer GA (eds), Durum wheat breeding: current approaches and future strategies, Ed. Vol. I **Howorth Press**, New York, pp 3–62

MacQueen, J. B. (1967) Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. **University of California Press**. pp. 281–297.

Major, D.J. & Kiniry, J.R. (1991) Predicting daylength effects on phenological processes. In T. Hodges, ed. Predicting crop phenology, p. 15-28. Boca Raton, FL, USA, **CRC Press**.

Manel, S., Bellemain, E., Swenson, JE. and Francois, O. (2004) Assumed and inferred spatial structure of populations : the Scandinavian brown bears revisited. **Molecular Ecology**, 13 : 1327-1331.

Manel, S., Berthoud, F., Bellemain, E. et al. (2007) A new individual-based spatial approach for identifying genetic discontinuities in natural populations. **Molecular Ecology**, 16, 2031–2043.

Manel, S., Schwartz, M., Luikart, G. and Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. **Trends in Ecology & Evolution**, 18 (4):189-197.

Manni, F., Guerard, E. and Heyer, E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation : how barriers can be detected by "Monmonier's algorithm". **Human Biology**, 76 : 173-190.

Martin, J.H., Leonard, W.H. & Stamp, D.L. (1976) Principles of field crop production. **New York, NY, USA, Macmillan**. 1118 pp.

Matula, D. W. & Sokal, R. R., 1980. "Properties of Gabriel graphs relevant to geographic variation research and clustering of points in the plane", **Geogr. Anal.**, 12 (3): 205–222.

McDonald, N.H. (1984) A Multi Media Approach to the User Interface", Human Factors and Interactive Computer Systems, edited by Y. Vissiliou, **Ablex Publishing Co**. pp. 105-116.

McVean, G. (2009) A Genealogical Interpretation of Principal Components Analysis. **PLoS Genet**, 5(10.

Melnikova, N.V., Mitrofanova, OP., Liapounova, OA. and Kudryavtsev, AM. (2010) Global diversity of durum wheat Triticum durum Desf. for alleles of gliadin-coding loci. **Russ. J. Genet.**, 46 (No. 1) 43e49.

Menozzi, P, Piazza, A. and Cavalli-Sforza, LL. (1978) Synthetic maps of human gene frequencies in Europeans. **Science**, 201 : 786-792.

Merah, O., Deléens, E. and Monneveux, P. (1999) Grain yield, carbon isotope discrimination, mineral and silicon content in durum wheat under different precipitation regimes. **Physiologia Plantarum**, 107:387–394.

Merah, O., Deléens, E., Souyris, I. and Monneveux, P. (2001) Ash content might predict carbon isotope discrimination and grain yield in durum wheat. **New Phytologist**, 149:275–282.

Mondini, L., Noorani, A. and Pagnotta, M. (2009) Assessing Plant Genetic Diversity by Molecular Tools. **Diversity**, 1: 19-35.

Monmonier, M. (1973) Maximum-difference barriers: an alternative numerical regionalization method. **Geographical Analysis**, 3 : 245-261.

Monneveux, P., Reynolds, MP., Trethowan, R., Peña, J. and Zapata, F. (2004) Carbon isotope discrimination, leaf ash content and grain yield in bread and durum wheat grown under full-irrigated conditions. **Journal of Agronomy and Crop Science**, 190:389–394.

Monneveux, P., Sánchez, C. and Tiessen, A (2008) Future progress in drought tolerance in maize needs new secondary traits and cross combinations. **Journal of Agricultural Science**, 146:1–14.

Monneveux, P., Sheshshayee, MS. Akhter, J. and Ribaut, JM. (2007) Using carbon isotope discrimination to select maize (Zea mays L.) inbred lines and hybrids for drought tolerance. **Plant Science**, 173:390–396.

Moraguees, M., Zarco-Hernandez, J., Moralejo, M.A. and Royo, C. (2006) Genetic diversity of glutenin protein subunits composition in durum wheat landraces from the Mediterranean basin. **Genetic Ressources and Crop Evo- lution**, 53: 993-1002.

Moragues, M., Garcia Del Moral, LF., Moralejo, M. and Royo, C. (2006) Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin: II. Biomass production and allocation. **Field Crops Res**, 95:182–193

Moragues, M., Garcia Del Moral, LF., Moralejo, M. and Royo, C. (2006) Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin: I. Yield components. **Field Crops Res**, 95:194–205

Moran, PAP. (1948) The interpretation of statistical maps. **Journal of the Royal Statistical Society**, 10 : 243-251.

Moran, PAP. (1950) Notes on continuous stochastic phenomena. **Biometrika**, 37 : 17-23.

Morris, R. and Sears, ER. (1967) The cytogenetics of wheat and its relatives. **Wheat and Wheat Improvement**, 19-87.

Motawaj, J. (2007) Genetic Gain in Morpho-Physiological Traits Related to Drought Tolerance in Durum Wheat, **PhD Thesis, Aleppo University**, Aleppo, Syria.

Mourant, A., (1954) The distribution of the human blood groups. **Blackwell Scientific**, Oxford.

Mueller-Warrant, G. W., Whittaker, G. W. & Young, W. C. (2008) GIS analysis of spatial clustering and temporal change in weeds of grass seed crops. **Weed Science**, 56(5), 647-669.

Nachit, MM., Baum, M. Impiglia, A. and Ketata, H. (1995) Studies on some grain quality traits in durum wheat grown in Mediterranean environments. Proc. Seminar on Durum Wheat Quality in the Mediterranean regions, Zargoza 17-19 Nov 1995. **Options Mediterraneennes Serie A**, n°22: 181-188, ICARDA/ CIHEAM/ CIMMYT.

Nachit, M. M. Elouafi, I. Pagnotta, M.A. Saleh, A. Lacono, E. Labhilili, M. Asbati, A. Azrak, M. Hazzam, H. Benscher, D. Khairallah, M. Ribaut, J.M. Tanzarella, O.A. Porceddu, E. and Sorrels, M.E. (2001) Molecular Linkage Map for an Intraspecific Recombinant Inbred Population of Durum Wheat (Triticum turgidum L. var. durum). **Theor. Appl. Genet.**, 102, 177–186.

Nachit, M.M. (1992) Durum Wheat Breeding for Mediterranean Dryland of North Africa and West Asia. Paper presented at Durum Wheat Workshop "**Discussion on Durum Wheat: Challenges and Opportunity**". CIMMYT, Ciudad Obregon, Mexico. March 23 - 25, 1992, p.14-27.

Nachit, M.M. (1998) Durum Breeding Research to Improve Dryland Productivity in the Mediterranean Region. **SEWANA**, Durum Research Network. p. 1-15. ICARDA-021/500 Sep 1998. Aleppo, Syria.

Nachit, M.M., Sorrells, M.E., Zobel, R.W., Gauch, H.G., Fischer, R.A., and Coffman, W.R. (1992a) Association of Morpho-Physiological Traits with Grain Yield and Genotype-Environment Interaction in Durum Wheat. I**. J. Genet. & Breed.**, 46: 50-55.

Nachit, M.M., Sorrells, M.E., Zobel, R.W., Gauch, H.G., Jr., Fischer, R.A., and Coffman, W.R. (1992b) Association of Environmental Variables with Sites' Mean Grain Yield and Genotype-Environment Interaction in Durum Wheat. II. **J. Genet. and Breed.**, 46: 41-49.

Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**, 89: 583-590.

Nei, M. (1972) Genetic distance between populations. **Am. Nat.**, 106:283-292

Nei, M. (1973) Analysis of gene diversity in subdivided populations. **Proc. Natl. Acad. Sci. USA**, 70: 3321–3323.

Nei, M., Tajima, F. and Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. **J.Mol. Evol.**, 19: 153–170.

Nordborg, M. and Tavaré, S. (2002) Linkage disequilibrium: what history has to tell us. **Trends Genet.**, 18:83-90.

Owen, P.C. (1971) Responses of a semi-dwarf wheat to temperatures representing a tropical dry season. II. Extreme temperatures. **Exp. Agric.**, 7: 43-47.

Pagnotta, M., Impiglia, A. Tanzarella, O., Nachit, M. and Porceddu, E. (2005) Genetic variation of the durum wheat landrace Haurani from different agro-ecological regions. **Genetic Resources and Crop Evolution**, Volume 51, Number 8, January 2005 , pp. 863-869(7).

Pakniyat, H., Powell, W., Baird, E., Handley, L.L., Robinson, D., Scrimgeour, C.M., Nevo, E., Hackett, C.A., Caligari, P.D.S. and Forster, B.P. (1997) AFLP variation in wild barley (Hordeum C. Koch) with reference to salt tolerance and associated ecogeography. **Genome**, 3 (40):332-341.

Patterson, N., Price, A.L. and Reich, D. (2006) Popula- tion structure and eigenanalysis. **PLoS Genetics**, 2: 2074- 2093.

Payne, R.W., Murray, D.A., Harding, S.A., Baird, D.B. & Soutar, D.M. (2009). **GenStat for Windows (12th Edition) Introduction**. VSN International, Hemel Hempstead.

Pearson, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. **Philosophical Magazine**, 2 (11): 559–572.

Pecetti, L. and Nachit, M.M. (1993) Phenotypic Variation of Durum Wheat Landraces from Morocco and Influence of Some Collecting Site Features. **Agriccoltura Mediterranea**, 123, 243-251.

Pecetti, L., Damania A.B. and Jana, S. (1992) Practical problems in large-scale germplasm evaluation. A case study in durum wheat. **Genet. Res. Newsl.**, 88/89: 5-10.

Peleg, Z., Fahima, T., Krugman, T., Abbo, S. & Saranga, Y. (2008) Genetic structure of natural wild emmer wheat populations as reflected by transcribed versus anonymous SSR markers. **Genome**, 51, 187-195.

Perkins, J.M. & Jinks, J.L. (1968) Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. **Heredity**, 23: 339-356.

Peterson, R.G (1985) Augmented designs for preliminary yield trials (revised). **Rachis**, 4: 27-32

Peterson, R.F. (1965) Wheat: botany, cultivation and utilisation. **London, Leonard Hill**. 448 pp.

Porceddu, E., Perrino, P. and Olita, G. (1975) Preliminary information on an Ethiopian wheat germplasm collection mission. In: Proc. Symp. Genetics and Breeding of Durum Wheat, **Scarascia Mugnozza, G.T. (ed.)**. University of Bari, Bari, Italy, pp. 181-200.

Prevosti, A., Ocana, J. and Alonzo, G. (1975) Distances between population for Drosophila subobscura based on chromosome arrangement frequencies. **Theorical Applied Genetics**, 45:231-241.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome- wide association studies. **Nature Genetics**, 38, 904-909.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. **Genetics**, 155, 945-959.

Proche, A. (2006) Latitudinal and longitudinal barriers in global biogeography. **Biology Letters**, 2: 69 – 72.

Ramachandran, S., Deshpande, O., Roseman, CC., NA Rosenberg, NA., MW Feldman,MW. and Cavalli-Sforza, LL. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. **Proceedings of the National Academy of Sciences USA**, 102: 15942-15947.

Reimer, S., Pozniak, C.J. Clarke, F.R. Clarke, J.M. Somers, D.J. Knox, R.E. and Singh, A.K. (2008) Association Mapping of Yellow Pigment in An Elite Collection of Durum Wheat cultivars and Breeding Lines. **Genome**, 51, 1016-1025.

Reiner, D. et al. (1984) The Database Design and Evaluation Workbench (DDEW) Project at CCA. **Bulletin of IEEE Technical Committee on Database Engineering**, 7(4) pp. 10-15,

Rejesus, R.M., Smale, M. and Van Ginkel, M. (1996) Wheat Breeders' Perspectives on Genetic Diversity and Germplasm Use: Findings from an International Survey. **Plant Varieties and Seeds**, Vol. 9, No. 3 (1996): 129-147.

Röder, M.S. Korzun, V. Wendehake, K. Plaschke, J. Tixer, M. H. Leroy, P. and Ganal, M.W. (1998) A Microsatellite Map of Wheat. **Genetics**, 149, 2007-2023.

Röder, M.S. Plaschke, J. König, S.U. Börner, A. Sorrells, M.E. Tanksley, S.D. and Ganal, M.W. (1995) Abundance Variability and Chromosomal Location of Microsatellites in Wheat. **Mol. Gen. Genet.**, 246, 327-333.

Roelfs, A.P. (1988) Advances and understanding of rust resistance. Lecture presented at Centro de Investigaciones Agricolas del Noroeste (**CIANO**), Mexico.

Rodgers, J.S. (1972) Measures of genetic similarity and genetic distance. In: Studies in Genetics VII**. University of Texas Publication 7213**, Austin, Texas, p. 145-153.

Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, PR., Roose, ML., Svensson, JT., Stein, N., Varshney, RK., Marshall, DF., Grainer, A., Close, TJ. and Waugh, R. (2006) Recent history of arti?cial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. **Proc Natl Acad Sci.**, 103:18656–18661.

Saini, H.S. & Aspinal, D. (1982) Abnormal sporogenesis in wheat (Triticum aestivum L.) induced by short periods of high temperature. **Ann. Bot.**, 49: 835-846.

Sanghvi, L. D. (1953) Comparison of genetical and morphological methods for a study of biological differences. **Am. J. Phys. Anthrop.**, 11: 385-404.

Schilling, AS., Abaye, AO., Griffeya, CA., Branna, DE., Alleya, MM. and TH Pridgena, TH. (2003) Adaptation and Performance of Winter Durum Wheat in Virginia**. Agron J.**, 95: 642-651.

Shriver, M.D., Jin, L., Boerwinkle, E., Deka, R., and Ferrell, R.E. (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. **Mol. Biol. Evol.**, 12(5): 914–920.

Shukla, G.K. (1972) Some statistical aspects of partitioning genotype-environmental components of variability. **Heredity**, 29: 237-245.

Simmons, S.R. & Crookston, R.K. (1979) Rate and duration of growth of kernels formed at specific florets in spikelets of spring wheat**. Crop Sci.**, 19: 690-693.

Skorppa, T. (1984) A critical evaluation of methods available to estimate the genotype x environment interaction. In Proceedings from a conference on genotype x environment interaction, 23–27 August 1982, Uppsala, Sweden. **Studia For. Suecia**, 166: 3–14.

Skøt, L., Hamilton, N.R.S., Mizen, S., et al. (2002) Molecular genecology of temperature response in Lolium perenne: 2. association of AFLP markers with ecogeography. **Molecular Ecology**, 9 (11):1865-1876.

Slafer, G.A. & Rawson, H.M., (1994) Sensitivity of wheat phasic development to major environmental factors: a reexamination of some assumptions made by physiologists and modellers. **Austr. J. Plant Physiol.**, 21: 393-426.

Slafer, G.A., Andrade, F.H. & Satorre, E.H. (1990) Genetic improvement effects on pre-anthesis physiological attributes related to wheat grain yield. **Fields Crops Res.**, 23: 255-263.

Slafer, G.A., Calderini, D.F. & Miralles, D.J. (1996) Yield components and compensation in wheat: opportunities for further increasing yield potential. In M.P. Reynolds, S. Rajaram & A. McNab, eds. Increasing yield potential in wheat: breaking the barriers, p.101-133. Mexico, DF, CIMMYT.

Slatkin, M. (1985) Gene flow in natural populations. **Annual Reviews of Ecology and Systematics**, 16 : 393-430

Smouse, P. and Peakall, R. (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. **Heredity**, 82, 561-573.

Sokal, R. and Wartenberg, D. (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. **Genetics**, 105, 219-237.

Sokal, R., Smouse, P. and Neel, J. (1986) The genetic structure of a tribal population, the Yanomama Indians. XV. Patterns inferred by autocorrelation analysis. **Genetics**, 114, 259-287.

Spagnoletti Zeuli, P.L.M., C.DePace & Porceddu, E. (1984) Variation in durum wheat populations from three geographical origins. I. Material and spike characteristics. **Euphytica**, 33: 563–575.

Spear, S.F., Peterson, C.R., Matocq, M.D. and Storfer, A. (2005) Landscape genetics of the blotched tiger salamander (Ambystoma tigrinum melanostictum). **Molecular Ecology**, 14:2553–2564.

Stakman, E.C. & Harrar, J.G. (1957) Principles of plant pathology. New York, NY, USA. **The Ronald Press Company**.

Stefany, P. (1993) Vernalisation requirement and response to day length in guiding development in wheat. **Wheat Special Report No. 22**, Mexico, DF, CIMMYT.

Tesemma, T., Tsegaye S., Belay G., Bechere E., Mitiku D. (1998) Stability of performance of tetraploid wheat landraces in the Ethiopian highland. **Euphytica**, 102: 301–308

Thioulouse, J., Chesse,l D. and Champely, S. (1995) Multivariate analysis of spatial patterns : a unified approach to local and global structures. **Environmental and Ecological Statistics**, 2: 1-14.

Torre, F., Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. **Rev. Statistique Appl.**, 43: 109–121.

Trione, E.J. & Metzger, R.J. (1970) Wheat and barley vernalisation in a precise temperature gradient. **Crop Sci.**, 10: 390-392.

Van Slageren, M.W. (1994) Wild wheats: a monograph of Aegilops L. and Amblyopyrum(Jaub. & Spach) Eig (Poaceae). **ICARDA/Wageningen Agricultural University Papers**, 94(7). i-xiv, 1-512.

Vavilov, N.I. (1951) The origin, variation, immunity and breeding of cultivated plants. **Chron. Bot.**, 13:1-364.

Wagner, H.H. and M.J. Fortin (2005) Spatial analysis of landscapes: concepts and statistics. **Ecology**, 86: 1975-1987.

Wang, H.Y., Wei, Y.M., Yan, Z.H. and Zheng, Y.L. (2007) EST-SSR DNA polymorphism in durum wheat (Triticum aestivum L.) collections. **Journal of Applied Genetics**, vol. 48, no. 1, p. 37-42.

Wartenberg, DE (1985) Multivariate spatial correlations: a method for exploratory geographical analysis. **Geographical Analysis**, 17: 263-283.

Watts, P., Rouquette, J., Saccheri, J., Kemp, S. and Thompson, D. (2004) Molecular and ecological evidence for small-scale isolation by distance in an endangered damselfly, Coenagrion mercurial. **Molecular Ecology**, 13 (10):2931-2945.

Weir, BS. and Cockerham, CC. (1984) Estimating F-Statistics for the analysis of population structure. **Evolution**, 38: 1358–1370.

Wong, H. and Kuo, H. (1982) GUIDE: A Graphical User Interface for Database Exploration. **Proceedings of the Conference on very large databases**. pp. 22-32.

Wricke, G. (1962) Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. Z. **Pflanzenzüchtg**., 47: 92-96.

Wright, S. (1943). Isolation by distance. **Genetics**, 28 : 114-138.

Wright, S. (1920) The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. **Proc. Natl. Acad. Sci. USA**, 6:320-32.

Wright, S. (1922) Coefficients of inbreeding and relation- ship. **American Nature**, 56: 330-338.

Wright, S. (1931) Evolution in mendelian populations. **Genetics**, 16:97-159.

Wu, ST. (1989) QPF: A Versatile Query Language for a Knowledge Based Geographical Information System. **Int. J. of Geographical Information Systems**, Vol. 3 No. 1.

Wuest, S.B. & Cassman, K.G. (1992) Fertiliser-nitrogen use efficiency of irrigated wheat. I. Uptake efficiency of pre-plant versus late-season application. **Agron. J.**, 84: 682-688.

Yu, J. & Buckler, ES. (2006) Genetic association mapping and genome organization of maize. **Current Opinions in Biotechnology**, 17:155-160.

Yu, J., Pressoir, W.H., Briggs, I., Vroh, M., Yamasaki, J.F., Doebley, M.D., McMullen, B.S., Gaut, D.M., Nielsen, J.B., Holland, S. & Buckler, E.S. (2006) A Unified General-Model Method for Association Mapping that Accounts for Multiple Levels of Relatedness. **Nat. Genet.**, 38: 203–208.

Yu, Q., Elizondo, S. & Bi, X. (2006) Structural analyses of Sum1-1p-dependent Transcriptionally silent chromatin in Saccharomyces cerevisiae. **J. Mol. Biol.**, 356: 1082- 1092.

Zadoks, J.C., Chang, T.T. & Konzak, C.F. (1974) A decimal code for the growth stages of cereals. **Weed Res.**, 14: 415-421.

Zarkti, H., Ouabbou, H., Hilali, A. & Udupa, S.M. (2010) Detection of genetic diversity in Moroccan durum wheat accessions using agro-morphological traits and microsatellite markers. **African Journal of Agricultural Research**, 5: 1837-1844.

Zohary, D, & Feldman, M. (1962) Hybridization between amphidiploids and the evolution of polyploids in the wheat (*Ægilops*-Triticum) group. **Evolution**,16: 4461.

Zou, Z.T. & Yang W.Y. (1995) Development of wheat germplasm research in Sichuean province. **Crop Genetic Resources**, 2: 19–20.

Glossary

Allele: (Gr. allelon, of one another, mutually each other); allelomorph (adj: allelic, allelomorphic). One of a pair, or series, of variant forms of a gene that occur at a given locus in a chromosome. Alleles are symbolized with the same basic symbol (e.g., B for dominant and b for recessive); B1, B2, ..., Bn for n additive alleles at a locus). In a normal diploid cell there are two alleles of any one gene (one from each parent), which occupy the same relative position (locus) on homologous chromosomes. Within a population there may be more than two alleles of a gene.

Allele frequency: The number of copies of an allele in a population, expressed as a proportion of the total number of copies of all alleles at a locus in a population.

Biodiversity: 1. The variety of species (species diversity) or other taxa of animals, microorganisms and plants in a natural community or habitat, or of communities in a particular environment (ecological diversity), or of genetic variation in a species (genetic diversity, q.v.). The maintenance of a high level of biodiversity is important for the stability of ecosystems. 2. The variety of life in all its forms, levels and combinations, encompassing genetic diversity, species diversity and ecosystem diversity.

Biotechnology: 1. The use of biological processes or organisms for the production of materials and services of benefit to humankind. Biotechnology includes the use of techniques for the improvement of the characteristics of economically important plants and animals and for the development of micro-organisms to act on the environment. 2. The scientific manipulation of living organisms, especially at the molecular genetic level, to produce new products, such as hormones, vaccines or monoclonal antibodies.

Breeding: The process of sexual reproduction and production of offspring.

Centers of origin: The locations in the world where particular domesticated plants originated. These areas show the highest variation, and are rich in wild alleles.

Chlorophyll: (Gr. chloros, green + phyllon, leaf) One of the two pigments responsible for the green color of most plants. It is essential in the absorption of light energy for photosynthesis.

Chromosome: (Gr. chroma, color + soma, body) 1. A single DNA molecule, a tightly coiled strand of DNA, condensed into a compact structure in vivo by complexing with accessory histones or histone-like proteins. 2. A group of nuclear bodies containing genes which are largely responsible for the differentiation and activity of a eukaryotic cell; one of the bodies into which the nucleus resolves itself at the beginning of mitosis and from which it is derived at the end of mitosis. Chromosomes contain most of the cell's DNA. Chromosomes exist in pairs in eukaryotes – one paternal (from the male parent) and one maternal (from the female parent). Each eukaryotic species has a characteristic number of chromosomes. Bacterial and viral cells contain only a single chromosome, consisting of a single or double strand of DNA or, in some viruses, RNA, without histones.

Co-dominance: The situation in which both alleles in a heterozygous individual are expressed, so that the phenotype of heterozygotes incorporates the phenotypic effect of each allele. For example, roan coat color in cattle results from a mixture of red hairs and white hairs, caused by heterozygosity for the red allele and the white allele. Also, protein polymorphisms and microsatellites show co-dominance: heterozygotes have two bands, whereas homozygotes have only one band.

Co-dominant alleles: Alleles that produce independent effects when in the heterozygous condition.

Correlation: A statistical association between variables.

Cultivar: (from cultivated + variety) (abbr: cv.) A category of plants that are, firstly, below the level of a sub-species taxonomically, and, secondly, found only in cultivation. It is an international term denoting certain cultivated plants that are clearly distinguishable from others by stated characteristics and that retain their distinguishing characters when reproduced under specific conditions.

Database: One or more structured sets of persistent data, managed and stored as a unit and generally associated with software to update and query the data. A simple database might be a single file with many records, each of which references the same set of fields. A GIS database includes data about the spatial locations and shapes of geographic features recorded as points, lines, areas, pixels, grid cells, or TINs, as well as their attributes.

DEM (Digital Elevation Model): Represents a topographic surface using a continuous array of elevation values, referenced to a common datum. DEMs are used typically to represent terrain relief.

Diploid: (Gr. diploos, double + oides, like) 1. The status of having two complete sets of chromosomes, most commonly one set of paternal origin and the other of maternal origin. 2. An organism or cell with a double set (2n) of chromosomes (most commonly one of paternal origin, and the other of maternal origin), or referring to an individual containing a double set of chromosomes per cell. Somatic tissues of higher plants and animals are ordinarily diploid in chromosome constitution, in contrast with the haploid gametes.

DNA: (deoxyribonucleic acid; formerly spelt desoxyribonucleic acid) The long chain of molecules in most cells that carries the genetic message and controls all cellular functions in most forms of life. The information-carrying genetic material that comprises the genes. DNA is a macro-molecule composed of a long chain of deoxyribonucleotides joined by phospho-diester linkages. Each deoxyribonucleotide contains a phosphate group, the fivecarbon sugar 2-deoxribose, and a nitrogen-containing base. The genetic material of most organisms and organelles so far examined is double-stranded DNA; a number of viral genomes consist of single-stranded DNA or single-or double-stranded RNA. In double-stranded DNA, the two strands run in opposite (anti-parallel) directions and are coiled round one another in a double helix. Purine bases on one strand specifically hydrogen bond with pyrimidine bases on the other strand, according to the Watson-Crick rules (A pairs with T; G pairs with C). Hence a constant width for the double helix of 20 Å (2.0 nm) is maintained. In the B-form, DNA adopts a right-handed helical conformation, with each chain making a complete turn every 34 Å (3.4 nm), or once every ten bases.

Dominant: 1. Describing an allele whose effect with respect to a particular trait is the same in heterozygotes as in homozygotes. The opposite is recessive. 2. Describing the most conspicuously abundant and characteristic species of a community. 3. Describing an animal that is allowed priority in access to food, mates, etc., by others of its species because of its success in previous aggressive encounters.

Environment: The aggregate of all the external conditions and influences affecting the life and development of an organism.

Epistasis: Interaction between genes at different loci, e.g., one gene suppresses the effect of another gene that is situated at a different locus. Suppressed genes are said to be hypostatic. Dominance is associated with members of allelic pairs, whereas epistasis is interaction among products of non-alleles.

ESRI (Environmental Systems Research Institute): The largest GIS software company, and the maker of ArcView 3.x, ArcINFO and ArcGIS.

GIS (Geographic Information Systems): A computer system for capturing, storing, checking, integrating, manipulating, analyzing and displaying data related to positions on the Earth's surface. Typically, a Geographical Information System (or Spatial Information System) is used for handling maps of one kind or another. These might be represented as several different layers where each layer holds data about a particular kind of feature. Each feature is linked to a position on the graphical image of a map.

Geostatistics: A class of statistics used to analyze and predict the values associated with spatial or spatio-temporal phenomena. Geostatistics provides a means of exploring spatial data and generating continuous surfaces from selected sampled data points.

Gene bank: 1. The physical location where collections of genetic material in the form of seeds, tissues or reproductive cells of plants or animals are stored. 2. Field gene bank: A facility established for the ex situ storage and maintenance, using horticultural techniques, of individual plants. Used for species whose seeds are recalcitrant, or for clonally propagated species of agricultural importance, e.g. apple varieties. 3. A collection of cloned DNA fragments from a single genome. Ideally the bank should contain cloned representatives of all the DNA sequences in the genome.

Gene conservation; genetic resources conservation: The conservation of species, populations, individuals or parts of individuals, by in situ or ex situ methods, to provide a diversity of genetic materials for present and future generations.

Gene flow: The spread of genes from one breeding population to another (usually) related populations by migration, possibly to changes in allele frequency.

Gene pool: 1. The total genetic information in all the genes in a breeding population at a given time. 2. In PGR: Use is made of the concept of primary, secondary and tertiary gene pools. In general, members of a primary gene pool are inter-fertile; those of the secondary gene pool can cross with the primary gene pool under special circumstances; with the tertiary gene pool, extreme techniques are required to achieve crossing.

Genetic distance: A measure of the genetic similarity between any pair of populations. Such distance may be based on phenotypic traits, allele frequencies or DNA sequences. For example, genetic distance between two populations having the same allele frequencies at a particular locus, and based solely on that locus, is zero. The distance for one locus is maximum when the two populations are fixed for different alleles. When allele frequencies are estimated for many loci, the genetic distance is obtained by averaging over these loci.

Genetic diversity: The heritable variation within and among populations which is created, enhanced or maintained by evolutionary forces.

Genetic drift: Change in allele frequency from one generation to another within a population, due to the sampling of finite numbers of genes that is inevitable in all real (finite) populations. The smaller the population, the greater is the genetic drift. Sooner or later (depending on the size of the population), genetic drift results in loss of alleles from a population, and hence leads to a loss of genetic variation. Because of this, the minimization of genetic drift is an important consideration for conservation of genetic resources.

Genetic heterogeneity: The situation in which different mutant genes produce the same phenotype.

Genetic mapping: Determining the linear order of genes and/or DNA markers along a chromosome.

Genetic marker: A DNA sequence used to "mark" or track a particular location (locus) on a particular chromosome.

Genetic variation: Differences between individuals attributable to differences in genotypes.

Genome: 1. The entire complement of genetic material (genes + noncoding sequences) present in each cell of an organism, or in a virus or organelle. 2. A complete set of chromosomes (hence of genes) inherited as a (haploid) unit from one parent.

Genotype: (from gene + type) 1. The genetic constitution (gene makeup) of an organism. 2. The pair of alleles at a particular locus, e.g., Aa or aa. 3. The sum total of all pairs of alleles at all loci that contribute to the expression of a quantitative trait.

Germplasm: 1. The genetic material that forms the physical basis of hereditary and which is transmitted from one generation to the next by means of the germ cells. 2. An individual or clone representing a type, species or culture that may be held in a repository for agronomic, historic or other reasons.

Gluten: A mixture of two seed storage protein classes, gliadin and glutenin, found in the endosperm of cereal (particularly wheat) grain. High levels of gluten impart elasticity to dough, and thus the composition of wheat glutens largely determines whether a specific flour is suitable for biscuit or bread making. Sensitivity of the lining of the intestine to gluten in some humans results in coeliac disease, a condition that requires a gluten-free diet.

Haplotype: 1. A group of alleles, each from a different locus in the same region of a chromosome, that exist in the same double helix.

Hardy-Weinberg equilibrium: The frequencies of genotypes at a locus resulting from random mating at that locus; for two alleles, A1 and A2, with respective frequencies p and q, the Hardy-Weinberg equilibrium frequencies are p2 A1A1, 2pq A1A2, q2 A2A2. Despite the simplifying assumptions required to predict these frequencies, most loci in most populations are in Hardy- Weinberg equilibrium. Thus the Hardy-Weinberg law, which predicts these frequencies, is one of the great unifying themes of biology.

Heritability: In the narrow sense: 1. the proportion of phenotypic superiority of parents that is seen in their offspring; 2. the proportion of the total phenotypic variation due to variation in breeding values. In the broad sense: the proportion of the total phenotypic variation due to genetic variation. The degree to which a given trait is controlled by inheritance.

Heterozygote: (adj: heterozygous) (Gr. heteros, different + zygon, yoke) An individual that has different alleles at the same locus in its two homologous chromosomes.

Homozygote: An individual that has two copies of the same allele for a given gene on its two homologous chromosomes. The condition is termed "homozygous". Opposite: heterozygote.

Linkage: The tendency of non-allelic genes to be inherited together more than would be expected if they were assorting independently. Linkage exists between two loci when they are located sufficiently close on the same chromosome that some gametes are produced without crossing-over occurring between the two loci.

Maps: Graphic representation of the physical features (natural, artificial, or both) of a part or the whole of the Earth's surface, by means of signs and symbols or photographic imagery, at an established scale, on a specified projection, and with the means of orientation indicated.

Marker: An identifiable DNA sequence that facilitates the study of inheritance of a trait or a gene. Such markers are used in mapping the order of genes along chromosomes and in following the inheritance of

particular genes: genes closely linked to the marker will generally be inherited with it. Markers must be readily identifiable in the phenotype, for instance by controlling an easily observable feature (such as eye color) or by being readily detectable by molecular means, e.g., microsatellite markers.

Marker-assisted selection (MAS): The use of DNA markers to increase the response to selection in a population. The markers will be closely linked to one or more quantitative trait loci.

Mean: In statistics, the arithmetic average; the sum of all measurements or values in a sample divided by the sample size.

Median: In a set of measurements, the central value above and below which there are an equal number of measurements.

Mendelian population: A natural, interbreeding unit of sexually reproducing plants or animals sharing a common gene pool.

Mendelism: The theory of heredity that forms the basis of classical genetics, proposed by Gregor Mendel in 1866 and formulated in two laws.

Mendel's Laws: Two laws summarizing Gregor Mendel's theory of inheritance. The Law of Segregation states that each hereditary characteristic is controlled by two 'factors' (now called alleles), segregate and pass into separate germ cells. The Law of Independent Assortment states that pairs of 'factors' segregate independently of each other when germ cells are formed.

Microsatellite: Segment of DNA characterized by the occurrence of a variable number of copies (from a few up to 30 or so) of a sequence of around 5 or fewer bases (called a repeat unit). A typical microsatellite is the repeat unit AC, which occurs at approximately 100 000 different sites in a typical mammalian genome. At any one site (locus), there are usually several different "alleles," each identifiable according to the number of repeat units. These alleles can be detected by PCR, using primers designed from the unique sequence that is located on either side of the microsatellite. When the PCR product is run on an electrophoretic gel, alleles are seen to differ in length in units equal to the size of the repeat unit. Microsatellites have been the standard DNA marker: they are easily detectable by PCR, and they tend to be evenly located throughout the genome. Thousands have been mapped in many different species.

Molecular biology: The area of knowledge concerned with the molecular aspects of organisms and their cells.

Molecular genetics: The area of knowledge concerned with the genetic aspects of molecular biology, especially with DNA, RNA and protein molecules.

Multiple alleles: The existence of more than two alleles at a locus in a population.

Phenotype: (Gr. phaneros, showing + type). The visible appearance or set of traits of an organism resulting from the combined action of genotype and environment.

Plant genetic resources (PGR): Defined in the International Undertaking on Plant Genetic Resources (FAO, 1983) to mean the reproductive or vegetative propagating material of the following categories of plants: (i) cultivated varieties (cultivars) in current use and newly developed varieties; (ii) obsolete cultivars; (iii) primitive cultivars (landraces); (iv) wild and weed species, near relatives of cultivated varieties; and (v) special genetic stocks (including elite and current breeder's lines and mutants).

Population genetics: The branch of genetics that deals with frequencies of alleles and genotypes in breeding populations.

Quantitative genetics: The area of genetics concerned with the inheritance of continuously-varying traits. Most practical improvement programs involve the application of quantitative genetics.

Quantitative trait: A measurable trait that shows continuous variation; a trait that can not be classified into a few discrete classes.

Quantitative trait locus (QTL) : A locus that affects a quantitative trait. The plural form (quantitative trait loci) is also abbreviated as QTL.

Spatial data: Any information about the location and shape of, and relationships among, geographic features. This includes remotely sensed data as well as map data.

Spike: (L. spica, an ear of grain) An inflorescence in which the main axis is elongated and the flowers are sessile.

Spikelet: (L. spica, an ear of grain + diminutive ending -let) The unit of inflorescence in grasses; a small group of grass flowers.

Standard deviation: A statistical measure of variability in a population of individuals or in a set of data; the square root of the variance.

Standard error: A statistical measure of variation in a population of means, used to indicate how well sample estimates represent population parameters.

Shapefile: A vector file format for storing the location, shape, and attributes of geographic features.

Stress: Non-optimal conditions for growth. Stresses may be imposed by biotic (pathogens, pests) or abiotic (environment, such as heat, drought etc.) factors.

Tetraploid: An organism whose cells contain four haploid (4x) sets of chromosomes.

Variety: A naturally occurring subdivision of a species, with distinct morphological characters and given a Latin name according to the rules of the International Code of Nomenclature. A taxonomic variety is known by the first validly published name applied to it so that nomenclature tends to be stable.

Suplementary table: Landraces information

Name	Collection site	Origin	Type	Collection date	province	Altitude	Longitude	Latitude
ICDW20036	Guelmine	MAR	LA	1985/05/03	Tiznit	200	W10 04	N28 56
ICDW20037	Near Asrir	MAR	LA	1985/05/03	Tiznit	200	W010 00	N28 55 41
ICDW20038	5 km S of Guelmine	MAR	LA	1985/05/03	Tiznit	200	W10 04	N28 55
ICDW20039	Tata main oasis	MAR	LA	1985/05/05	Tiznit	700	W08 00	N29 46
ICDW20041	Kasba-ej-Joua, village stack	MAR	LA	1985/05/05	Tiznit	600	W07 38	N29 50
ICDW20042	Tanskit	MAR	LA	1985/05/07	Ouarzazate	850	W006 12 02	N30 41 33
ICDW20043	Just W of Tinejdad	MAR	LA	1985/05/08	Er Rachidia	900	W005 00 54	N31 30 54
ICDW20045	Mellah	MAR	LA	1985/05/08	Beni Mellal	800	W006 48 51	N31 58 48
ICDW20046	Fezna	MAR	LA	1985/05/08	Er Rachidia	740	W04 28	N31 32
ICDW20047	3 km S of Aoufouss;	MAR	LA	1985/05/09	Er	750	W04 10	N31 39

	outside main oasis				Rachidia			
ICDW20048	5 km E of Boulaouane	MAR	CV	1985/05/12	El Jadida	150	W08 03	N32 59
ICDW20050	15 km W of Sidi Bennour	MAR	CV	1985/05/12	El Jadida	100	W08 25	N32 59
ICDW20052	Akermould	MAR	LA	1985/05/13	Agadir	1	W09 37	N31 40
ICDW20053	Ain-el-Hajer, near Akermoud	MAR	LA	1985/05/13	Agadir	1	W09 37	N31 40
ICDW20054	Ain-el-Hajer, near Akermoud	MAR	LA	1985/05/13	Agadir	1	W09 37	N31 40
ICDW20055	2 km E of Smimou	MAR	LA	1985/05/13	Tiznit	300	W09 07 33	N30 47 05
ICDW20056	Tnine Sidi el Yamani	MAR	LA	1985/05/13	Tetouan	400	W05 56	N35 23
ICDW20057	Tnine Sidi el Yamani	MAR	LA	1985/05/13	Tetouan	300	W005 47 29	N35 22
ICDW20058	Ounara	MAR	LA	1985/05/13	Agadir	150	W009 43 06	N31 20 47
ICDW20059	15 km N of Chichaoua	MAR	LA	1985/05/14	Marrakech	200	W008 46 53	N31 33 04
ICDW20060	30 km S of Chemaia	MAR	LA	1985/05/14	Agadir	200	W08 37	N31 49
ICDW20061	Chemaia	MAR	LA	1985/05/14	Agadir	300	W08 38	N32 05
ICDW20062	10 km W of Ben Guerir	MAR	LA	1985/05/14	Marrakech	400	W07 59	N32 19
ICDW20063	50 km S of Borouj	MAR	LA	1985/05/14	Marrakech	300	W 07 10	N32 30
ICDW20064	20 km N of Oued Zem	MAR	LA	1985/05/14	Khouribga	750	W06 33	N32 55
ICDW20065	El-Kbab	MAR	LA	1985/05/15	Khenifra	900	W05 31	N32 44
ICDW20066	Mengoub	MAR	LA	1985/05/16	Bouarfa	900	W02 21	N32 15
ICDW20067	Figuig oasis	MAR	LA	1985/05/16	Bouarfa	800	W01 15	N32 10
ICDW20068	Ahfir	MAR	LA	1985/05/17	Oujda	200	W02 14 18	N35 03 59
ICDW20069	Zaio	MAR	CV	1985/05/17	Oujda	150	W02 44	N34 57
ICDW20070	Oulda Berrehil; just W of Aoulouz	MAR	LA	1985/07/06	Tiznit	500	W008 09 20	N30 42 06
ICDW20071	Oulda Berrehil; just W of Aoulouz	MAR	LA	1985/07/06	Tiznit	500	W008 09 20	N30 42 06
ICDW20072	Oulda Berrehil; just W of Aoulouz	MAR	LA	1985/07/06	Tiznit	500	W008 09 20	N30 42 06
ICDW20073	Tessouert; 10 km SW of Ijoukak	MAR	LA	1985/07/06	Marrakech	1400	W08 06	N30 58
ICDW20074	Ait Barka near Toufilat	MAR	LA	1985/07/07	Tiznit	1400	W09 09	N30 07
ICDW20075	15 km N of Tedders	MAR	LA	1985/07/07	Meknes	1400	W06 17	N33 42
ICDW20076	15 km N of Tedders	MAR	LA	1985/07/07	Meknes	1400	W06 17	N33 42
ICDW20077	Tizouggart; 10 km N of Tedders	MAR	LA	1985/07/07	Meknes	1500	W06 17	N33 40
ICDW20078	Aguelmous near Agoudal	MAR	LA	1985/07/07	Khenifra	2200	W007 22 48	N31 15 36
ICDW20079	Isfoutelil Oasis; 7 km NW of Ourzazat	MAR	LA	1985/07/07	Ouarzazate	1300	W06 51	N30 58
ICDW20080	5 km S of Ait Hani	MAR	LA	1985/07/08	Er Rachidia	1900	W005 29 53	N31 45 09
ICDW20081	Ait Hani; flat	MAR	LA	1985/07/08	Er Rachidia	2000	W05 30	N31 48
ICDW20082	16 km N of Imilchil	MAR	LA	1985/07/08	Beni	2000	W05 40	N32 14

					Mellal			
ICDW20083	near Imilchil, in protected forest area, by stream	MAR	LA	1985/07/08	Er Rachidia	1900	W05 40	N32 08
ICDW20084	ca. 5 km N of site 154	MAR	LA	1985/07/08	Er Rachidia	1600	W005 38 53	N32 16 23
ICDW20085	Arhbala	MAR	LA	1985/07/08	Beni Mellal	1700	W05 39	N32 29
ICDW20086	ca. 20 km W of Arhbala, in clearing in oak forest	MAR	LA	1985/07/08	Beni Mellal	1700	W05 45	N32 29
ICDW20087	El-Ksiba	MAR	LA	1985/07/08	Tanger	1400	W05 56	N35 41
ICDW20088	near Ouaouizarht	MAR	LA	1985/07/09	Beni Mellal	1300	W06 21 02	N32 09 59
ICDW20089	Just S of Ouaouizarht	MAR	LA	1985/07/09	Beni Mellal	1000	W06 21 02	N32 09 59
ICDW20090	S of Ouaouizarht (S of site 161)	MAR	LA	1985/07/09	Beni Mellal	1000	W06 21 02	N32 09 59
ICDW20091	Ait Simour	MAR	LA	1985/07/09	Marrakech	1650	W09 13	N31 22
ICDW20092	25 km N of Tilouguitte	MAR	LA	1985/07/09	Beni Mellal	1850	W005 39 36	N32 16 12
ICDW20093	Tilouguit	MAR	LA	1985/07/09	Beni Mellal	1400	W06 13	N32 02
ICDW20094	10 km N of Zaouia Ahansal	MAR	LA	1985/07/09	Beni Mellal	1200	W06 07	N31 57
ICDW20095	2 km W of Zaouia Ahansal	MAR	LA	1985/07/09	Beni Mellal	1300	W06 08	N31 51
ICDW20096	Just N of Zaouia	MAR	LA	1985/07/09	Tetouan	1600	W005 05 28	N35 16 59
ICDW20097	Zaouia	MAR	LA	1985/07/09	Tetouan	1500	W005 05 32	N35 16 59
ICDW20098	near Azilal	MAR	LA	1985/07/10	Beni Mellal	1400	W06 35	N31 59
ICDW20099	AitTagelou	MAR	LA	1985/07/10	Ouarzazate	1200	W05 20	N30 37
ICDW20100	Ouzoud	MAR	LA	1985/07/10	Beni Mellal	900	W06 47	N32 02
ICDW20101	Just E of Tanannt	MAR	LA	1985/07/10	Marrakech	1200	W06 52	N31 50
ICDW20102	Ait-Ourir	MAR	LA	1985/07/10	Marrakech	700	W07 38	N31 33
ICDW20103	Jamait Agoumat	MAR	LA	1985/07/10	Ouarzazate	1000	W06 35	N31 01
ICDW20104	Ouriki; Marrakech oasis	MAR	LA	1985/07/10	Marrakech	900	W08 00	N31 49
ICDW20105	Sidi Abbou	MAR	LA	1985/07/13	Meknes	800	W06 14	N33 28
ICDW20106	Ouelmes (Oulmes)	MAR	LA	1985/07/13	Meknes	1100	W005 59 52	N33 25 48
ICDW20107	Ouelmes (Oulmes)	MAR	LA	1985/07/13	Meknes	1100	W005 59 52	N33 25 48
ICDW20108	Aguelmous; 30 km SE of Ouelmes	MAR	LA	1985/07/13	Meknes	1200	W05 57	N33 25
ICDW20110	ca. 25 km E of Khenifra	MAR	LA	1985/07/13	Khenifra	1400	W05 58	N33 00
ICDW20111	Ain Roubea; 2-3 km E of Khenifra	MAR	LA	1985/07/13	Khenifra	1400	W05 39	N33 00
ICDW20112	just past Agouelmane springs	MAR	LA	1985/07/13	Tetouan	1500	W05 24	N35 24
ICDW20113	15 km S of Ain Leuh	MAR	LA	1985/07/13	Khenifra	1700	W05 23	N33 09
ICDW20114	Boulemane	MAR	LA	1985/07/14	Fes	1700	W04 45	N33 22

ICDW20115	Ifrane	MAR	LA	1985/07/14	Khenifra	1750	W05 10	N33 31
ICDW20116	Ait Makhlouf	MAR	LA	1985/07/14	Fes	1400	W04 20	N33 26
ICDW20118	Teggour oasis; S of Moyen Atlas	MAR	LA	1985/07/14	Taza	800	W03 50	N33 10
ICDW20119	Taddint Oasis; 25 km SW of Ouled El-Haj	MAR	LA	1985/07/14	Ouarzazate	700	W06 02	N30 14
ICDW20120	Tarileet; 20 km SW of Midar	MAR	LA	1985/07/15	Oujda	1000	W03 30	N34 55
ICDW20121	Tizi Ousli	MAR	LA	1985/07/15	Taza	1300	W03 47	N34 46
ICDW20122	Tizi Ousli	MAR	LA	1985/07/15	Taza	1300	W003 47 23	N34 45 55
ICDW20123	Just N of Aknoul	MAR	LA	1985/07/15	Taza	1200	W03 49	N34 43
ICDW20124	Aknoul	MAR	LA	1985/07/15	Taza	1200	W03 49	N34 43
ICDW20125	near Boured	MAR	LA	1985/07/15	Taza	1400	W04 06	N34 45
ICDW20126	Nahnach; between Boured & Tahar Souk	MAR	LA	1985/07/15	Taza	1100	W04 06	N34 45
ICDW20127	between Boured and Taher Souk	MAR	LA	1985/07/15	Taza	1100	W04 06	N34 45
ICDW20128	Ain al Beida; near Tahar Souk	MAR	LA	1985/07/15	Taza	1100	W04 08	N34 26
ICDW20129	Ain lemn; 10 km W of Taher Souk	MAR	LA	1985/07/15	Taza	500	W04 09	N34 22
ICDW20130	15 km W of Taher Souk	MAR	LA	1985/07/15	Taza	500	W04 12	N34 21
ICDW20131	Imarzen; 5 km N of Taounate	MAR	LA	1985/07/16	Fes	600	W004 37 58	N34 26
ICDW20132	30 km SE of El-Jebha	MAR	LA	1985/07/16	Al Hoceima	1400	W04 29	N34 57
ICDW20133	Bou Ahmed, river flood plain	MAR	LA	1985/05/08	Tetouan	1	W04 58	N35 19
ICDW20134	near Sebta Beni Zarfet, hill by sea	MAR	LA	1985/07/17	Tetouan	100	W05 50 09	N35 15 58
ICDW20135	10 km W of Sebt Beni Zerfet	MAR	LA	1985/07/17	Tetouan	200	W005 51 31	N35 15 34
ICDW20136	Ksar Sghir	MAR	LA	1985/07/17	Tetouan	1	W05 34	N35 50
ICDW20137	10 km E of Tinejdad	MAR	LA	1985/05/08	Er Rachidia	900	W004 57	N31 31 48
ICDW20138	Iguirene Brahim ou Brahim	MAR	LA	1985/07/17	Tiznit	800	W10 05	N29 17
ICDW20139	Oulda Berrehil; just W of Aoulouz	MAR	LA	1985/07/06	Tiznit	500	W008 09 20	N30 42 06
ICDW20140	S of Ouaouizarht (S of site 161)	MAR	LA	1985/07/09	Beni Mellal	1000	W06 21 02	N32 09 59
ICDW20141	Almis de Marmoucha	MAR	LA	1985/07/14	Fes	1700	W04 08 47	N33 20 05
Sourie	6 km W Al Hafa	SYR	LA	1987/06/05	Lattakia	90	E35 57 50	N35 33 15
Sourie haririe	8 km W Silifreh	SYR	LA	1987/06/05	Lattakia	880	E36 06 22	N35 35 37
Souedie	El Morioniaet	SYR	LA	1987/06/05	Lattakia	640	E36 07	N35 42
Sourie	Al Hamam	SYR	LA	1987/06/05	Hama	330	E36 15	N35 34
Souedie	Jabal Al Ghab	SYR	LA	1987/06/05	Hama	700	E36 14	N35 33
Souedie	5 km down Jeb Ahmar	SYR	LA	1987/06/05	Hama	730	E036 13 25	N35 37

								04
Ahmar	Kasab	SYR	LA	1987/06/05	Lattakia	650	E35 59 00	N35 55 45
Souedie	Al Meshrefe	SYR	LA	1987/06/05	Lattakia	310	E035 54 36	N35 52 48
Souedie	Zghreirien	SYR	LA	1987/06/05	Lattakia	40	E35 53 58	N35 44 00
Baladi	Shabat Lieh	SYR	LA	1987/06/05	Lattakia	80	E35 49 45	N35 41 30
Tunsie	Burj El-Kasab	SYR	LA	1987/06/05	Lattakia	30	E35 47 00	N35 36 25
Souedie	Bahlulieh	SYR	LA	1987/06/05	Lattakia	45	E35 57 30	N35 38 00
Tunsie	Khan Zaarur	SYR	LA	1987/06/05	Lattakia	60	E36 02 13	N35 40 05
Haririe	Awienat	SYR	LA	1987/06/05	Lattakia	150	E36 05 35	N35 43 10
Haririe	1 km E Bdama	SYR	LA	1987/06/05	Idlib	360	E36 12 20	N35 48 15
Hamari	2 km S Jisr El Shughour	SYR	LA	1987/06/05	Idlib	475	E36 18 44	N35 47 24
Souedie	2 km S Jisr El Shughour	SYR	LA	1987/06/05	Idlib	475	E36 18 44	N35 47 24
Biadi	Sha'ieraat	SYR	LA	1987/06/05	Homs	800	E37 00	N34 29
Biadi	30 km SE Ka'a Luly	SYR	LA	1987/06/05	Homs	520	E39 15 51	N34 34 40
Biadi	Palmyra; 30 km SE Sukhnah	SYR	LA	1987/06/05	Homs	450	E39 03	N34 43
Biadi	Palmyra; 21 km SE Sukhnah	SYR	LA	1987/06/05	Homs	530	E39 02 00	N34 45 20
ID318	6 km East Palmyra	SYR	LA	1987/06/05	Homs	460	E38 20 13	N34 34 01
Biadi	Hamam	SYR	LA	1987/06/05	Raqqa	320	E38 46 30	N35 54 10
Biadi	Mansura; 12 km South	SYR	LA	1987/06/05	Raqqa	330	E38 44 30	N35 50 35
Hamari	Safsafeh	SYR	LA	1987/06/05	Tartous	160	E36 03 25	N34 44 00
Baladi	Karfas	SYR	LA	1987/06/05	Tartous	510	E36 07 20	N34 57 20
Baladi	Dahr Al Mahshleh	SYR	LA	1987/06/05	Tartous	380	E36 03 00	N34 54 45
Baladi akraa	Askabouli	SYR	LA	1987/06/05	Tartous	190	E35 56 10	N34 54 50
Baladi hreidini souri	Brmaneh Road	SYR	LA	1987/06/05	Tartous	520	E36 08 45	N35 00 00
Hamari	Ram Al Aoz	SYR	LA	1987/06/05	Homs	610	E36 31 05	N34 44
Biadi	Jnan	SYR	LA	1987/06/05	Hama	350	E36 50 10	N35 04 45
Sherieh	Al Swireh	SYR	LA	1987/06/05	Homs	670	E36 28 00	N34 44 55
Hamari abasie	Al Swireh	SYR	LA	1987/06/05	Homs	670	E36 28 00	N34 44 55

Abasie	Arqayah	SYR	LA	1987/06/05	Homs	560	E36 28 40	N34 48 50
Souedi	Fahel	SYR	LA	1987/06/05	Homs	590	E36 24 25	N34 50 50
Souedi	Tall Douw	SYR	LA	1987/06/05	Homs	410	E36 31 30	N34 52 40
Souedi abasie	Tall Douw	SYR	LA	1987/06/05	Homs	410	E36 31 30	N34 52 40
Biadi hamari	Kafr Nan	SYR	LA	1987/06/05	Homs	450	E36 38 30	N34 53 15
Baladi	Tall Hasan Basha	SYR	LA	1987/06/05	Homs	550	E37 04 15	N34 54 45
Baladi	Tall Jadid	SYR	LA	1987/06/05	Homs	640	E37 15 00	N34 55 45
Baladi	Tall Al Tot	SYR	LA	1987/06/05	Hama	550	E37 08 56	N34 59 01
Biadi	Moshrefe	SYR	LA	1987/06/05	Homs	520	E36 51 55	N34 50 00
Souedi	Al Mentar	SYR	LA	1987/06/05	Idlib	350	E036 26 24	N36 08 42
Souedi	Frikeh	SYR	LA	1987/06/05	Idlib	300	E36 21 50	N35 45 30
Chamie	Baglied	SYR	LA	1987/06/05	Idlib	280	E36 28	N36 08
Souedie	Armanaz	SYR	LA	1987/06/05	Idlib	260	E36 30 10	N36 05 00
Souedi	Hafasraja	SYR	LA	1987/06/05	Idlib	500	E36 31 30	N36 01 00
Haurani	Al Ra'i	SYR	LA	1987/06/05	Aleppo	610	E37 26 55	N36 37 00
Haurani	Susnabat	SYR	LA	1987/06/05	Aleppo	540	E37 28 55	N36 28 50
Jori abiad	Deir Qaq	SYR	LA	1987/06/05	Aleppo	470	E37 26 52	N36 18 50
Baladi biadi	Mare'	SYR	LA	1987/06/05	Aleppo	490	E37 12 00	N36 29 20
Hamari	Azaz	SYR	LA	1987/06/05	Aleppo	590	E37 03 21	N36 35 08
Hamari	Katmeh	SYR	LA	1987/06/05	Aleppo	600	E36 57 15	N36 35 25
Baladi	Abbeen	SYR	LA	1987/06/05	Aleppo	490	E36 59 30	N36 28 15
Baladi	Kifin	SYR	LA	1987/06/05	Aleppo	480	E37 01 50	N36 24 50
Souedie	Barisha	SYR	LA	1987/06/05	Idlib	610	E36 38 00	N36 11 00
Souedie	Sarmada	SYR	LA	1987/06/05	Idlib	440	E36 42 48	N36 11 27
Baladi	Tall Hasil	SYR	LA	1987/06/05	Aleppo		E37 18 22	N36 08 05
Baladi	Rasm Al Sheikh	SYR	LA	1987/06/05	Aleppo	540	E37 14 40	N36 01 45
Biadi	Blass	SYR	LA	1987/06/05	Aleppo	440	E37 09 40	N36 00 20
Baladi	Shukidleh	SYR	LA	1987/06/05	Aleppo	420	E37 05 20	N36 01

								20
Shihani	Tall Nasri	SYR	LA	1987/06/05	Al Hasakah	410	E40 21 56	N36 37 03
Shihani	Rehikeh	SYR	LA	1987/06/05	Al Hasakah	540	E40 46	N37 04
Sin al jamal	Karam Koq	SYR	LA	1987/06/05	Al Hasakah	560	E40 49 34	N37 06 19
Shihani	Salam Alek	SYR	LA	1987/06/05	Al Hasakah	560	E40 33 00	N37 01 09
Biadi	Al Asadia village	SYR	LA	1987/06/05	Al Hasakah	520	E40 20 20	N36 56 00
Shihani	Kherbet Al Jamal	SYR	LA	1987/06/05	Al Hasakah	420	E40 51 06	N36 45 03
Sin al jamal	Tall Khas	SYR	LA	1987/06/05	Al Hasakah	490	E40 37 15	N36 57 55
Shihani	Ghweitly	SYR	LA	1987/06/05	Al Hasakah	460	E40 40	N36 55
Halabi	Tall Bedar village	SYR	LA	1987/06/05	Al Hasakah	420	E40 34 56	N36 44 12
Shihani kandahari	Tall Bedar village	SYR	LA	1987/06/05	Al Hasakah	420	E40 34 56	N36 44 12
Jouda	Salu Regional Research Station	SYR	LA	1987/06/05	Der Ezzor	230	E40 20 19	N35 08 35
Mouserieh	Salu Regional Research Station	SYR	LA	1987/06/05	Der Ezzor	230	E40 20 19	N35 08 35
Biadi	Salu Regional Research Station	SYR	LA	1987/06/05	Der Ezzor	230	E40 20 19	N35 08 35
Haurani	Kamuneh	SYR	LA	1987/06/05	Damascus	750	E36 14 40	N33 14 30
Haurani Salamie	Moadamieh	SYR	LA	1987/06/05	Damascus	860	E36 38 15	N33 44 25
Haurani	Al Hakef	SYR	LA	1987/06/05	Sweida	870	E36 42 25	N33 00 05
Haurani	Al Sura Al Kubra; 11 km E the village	SYR	LA	1987/06/05	Sweida	750	E36 39	N33 08
Haurani	Nawa village	SYR	LA	1987/06/05	Dar'a	670	E36 02 30	N32 53 35
Haurani	Ghabagheb	SYR	LA	1987/06/05	Dar'a	780	E36 14 00	N33 10 15
Haurani	Trunje	SYR	LA	1987/06/05	Qunaytirah	1060	E35 51 00	N33 13 55
Haurani	Danun; 6 km south	SYR	LA	1987/06/05	Damascus	800	E36 13 40	N33 16 40
Haurani	Jeb Al Safa	SYR	LA	1987/06/05	Damascus	750	E36 18 30	N33 14 00
Haurani	Jadal	SYR	LA	1987/06/05	Sweida	780	E36 22 40	N32 58 20
Haurani	Orika	SYR	LA	1987/06/05	Dar'a	850	E36 28	N32 23
Haurani	Bosra	SYR	LA	1987/06/05	Dar'a	950	E36 29 36	N32 31 16
Haurani	Bosra; 6 km east	SYR	LA	1987/06/05	Sweida	780	E36 33 16	N32 30 36
Haurani	Welgha	SYR	LA	1987/06/05	Sweida	930	E36 31 15	N32 44 40

Haurani	Breeka village	SYR	LA	1987/06/05	Sweida	990	E36 34 10	N32 50 30
Zaraa	Rudimma	SYR	LA	1987/06/05	Sweida	840	E36 34 35	N33 01 15

1