DISSECTING SEED SHAPE IN LENTIL (*LENS CULINARIS* MEDIK.) WITH HIGH-THROUGHPUT PHENOTYPING AND GENOME-WIDE ASSOCIATION STUDIES

A Thesis Submitted to the College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements for the Degree of Master of Sciences
In the Department of Plant Sciences
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

By
Sebastian I. Olivares Herrera

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department of Plant Sciences or the Dean of the College of Agriculture and Bioresources in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Plant Sciences
College of Agriculture and Bioresources
University of Saskatchewan
Room 4D36, Agriculture Building 51 Campus Drive
Saskatoon, Saskatchewan S7N 5A8 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Seed shape is an important trait for dehulling and splitting during the milling process of lentil. A high-throughput image-based phenotyping system was previously developed to provide direct metrics from single seeds. This study aimed to take this new, and more precise system, to phenotype and dissect lentil seed shape and identify new genetic markers for the lentil breeding program. To accomplish this, seed from 324 diverse genotypes that had been grown at six site-years in two major lentil growing environments (temperate and Mediterranean) were imaged. Significant differences were identified between seeds from temperate and Mediterranean macro-environments for four seed shape parameters: diameter, circularity (measure of uniformity of the seed edges), height (seed thickness), and plumpness. Each seed parameter had high heritability, suggesting a low environmental influence, which is also supported by an absence of significant correlations with phenological parameters or temperature. A stability analysis of the seed shape traits across the six site-years revealed that seed height has fewer stable lines when compared with circularity, diameter, and plumpness and that for each of these traits there is a range of stability, making some genotypes more stable than others across site-years. Genome-wide association studies were used to identify QTL for most seed shape traits, with diameter having the most consistent identification of significant QTL. The new phenotyping system used in this study coupled with multi-locus GWAS models achieved similar results when compared with previous studies, although it can help breeding programs identify new candidate genes with improved precision by extracting multiple traits like seed shape and seed color simultaneously.

Acknowledgements

I would like to express most sincere gratitude to my supervisor Dr. Kirstin Bett. Her support and guidance throughout my entire experience as a graduate student has helped to drive my academic development and aided tremendously in the completion of this thesis. I am grateful to my advisory committee, Dr. Scott Noble, Dr. Aaron Beattie, and Dr. Rosalind Bueckert, for their valuable suggestions and constructive criticism throughout this research.

I would like to sincerely thank Derek Wright, Keith Halcro and Sandesh Neupane for all their suggestions and assistance with data analysis and phenotyping with BELT. I would also like to thank my fellow graduate students for their support, suggestions, and friendship throughout this experience.

I sincerely acknowledge the financial support provided for this research by the Saskatchewan Pulse Growers, Genome Prairie, Genome Canada, the Government of Saskatchewan, Canada First Research Excellence Fund, Western Grains Research Foundation, Application of Genomics to Innovation in the Lentil Economy (AGILE) and Enhancing the Value of Lentil Variation for Ecosystem Survival (EVOLVES). I would also like to show my most sincere appreciation to the family of Roderick Alan McLean for the support I received through the Roderick Alan McLean Postgraduate Scholarship in Crop Science during my studies.

Dedication

To my family, friends, and peers for their support and encouragement throughout my graduate studies. To my grandpa and my grandmas, for instilling the love and appreciation of life and nature which has led to my interest in science.

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

AGILE	Application of Genomics to Innovation in the Lentil Economy
AGILE LDP	Lentil diversity panel develop in the AGILE project
AM	Association mapping
BELT	Portable imaging system
BLINK	Bayesian information and LD iteratively nested keyway
DTE	Days to emergence
DTF	Days to flowering
DTM	Days to maturity
DTS	Days to swollen pods
FarmCPU	Fixed and random model circulating probability unification
GLM	General linear model

ICARDA	International Centre for Agricultural Research in the Dry Areas
LD	Linkage disequilibrium
MLM	Mixed linear model
MLMM	Multi-locus mixed model
PCA	Principal component analysis
QTL	Quantitative trait loci
REP	Reproductive period
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
TFT	Thermal flowering time
TRT	Thermal reproductive time
VEG	Vegetative period
GWAS	Genome-wide association study

1 Introduction

Lentil (*Lens culinaris* spp. *culinaris* Medik.) is a crop consumed for its high levels of protein, fiber, complex carbohydrates, vitamins, and minerals. For almost 40 years lentil has been the second most produced pulse crop in Canada after dry peas, reaching more than 2.8 million tonnes of production in 2020 (FAOSTAT 2023). Quality of seeds is a key aspect to lentil marketing and seed shape is among the pivotal parameters of quality in lentil markets. Seed shape is understood to include seed diameter, mass, thickness, plumpness, and roundness, and has been proved to be important in cooking, and dehulling, and splitting during milling processes (Shahin et al. 2012; Subedi et al. 2018). Environment plays an important role in seed shape control; although it is important to assess these parameters in conjunction with phenology which provides useful data of plants adaptation to climatic variables (Alghamdi et al. 2013; Fedoruk et al. 2013; Wright et al. 2020).

Diverse genetic resources are needed to enhance genetic variability in size and shape to release new varieties of importance to the lentil market (Khazaei et al. 2016; Al-Khayri et al. 2019). To capture new genes and/or alleles, diversity panels can be used to represent a range of genetic diversity in crops and have been a useful tool for plant breeders to manage and use germplasm for crop improvement (Wang et al. 2014). A diverse selection of seeds from 324 accessions (AGILE LDP) grown in Saskatchewan, Italy, Spain, and Morocco in 2017-2019 and then returned to Saskatoon were available for seed quality analyses. The diversity captured with this set of germplasm and from different macro-environments makes the scope of research bigger and offers the opportunity for more robust results. The differences between the macro-environments come from daylength, daily temperatures and the season in which plants were growing - there are some winter-growing macro-environments, like locations from the Mediterranean, and the others are summer-grown temperate environments (Wright et al. 2020). There are changes in the phenological events such as flowering and maturity across these locations due to the changes in daylength and temperature. These events have been linked to seed growth that impacts directly seed size and shape, especially during early growth (Li et al. 2019).

A genome-wide association study (GWAS) is an association mapping (AM) technique that uses single or multiple loci tests to associate target traits with markers (Singh and Singh 2015a). The diversity panel has been previously used in a genome-wide association study (GWAS) for

phenological parameters and is available to help interpret results. Accurate seed phenotyping is a prerequisite to a valid genetic conclusion. High-throughput genotyping and phenotyping have enabled genomics-assisted breeding to help accelerate generation of new varieties (e.g., Leng et al. 2017). A high-throughput image-based phenotyping system was developed by Halcro et al. (2020), which includes an imaging system, BELT, and the image processing software, phenoSEED. This combination can be used to phenotype seed shape based on top and side views of single seeds and provides metrics describing the sample based on the averages of the single seed data.

The main objective of this project was to use this new method of phenotyping seed shape to better understand the genetics of seed shape and how it interacts with the growing environment using the AGILE LDP seeds from different growing regions.

1.1 Hypotheses

1. There will be insignificant variance in seed shape phenotypes across replications within a site-year.

2. Significant correlations will exist among seed shape traits, and between seed shape traits and phenological and environmental parameters.

3. Most of the accessions will be stable across all site-years for each seed shape trait.

4. New genomic regions, significantly associated with seed shape, will be detected via GWAS.

1.2 Objectives

1. Phenotype seed shape of the AGILE LDP using BELT and phenoSEED.

2. Evaluate variation in seed shape in the lentil diversity panel (AGILE LDP) across replications, genotypes, and site-years to establish if it is possible to use only a single replication to represent each site-year.

3. Identify the main environmental and phenological parameters that are correlated with seed shape parameters.

4. Determine the stability of the genotypes across macro-environments and the heritability of seed shape traits.

5. Identify clusters based on seed shape traits from each site-year and determine what seed characteristics determine each cluster.

6. Identify genomic regions involved in seed shape control via a GWAS.

2 Literature review

2.1 Lentil Origin and Domestication

Lentil was among the first domesticated grain legumes; originating in the Middle East (Ladizinsky, 1993; Zohary, 1999). Archaeological studies have suggested that lentils were first domesticated in the agricultural communities of the Fertile Crescent within the Middle East (from today's southeast Turkey and northeast Iran to northeast Jordan and northwest Egypt) around 5000 B.C. and then lentils spread west to Greece and east to India (Ferguson et al. 1998; Erskine and Sarker, 2004; Erskine et al. 2009). Then lentils spread to Central Asia and Mediterranean regions and finally arrived in North America to the northwest USA in 1930s and to Canada in the Richlea, Saskatchewan area in 1969 (Muehlbauer et al. 1992; Slinkard and Vandenberg 1995). Lentil has been classified into two major groups: microsperma and macrosperma. The microsperma seed type was characterized with a diameter of 2 to 6 mm. The macrosperma seed type was characterized by a diameter of 6 to 9 mm (Barulina 1930; Sandhu and Singh, 2007). Today, lentil is grown in three main climatic regions or macro-environments around the world: Mediterranean, sub-tropical, and temperate, with contrasting daylength and temperature regimes during the crop life cycle (Tullu et al. 2011; Khazaei et al. 2016).

2.2 Lentil Seed Quality

Seed quality improvement is a pivotal objective for lentil breeders. Seed quality can be determined through the assessment of characteristics such as cooking time, colour, shape, taste, and absence of defective seeds (Tiwari et al. 2021). Seed shape comprises diameter, mass, thickness, plumpness, and roundness of the seeds.

Seed diameter has a direct and positive correlation with cooking time (Erskine et al. 1985; Hamdi et al. 1991). Seed quality traits including seed size and shape have been reported to be among the main influences in milling yield in pulse crops (Wood and Malcolmson, 2011). The milling process in pulses usually consists of dehulling, or removal of seed coat, and splitting of cotyledons. After the milling process, there could be two cotyledons (splits) or unsplit cotyledons (footballs); both products are also referred to as Dhal (Wood and Malcolmson, 2011; Vishwakarma et al. 2018; Subedi et al. 2018). Uniformity in size and rounder seeds are the ideal traits for the milling process (Wood et al. 2008). Milling quality of lentil is important to the lentil industry due to the importance of milled red lentil (Dhal) in the eastern Mediterranean regions

(Egypt, Syria, and Turkey) and, in one of the main markets for Canadian lentil producers, India (Vandenberg, 2009; Subedi et al. 2018; FAOSTAT, 2023).

2.3 Environmental and Phenological Factors Involved in Seed Quality

For grain crops, seed size has been linked to photoperiod, precipitation, and temperature during the growing season, among other environmental factors. In general, these environmental factors directly impact morphological traits from the vegetative or reproductive organs and indirectly seed traits. Environmental factors play a pivotal role in triggering the molecular signals for the transition from vegetative period (VEG) to reproductive period (REP), these developmental phases have an impact in seed size, maturity, filling, and yield (Roberts et al. 1986; Gupta et al. 2006).

Understanding seed development is key to dissect seed shape genetically and elucidate how environment modifies development. Seed development in legumes starts with tissue-specific differentiation during early seed growth when the maternal seed coat regulates the cell division phase by modification of control and supply of nutrients to the embryo (Li et al. 2019). After the cell division phase, the embryo enters a transition phase (from cell division into cell expansion) in which the embryo transitions from a non-differentiated tissue into a storage organ, during this phase embryos become photosynthetically active, and the metabolism of sugars plays a huge role in the initiation of the storage process. Finally, during cell expansion environment controls carbohydrate partitioning allowing seed filling (seed thickness) (Weber et al. 2005; Domoney et al. 2006).

Lentils are categorized according to their geographic origin. Khazaei et al. (2016) used genome-wide single nucleotide polymorphisms (SNPs) to categorize lentil environments into three major macro-environments: subtropical savannah (South Asia), Mediterranean, and temperate. Wright et al. (2020) described the macro-environments used for the present study: Temperate macro-environment is characterized by warm temperatures (mean daily temperatures of 15°C to 25°C) and long days (12.7 to 16.6 hr) during summer; mediterranean macro-environment lentils are seeded during winter, with emergence happening with cool temperatures (under 15°C) and short daylengths (9.1 to 14.9 hr), with increasing temperatures and daylength after spring equinox; and south Asian environments seed lentils during winter too, although emergence happens with warm temperatures (exceeding 25°C) and short days (10.2 to 12.9 hr).

Environment impacts seed shape parameters in legumes. Increased temperatures and drought stress have been reported to alter the filling process from grain seeds depending on the stage of development in which the stress is made and the amount of time it persists (Coventry et al. 2003). Heat stress can alter starch accumulation and sucrose synthesis (particularly important during cell expansion), although the major impact on reduced seed size is attributed to a low photosynthetic rate during the seed filling on heat stress conditions (Sita et al. 2017; Sita et al. 2018). Low temperatures decrease the rate and duration of seed filling, increasing seed and pod abortion to finally have smaller seeds for legumes (Nayyar et al. 2007; Kaur et al. 2008). Both precipitation and temperature have been reported to influence traits like dehulling, mainly due to the influence of variation in canopy humidity at key stages of seed production in a context of a Temperate environment (Subedi et al. 2018). Fedoruk et al. 2013 reported that a higher rates of precipitation, particularly after flowering, can increase seed thickness in lentil. Fedoruk et al. (2013) reported high broad sense heritability estimates for seed diameter and plumpness (H^2=0.92 and H^2=0.94, respectively) and moderate heritability estimates for seed thickness (H^2=0.60) in lentil grown in SK. These heritability estimates indicate that environmental factors do not play a large role in seed size and shape within temperate environments, suggesting that lentil has a higher heritability for seed shape and size in comparison to other legumes.

Phenology is the study of life cycle stages, especially with their timing and relationships with biotic and abiotic causes (Schwartz 2013). Flowering time is a phenological event, a complex trait in plant development governed by multiple genes. Previous studies in legumes and other species, have shown that flowering time molecular pathways are responsive to environmental factors (Weller and Ortega, 2015; Xu et al. 2016; Jeyakumar et al. 2020). Temperature and photoperiod play an important role in lentil developmental stages. Warmer temperatures promote early start of flowering and end of VEG, moreover longer photoperiods promote an earlier start to flowering and the end of VEG (Summerfield et al. 1985). Wright et al. (2020) described days to flowering (DTF) in lentil to be highly influenced by the environment and to be a good phenological parameter to explain adaptation, but their results also suggest that there may be other environmental factors at play. Previous studies have linked seed size with DTF and seed weight (Hovav et al. 2003). Particularly during flowering, changes pre-anthesis can alter the assimilates that are partitioned to the seed during development, processes post-anthesis can affect seed maturation or filling, both processes can, therefore, change seed size (Gupta et al. 2006).

2.4 Genetic Diversity and Quality Traits in the Lentil Breeding Program

There are 43,214 lentil accessions in genebanks worldwide. The International Centre for Agricultural Research in the Dry Areas (ICARDA) has the biggest collection of germplasm accessions with nearly 29 % of accessions both available and not available to the public (Crop Trust, 2020). A successful breeding program depends on genetic biodiversity coming from genebanks, with the purpose of finding useful genes or alleles (Mascher et al. 2019).

The lentil breeding program in Canada has used a Systematic Introduction of New Germplasm (SINGing) system for more than 20 years, to introduce non-adapted germplasm with the purpose of identifying lines with traits of interest and to overcome the adaptation barrier between macro-environments. The non-adapted lines are crossed with adapted breeding lines, and their progeny are crossed again to get a representation of no more than 12.5 % of the pedigree, and then the germplasm is introduced in the breeding program for evaluation (Tullu et al. 2011).

Lentil germplasm movement for introduction from one macro-environment to another is difficult due to photoperiod and temperature adaptation (Tullu et al. 2011; Wright et al. 2020). There are some discrepancies regarding the relationship between the macro-environment of origin and the diversity of lentil accessions. Dissanayake et al. (2020) and Khazaei et al. (2016), reported a relationship between geographical origin and diversity of lentil accessions. Lombardi et al. (2014), to the contrary, reported a weak relationship among geographical zones and genetic diversity of lentil. A wide diversity of traits has been reported from the diversity of accessions available, including responses and tolerance to biotic and abiotic stresses (Furman et al. 2009).

Seed quality traits have become an important objective for lentil breeders due to their importance in the market and having genetic markers would help to save time in the breeding process. In the past, lentil has been genotyped with simple sequence repeats (SSR) and these have been associated with seed size and weight (Verma et al. 2015). Now with the development of high-throughput genotyping, SNPs associated with shape and color have been identified opening doors for high-throughput genotyping (Fedoruk et al. 2013; Sharpe et al. 2013). QTL have been identified and mapped for seed size and flowering time in single regions (Fedoruk et al. 2013 in SK; Polanco et al. 2019 in Spain). Seed size and shape QTL have also been associated with milling quality in SK (Subedi et al. 2018).

2.5 Phenotyping

The first step in the analysis of a trait in a genetics study is robust phenotyping, which helps to identify and understand the interaction between plant genotype, environment, and crop management (Lobos et al. 2017). Phenotyping involves measuring plant growth, architecture, and composition at different biological scales - from organs to canopy (Fiorani and Schurr, 2013). Accurate phenotyping of plant populations is a bottleneck and a demanding activity for plant breeding programs to accelerate the development of new and improved cultivars. In this scenario, a twenty-first century approach comes to the game as a systematic genome-wide study of phenotypes, known as phenomics (Singh and Singh, 2015b). Phenomics lets the phenotyping process operate at multiple levels of organization to get a more complete picture of the phenotype. This approach generates big quantities of data, making the process of phenotyping more robust (Lobos et al. 2017). The main advantages of phenomics are accurate, fast, precise, and large-scale data collection.

Seed shape measurements have traditionally been estimated by passing a set weight of seed sample through round-hole and slotted-hole sieves and then seed shape parameters such as diameter, thickness and plumpness are estimated based on equations (Hossain et al. 2010; Fedoruk et al. 2013) Even though this approach is helpful, it is not close to being accurate in terms of the actual size and shape of the seed (Hossain et al. 2010).

Visible light imaging technology uses digital cameras and an automatic imaging system to provide phenotypic information (Awada et al. 2018). Digital imaging offers the advantages of being a non-invasive and non-destructive procedure, relatively easy to perform, and different kinds of data can be captured within a single image. The digital images can be stored so that the actual raw data can be analyzed with different tools as they become available (Singh and Singh, 2015b). Tanabata et al. (2012) developed a system to take top view images using rice as a first subject of study, however, because the camera takes a picture of multiple seeds at a time, some precision is lost in the process. Shahin et al. (2012) developed a system using two cameras to capture both top and side views in lentil, however the seeds need to be placed manually which takes a lot of time especially when working with large number of samples and with dirty samples. The development of BELT and phenoSEED, a high-throughput imaging system (Halcro et al. 2020), allows one to directly measure seed shape parameters through digital imaging. This

system is designed to take high-resolution images of single seeds, including their side profile view, as they travel along a conveyor belt. Then from these high-quality images the software phenoSEED extracts the shape metrics and provides the data to be analyzed.

2.6 Association Mapping

Association mapping (AM) is an approach widely used to identify the relevance of genetic loci on the expression of one or multiple traits (Álvarez et al. 2014). AM exploits the linkage disequilibrium (LD) that comes from hybridization between parental and progeny lines and the historical LD between those populations which leads to a higher QTL resolution (Singh and Singh, 2015a). AM allows high-resolution mapping of traits depending on the LD of the genomic region, the crop used for the study, the type of trait and marker density. AM is an important tool that allows one to discover associations between molecular markers and traits of interest promoting the acceleration of breeding programs as a pre-breeding activity (Álvarez et al. 2014).

A genome-wide association study (GWAS) is a statistical genetics method that is useful to identify candidate genes controlling target traits through scanning of the whole genome, in other words, it helps to understand the association between genetic markers (genotype) and traits of interest (phenotype) and supports the assessment of genes involved in complex traits (Rafalski, 2010; Alqudah et al. 2020). Historical recombinations over generations are key to understand LD decay and to elucidate the number of markers required to saturate the genome for GWAS (Earp and Goode, 2017; Alqudah et al. 2020). The most important steps for a successful GWAS are phenotyping, genotyping and then performing the actual association analysis. Before performing the GWAS it is important to use a diverse collection of accessions and check population structure, and after performing the GWAS, the statistical model and the marker-trait association tests will define QTL. Candidate gene identification can be done if a genome assembly is available (Alqudah et al. 2020). GWAS have been used to identify markers and genes associated with seed shape in pulses (e.g., *Glycine max* – Shao et al. 2022; *Phaseolus vulgaris* – Giordani et al. 2022; *Pisum sativum* – Gali et al. 2019) and other species (e.g., *Oryza sativa* – Ponce et al. 2020; *Sorghum bicolor* – Sakamoto et al. 2019; *Zea mays* – Qu et al. 2022). There are different statistical methods of GWAS, each method is different and could potentially identify loci associated with a trait.

The first methods developed for GWAS, attempted to account for population structure to mitigate false positives. The general linear model (GLM) assigns individuals to subpopulations, and this subpopulation memberships are added as cofactors to account for population structure (Pritchard et al. 2000). Then a huge improvement came with the mixed linear model (MLM), this model takes population structure and kinship into account (Yu et al. 2006). By increasing the power to account for false positives, false negatives increased the decay of the statistical power of GWAS, so after the development of MLM, improvements came to improve the model, these improvements helped to build the next generation of GWAS, the multi-locus methods (Klasen et al. 2016; Tibbs Cortes et al. 2021).

Multi-locus GWAS methods were developed to improve the statistical power over single-locus methods by simultaneously incorporating multiple markers as covariates. This improvement was first achieved with the multi-locus mixed model (MLMM) (Segura et al. 2012), then two GWAS methods were build upon MLMM, these are the fixed and random model circulating probability unification (FarmCPU) (Liu et al. 2016) and Bayesian information and LD iteratively nested keyway (BLINK) (Huang et al. 2019). Single-locus statistical methods for GWAS (GLM and MLM) are simple and work fast but, they assume that only one QTL and a single gene has an effect. The multi-locus GWAS approaches (MLMM, FarmCPU and BLINK) are more powerful for complex traits controlled by several large-effect loci (Tibbs Cortes et al. 2021). However, some studies suggest that employing integrated single-locus and multi-locus models aids with verification of the significance of the markers (Chang et al. 2018; Xu et al. 2018; Li et al. 2018; Soto-Cerda et al. 2021; Muhammad et al. 2021).

Genetic studies for seed shape and size in lentil have struggled throughout time with the diversity of populations, phenotyping, and the restraints of statistical methods. The first QTL identified for seed size was described by Fratini et al. (2007) by using various older type molecular markers to map QTL in a bi-parental population in which seed size was variable but the QTL only explained a relatively low amount of the phenotypic variance. Fedoruk et al. (2013) improved the genotyping by using SNPs and a recombinant inbred line (RIL) population for QTL mapping of seed shape parameters in a population grown in SK. They identified three markers associated with seed diameter and plumpness and seven markers associated with seed thickness. They found that all QTL for seed diameter and plumpness were significant across all site-years, however seed

thickness had only three that were significant in multiple site-years. These results allowed them to link seed shape and size genetics with seed development and how environment would be more linked to seed thickness than seed diameter. Verma et al. (2015) used SSRs to construct a QTL map of a RIL population, which helped identify one QTL for seed size. Polanco et al. (2019) described three QTL for seed size using SNP QTL mapping in a bi-parental RIL population. Khazaei et al. (2018) used GWAS and linkage mapping to find two markers for seed diameter, three markers for seed thickness and two markers for seed plumpness in a panel of diverse lentil germplasm grown in SK. The use of a different and more diverse population for analysis helped in the discovery of three new SNP markers. These previous studies identified QTL using the above-mentioned traditional seed size and shape phenotyping in SK.

The following study will have more accurate results through high-throughput phenotyping by taking measurements of single seeds from samples with nearly 200 seed samples per genotype. Also, it will be possible to evaluate additional site-years representing different macro-environments with an increased number of markers available with the purpose of better defining the loci involved in seed shape in lentil.

3 Materials and Methods

3.1 Seed sources

A lentil diversity panel (AGILE LDP) consisting of of 324 lentil accessions from around the world was used for this study (https://knowpulse.usask.ca/Lentil-Diversity-Panel).The accessions represent breeding materials from the genebanks of the Crop Development Center (CDC) of the University of Saskatchewan, International Center for Agricultural Research in the Dry Areas (ICARDA), United States Department of Agriculture (USDA) and Plant Gene Resources Canada (PGRC). The AGILE LDP was grown in a randomized lattice square design with 3 replications in site-years in Italy, Morocco and Spain representing the Mediterranean macro-environment, and in Rosthern and Sutherland representing the temperate macro-environment over 2017-2019 (Wright et al. 2020). Seed harvested from these site-years were made available for this project.

A subset of 20 accessions (representing different origins and sizes) was selected from the Morocco 2017 site-year (Mediterranean macro-environment), and the Rosthern and Sutherland 2017 site-years (temperate macro-environment) and were phenotyped across all three replications at each site-year and analyzed to determine if replications within site-years were needed for further experiments.

3.2 Phenotyping

Seed shape was measured using a high-throughput image-based phenotyping system developed by Halcro et al. (2020), which includes the imaging system BELT, and the image processing software, phenoSEED. Approximately 200 seeds per sample were loaded into the BELT system through a feeder and then a conveyor moved the samples into an imaging chamber to be automatically imaged with top and side views. After this process, images were manually checked and filtered to be sure there is only one seed per image. Finally, the phenoSEED software was used to extract quantifiable traits related to seed shape such as height (i.e., thickness), diameter, and circularity. The seed shape descriptor circularity, which can be used to quantify how smooth are the edges of the seed, was calculated from the directly measured traits area and perimeter (Equation 3.1). Seed plumpness was calculated using equation 3.2 by taking the ratio of seed height to seed diameter.

$$\text{Circularity} = \frac{4*\pi*\text{Area}}{\text{Perimeter}^2} \quad \text{[Equation 3. 1]}$$

$$\text{Plumpness} = \frac{\text{Seed Height}}{\text{Seed Diameter}} \quad \text{[Equation 3. 2]}$$

Seed shape phenotypic data were combined with phenological data for the respective site-years. These datasets were accessed from the Knowpulse web portal (https://knowpulse.usask.ca/phenotypes/trait-search/Lens). Days to emergence (DTE) and days to flowering (DTF) had been recorded on a plot basis when 10 % of the plants had the first visible seedling stem or leaves and the first open flower respectively. Days to maturity (DTM) had been recorded as the day when 10 % of the plants had 50 % pod maturity. Days to swollen pods (DTS) had been recorded on a plot basis when 10 % of the plants had one swollen pod. Thermal flowering time (TFT) was calculated by using daily mean temperature from seeding to flowering and thermal reroductive time (TRT) is calculated using the daily mean temperature during REP (Wright et al. 2020).

3.3 Data Analyses

The analyses for the preliminary study and the full study were done using R Studio version 1.2.5033 (RStudio Team 2020). For the preliminary study with a subset of 20 genotypes, a Kruskal-Wallis analysis was done on the means from the approximately 200 seeds per genotype to determine if there were significant differences between replications within each of the 3 site-years sampled. The full dataset, from one replication per site-year, was analyzed to evaluate range, mean, standard deviation and broad-sense heritability from six site-years. All data was checked for normality statistically with Shapiro-Wilk test, and visually with Q-Q plots and histograms. Logarithmic and square-root transformations were performed on all data before using non-parametric analyses. The data visualization tools used to check distribution, normality and outliers were performed with the 'ggplot2' package in R (Wickham 2016) and 'rcompanion' package in R (Mangiafico 2023) for the preliminary and full study. Heritability of each trait between macro-environments was calculated using the 'variability' package in R (Popat et al. 2020) for the full study.

3.3.1 Non-parametric stability analysis

Three sets of non-parametric statistics were used to estimate stability in this study. One of them (Huehn, 1990; Nassar and Hühn, 1987) consisted of four parameters ($S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$)

combining the mean of the trait to analyze stability. The $S_i^{(1)}$ parameter measures the mean absolute rank difference of a genotype over *n* environments, with $S_i^{(1)}$ equals zero for the most stable genotype. $S_i^{(2)}$ gives the variance between the ranks over *n* environments, with values of zero indicating the most stable genotypes. $S_i^{(3)}$ and $S_i^{(6)}$ are the sum of the absolute deviations and sum of square of ranks, respectively.

Another measurement of stability is Kang's rank-sum (RS), where both target trait and stability parameter have an equal weight of one, then ranks of target trait and stability parameter are added for each genotype. The genotype with the lowest rank-sum is the most stable.

Finally, Thennarasu (1995) proposed four parameters ($NP^{(1)}$, $NP^{(2)}$, $NP^{(3)}$ and $NP^{(4)}$), which are based on the ranks of adjusted means genotypes in *n* environments. Low values of this statistic indicate high stability.

All non-parametric statistics were calculated with STABILITYSOFT (Pour-Aboughadareh et al. 2019). Spearman's correlation was calculated and plotted to statistically and visually correlate the stability parameters analyzed with seed shape traits (circularity, diameter, height, and plumpness) to select the stability parameter most correlated to every seed shape trait, the selected stability parameter was visualized with the 'ggridges' (Wilke 2022) package in R.

3.3.2 Correlations of seed shape traits with phenological and environmental parameters

Correlations calculated with the Spearman's method between the phenological and environmental parameters and lentil seed shape traits because the parameters and seed traits are quantitative variables, the calculations were performed using R studio version 1.2.5033 (RStudio Team 2020). Data visualization was performed using the 'corrplot' (Wei and Simko 2021) and 'psych' (Revelle 2023) packages in R.

3.3.3 Hierarchical clustering

To prepare data for clustering, the data from all seed shape parameters: circularity, diameter, height, and plumpness were scaled with the scale function from R software to standardize data, then the missing data were imputed with mean values. Clustering method was evaluated with 'clValid' R package (Brock et al. 2008). Number of clusters was assessed with 'NbClust' R package (Charrad et al. 2014). Principal component analysis (PCA) and hierarchical k-means clustering were performed using the 'FactoMineR' R package (Lê et al. 2008) for each trait and

site-year. Appropriate number of clusters was assessed with ‘FactoMineR’ R package (Lê et al. 2008) and ‘dendextend’ R package (Galili 2015).

3.3.4 Genome-wide association study

Genotypic data for the 324 accessions (Haile et al. 2020), derived using an exome capture array (Ogutcen et al. 2018), was accessed from KnowPulse as the AGILE LDP Exome Capture SNP Set (https://knowpulse.usask.ca/filter_vcf) and used for association analyses.

Analyses were conducted using four different models: MLM, MLMM, farmCPU and BLINK, all integrated in the R package GAPIT (Lipka et al. 2012). In addition to the general run of target traits, seed shape traits were used as covariates to help highlight peaks or to discover new ones. Then a comparison of the results of all the models was carried out to find common peaks and to have more robust results. Visualization was carried out with the R package ‘gwaspr’ (Wright 2022) GWAS models were filtered out of the visualization of significant QTL when they showed high number of false positive or negative data in GWAS Q-Q plots.

4 Results

4.1 Distribution of seed shape phenotypes

There were differences in seed shape trait distributions across macro-environments (Figure 4.1; Appendix Table B). Diameter and height had the highest mean values in temperate site-years, and plumpness had slightly higher values in these site-years too, while circularity remained uniform across all site-years. Temperate site-years had the maximum values for height and diameter and Mediterranean site-years had the minimum values for height. Sutherland 2018 had the highest mean values and maximum values for height and diameter in comparison to the other site-years. For plumpness, Sutherland 2017 had both the minimum value and the maximum value out of all the site-years.

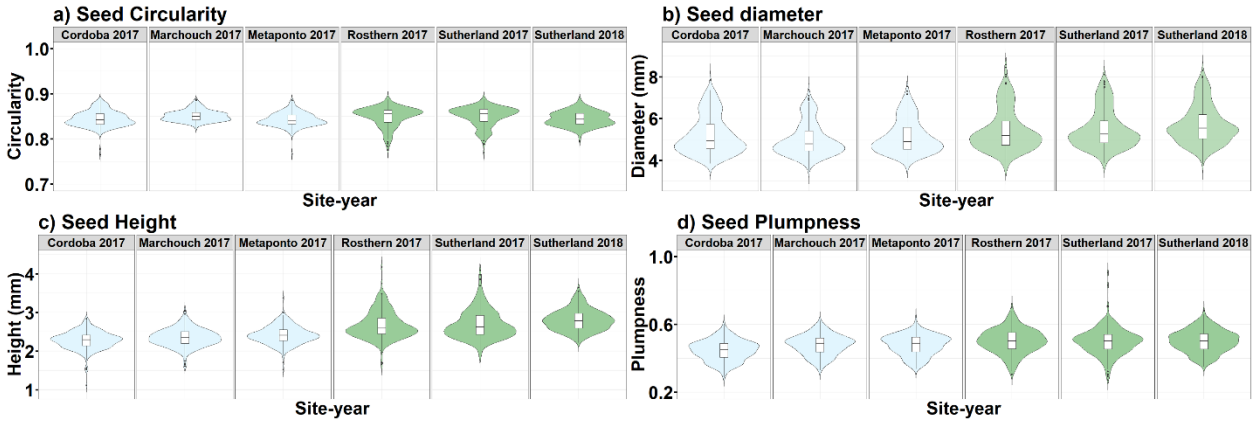

Figure 4.1. Variation in average shape parameters per accession - a) Seed circularity, b) seed diameter, c) seed height d) seed plumpness, for 324 lentil genotypes grown at 6 site-years (Mediterranean: Cordoba 2017, Marchouch 2017 and Metaponto 2017; Temperate: Rosthern 2017, Sutherland 2017 and Sutherland 2018). Blue color represents the Mediterranean site-years and green the temperate site-years. The width of the plots shows the density of the distributions. Individuals falling outside the range of the whiskers are represented as dots.

Broad sense heritability estimates between macro-environments (Appendix Table B) showed that diameter was the most heritable trait at 0.95, circularity and plumpness were also fairly high - 0.83 and 0.86, respectively, and height had the lowest heritability at 0.73.

4.2 The role of phenology and environment in final seed shape

Daylength, temperature and precipitation were substantially different across the macro-environments (Fig. 4.2). Reproductive period (REP) started earlier for temperate site-years than in Mediterranean site-years. Mediterranean site-years are characterized by short days, with a gradual increase in day length and temperatures throughout the REP. In the temperate site-years,

day lengths were long, and the mean temperatures oscillated in a uniform fashion throughout the REP. REP was shorter in temperate site-years and started earlier than in Mediterranean site-years. Marchouch 2017 was the driest site-year during the REP and Cordoba had the most precipitation during the REP of all the site-years. Sutherland 2018 had higher precipitation in some days than the other temperate site-years in the first half of the REP.

Among seed shape parameters, diameter and height have a positive significant correlation. Plumpness is slightly positively correlated to height in most site-years and negatively correlated to diameter across all site-years. (Fig. 4.3). Sutherland 2018, however, had higher correlation coefficients between diameter and phenological parameters in comparison to all the other site-years – a positive correlation of 0.53 with DTS and a moderately positive correlation of 0.44 and 0.41 with DTM and REP_P, respectively. DTE, on the contrary, had a moderately negative correlation (-0.46) with diameter. Plumpness had a negative correlation (-0.52) with DTS. Plumpness in Marchouch 2017 had a moderately negative correlation of -0.40 and -0.49 with DTS and DTM, respectively, and the highest negative correlation coefficient among the Mediterranean macro-environment.

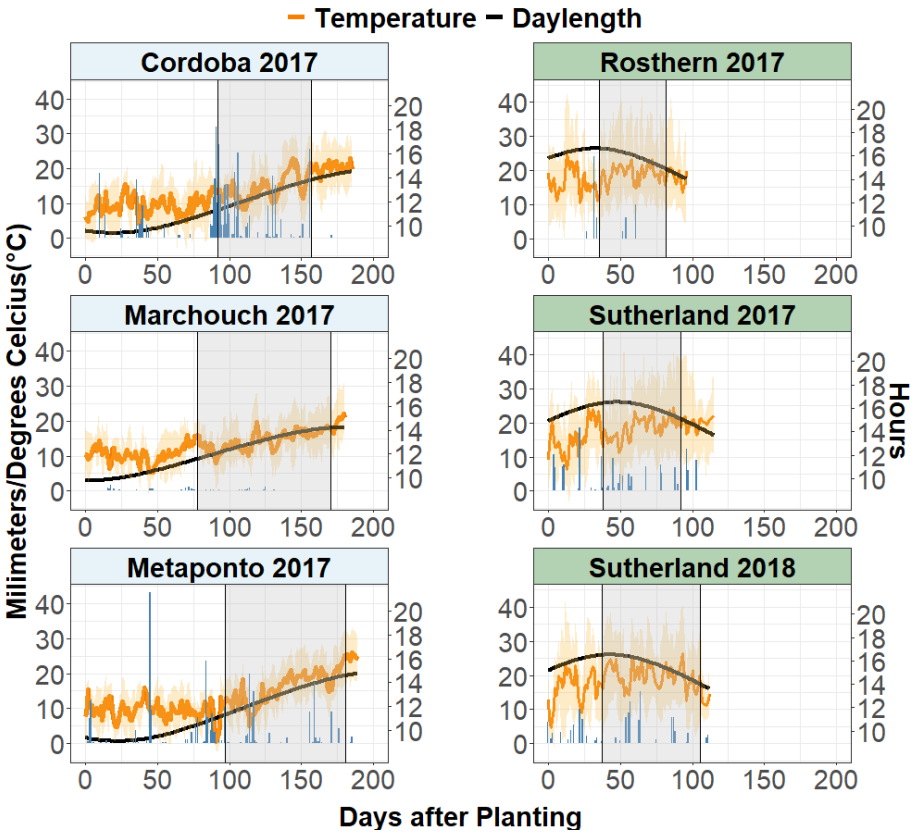

Figure 4.2 Environmental factors and reproductive period from each site-year. Daily mean temperature (orange line), day length (black line) and precipitation (blue bars) from seeding to maturity across six site-years. The shaded orange ribbon represents the daily minimum to maximum temperatures. The grey shaded areas correspond to the reproductive period.

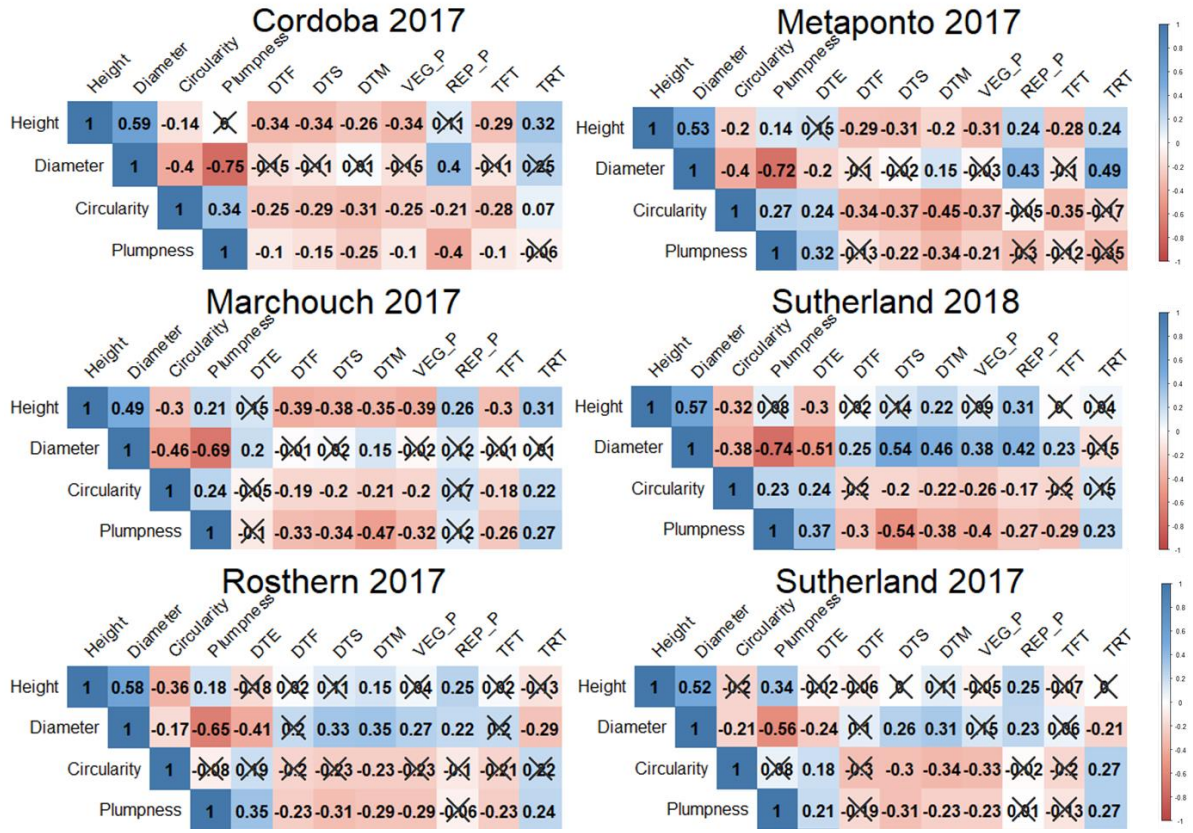

Figure 4.3. Correlations among environmental and phenological factors and seed shape traits from 324 genotypes grown in six site-years: Cordoba 2017, Marchouch 2017, Metaponto 2017, Rosthern 2017, Sutherland 2017, Sutherland 2018. DTE: days to emergence, DTF: days to flowering, DTS: days to swollen pods, DTM: days to maturity, VEG_P: vegetative period, REP_P: reproductive period, TFT: thermal flowering time and TRT: thermal reproductive time. The numbers inside the squares show the Spearman's coefficient of correlation. Where the correlation is not significant, squares are crossed out. Blue indicates a positive correlation, red a negative correlation and white no correlation. Correlations are considered significant when p >0.05.

Stability analyses help measure the stability of accessions across multiple environments. The Spearman correlations showed that $S^{(6)}$, $NP^{(2)}$,$NP^{(3)}$ and $NP^{(4)}$ have a significant negative correlation with the seed shape parameters (Appendix Fig. C.1), meaning that genotypes with larger dimensions of these traits are more stable. Furthermore, $S^{(6)}$ had a positive significant correlation with $S^{(1)}$, $S^{(2)}$, $S^{(3)}$, $NP^{(2)}$, $NP^{(3)}$ and $NP^{(4)}$ for all traits, which indicates that each of these stability measures can be helpful to evaluate the stability of circularity, diameter, height and plumpness. The stability measures calculated as in Pour-Aboughadareh et al. (2019), revealed

that height and plumpness had a wider distribution of stability scores in comparison with those for DTF and DTS (Fig. 4.4; Appendix C). Diameter and circularity had a narrower distribution, that was skewed towards higher stability. Diameter had more genotypes closer to zero than the other traits.

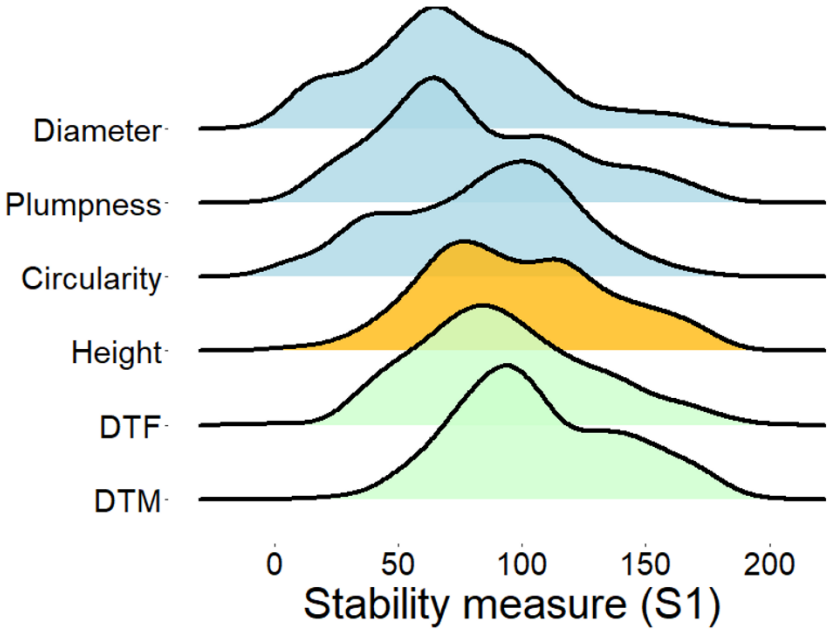

Figure 4.4. Variations in stability across 324 genotypes grown at six different site-years. Based on $S^{(1)}$ stability measure, for - circularity, diameter, height, plumpness, days to flowering time (DTF) and days to swollen pods (DTS). Seed shape traits with genotypes with the lowest stability values (blue), seed shape trait with genotypes with higher stability values (orange) and phenological traits as a representative control (green).

The PCA done using seed shape traits (circularity, diameter, height, and plumpness) across the six site-years (Fig. 4.5 A) resulted in three main clusters. Cluster 1 is mainly composed of red cotyledon genotypes with the smallest diameter, and cluster 3 is mostly yellow cotyledon genotypes with the largest diameter. Cluster 2 is a mix of genotypes with red or yellow cotyledons with a range of diameters. PC1 trends with seed diameter and represents 41% of the variability seed in shape (Fig. 4.5 B). PC2 trends with height and plumpness and represent 17.3% of the variability in seed shape. (Fig. 4.5 C). When the other combinations of PCs are plotted (PC2 with PC3 and PC1 with PC3), no new patterns are identified.

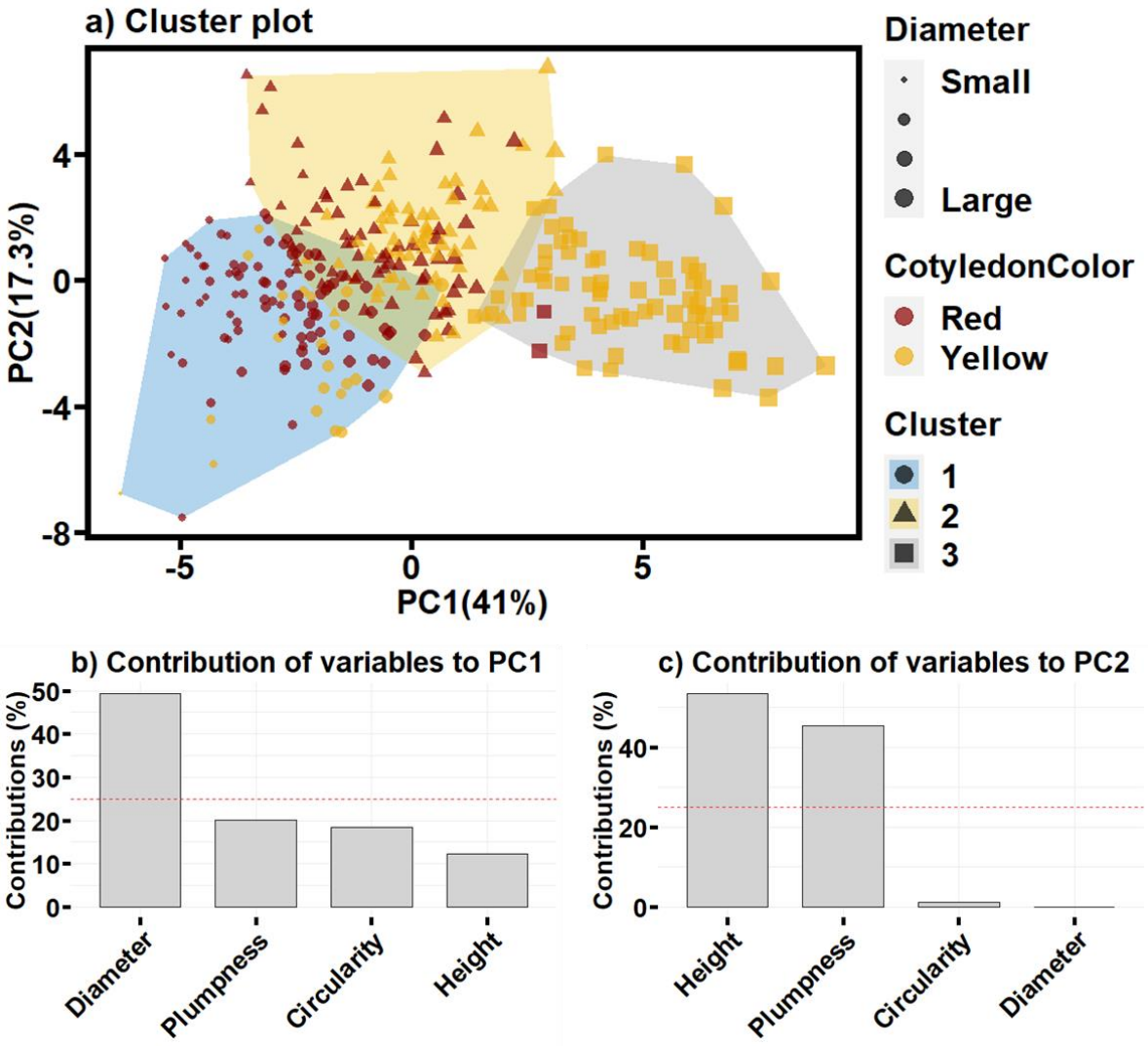

Figure 4.5. Hierarchical clustering using principal components of scaled seed shape parameters. A) Clustering plot for seed circularity, diameter, height, and plumpness from six site-years: Cordoba 2017, Marchouch 2017, Metaponto 2017, Rosthern 2017, Sutherland 2017, and Sutherland 2018. Shapes represent the three clusters. Point size represents seed diameter. Color represents cotyledon color. **B)** Contribution of seed shape traits to PC1, the red dashed line on the graph above indicates the expected average contribution of variables. **C)** Contribution of seed shape traits to PC2, the red dashed line on the graph above indicates the expected average contribution of variables.

The distribution of seed shape parameters (circularity, diameter, height, and plumpness) separated by cluster group for each site-year (Fig. 4.6) showed that in general cluster 1 had accessions with smaller height and diameter and higher circularity scores than the other clusters. Cluster 3 was composed of large and less plump seeds than the other 2 clusters. Cluster 2 was an

intermediate between the other 2 clusters and has accessions with less circularity like cluster 3 and higher plumpness like cluster 1.

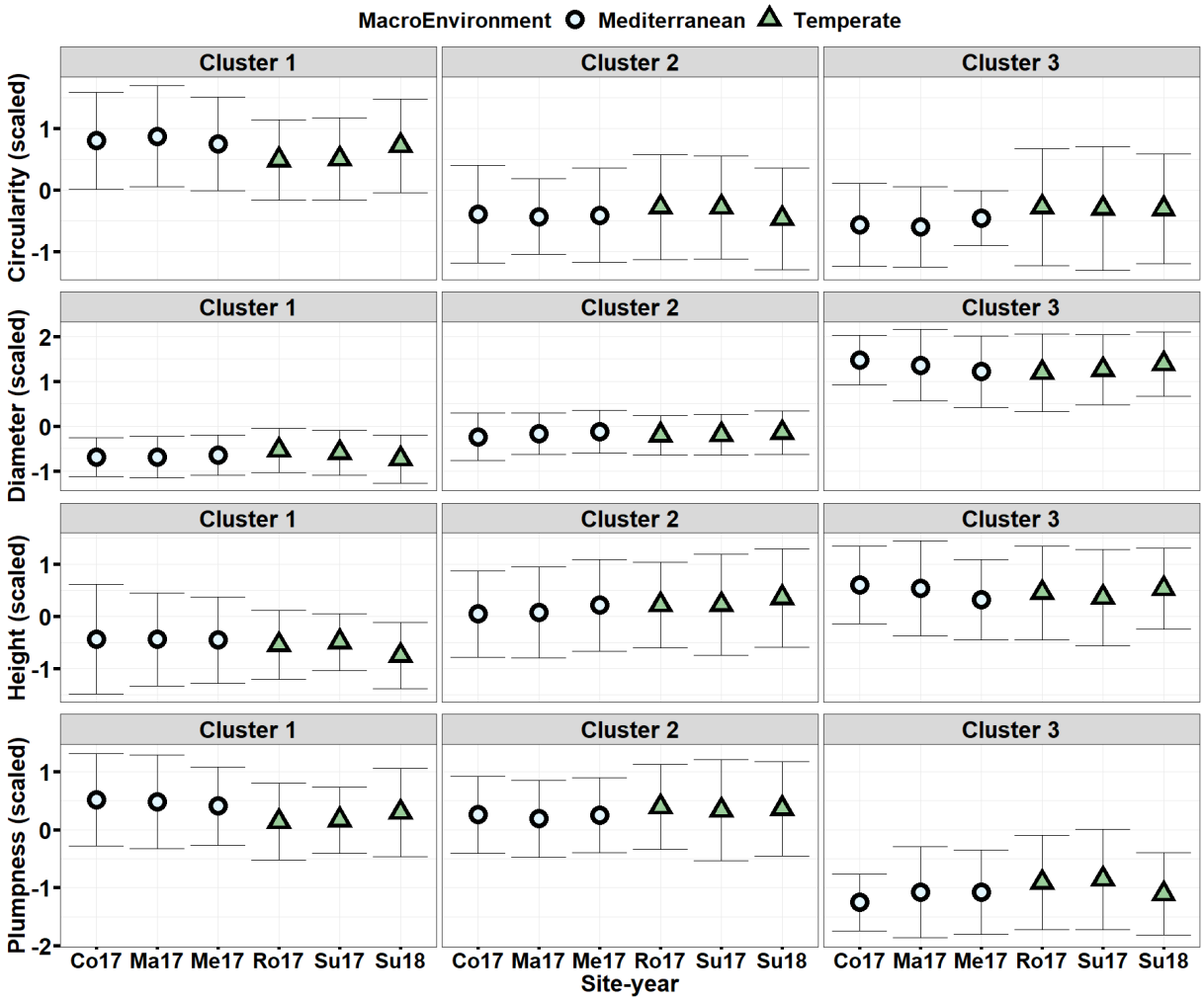

Figure 4.6. Mean scaled seed shape parameters - a) Seed circularity, b) seed diameter, c) seed height d) seed plumpness, for each of the three cluster groups across all field trials: Rosthern, Canada 2017 (Ro17), Sutherland, Canada 207, and 2018 (Su17, Su18), Metaponto, Italy 2017 (Me17), Marchouch, Morocco 2017 (Ma17), Cordoba, Spain 2017 (Co17). Shapes represent macro-environments: Mediterranean (circles) and temperate (triangles).

4.3 Genomic regions associated with seed shape parameters.

Based on the Manhattan plots and their associated Q-Q plots (Appendices D-K) the GWAS models had a good fit for analyses with a few exceptions where FarmCPU and BLINK did not fit

and had a lot of false positives (Q-Q plots Figures K.3 and K.6), so those cases were filtered out. Significant QTL were identified for all traits and results by site-years are plotted in Fig. 4.7. Strong patterns based on when several site-years identified the same QTL are marked with black vertical lines. Three sets of QTL were found to be associated with diameter, the first is the strongest pattern and is on chromosome 1. The other two sets of QTL had less prominent patterns, the second set was identified in chromosome 2 in all Mediterranean site-years plus Sutherland 2017, and the third set of QTL were identified in chromosome 4. For circularity, QTL were identified on chromosome 2 for all Mediterranean site-years. Finally, for plumpness, QTL were identified on chromosome 4 for all Mediterranean site-years.

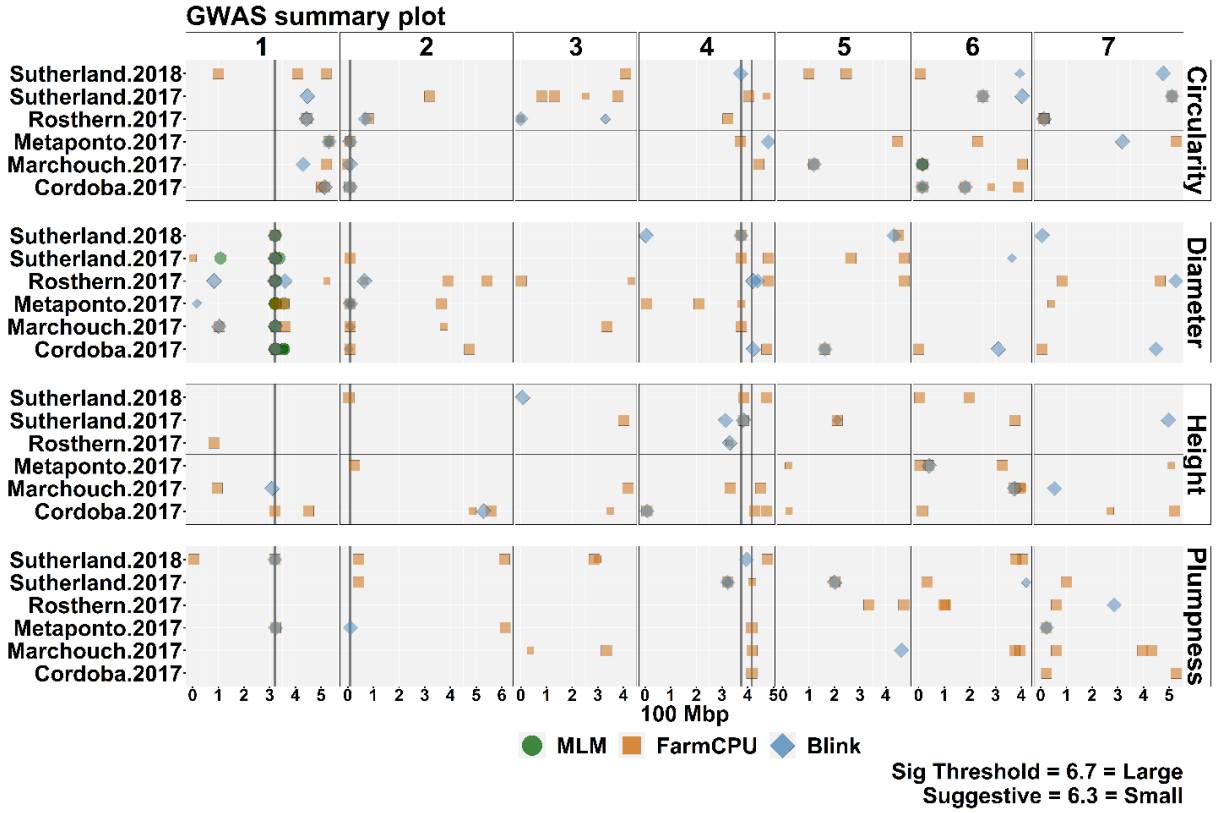

Figure 4.7. Summary plots of significant QTL for diameter, circularity, height, and plumpness, based on data from 324 genotypes grown in the field in Metaponto 2017, Morocco 2017, Spain 2017, Rosthern 2017, Sutherland 2017 and Sutherland 2018. The x-axis represents lentil chromosomes, and the Y axis the site-years and seed shape traits. Boxes along the y-axis separate the traits. Horizontal lines within the trait boxes split site-years into the two macro-environments: Temperate (top site-years) and Mediterranean (bottom site-years). The shapes within the plot show markers detected above a significance threshold [-log10(P)>6.7] for larger shapes and a significance threshold [-log10(P)>6.3] for smaller shapes.

Adding covariates helped clarify several QTL and revealed additional ones. For seed circularity, the QTL previously found in chromosome 2 only for Mediterranean site-years (Fig. 4.7) was also identified when diameter, height or plumpness were added as covariates to the GWAS (Fig. 4.8 B-D). Two new sets of QTL were found on chromosome 6 for circularity when diameter and plumpness were added as covariates (Fig. 4.8 B and D), these QTL were not present in the GWAS without covariates (Fig.4.7). In addition, QTL were identified for 2 Mediterranean site-years (Cordoba 2017 and Marchouch 2017) and a temperate site-year (Sutherland 2018) when diameter was added as covariate. Moreover, when plumpness was added as covariate, two sets of QTL were identified, the first one had all Mediterranean site-years plus Sutherland 2018 from the temperate site-years, the second set was only identified in all Mediterranean site-years.

The QTL found for diameter on chromosome (Fig. 4.7) was also identified when circularity, height or plumpness were added as covariates to the GWAS (Fig. 4.8 A, C and D). In addition, the QTL found on chromosome 2 was also identified when circularity, height and plumpness were added as covariates, with circularity having the strongest pattern by having QTL in all Mediterranean site-years plus Sutherland 2018, whereas when height or plumpness were used as covariates only the Mediterranean site-years had QTL detected.

For plumpness, 2 sets of QTL were identified, the first on chromosome 1 and the second on chromosome 2 when circularity, diameter or height were added as covariates (Fig. 4.8 A, B and C), except when diameter was added as covariate in which case only the set from chromosome 2 is present and with fewer site-years showing QTL (Fig. 4.8 B).

No QTL were identified for height when circularity, diameter and height were used as covariates for the GWAS (Fig. 4.8 A-C), however when plumpness was added, new QTL were found for height (Fig. 4.8 C) in similar position to those found for plumpness on chromosome 2 for the other seed shape traits. The QTL found on chromosome 4 identified in Fig. 4.7 disappeared for diameter when circularity, height or plumpness were added as covariates (Fig. 4.8 A, C and D) and disappeared for plumpness when circularity, diameter or height were added as covariates (Fig. 4.8 A-C).

The QTL found for circularity on chromosome 2 (Fig. 4.7 and Fig. 4.8 B-C) was also identified when DTF, TFT and TRT were added as covariates (Fig. 4.9 A, C and D), however when DTS

was added as covariate, the QTL for Metaponto 2017 disappeared, and a QTL appeared for Sutherland 2018. New QTL were found on chromosome 5 for circularity when TFT was added as covariate (Fig. 4.9 C). Another QTL was found on chromosome 6 for circularity when DTF, DTS or TFT were added as covariates, with all the Mediterranean site-years, however when DTS or TFT were added as covariates Sutherland 2018 was also included in the set. In addition, using TFT as a covariate resulted in the strongest set of significant QTL with all peaks above a significant threshold of LOD 6.7.

For diameter, the sets of QTL found on chromosomes 1 and 2 were similarly identified across all phenological and environmental traits (Fig. 4.9). The set identified on chromosome 1 remained the strongest, as in Fig.4.7, and independent of the covariates added, as in Fig. 4.8, where the seed shape traits are added as covariates. In the case of the set identified on chromosome 2 with all phenological and environmental traits used as covariates (Fig. 4.9), DTS had the strongest pattern showing QTL across all site-years (Fig. 4.9 B). The set of QTL found for diameter on chromosome 4 (Fig. 4.7) were lost when any of the phenological traits were used as covariate (Fig. 4.9 A-D).

A set of QTL was found on chromosome 2 for plumpness when any of the phenological traits are added as covariates (4.9 A-D), with more QTL identified for Mediterranean site-years. No sets of QTL were identified for height when seed shape traits were used as covariates for the GWAS.

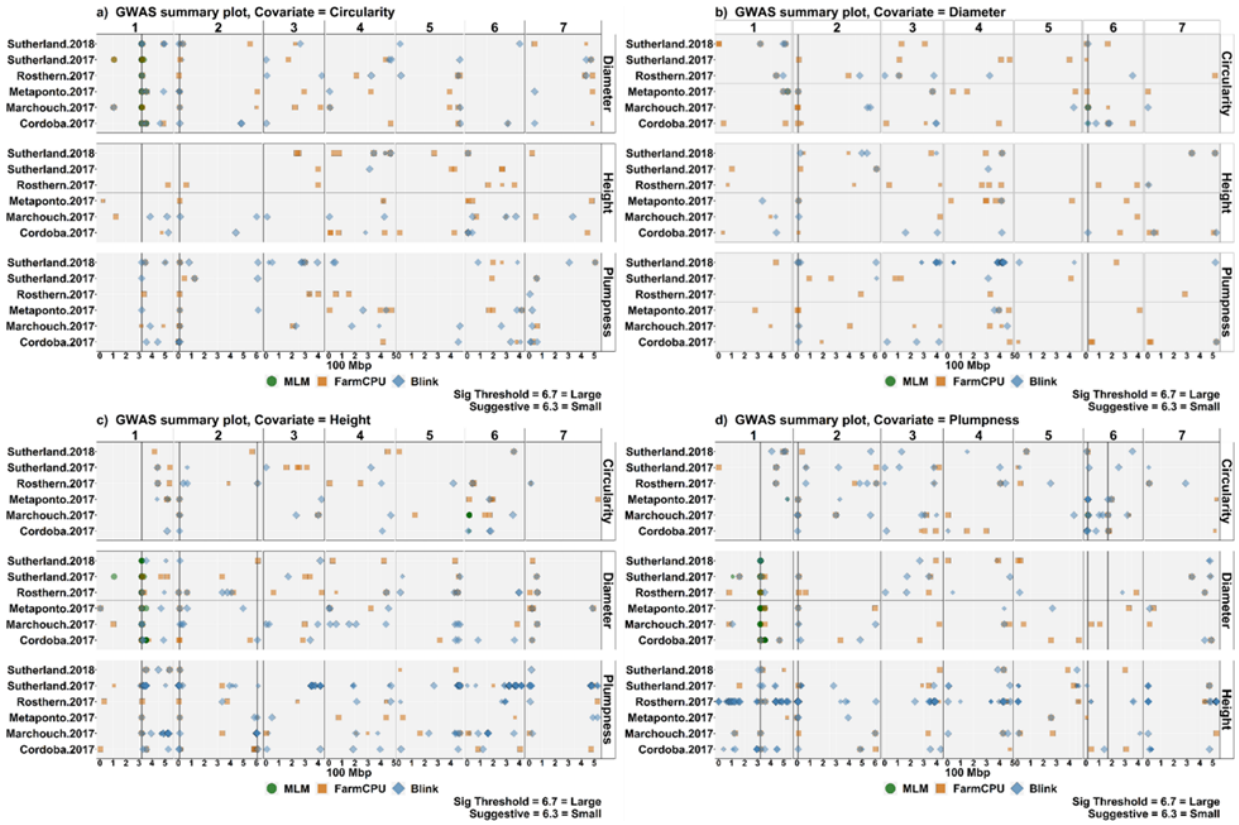

Figure 4.8. Summary plots of significant QTL for diameter, circularity, height, and plumpness with seed shape traits added as covariates. Based on data from 324 genotypes grown in the field in Metaponto 2017, Morocco 2017, Spain 2017, Rosthern 2017, Sutherland 2017 and Sutherland 2018. A) Circularity, B) diameter, C) height, D) plumpness, were added as covariates to the GWAS model. The x-axis represents lentil chromosomes, and the Y axis the site-years and seed shape traits. Boxes in the Y axis separate the traits. Horizontal lines split site-years into the two macro-environments: Temperate (top site-years) and Mediterranean (bottom site-years. The shapes within the plot show markers detected above a significance threshold [-log10(P)>6.7] for larger shapes and a significance threshold [-log10(P)>6.3] for smaller shapes.

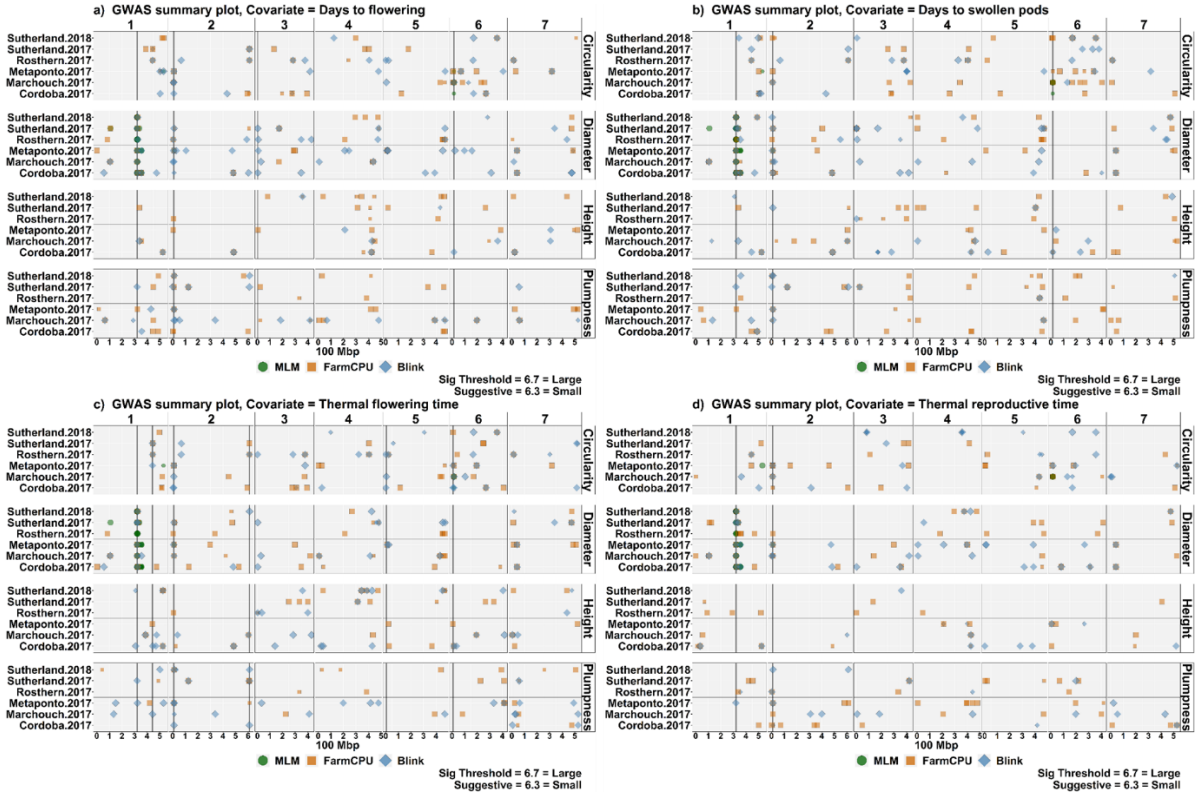

Figure 4.9. Summary plots of significant QTL for diameter, circularity, height, and plumpness with phenological and environmental parameters added as covariates. Based on data from 324 genotypes grown in the field in Metaponto 2017, Morocco 2017, Spain 2017, Rosthern 2017, Sutherland 2017 and Sutherland 2018. A) days to flowering, B) days to swollen pods, C) thermal flowering time and D) thermal reproductive time were added as covariates to the GWAS model. The x-axis represents lentil chromosomes, and the Y axis the site-years and seed shape traits. Boxes in the Y axis separate the traits. Horizontal lines split site-years into the two macro-environments: Temperate (top site-years) and Mediterranean (bottom site-years. The shapes within the plot show markers detected above a significance threshold [-log10(P)>6.7] for larger shapes and a significance threshold [-log10(P)>6.3] for smaller shapes.

The summary plots in Fig. 4.7, 4.8 and 4.9 show QTL for diameter across all site-years detected by most GWAS models. One of them, present in most site-years has a peak marker at Lcu.2RBY.Chr1p320538401 and was highly relevant in all site-years (Fig. 4.7, 4.8 and 4.9). Comparing the diameter of lines carrying the different alleles at this locus shows a similar large effect pattern across all site-years (Fig.4.9).

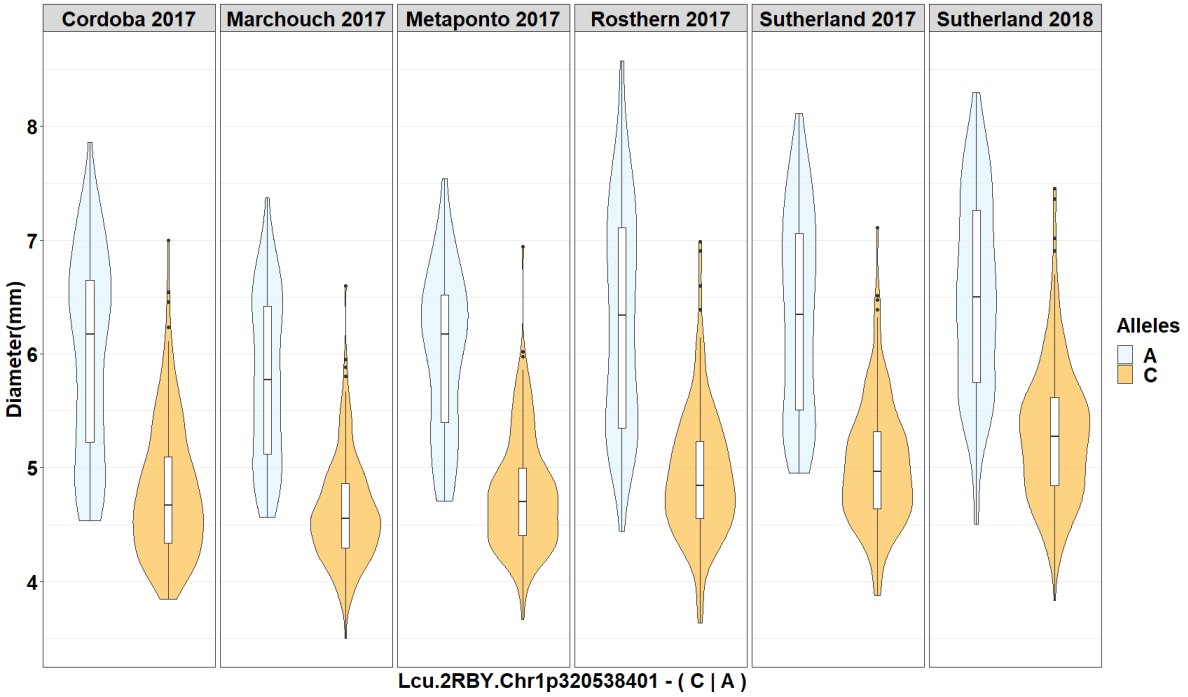

Figure 4.10. Allelic effects of marker Lcu.2RBY.Chr1p320538401 on seed diameter of 324 genotypes grown in the field in Metaponto 2017, Morocco 2017, Spain 2017, Rosthern 2017, Sutherland 2017 and Sutherland 2018. The x axis represents the alleles from marker Lcu.2RBY.Chr1p320538401 and the Y axis is seed diameter.

5 Discussion

A lentil diversity panel (AGILE LDP) had been grown under contrasting environments, temperate and Mediterranean (Wright et al. 2020), and seed was available to explore the effect of genetics, environment, and phenology on phenotype. A new high-throughput phenotyping platform was tested with the hope that it would increase the accuracy of QTL discovery. There was insignificant variance in seed shape phenotypes across replications within site-years in the pre-analysis of the present study (Appendix Fig. A. 1-4 and Table A), which allowed an increase in the scope of the research to more site-years from the two different macro-environments.

Significant correlations were identified between seed shape traits, and seed shape traits and phenological and environmental parameters. Seed shape traits had high correlations amongst themselves; especially in the case of diameter and plumpness in which case a strong negative correlation was found. However, when correlating seed shape traits and phenological and environmental parameters, only low and insignificant correlations were identified. The stability of genotypes for each trait (circularity, diameter, height, and plumpness) across multiple site-years was dynamic, and genotypes had a wide range of distribution for stability in each trait.

Clustering of lentil seed grouped genotypes into 3 clusters, one with mainly larger seed, another with mainly smaller seeds and a cluster with mixed seed diameters, showing that in terms of seed size there is a spectrum. Finally new genomic regions were significantly associated with seed shape some QTL were more likely the same as reported in previous studies and some QTL were newly introduced in this study. These identified QTL may be of used for marker identification, which are of great importance in breeding programs to skip steps of the breeding process, which helps reducing the amount of time and money spent for selection of traits.

Lentil is currently grown in mostly Temperate, Mediterranean, and South Asian environments (Tullu et al. 2011; Khazaei et al. 2016). However, the exhaustive breeding of domesticated lentil within each of these regions based on a relatively narrow genetic base, has limited the options for selection in plant breeding programs (Govindaraj et al. 2015; Khazaei et al. 2016). Genetic diversity can be improved using diverse germplasm (Guerra-Garcia et al. 2021), and use of high-throughput phenotyping (Nielsen et al. 2022) to test adaptation and stability of traits across different site-years, these tools will facilitate the use of exotic germplasm. The BELT platform could be used to directly measure seed diameter and height (seed thickness) and calculate seed

plumpness (the ratio of seed height to seed diameter) and seed circularity with the phenoSEED software (Halcro et al. 2020). Seed circularity is among the shape descriptors introduced with the new phenotyping platform and quantifies how smooth the edges of the seed are which can be useful to describe how uniform the seeds are. This could be an important trait to improve dehulling yield.

Environmental factors were different between macro-environments (Fig. 4.2), mainly because lentils are seeded at different times of the year, and this means the Mediterranean site-years have longer and cooler growing seasons when compared with temperate site-years. Mediterranean site-years lentils are seeded during winter with cold and short days at the beginning and longer and warmer days as the season progresses (Fig. 4.2). Precipitation was prominent in Cordoba and Metaponto in the Mediterranean macro-environment with Marchouch having a dry year, with almost no precipitation (Fig. 4.2). Temperate Sutherland site-years have similar precipitation patterns and more days of rain in comparison with Rosthern (Fig. 4.2). The reproductive window (first day of DTF and last day of DTS) was wider in the Mediterranean site-years (Fig. 4.2).

The results for heritability show that diameter had the highest heritability (0.95), circularity and plumpness had a similar heritability (0.83 and 0.86 respectively), and height (0.73) had the lowest heritability. These results are comparable to those from Fedoruk et al. (2013), although their plumpness heritability estimate was higher (0.94) than the one from the present study, this difference my be due to the increase in environments tested in this study. The difference in heritability between seed height and all the other seed shape traits may originate from the difference in environments from site to site during seed developmental stages in which seed diameter and seed height happen. Diameter expansion happens during seed cell division and seed height increases during cell expansion. Cell division has been described to happen faster than cell expansion, suggesting that cell expansion is more exposed to the environment. Thus, making cell division largely insensitive to environment and cell expansion sensitive to environmental factors (Weber et al. 2005; Domoney et al. 2006).

Significant correlations were identified between seed shape traits, significant low to moderate correlations between seed shape traits and phenological parameters (DTE, DTF, DTS, DTM, VEG and REP) and an no correlations were observed between seed shape traits and environmental factors (TFT and TRT). All seed shape traits: circularity, diameter, height, and

plumpness, across two different macro-environments, showed significant differences among site-years (Figure 4.1; Table B), but the differences were most noticeable when comparing the two macro-environments. Average seed height had a larger difference between macro-environments when compared with all the other traits (Appendix Table B). Khazaei et al. (2018), previously reported similar ranges for seed diameter, height, and plumpness for lentil seeds grown in SK, although the averages presented in this study (Appendix Table B), show a slightly wider range in all seed shape traits. This could be because this study used a wider range of genotypes and Mediterranean site-years as well as temperate site-years. The correlations between seed diameter and seed plumpness (Fig. 4.3) had a strong negative correlation supporting previous reports (Shahin et al. 2012; Fedoruk et al. 2013; Khazaei et al. 2018), demonstrating that smaller lentil seeds are typically plumper than large seeds.

Environmental factors like temperature, day length and precipitation do not show a clear pattern that may explain the small but significant differences in seed shape across site-years (Fig. 4.2). This observation is reinforced by the absence of a significant correlation between seed shape parameters and TFT and TRT. Choukri et al. (2022), exposed lentil plants to heat stress and seeds had a slight increase in seed diameter but no differences in seed circularity and height when compared with control. These differences in results may be because of the differences in the experiments. Choukri et al. (2022) used a narrower number of genotypes and plant were seeded under greenhouse conditions. Studies in other pulses have similar outcomes as this study, in chickpea for instance, seed shape showed high heritability and low genotype-by-environment interaction, suggesting high stability (Gan et al. 2006, Hossain et al. 2010). Conversely, a study in common bean showed that different site-years highly affected seed size, however, factors like sample size and phenotyping platform may play a role in this observed difference (Lei et al. 2020).

Phenological parameters yielded some low to moderate correlations with seed shape parameters. REP was significantly different between macro-environments (Fig. 4.3); however, differences in seed shape did not correlate significantly with either VEG or REP (Fig. 4.1 and 4.3). Fedoruk et al. (2013), working with a bi-parental population, showed that DTF had a low significant correlation (~ R^2= 0.20) with all seed shape parameters (diameter, height, and plumpness) among different site-years in most SK site-years. In addition, Fedoruk et al (2013) identified a QTL for

DTF at the same locus as QTL for seed diameter and plumpness. The results from the current study show no significant correlation between DTF and diameter, except for Sutherland 2018 (Fig. 4.3). Fedoruk et al. (2013) found that seed height was negatively correlated with DTF in all their SK site-years, but these current AGILE LDP results show that this trend was only detected for Mediterranean site-years. Changes in the patterns observed by Fedoruk et al. (2013) and the present study may have come from the different number and diversity of germplasm tested and the number of site-years evaluated. In other legumes, like common bean, phenological parameters like DTF also had low or no correlation with seed shape parameters (González et al. 2016), and chickpea seed shape parameters had low correlations with DTF and DTM (Sundaram et al. 2019).

Stability measures how stably genotypes or accessions perform across different environments; in other words, the measure of the constancy of features, over periods of time or across various spatial environments (Urruty et al. 2016; Reckling et al. 2021). Stability analyses (Fig. 4.4) showed the overall distribution of stability from each of the 324 accessions for each trait and evaluate which traits have more accessions closer to more stable values. In this study, the S^1 non-parametric statistic was calculated as the mean of the absolute rank differences of an accession over all site-years. The data for the AGILE LDP had a non-normal distribution even after transformation, so non-parametric methods for determining stability were considered more appropriate. The results demonstrated that there is a range of stability values of genotypes for each trait across site-years. Diameter and circularity had similar patterns in their stability scores, showing slightly more stable accessions (lower stability scores) and fewer accession with unstable ones. Height, on the other hand, had fewer stable accessions. Usually, parametric, and non-parametric stability analyses are used to estimate yield and yield related traits, and to a lesser extent to estimate other quantitative traits, related to quality for instance (Huehn 1990; van Eeuwijk et al. 2016; Kyratzis et al. 2022). Non-parametric stability analyses have been used in lentil to assess stability of yield (Sabaghnia et al. 2006; Sabaghnia et al. 2014) but this is the first time to my knowledge that it has been used to look at a quality trait.

Genotypes that are less influenced by the environment can be selected when looking at stability scores, however there were no genotypes that were consistently among the most stable for all traits. Fig. 4.4 shows a more general picture of what is happening with accessions, and it shows that some of them (less stable, i.e., with larger stability scores) should be discounted when

selecting for that particular trait. Stability results were visualized as distributions (Fig. 4.4) to demonstrate the magnitude of differences among individuals within the AGILE LDP for each trait, other studies, mainly looking at yield stability in different plant models including lentil have used fewer genotypes for evaluation (Abera et al. 2006; Karimizadeh et al. 2008; Ahmadi et al. 2015; Abdipour et al. 2017). By using fewer genotypes to analyze stability, stability results are typically shown in a table (Sabaghnia et al. 2006), however, since the AGILE LDP consists of many genotypes, results had to be visualized in a new way to be able to extract information about the general pattern of the stability of all genotypes for each trait.

Hierarchical clustering and PCA have been used together in phenotypic studies to cluster different types of data and to visualize data globally (e.g., Chen et al. 2014; Wright et al. 2020). PCA is useful for dimensionality reduction of data and hierarchical clustering groups data into clusters by categorizing data into hierarchical tree structures (dendrograms) (Jafarzadegan et al. 2019). Hierarchical clustering has become popular mainly because, unlike other types of clustering methods like k-means, the number of clusters arises from the clustering process itself by assessment of dendrograms (Jain et al. 1999; Karna and Gibert, 2022). Kumar et al. (2018) used hierarchical clustering to partition different *Lens* species into clusters to select for accessions enriched in minerals. Two studies identified clusters based on hierarchical clustering in seed weight (associated with seed shape) (Choukri et al. 2022; Tripathi et al. 2022), even though this can be an approximation to cluster seed shape, seed shape only differentiates large versus small seeds but does not indicate actual seed shape, this was showed when Fratini et al. 2007 found a significant but low correlations ($r = 0.34$) between the two traits. Other traits like plant height, grain yield or phenological traits such as early or late maturity and flowering (Dehghani et al. 2008; Wright et al. 2020; Tripathi et al. 2022) have also been studied using this method.

Hierarchical clustering (Fig. 4.5) helped visualize some of the relevant characteristics of the 3 clusters that were identified in the AGILE LDP based on circularity, diameter, height, and plumpness. Cluster 1 consisted of mainly small seeds with red cotyledons, however some small, yellow cotyledon lentils were also found in cluster 1. Cluster 3 consisted mainly of large, yellow cotyledon seeds. Cluster 2 contained a mixture of cotyledon color and diameter of seeds. Even though only two marked clusters were expected, the clustering plot (Fig. 4.5) showed that there is a spectrum and that lentil seeds have more than the two size and shape groups suggested by the

categories macro- and microsperma. Tripathi et al. (2022) used hierarchical clustering analysis with a diverse set of lentil germplasm also, however they only investigated seed diameter and did not look at the role of cotyledon color. They identified 12 clusters with some of them having the largest and smallest seeded lines and all the rest having a range of seed diameters. They probably yielded more clusters because they used a larger number of genotypes comprising indigenous and exotic collections.

The average of the stability scores of all seed shape traits for each cluster were calculated to learn which clusters would have more stable scores for each of the seed shape traits (Appendix Table C). The stability scores found for cluster 1 were the most stable for all traits except for plumpness, however, when plotting the distribution of the stability values of seed shape traits separately for each cluster (Appendix Fig. B.1) a wide distribution was found, which follows the same pattern as the distribution plots for stability (Fig. 4.4). Although there are some exceptions, seed diameter in cluster 1 (Appendix Fig. B.1 B) and seed plumpness in cluster 3 (Appendix Fig. B.1 D) had more stable accessions with a narrower distribution of stability scores, suggesting that when selecting for diameter or plumpness, genotypes from cluster 1 or 3 can be selected for more stability. Other studies in pulses like bean, chickpea and faba bean have studied the concepts of stability and clustering together, and have identified clusters with the most stable genotypes, although in all cases they use parametric stability analyses (Atta and Shah, 2009; Burbano-Erazo et al. 2021; Gela et al. 2023).

The GWAS approach has been widely used to detect associations between genotype (alleles) and phenotypic traits. It is a powerful tool that in the last 16 years has been useful in dissecting the genotype for variation of complex phenotypes including agronomic traits in plants (e.g., Huang et al. 2010, Tian et al. 2011, Horton et al. 2012, Tian et al. 2020). QTL were identified with the GWAS for all seed shape parameters. In the GWAS without any covariates, the most significant set of QTL detected was for diameter, on chromosome 1, that was consistent across all site-years and held up even when covariates were added (Fig.4.7, 4.8 and 4.9). In other cases, like with the set of QTL detected in chromosome 2 and 4 for diameter, QTL are detected only in some site-years (Fig.4.7, 4.8 and 4.9). Fedoruk et al. (2013) and Khazaei et al. (2018) also identified QTL for lentil seed diameter on chromosome 1. Fedoruk et al. (2013) also found QTL on chromosome 2 for different temperate site-years. These previously described QTL are most likely found within

the set of QTL identified in this study for diameter in those chromosomes (Fig. 4.7). However, both of these previous studies also identified QTL for diameter on chromosome 7, unlike the present study. Conversely, in this study a new set of QTL was identified on chromosome 4 for diameter for 4 site-years from different macro-environments.

In the case of the other seed shape traits, both Fedoruk et al. (2013) and Khazaei et al. (2018) found QTL for seed height and plumpness on different chromosomes when comparing both studies, Fedoruk et al. (2013) found more markers in chromosomes not found by Khazaei et al. (2018), in the present study the findings match for plumpness but only when covariates are added, whereas height is only matches the results from Fedoruk et al. (2013) for chromosome 2, however this study identified new QTL found on chromosome 4 in several site-years. Shahin et al. (2012) and Fedoruk et al. (2013) found a very strong negative correlation between seed diameter and seed plumpness, this translates to seeds having more plumpness when they are smaller. Khazaei et al. 2018 suggested that there are genomic regions associated with plumpness that are independent of diameter, however in this study, no QTL for plumpness were identified that were independent from diameter because when diameter is taken out of the equation (used as covariate) the only QTL that were identified (in chromosome 1) for plumpness, had QTL identified in less site-years when compared with other experiments using seed shape traits or phenological and environmental parameters. Circularity is a new shape descriptor that was not described when previous studies identifying QTL for seed shape were published, so the results presented here are new.

The previous studies identifying QTL for seed shape in lentil (Fedoruk et al. 2013; Khazaei et al. 2018) reaffirm some of the QTL identified in the present study. The differences in the results from this study and Fedoruk et al. 2013 and Khazaei et al. 2018 may come from the germplasm used, the site-years, as well as the method used to genotype and phenotype. Fedoruk et al. (2013) used QTL mapping on one RIL population as germplasm and only temperate site-years over two years, and Khazaei et al. (2018) used AM on 138 diverse lines as germplasm and only temperate site-years over two years. This study used the AGILE LDP with 324 lines and increased the environmental scope with 6 site-years from two different macro-environments. Genotyping and phenotyping were also different. The present study had more SNP coverage of the genome when compared with the previous studies, allowing access to the entire set of exons in the lentil

genome, which are the main target for marker assisted selection in breeding programs (Ogutcen et al. 2018). Both previous studies used traditional phenotyping with sieves with 1/16" increments, whereas the BELT platform directly measure seeds with more precision. Moreover, in the present study seed shape, phenological and environmental covariates were added to explore the effect of these traits on the others.

Significant QTL were also identified for other seed shape parameters but for only one macro-environment or when covariates like other seed shape traits, phenological or environmental parameters were added to the model. The addition of covariates is vital for GWAS experiments because it helps with the filtering of false positives arising from other variables coming from other phenotypic traits, e.g., phenological, or environmental traits, they can be understood as noise that is taken away from the analysis to evaluate if QTL are appearing or disappearing when compared with the GWAS experiment without covariates (Listgarten et al. 2013; Llinares-López et al. 2017). For instance, a shape QTL that is caused by a flowering time gene will disappear when a flowering parameter is added as a covariate.

Covariates added helped identifying new sets of QTL (Fig. 4.8 and 4.9) in comparison with the GWAS without any covariates (Fig. 4.7) as they account for variability due to the covariate, letting the true variability in the trait of interest be revealed independently. The QTL found for circularity on chromosome 2 was present in most GWAS experiments in Mediterranean site-years. When diameter, plumpness, DTF, DTS or TRT, were added as covariates, QTL were also found on chromosome 6 for Mediterranean site-years (Fig. 4.8 and 4.9).

The QTL found for diameter on chromosome 1 was present in all GWAS experiments and it was detected by most GWAS models, it was the most robust QTL detected, suggesting that none of the seed shape traits, phenological or environmental parameters caused difficulties in the identification of these QTL. Moreover, Fig. 4.10 shows one of the diameter QTL identified on chromosome 1 had a large effect in all site-years. The QTL found on chromosome 2 for diameter were found in all experiments however when the environmental covariates are added (TFT and TRT) in Fig. 4.9, the number of site-years were less than in any other experiment, suggesting that temperature may have an influence on these QTL for diameter. This may be in line with the results from Choukri et al. (2022) in which they found that higher temperatures yielded more seed diameter, although this does not match with the correlation results from the present study

(Fig. 4.3), because the correlations between diameter and temperature were not existent or significantly weak with Spearman correlation coefficients around 0.20.

When seed shape traits and environmental and phenological parameters were used as covariates, new QTL were found on chromosomes 1 and 2 for plumpness. The QTL on chromosome 2 were identified across all GWAS experiments, and the QTL on chromosome 1 was only identified when circularity or height were added as covariates (Fig. 4.8 A and C). In most site-years there were low but significant positive or non-significant correlations between plumpness and circularity and height, however in Rosthern 2017 and especially in Sutherland 2017 the correlations were stronger. Other QTL were found for plumpness on chromosome 2 but only when height or TFT were used as covariates (Fig. 4.8 C and Fig. 4.9 C, respectively). This may suggest that both height and TFT are interfering with the identification of this QTL.

For height, only when diameter and plumpness were added as covariates were QTL identified on chromosome 2. This may suggest that this set of QTL are associated with height but diameter may be interfering with their detection. This could be happening because both diameter and height are used to calculate plumpness, moreover, diameter and height had significant, moderate, and positive correlations across all site-years (Fig. 4.3). New QTL were identified when compared with previous results from Khazaei et al. (2018) and Fedoruk et al. (2013) who found QTL for height on chromosome 2. The weak association of QTL with height, probably comes from the lower heritability in comparison with the other seed shape traits. Diameter, on the other hand, had the highest heritability and yielded several QTL across all site-years on chromosome 1.

In many cases, QTL were identified more consistently in Mediterranean site-years, for instance, the QTL found for plumpness on chromosome 2 were detected mainly in Mediterranean site-years when diameter, height or TFT were added as covariates. This difference between macro-environments may arise from environmental factors not measured in this study such as radiation, which is an important factor of the transition of plants into flowering time, and which is crucial for seed development (Dotto et al. 2018; Osnato et al. 2022).

With the addition of BELT as a phenotyping method, it was possible to assess seed shape in a new way with new shape traits and improved accuracy. The BELT platform has allowed lentil seed shape to be phenotype and the trait to be dissected into several seed shape traits such as seed circularity, diameter, height, and plumpness. The most common approach to digitally phenotype

seeds is to take one single picture of multiple seeds in it to directly measure traits like diameter, height, or area. Narendra and Abdorrazzaghi (2013) used a computer camera and took pictures of multiple seeds per image. LeMasurier et al. (2014) used a commercial platform that takes the picture of multiple seeds and then cuts the image into single seed images. More recently Choukri et al. (2022) used another commercial platform that uses a camera, and it also takes a picture of multiple seeds. However, when multiple seeds are phenotyped together, the spatial resolution can be lower when compared to single seed imaging (Halcro et al. 2020). When comparing this phenotyping platform with other methods previously used to phenotype seed shape in lentil (Fedoruk et al. 2013; Khazaei et al. 2018), there was an improvement in precision of measurements. It also increased the scope of the phenotyping process by not only extracting seed shape traits but also color which is being investigated separately.

6 Conclusions and Future Work

The first objective of this study was to phenotype lentil seed shape using the new high-throughput phenotyping platform BELT and the second objective was to identify variations in seed shape parameters across different macro-environments. The results confirmed the first hypothesis, there was insignificant variance of seed shape across replication within a site-year. The third objective was to identify the possible environmental and phenological correlations with seed shape variation. The fourth objective was to determine the stability of the genotypes across macro-environments and the heritability of seed shape traits. The fifth objective was to identify clusters based on seed shape traits from each site-year and determine what seed characteristics determine each cluster. The second hypothesis was refuted, because not all correlations were significant. In addition, the third hypothesis, was also refuted because the results showed a spectrum of stability scores of accessions for all seed shape traits. The sixth objective was to identify new QTL which could be used to develop genetic markers as important tools for lentil seed shape selection in breeding programs. The results confirmed the fifth hypothesis, new genomic regions, significantly associated with seed shape, were detected via GWAS.

The overall high heritability of lentil seed shape parameters for each macro-environment and the absence of major correlations between shape parameters with both environmental and phenological factors suggests that lentil seed shape is stable across different macro-environments and site-years, this is supported by the results in stability and clustering. The clustering analysis coupled with the results from the stability analysis showed that cluster 1, composed mainly of red lentils, is the most stable cluster and could be the most appropriate set of accessions to use in breeding programs. The BELT platform helped to identify new QTL for the four lentil seed shape parameters evaluated in this study with seed diameter having the strongest association with QTL across all site-years.

The BELT phenotyping platform can provide information about seed color at the same time as it extracts information from seed shape. In the future, this phenotyping platform could be used to extract new seed shape traits that may be of economic importance, like circularity for instance. It could also be possible to map the testa patterns present in some genotypes.

Future work would involve improving the BELT platform to make the process of phenotyping even faster and the addition of new technologies that could extract other seed traits such as seed

coat thickness. In addition, further work is needed to validate the QTL identified in this study with the purpose of their application in lentil breeding programs.

7 References

Abdipour, M., Vaezi, B., Younessi-Hamzekhanlu, M. 2017. Nonparametric phenotypic stability analysis in advanced barley (Hordeum vulgare L.) genotypes. J. Crop Sci. Biotechnol. 20: 305–314. https://doi.org/10.1007/s12892-017-0050-0.

Abera, W., Labuschagne, M., & Maartens, H. 2006. Evaluation of maize genotypes using parametric and non-parametric stability estimates. Cereal Research Communications, 34(2/3), 925–931. http://www.jstor.org/stable/23789298.

Ahmadi, J., B. Vaezi, A. Shaabani, K. Khademi, S. Fabriki-Ourang, and A. PourAboughadareh. 2015. Non-parametric measures for yield stability in grass pea (Lathyrus sativus L.) advanced lines in semi warm regions. Journal of Agricultural Science and Technology 17: 1825–1838.

Alghamdi, S.S., Khan, A.M., Ammar, M.H., El-Harty, E.H., Migdadi, H.M., El-Khalik, S.M., Al-Shameri, A.M., Javed, M.M., AL-Faifi, S.A. 2013. Phenological, nutritional and molecular diversity assessment among 35 introduced lentil (Lens culinaris Medik.) genotypes grown in Saudi Arabia. Int. J. Mol. Sci. 15(1): 277-295.

Al-Khayri, J.M., Jain, S.M., Johnson, D.V. 2019. Advances in plant breeding strategies: industrial and food crops. Springer Nature 744p.

Alqudah, A.M., Sallam, A., Stephen Baenziger, P., Börner, A. 2020. GWAS: Fast-Forwarding Gene Identification in Temperate Cereals: Barley as a Case Study- A review. Journal of Advanced Research. 22: 119-135 doi:10.1016/j.jare.2019.10.013.

Álvarez, M.F., Mosquera, T., Blair, M.W. 2014. The Use of Association Genetics Approaches in Plant Breeding. Plant Breeding Reviews: Volume 38, 17–68. doi:10.1002/9781118916865.ch02.

Atta, B. M., Shah, T. M. 2009. Stability analysis of elite chickpea genotypes tested under diverse environments. Australian Journal of Crop Science, 3(5), 249-256.

Awada, L., Phillips, P.W.B., Smyth, S.J. 2018. The adoption of automated phenotyping by plant

breeders. Euphytica. 214, 148. https://doi.org/10.1007/s10681-018-2226-z.

Barulina, O.H. 1930. Lentil of USSR and other countries. Bull. Appl. Bot. Genet. Pl. Breed. 40: 1-319.

Brock, G., Pihur, V., Datta, S., Datta, S. 2008. clValid: An R package for cluster validation. Journal of Statistical Software. 25(4): 1-22. https://doi.org/10.18637/jss.v025.i04.

Burbano-Erazo, E., León-Pacheco, R.I., Cordero-Cordero, C.C., López-Hernández, F., Cortés, A.J., Tofiño-Rivera, A.P. 2021. Multi-Environment Yield Components in Advanced Common Bean (*Phaseolus vulgaris* L.) × Tepary Bean (*P. acutifolius* A. Gray) Interspecific Lines for Heat and Drought Tolerance. Agronomy. 11(10):1978. https://doi.org/10.3390/agronomy11101978.

Chang, F., Guo, C., Suo, F., Zhang, J., Wang, Z., Kong, J., He, Q., Sharmin R.A., & Zhao, T. 2018. Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. Front. Plant Sci, 9,1184. doi: 10.3389/fpls.2018.01184.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6): 1–36. https://doi.org/10.18637/jss.v061.i06.

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., Klukas, C. 2014. Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis. The Plant Cell, 26(12): 4636-4655. https://doi.org/10.1105/tpc.114.129601.

Choukri, H., El Haddad, N., Aloui, K., Hejjaoui, K., El-Baouchi, A., Smouni, A., Thavarajah, D., Maalouf, F., Kumar, S. 2022. Effect of High Temperature Stress During the Reproductive Stage on Grain Yield and Nutritional Quality of Lentil (Lens culinaris Medikus). Front. Nutr. 9:857469. doi: 10.3389/fnut.2022.857469.

Coventry, S.J., Barr, A.J., Eglinton, J.K., McDonald, G.K. 2003. The determinants and genome locations influencing grain weight and size in barley (*Hordeum vulgare* L.). Aust. J. Agric.

Res. 54: 1103-1115. doi: 10.1071/AR02194.

Crop Trust (2020) https://www.croptrust.org/. Accessed July 2020.

Dehghani, H., Sabaghpour, S. H., Sabaghnia, N. 2008. Genotype x environment interaction for grain yield of some lentil genotypes and relationship among univariate stability statistics. Spanish Journal of Agricultural Research, 6(3), 385-394.

Dissanayake, R., Braich, S., Cogan, N.O.I., Smith, K., Kaur S. 2020. Characterization of genetic and allelic diversity amongst cultivated and wild lentil accessions for germplasm enhancement. Front. Genet. 11:546. doi: 10.3389/fgene.2020.00546.

Domoney, C., Duc, G., Ellis, T. H., Ferrándiz, C., Firnhaber, C., Gallardo, K., Hofer, J., Kopka, J., Küster, H., Madueño, F., Munier-Jolain, N.G., Mayer, K., Thompson, R., Udvardi, M., Salon, C. 2006. Genetic and genomic analysis of legume flowers and seeds. Curr. Opin. Plant Biol. 9(2):133-141. doi: 10.1016/j.pbi.2006.01.014.

Dotto, M., Gómez, M.S., Soto, M.S., Casati, P. 2018. UV-B radiation delays flowering time through changes in the PRC2 complex activity and miR156 levels in Arabidopsis thaliana. Plant Cell Environ. 2018 Jun;41(6):1394-1406. doi: 10.1111/pce.13166.

Earp, M.A., Goode, E.L. 2017. Linkage Disequilibrium. In: Schwab, M. (eds) Encyclopedia of Cancer. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-46875-3_3368

Erskine, W., Williams, P.C., Nakkoul, H. 1985. Genetic and environmental variation in the seed size, protein, yield, and cooking quality of lentils. Field Crop Res. 12: 153-161. https://doi.org/10.1016/0378-4290(85)90061-9.

Erskine, W., Sarker, A. 2004. Lentil. In: Corke, H, Walker, C.E. (ed) Encyclopedia of grain scienes. Elsevier, London, UK, pp 142-150.

Erskine, W., Muehlbauer, F., Sarker, A., Sharma, B. 2009. The lentil: botany, production and uses. Edited by W. Erskine, F. Muehlbauer, A. Sarker, B. Sharma. CAB International.

FAOSTAT (2023) http://faostat.fao.org Accessed June 2023.

Fedoruk, M.J., Vandenberg, A, Bett, K.E. 2013. Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. The plant genome. 6(3). doi:10.3835/plantgenome2013.05.0012.

Ferguson, M.E., Robertson, L.D., Ford-Lloyd, B.V. 1998. Contrasting genetic variation amongst lentil landraces from different geographical origins. Euphytica. 102: 265-173. https://doi.org/10.1023/A:1018331432580.

Fiorani, F., Schurr, U. 2013. Future Scenarios for Plant Phenotyping. Annual Review of Plant Biology. 64(1), 267–291. doi:10.1146/annurev-arplant-050312-.

Fratini, R., Durán, Y., García, P., & Pérez de la Vega, M. 2007. Identification of quantitative trait loci (QTL) for plant structure, growth habit and yield in lentil. Spanish J. Agricul. Res, 5(3), 348–356. https://doi.org/10.5424/sjar/2007053-255.

Furman, B.J., Coyne, C., Redden, B., Sharma, S.K., Vishnyakova, M. 2009. Genetic resources: collection, characterization, conservation and documentation. In: W. Erskine and F.J. Muehlbauer, editors, The lentil: Botany, production and uses. CAB International, Wallingford, UK. p. 391–424.

Gali, K.K., Sackville, A., Tafesse, E.G., Lachagari, V.B.R., Mcphee, K., Hybl, M., Mikić, A., Smýkal, P., McGee, R., Burstin, J., Domoney, C., Ellis, T.H.N., Tar'an, B., & Warketin, T.D. 2019. Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.) Front. Plant Sci, 10,1538. doi: 10.3389/fpls.2019.01538.

Galili, T. 2015. "dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering." Bioinformatics. doi:10.1093/bioinformatics/btv428, https://academic.oup.com/bioinformatics/article-pdf/31/22/3718/17122682/btv428.pdf, https://academic.oup.com/bioinformatics/article/31/22/3718/240978/dendextend-an-R-package-for-visualizing-adjusting.

Gan, Y., Jayakumar, P., Zentner, R., McDonald, C. 2006. Selection for seed size and its impact on grain yield and quality in kabuli chickpea. Canadian Journal of Plant Science. 86(2): 345-352. https://doi.org/10.4141/P05-113.

Gela, T. S., Khazaei, H., Podder, R., Vandenberg, A. 2023. Dissection of genotype-by-environment interaction and simultaneous selection for grain yield and stability in faba bean (*Vicia faba* L.). Agronomy Journal, 115, 474– 488. https://doi-org.cyber.usask.ca/10.1002/agj2.21268

Giordani, W., Castro Gama, H., Fernano Chiorato, A., Franco Garcia A. A., & Carneiro Viera, M.L. 2022. Genome-wide association studies dissect the genetic architecture of seed shape and size in common bean. G3 genes|Genomes|Genetics, 12(4): jkac048. https://doi.org/10.1093/g3journal/jkac048.

González, A., Yuste-Lisbona, F., Saburido, S., Bretones, S., De Ron, A., Lozano, R., Santalla, M. 2016. Major Contribution of Flowering Time and Vegetative Growth to Plant Production in Common Bean as Deduced from a Comparative Genetic Mapping. Front. Plant Sci. 7:1940. doi: 10.3389/fpls.2016.01940.

Govindaraj, M., Vetriventhan, M., Srinivasan, M. 2015. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int. 2015:431487. doi: 10.1155/2015/431487.

Guerra-García, A., Gioia, T., von Wettberg, E., Logozzo, G., Papa, R., Bitocchi, E., Bett, K. E. 2021. Intelligent characterization of lentil genetic resources: Evolutionary history, genetic diversity of germplasm, and the need for well-represented collections. Current Protocols, 1, e134. doi: 10.1002/cpz1.134.

Gupta, P.K., Rustgi, S., Kumar, N. 2006. Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. Genome. 49(6): 565-571. doi:10.1139/g06-063.

Haile, T., Heidecker, T., Wright, D., Neupane, S., Ramsay, L., Vandenberg, A., Bett, K.E. 2020. Genomic selection for lentil breeding: empirical evidence. Plant Genome. 13(1): e20002. doi: 10.1002/tpg2.20002

Halcro, K., McNabb, K., Lockinger, A., Socquet-Juglard, D., Bett, K.E., Noble, S.D. 2020. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. Plant

Methods 16, 49. https://doi.org/10.1186/s13007-020-00591-8

Hamdi, A., Erskine, W., Gates, P. 1991. Relationships among economic characters in lentil. Euphytica. 57: 109-116. https://doi.org/10.1007/BF00023068.

Horton, M.W., Hancock, A.M., Huang, Y.S., Toomajian, C., Atwell, S., Auton, A., Muliyati, N.W., Platt, A., Sperone, F.G., Vilhjálmsson, B.J., Nordborg, M., Borevitz, J.O., Bergelson, J. 2012. Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat Genet. 44(2):212-6. doi: 10.1038/ng.1042.

Hossain, S., Ford, R., McNeil, D., Pittock, C., Panozzo, J.F. 2010. Development of a selection tool for seed shape and QTL analysis of seed shape with other morphological traits for selective breeding in chickpea (*Cicer arietinum* L.). Aus. J. Crop Sci. 4: 278-288.

Hovav, R., Upadhyaya, K.C., Beharav, A., & Abbo, S. 2003. Major flowering time gene and polygene effects on chickpea seed weight. Plant Breed, 122,539–541. doi: 10.1111/j.1439-0523.2003.00895.x.

Huang, M., Liu, X., Zhou, Y., Summers, R.M., Zhang, Z. 2019. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience. 8(2):giy154. doi: 10.1093/gigascience/giy154

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.F., Li, J., Han, B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 42(11):961-7. doi: 10.1038/ng.695.

Huehn, M. 1990. Nonparametric measures of phenotypic stability. Part 1: Theory. Euphytica 47, 189–194. https://doi.org/10.1007/BF00024241.

Jafarzadegan, M., Safi-Esfahani, F., Beheshti, Z. 2019. Combining hierarchical clustering approaches using the PCA method. Expert Systems with Applications, 137: 1-10, 10.1016/j.eswa.2019.06.064.

Jain, A.K., Murty, M.N., Flynn, P.J. 1999. Data clustering: a review. ACM Comput. Surv. (CSUR) 31(3):264–323. https://doi.org/10.1145/331499.331504.

Jeyakumar, J., Martin, J., Ali, A., Wang, W., & Thiruvengadam, M. 2020. Characterizing the Role of the miR156-SPL Network in Plant Development and Stress Response, Plants, 9, 1206. https://doi.org/10.3390/plants9091206.

Kaur, G., Kumar, S., Nayyar, H., & Upadhyaya, H.D. 2008. Cold stress Injury during the Pod-Filling Phase in Chickpea (Cicer arietinum L.): Effects on Quantitative and Qualitative Components of Seeds. Journal of Agronomy and Crop Science, 194, 457-464. https://doi.org/10.1111/j.1439-037X.2008.00336.x.

Karimizadeh, R., Safikhani Nasimi, M., Mohammadi, M., Seyyedi, F., Mahmoodi, A., Rostami, B. 2008. Determining rank and stability of lentil genotypes in rainfed condition by nonparametric statistics. JWSS-Isfahan University of Technology, 12(43), 93-102.

Karna, A., Gibert, K. 2022. Automatic identification of the number of clusters in hierarchical clustering. Neural Comput & Applic. 34: 119–134. https://doi.org/10.1007/s00521-021-05873-3.

Khazaei, H., Caron, C.T., Fedoruk, M., Diapari, M., Vandenberg, A., Coyne, C., McGee, R., Bett. K. 2016. Genetic diversity of cultivated lentil (Lens culinaris Medik.) and its relation to the world's agro-ecological zones. Front Plant Sci. 7:1093.

Khazaei, H., Fedoruk, M., Caron, C.T., Vandenberg, A., Bett, K.E. 2018. Single nucleotide polymorphism markers associated with seed quality characteristics of cultivated lentil. Plant genome. 11: 170051. doi: 10.3835/plantgenome2017.06.0051.

Klasen, J.R., Barbes, E., Meier, L., Meinshausen, N., Bühlmann, P., Koorneef, M., Busch, & W., Schneeberger, K. 2016. A multi-marker association method for genome-wide association studies without the need for population structure. Nat. Commun, 7,13299. https://doi.org/10.1038/ncomms13299.

Kumar, S., Choudhary, A., Rana, K., Sarker, A., Singh, M. 2018. Bio-fortification potential of global wild annual lentil core collection. PLoS ONE 13(1): e0191122.

https://doi.org/10.1371/journal.pone.0191122.

Kyratzis, A., Pallides, A., Katsiotis, A. 2022. Investigating Stability Parameters for Agronomic and Quality Traits of Durum Wheat Grown under Mediterranean Conditions. Agronomy, 12: 1774. https://doi.org/10.3390/agronomy12081774.

Ladizinsky, G. 1993. Lentil domestication: on the quality of evidence and arguments. Econ. Bot. 47: 60-64. https://doi.org/10.1007/BF02862206.

Lê, S., Josse, J., Husson, F. 2008. "FactoMineR: A Package for Multivariate Analysis." Journal of Statistical Software, 25(1), 1–18. doi:10.18637/jss.v025.i01.

Lei, L., Wang, L., Wang, S., Wu, J. 2020. Marker-Trait Association Analysis of Seed Traits in Accessions of Common Bean (Phaseolus vulgaris L.) in China. Front. Genet. 11:698. doi: 10.3389/fgene.2020.00698.

LeMasurier, L.S., Panozzo, J.F., Walker, C.K. 2014. A digital image analysis method for assessment of lentil size traits. J Food Eng. 128:72–8. https://doi.org/10.1016/j.jfoodeng.2013.12.018.

Leng, P.F., Luberstedt, T., Xu, M.L. 2017. Genomics-assisted breeding—a revolutionary strategy for crop improvement. J. Intregr. Agric.16: 2674-2685.

Li, C., Fu, Y., Sun, R., & Wang, Y. 2018. Single-locus and multi-locus genome-wide associaton studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutun* L.) Front. Plant Sci, 9,1083. doi: 10.3389/fpls.2018.01083.

Li, N., Xu, R., Li, Y. 2019. Molecular networks of seed size control in plants. Annu Rev Plant Biol. 70: 435-463. doi: 10.1146/annurev-arplant-050718-095851.

Lipka, A., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P., Gore, M., Buckler, E., Zhang, Z. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics. 28(18): 2397-2399. https://doi.org/10.1093/bioinformatics/bts444.

Listgarten, J., Lippert, C., Kang, E.Y., Xiang, J., Kadie, C.M., Heckerman, D. 2013. A powerful

and efficient set test for genetic markers that handles confounders. Bioinformatics. 29(12):1526-33. doi: 10.1093/bioinformatics/btt177.

Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z. 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet.12(2): e1005767. doi: 10.1371/journal.pgen.1005767. Erratum in: PLoS Genet. 12(3):e1005957.

Llinares-López, F., Papaxanthos, L., Bodenham, D., Roqueiro, D.; COPDGene Investigators; Borgwardt, K. 2017. Genome-wide genetic heterogeneity discovery with categorical covariates. Bioinformatics. 33(12):1820-1828. doi: 10.1093/bioinformatics/btx071.

Lobos, G. A., Camargo, A. V., del Pozo, A., Araus, J. L., Ortiz, R., Doonan, J. H. 2017. Editorial: Plant Phenotyping and Phenomics for Plant Breeding. Frontiers in Plant Science, 8.

Lombardi, M., Materne, M., Cogan, N.O.I., Rodda, M., Daetwyler, H.D., Slater, A.T., Forster, J.W., Kaur, S. 2014. Assessment of genetic variation within global collection of lentil (Lens culinarisMedik.) cultivars and landraces using SNP markers. BMC Genet. 15: 150. https://doi.org/10.1186/s12863-014-0150-3.

Mangiafico, S.S. 2023. rcompanion: Functions to Support Extension Education Program Evaluation. Rutgers Cooperative Extension, New Brunswick, New Jersey. version 2.4.30, https://CRAN.R-project.org/package=rcompanion/.

Mascher, M., Schreiber, M., Scholz, U. 2019. Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. Nat. Genet. 51: 1076-1081. https://www.nature.com/articles/s41588-019-0443-6.

Muehlbauer, F.J., Summerfield, R.J., Kaiser, W.J., Boerboom, C.M., Welsh-addux, M.M., Short, R.W. 1992. Principles and practice of lentil production. Electronic publication at https://www.ars.usda.gov/is/np/lentils/lentilsintro.htm.

Muhammad, A., Li, J., & Hu, W. 2021. Uncovering genomic regions controlling plant architectural traits in hexaploid wheat using different GWAS models. Sci. Rep, 11,6767. https://doi.org/10.1038/s41598-021-86127-z.

Narendra, V.G.; Abdorrazzaghi, M. 2013. An intelligent system for identification of Indian Lentil types using Artificial Neural Network (BPNN). *IOSR J.* Comput. Eng. 15: 54–60.

Nassar, R., & Hühn, M. 1987. Studies on Estimation of Phenotypic Stability: Tests of Significance for Nonparametric Measures of Phenotypic Stability. Biometrics, 43(1): 45–53. https://doi.org/10.2307/2531947

Nayyar, H., Kaur, G., Kumar, S., & Upadhyaya, H.D. 2007. Low temperature effect Effects during Seed Filling on Chickpea Genotypes (*Cicer arietinum* L.): Probing Mechanisms affecting Seed Reserves and Yield. Journal of Agronomy and Crop Science, 193, 336-344. https://doi.org/10.1111/j.1439-037X.2007.00269.x.

Nielsen, K.M.E., Duddu, H.S.N., Bett, K.E., Shirtliffe, S.J. 2022. UAV Image-Based Crop Growth Analysis of 3D-Reconstructed Crop Canopies. Plants. 11: 2691. https://doi.org/10.3390/plants11202691.

Ogutcen, E., Ramsay, L., von Wettberg, E.B., Bett, K.E. 2018. Capturing variation in Lens (Fabaceae): Development and utility of an exome capture array for lentil. Appl Plant Sci. 6(7):e01165. doi: 10.1002/aps3.1165.

Osnato, M., Cota, I., Nebhnani, P., Cereijo, U., Pelaz, S. 2022. Photoperiod Control of Plant Growth: Flowering Time Genes Beyond Flowering. Front. Plant Sci. 12:805635. doi: 10.3389/fpls.2021.805635.

Polanco, C., Sáenz de Miera, L. E., González, A. I., García, P., Fratini, R., Vaquero, F., Vences, F. J., Pérez de la Vega, M. 2019. Construction of a high-density interspecific (Lens culinaris x L. odemensis) genetic map based on functional markers for mapping morphological and agronomical traits, and QTLs affecting resistance to Ascochyta in lentil. PloS one, 14(3), e0214409. https://doi.org/10.1371/journal.pone.0214409.

Ponce, K., Zhang, Y., Guo, L., Leng, Y., Ye, G. 2020. Genome-wide association study of grain size traits in Indica rice multiparent advanced generation intercross (MAGIC) population. Front. Plant Sci. 11:395. doi: 10.3389/fpls.2020.00395.

Popat, R., Patel, R., Parmar, D. 2020. Variability: genetic variability analysis for plant breeding

research. R package version 0.1.1. http://CRAN.R- project.org/package= variability.

Pour-Aboughadareh, A., Yousefian, M., Moradkhani, H., Poczai, P., Siddique, K.H.M. 2019. STABILITYSOFT: A new online program to calculate parametric and non-parametric stability statistics for crop traits. Appl Plant Sci. 7(1): e01211. doi: 10.1002/aps3.1211.

Pritchard, J.K., Stephens, M., Donelly, P. 2000. Inference on population structure using multilocus genotype data. Genetics. 155: 945-959. https://doi.org/10.1093/genetics/155.2.945.

Qu, Z., Wu, Y., Hu, D., Li., T. Liang, H., Ye, F., Xue, J., Xu, S. 2022. Genome-wide association analysis for candidate genes contributing to kernel-related traits in maize. Front. Plant Sci. 13:872292. doi: 10.3389/fpls.2022.872292.

Rafalski, J.A. 2010. Association genetics in crop improvement. Current Opinion in Plant Biology. 13(2):174-180. DOI: 10.1016/j.pbi.2009.12.004.

Reckling, M., Ahrends, H., Chen, T.W., Eugster, W., Hadasch, S., Knapp, S., Laidig, F., Linstädter, A., Macholdt, J., Piepho, H., Schiffers, K., Döring, T.F. 2021. Methods of yield stability analysis in long-term field experiments. A review. Agron. Sustain. Dev. 41: 27. https://doi.org/10.1007/s13593-021-00681-4.

Revelle, W. 2023. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.3.6, https://CRAN.R-project.org/package=psych.

Roberts, E.H., Summerfield, R.J., Muehlbauer, F.J., Short, R.W. 1986. Flowering in lentil (Lens culinaris Medic.) the duration of the photoperiodic inductive phase as a function of accumulated daylength above the critical photoperiod. Annals of Botany. 58: 235-248.

RStudio Team 2020. RStudio: Integrated development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Sabaghnia, N., Dehghani, H., Sabaghpour, S.H. 2006. Nonparametric Methods for Interpreting Genotype × Environment Interaction of Lentil Genotypes. Crop Sci., 46: 1100-

1106. https://doi.org/10.2135/cropsci2005.06-0122.

Sabaghnia, N., R. Karimizadeh, and M. Mohammadi. 2014. Graphic analysis of yield stability in new improved lentil (Lens culinaris Medik.) genotypes using nonparametric statistics. Acta Agric. Slov. 103, 113-127. Doi: 10.14720/aas.2014.103.1.12.

Sakamoto, L., Kajiya-Kanegae, H., Noshita, K., Takanashi, H., Kobayashi, M., Kudo, T., Yano, K., Tokunaga, T., Tsutsumi, N., Iwata, H. 2019. Comparison of shape quantification methods for genomic prediction, and genome-wide association study of sorghum seed morphology. PLoS ONE 14(11): e0224695. https://doi.org/10.1371/journal.pone.0224695.

Sandhu, J.S., Singh, S. 2007. History and Origin. Pages 1-9 in Lentil. Springer Netherlands, Dordrecht. doi: 10.1007/978-1-4020-6313-8_1.

Schwartz, M.D. 2013. Introduction. In: Schwartz MD (ed). Phenology: an integrative environmental science. Springer, Netherlands, pp 3-7.

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q., Nordborg, M. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structures populations. Nat. Genet. 44: 825-830. https://doi.org/10.1038/ng.2314.

Shahin, M.A., Symons, S.J., Wang, N. 2012. Predicting dehulling efficiency of lentils based on seed size and shape characteristics measured with image analysis. Quality assurance and safety of crops & foods. 4(1): 9-16.

Shao, Z., Shao, J., Huo, X. 2022. Identification of closely associated SNPs and candidate genes with seed size and shape via deep re-sequencing GWAS in soybean. Theor. Appl. Genet. https://doi.org/10.1007/s00122-022-04116-w.

Sharpe, A.G., Ramsay, L., Sanderson, L., Fedoruk, M., Clarke, W.E., Li, R., Kagale, S., Vijayan, P., Vandenberg, A., Bett, K.E. 2013. Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. BMC Genomics. 14: 92. https://doi.org/10.1186/1471-2164-14-192

Singh, B.D., Singh, A.K. 2015a. Association Mapping. In: marker-assisted plant breeding:

principles and practices. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2316-0_8

Singh, B.D., Singh, A.K. 2015b. Phenomics. In: marker-assisted plant breeding: principles and practices. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2316-0_15.

Sita, K., Sehgal, A., HanumanthaRao, B., Nair, R. M., Vara Prasad, P. V., Kumar, S., Gaur, P. M., Farooq, M., Siddique, K., Varshney, R. K., Nayyar, H. 2017. Food Legumes and Rising Temperatures: Effects, Adaptive Functional Mechanisms Specific to Reproductive Growth Stage and Strategies to Improve Heat Tolerance. Front. in plant sci. 8, 1658. https://doi.org/10.3389/fpls.2017.01658.

Sita, K., Sehgal, A., Bhandari, K., Kumar, J., Kumar, S., Singh, S., Siddique K.H., Nayyar, H. 2018. Impacto f heat stress during seed filling on seed quality and seed yield in lentil (Lens culinaris Medikus) genotypes. J. Sci. Food Agric. 98(13): 5134-5141. doi: 10.1002/jsfa.9054

Slinkard, A.E., Vandenberg, A. 1995. Lentil. In: A.E. Slinkard and D.R. Knott (eds.). Harvest of gold. University extension press, University of Saskatchewan. p 191.

Soto-Cerda, B.J., Aravena, G., Cloutier, S. 2021. Genetic dissection of flowering time in flax (Linum usitatissimum L.) through single- and multi- locus genome-wide association studies. Mol. Genet. Genomics. 296(4): 877-891. doi: 10.1007/s00438-021-01785-y.

Subedi, M., Bett, K.E., Khazaei, H., Vandenberg, A. 2018. Genetic mapping of milling quality traits in lentil (*Lens culinaris* Medik.). Plant Genome. 11(2).

Summerfield, R.J., Roberts, E.H., Erskine, W., Ellis, R.H. 1985. Effects of temperature and photoperiod on flowering in lentils (*Lens culinaris* Medic.), Annals of Botany. 56(5): 659–671, https://doi.org/10.1093/oxfordjournals.aob.a087055.

Sundaram, P., Samineni, S., Sajja, S. B., Roy, C., Singh, S. P., Joshi, P., & Gaur, P. M. 2019. Inheritance and relationships of flowering time and seed size in kabuli chickpea. Euphytica, 215(9). doi:10.1007/s10681-019-2464-8.

Tanabata, T., Shibaya, T., Hori, K., Ebana, K., Yano, M. 2012. SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiology.

160(4): 1871-1880. https://doi.org/10.1104/pp.112.205120.

Thennarasu K. 1995. On certain non-parametric procedures for studying genotype environment interactions and yield stability. Indian Journal of Genetics 60: 433–9.

Tian, D., Wang, P., Tang, B., Teng, X., Li, C., Liu, X., Zou, D., Song, S., Zhang, Z. 2020. GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. Nucleic Acids Res. 48(D1): D927-D932. doi: 10.1093/nar/gkz828.

Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B., Buckler, E.S. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet. 43(2):159-62. doi: 10.1038/ng.746.

Tibbs Cortes, L., Zhang, Z., Yu, J. 2021. Status and prospects of genome-wide association studies in plants. Plant Genome. 14: e20077. https://doi.org/10.1002/tpg2.20077.

Tiwari, U., Singh Bawa, A. 2021. Pulses: quality standards and evaluation. In B.K. Tiwari, A. Gowen, B. McKenna (Eds.), Pulse Foods (pp. 351-368). Academic Press. https://doi.org/10.1016/B978-0-12-818184-3.00014-3.

Tripathi, K., Kumari, J., Gore, P., Mishra, D., Singh, A., Mishra, G., Gayacharan, C., Dikshit, H., Singh, N., Semwal, D., Mehra, R., Bhardwaj, R., Bansal, R., Rana, J., Kumar, A., Gupta, V., Singh, K., Sarker, A. 2022. Agro-Morphological Characterization of Lentil Germplasm of Indian National Genebank and Development of a Core Set for Efficient Utilization in Lentil Improvement Programs. Front. Plant Sci. 12:751429. doi: 10.3389/fpls.2021.751429.

Tullu, A., Diederichsen, A., Suvorova, G., Vandenberg, A. 2011. Genetic and genomic resources of lentil: status, use and prospects. Plant Genet. Resources. 9(1): 19-29. doi:10.1017/S1479262110000353.

Urruty, N., Tailliez-Lefebvre, D., Huyghe, C. 2016. Stability, robustness, vulnerability and resilience of agricultural systems. A review. Agron. Sustain. Dev. 36: 15. https://doi.org/10.1007/s13593-015-0347-5.

Vandenberg, A. 2009. Postharvest processing and value addition. In: W. Erskine and F.J. Muehlbauer, editors, The lentil: Botany, production and uses. CAB International, Wallingford, UK. p. 391–424.

van Eeuwijk, F.A., Bustos-Korts, D.V., Malosetti, M. 2016. What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions?. Crop Science, 56: 2119-2140. https://doi.org/10.2135/cropsci2015.06.0375.

Verma, P., Goyal, R., Chahota, R.K. Sharma T.R., Abdin, M.Z., Bhatia, S. 2015. Construction of a genetic linkage map and identification of QTLs for seed weight and seed size traits in lentil (*Lens culinaris* Medik.). PLoS ONE. 10(10): e0139666. https://doi.org/10.1371/journal.pone.0139666

Vishwakarma, R.K., Shivhare, U.S., Gupta, R.K., Yadav, D.N., Jaiswal, A., Prasad, P. 2018. Status of pulse milling processes and technologies: A review. Crit Rev Food Sci Nutr. 58(10):1615-1628. doi:10.1080/10408398.2016.1274956

Wang, Y, Upadhyaya, H.D., Kole, C. 2014. Genetics, genomics and breeding of sorghum. CRC Press 366p.

Weber, H., Borisjuk, L., Wobus, U. 2005. Molecular physiology of legume seed development. Annu. Rev. Plant Biol. 56: 253-279. doi: 10.1146/annurev.arplant.56.032604.144201.

Wei, T., Simko, V. 2021. R package 'corrplot': visualization of a correlation matrix (version 0.92). Available from https://github.com/taiyun/corrplot.

Weller, J. L., Ortega, R. 2015. Genetic control of flowering time in legumes. Frontiers in plant science, 6, 207. https://doi.org/10.3389/fpls.2015.00207

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag New York. [online]. Available: https://www.springer.com/gp/book/9783319242750 [2021 Feb. 1]

Wilke, C. 2022. ggridges: Ridgeline Plots in 'ggplot2'. R package version 0.5.4, https://wilkelab.org/ggridges/.

Wood, J.A., Knigths, E.J., Harden, S. 2008. Milling performance in desi-type chickpea (*Cicer arietinum* L.): effects of genotype, environment and seed size. J. Sci. Food Agric. 88, 108-115. doi:10.1002/jsfa.3053

Wood, J.A., Malcolmson, L.J. 2011. Pulse milling technologies. In: B. Tiwari, A. Gowen, and B. McKenna, editors, Pulse foods: Processing, quality, and nutraceutical applications. Academic Press, Cambridge, MA. p. 193–221.

Wright, D. 2022. gwaspr: GWAS plots with R. R package version 1.0. https://derekmichaelwright.github.io/gwaspr.

Wright, D., Neupane, S., Heidecker, T., Haile, T., Chan, C., Coyne, C., McGee, R., Udupa, S., Henkrar, F., Barilli, E., Rubiales, D., Gioia, T., Logozzo, G., Marzario, S., Mehra, R., Sarker, A., Dhakal, R., Anwar, B., Sarkar, D., Vandenberg, A., Bett, K. 2020. Understanding photothermal interactions will help expland production range and increase genetic diversity of lentils (*Lens culinaris* Medik.). Plants, People, Planet. 00: 1-11. https://doi.org/10.1002/ppp3.10158.

Xu, M., Hu, T., Zhao, J., Park, M. Y., Earley, K. W., Wu, G., Yang, L., Poethig, R. S. 2016. Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana. PLoS genetics, 12(8), e1006263. https://doi.org/10.1371/journal.pgen.1006263.

Xu, Y. Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., Xu, S., Yang, Z., Xu, C. 2018. Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. Front. Plant Sci. 9:1311. doi: 10.3389/fpls.2018.01311.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203-208. https://doi.org/10.1038/ng1702.

Zohary, D. 1999. Monophyletyc vs. polyphyletic origin of the crops on which agriculture was founded in the near east. Genet. Resour. Crop Evol. 46:133-142.

https://doi.org/10.1023/A:1008692912820.

8 Appendices

Appendix A

A pre-analysis to increase the scope of the research into six locations and two macro-environments was performed by assessing whether accessions had similar results when comparing replications. 20 accessions were used from 3 replicates and from three locations.

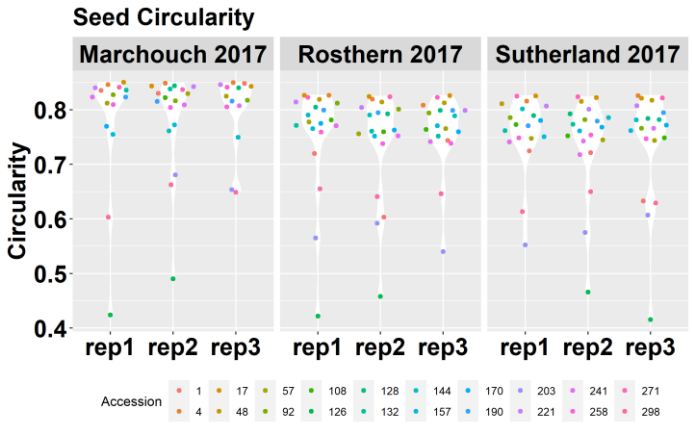

Figure A. 1. Mean seed circularity of a subset of 20 accessions selected from the AGILE LDP grown at Sutherland, Rosthern, and Morocco for 2017, by replication.

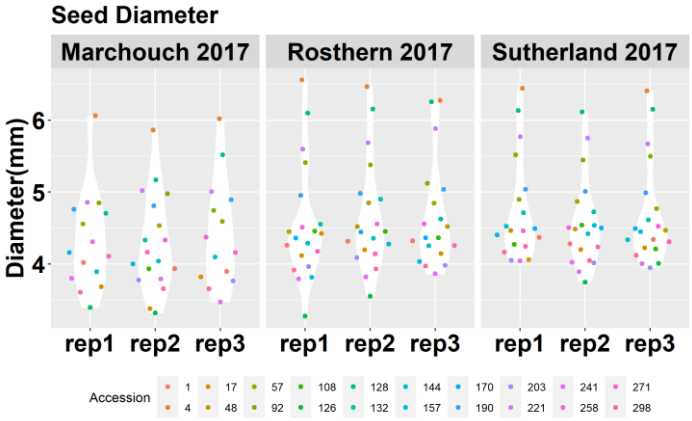

Figure A. 2. Mean seed diameter of a subset of 20 accessions selected from the AGILE LDP grown at Sutherland, Rosthern, and Morocco for 2017, by replication.

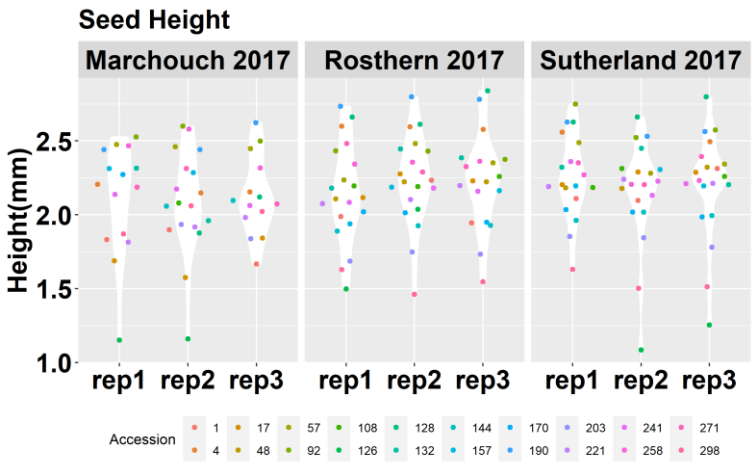

Figure A. 3. Mean seed height of a subset of 20 accessions selected from the AGILE LDP grown at Sutherland, Rosthern, and Morocco for 2017, by replication.

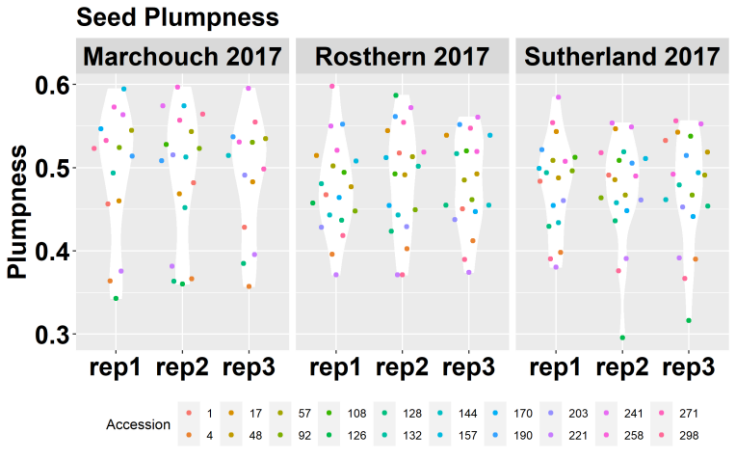

Figure A. 4. Mean seed plumpness of a subset of 20 genotypes selected from the AGILE LDP grown at Sutherland, Rosthern, and Morocco for 2017, by replication.

Table A. P-values from Kruskal-Wallis test for three replications within each location of seed shape parameters.

Seed size and shape parameter	Degrees of freedom	Sutherland	Degrees of freedom	Rosthern	Degrees of freedom	Morocco
		P-value		P-value		P-value
Height	2	0.762	2	0.522	2	0.927
Plumpness	2	0.970	2	0.852	2	0.953
Diameter	2	0.807	2	0.760	2	0.874
Circularity	2	0.764	2	0.649	2	0.870

Appendix B

Table B. Range, Mean (SD) and broad sense heritability of seed shape (circularity, diameter, height, and plumpness) from 324 lentil germplasm over locations and years.

Year	Location	Circularity		Diameter		Height		Plumpness	
		Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)
2017	Sutherland	0.76-0.88	0.85 (0.02)	3.87-8.11	5.49 (0.92)	1.90-4.08	2.69 (0.23)	0.27-0.90	0.49 (0.07)
	Rosthern	0.77-0.88	0.84 (0.02)	3.63-8.57	5.42 (0.98)	1.66-4.17	2.67 (0.35)	0.30-0.70	0.50 (0.07)
	Metaponto	0.76-0.88	0.84 (0.01)	3.66-7.53	5.13 (0.79)	1.52-3.37	2.41 (0.23)	0.33-0.63	0.47 (0.06)
	Cordoba	0.76-0.88	0.84 (0.01)	3.84-7.85	5.20 (0.90)	1.11-2.83	2.28 (0.22)	0.28-0.60	0.45 (0.06)
	Marchouch	0.82-0.88	0.85 (0.01)	3.50-7.37	4.99 (0.78)	1.61-3.06	2.34 (0.23)	0.32-0.62	0.47 (0.06)
2018	Sutherland	0.79-0.88	0.84 (0.01)	3.82-8.29	5.69 (0.90)	2.21-3.64	2.80 (0.27)	0.35-0.68	0.50 (0.06)
Heritability (Broad sense)		0.7500		0.9022		0.4482		0.7714	

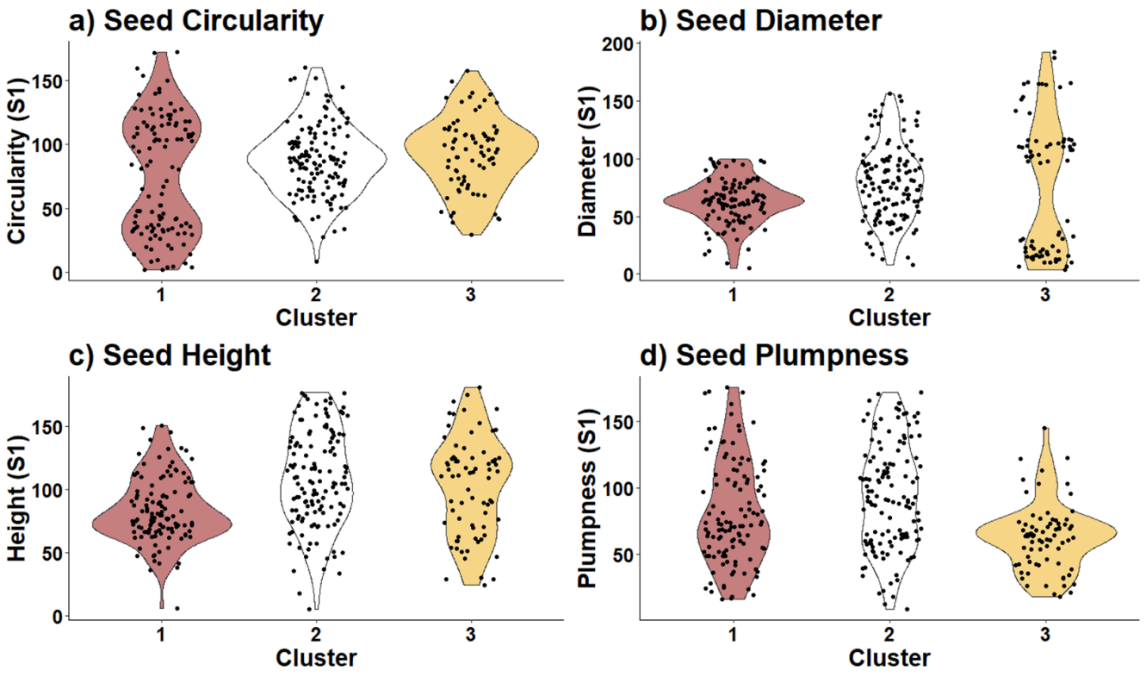

Figure B. 1. Distribution of stability values (S^1) of lentil seed shape traits per cluster. Color red and yellow represent cotyledon color and white represents the mixture of both.

Appendix C

Table C Mean of stability statistic (S^1) for seed shape traits and clusters.

Cluster	Circularity (S^1)	Diameter (S^1)	Height (S^1)	Plumpness (S^1)
1	75.17	62.76	83.83	80.75
2	87.52	79.11	108.27	93.66
3	92.85	75.16	102.63	62.55

Appendix D

Tables showing stability parameters of 324 lentil genotypes for each seed shape trait: circularity, diameter, height, and plumpness can be found in https://knowpulse.usask.ca/experiment/phenomics/High-throughput-phenotyping-of-LDP-with-BELT.

Appendix E - L

Manhattan plots from GWAS with covariates (circularity, diameter, height, plumpness, DTF, DTS, TFT and TRT) and without covariates for each seed trait in 6 site-years can be found in https://knowpulse.usask.ca/experiment/phenomics/High-throughput-phenotyping-of-LDP-with-BELT.

Appendix M

Summary plot of significant GWAS results of diameter, circularity, height, plumpness, and seed mass can be found in https://knowpulse.usask.ca/experiment/phenomics/High-throughput-phenotyping-of-LDP-with-BELT.