FLOWERING TIME RESPONSE OF DIVERSE LENTIL (*Lens culinaris* Medik.)
GERMPLASM GROWN IN MULTIPLE ENVIRONMENTS

A Thesis Submitted to the College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Plant Sciences
University of Saskatchewan

By **Sandesh Neupane**

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 or the Dean of the College 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 use of the material in this thesis, whole or part, should be addressed to:

Head of the Department of Plant Sciences
University of Saskatchewan
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

Adaptation of lentil (*Lens culinaris* Medik.) germplasm from one environment to another is a complex process and days to flower (DTF) is considered as the primary phenological stage determining the adaptation of genotypes. Studies revealed that temperature and photoperiod are major environmental factors defining DTF. This research was conducted with the objective of understanding the variation of DTF governed mainly by temperature and photoperiod and their interactions and identifying genomic regions and candidate genes or markers associated with DTF in specific environments. To accomplish this, 324 lentil genotypes were grown in three replications at ten field locations over two seasons in three major lentil growing macro-environments (Northern temperate, Mediterranean and South Asian). Results showed significant variation ($p<0.001$) in DTF among the genotypes (G), site-years (E), and genotype by site-year interaction (G × E). However, site-years was by far the most important determinant of DTF. In temperate site-years, the DTF variation occurred mainly because of the genotypic variability. Temperature was observed as the major factor defining DTF variation in Mediterranean site-years, whereas, the interaction between temperature and photoperiod was observed to be the determinant factor in South Asia. The effect of temperature on DTF variation was better described in the form of thermal flowering time (TFT) by considering 5°C base temperature in long day environments. A complete model to analyze the interaction effect of temperature and photoperiod in DTF variation could not be confirmed due to constraints associated with the critical photoperiod calculation. To identify candidate genes or genomic regions associated with DTF to a specific environment, association studies were conducted using a mixed linear model that included both relative kinship and population structure using 255,714 markers derived from an exome capture array, and phenotypic data of the same 324 genotypes. Association studies detected three quantitative trait loci (QTLs) for DTF on chromosome 2 and one on chromosome 5. A flowering time related candidate gene *ELF4* was identified at QTL qDTF.2-1 from the Bardiya 2016 and Jessore 2016 site-years. This gene may serve as a promising target for flowering time related studies in lentil in South Asia and may assist in improving the adaptation of lentil germplasm from a long day to short day situations.

ACKNOWLEDGEMENTS

DEDICATION

To my parents, wife, siblings and all family members. Thank you for your love, support and dedicated partnership for the success in my life.

TABLE OF CONTENTS

LIST OF EQUATIONS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AGILE	Application of Genomic Innovation into Lentil Economy
AMMI	Additive Main effect Multiplicative Interaction
ANOVA	Analysis of Variance
CDC	Crop Development Centre
CPP	Cumulative Photoperiod
DAS	Days After Seeding
DTE	Days to Emergence
DTF	Days to Flower
°C	Degree Celsius
FDR	False Discovery Rate
FAO	Food and Agriculture Organization of the United Nations
GAPIT	Genome Association and Prediction Integrated Tool
GWAS	Genome-Wide Association Studies
G × E	Genotype by Site-Year Interaction
GDD	Growing Degree Days
h	Hour
ICARDA	International Center for Agriculture Research in the Dry Areas
K	Kinship Matrix
LD	Linkage Disequilibrium
MLM	Mixed Linear Model
NPBGR	National Bureau of Plant Genetic Resources
PTT	Photo-Thermal Time
PGRC	Plant Gene Resources of Canada
SK	Saskatchewan
SNP	Single Nucleotide Polymorphism
Q	Structure Matrix
QTL	Quantitative Trait Loci
TFT	Thermal Flowering Time
USDA	United States Department of Agriculture
U of S	University of Saskatchewan

CHAPTER 1. INTRODUCTION

Lentil (*Lens culinaris* Medik.) is a self-pollinated annual pulse crop first domesticated around the Fertile Crescent during Neolithic period (Arumuganathan and Earle 1991; Ferguson et al. 1998; Ferguson and Robertson 1999; Sonnante et al. 2009). Lentils are high in dietary fibre, protein, vitamin B, and minerals; low in sodium, fat and calories; and free from cholesterol (Bhatty 1988). They are also an excellent source of complex carbohydrates, vegetable protein, and micro-nutrients (Salunkhe and Kadam 1989; USDA 2016). Hence, lentil is recognized as part of the solution to combat global food and nutritional insecurity.

Currently, lentil is grown in more than 50 countries around the world (FAOSTAT 2018) representing the Mediterranean, sub-tropical savannah (mostly South Asia) and temperate environments (Tullu et al. 2011) where temperature and daylength differ considerably during the growing season (Khazaei et al. 2016). There are 58,405 accessions of genus *Lens* (cultivated and wild species) currently housed in different gene banks around the world. These collections include landraces, breeding lines, advanced cultivars, and unknown mixtures (FAO 2010). However, the majority of the lentil breeding programs are based on only a fraction of total available genetic diversity (Khazaei et al. 2016). This is primarily due to the adaptation constraints of lentil germplasm from one environment when grown in a differing environment. Unadapted germplasm typically flowers at inappropriate times. Studies revealed that the adaptation problem is mostly due to temperature, photoperiod and their interaction effects (Erskine et al. 1990a, 1994).

Therefore, for sustainable lentil breeding anywhere in the world, understanding the flowering time (in the form of days to flower; DTF) response of diverse genotypes when grown in varied environments is essential for accessing additional genetic diversity. The systematic use of genetic variability through judicious use of diverse germplasm helps in maximizing genetic gain and, over time, productivity and economic value. Studies are underway to identify the genes controlling DTF differentiation in lentil, an activity which is essential for broadening lentil diversity around the globe. The results presented on this thesis are part of a larger study at the University of Saskatchewan which was envisioned to better understand the genetics underlying adaptation in a given growing environment and to determine regions of the genome that are key to selecting new breeding material for better adaptation.

1.1 **Hypotheses**

1) Temperature, photoperiod, and their interaction define the days to flower of lentil germplasm when grown in differing field environments.

2) A few key genes define the days to flower of lentil germplasm to a specific environment, and candidate genes that are key to days to flower in diverse environments can be identified through association studies.

1.2 **Objectives**

1) To characterize variation in days to flower of diverse lentil genotypes grown in field experiments with contrasting temperature and photoperiod regimes.

2) To use association studies to tag genes or identify markers that are associated with the days to flower of lentils in specific environments.

Two studies were conducted to test the above-listed hypotheses:

Study 1 identified the effects of temperature and photoperiod on DTF variation when diverse lentil germplasm was grown in three major lentil growing macro-environments.
Study 2 identified the genomic regions and/or potential candidate genes related to DTF of the lentil germplasm while growing on those three macro environments.

These studies are presented in the manuscript format in Chapters 3 and 4 of this thesis.

CHAPTER 2. LITERATURE REVIEW

2.1 Origin and Domestication

Lentil (*Lens culinaris* Medik.) is one of the oldest food crops grown in the world, along with einkorn, emmer, barley, linseed and pea (Harlan and Harlan 1992a, 1992b). Different archaeological studies suggested that lentils were first domesticated in the Fertile Crescent, from today's Jordan northward to Turkey and southwest to the Islamic Republic of Iran around 7000 B.C., and eventually moved both east and west during Neolithic times (Ferguson et al. 1998; Ferguson and Robertson 1999; Sonnante et al. 2009). It was originally believed that during the process of domestication and cultivation, lentil was divided into two major groups/types: microsperma and macrosperma. The microsperma lentil type was generally characterized by strongly pigmented flowers with almost spherical seeds, with red and yellow cotyledons and seed diameter of 2 to 6 mm (Barulina 1930; Erskine et al. 1985; Erskine 1996). The macrosperma lentil type was characterized having lighter pigmented flowers with flattened seeds, with yellow cotyledons and seed diameter of 6 to 9 mm (Barulina 1930; Sandhu and Singh 2007). Barulina (1930) also suggested microsperma and macrosperma types of lentils have specific growing locations, i.e., microsperma types are found in South Asian regions whereas macrosperma types are commonly produced in western Asia, Europe and North & South America. However, these assumptions are not valid today as lentil producers in regions such as Canada are growing both macro and microsperma type of lentils. Ferguson et al. (1998a) also reported that the microsperma and macrosperma types of lentils were not found to be associated with geography.

Lentil is currently grown in more than 50 countries around the world representing three major macro-environments of crop production; Mediterranean, Sub-Tropical Savannah (esp. South Asia), and the Temperate where day length and temperature are significantly different during the growing season (Tullu et al. 2011; Khazaei et al. 2016; FAOSTAT 2018).

2.2 Genetics and Diversity

Lentil is a diploid (2n = 2x = 14, ~4Gb), self-pollinated, annual pulse crop belonging to the Fabaceae family (Arumuganathan and Earle 1991; USDA 2016). *Lens culinaris* is the domesticated species of lentil which is believed to have originated from *L. orientalis* (Ladizinsky 1979; Ferguson et al. 1998).

Alo et al. (2011) proposed two distinct subgroups within the cultivated species based on their analysis of the sequence alignment of introns from 22 conserved genes identified in *M. truncatula*. They are domesticated culinaris-M and culinaris-m; culinaris-M are broadly distributed to the Mediterranean, and northern and eastern Africa whereas culinaris-m is distributed to east Africa and some parts of Europe, and a mixture of both are found in the Fertile Crescent, the center of origin.

According to the Food and Agriculture Organization of United Nations (FAO 2010), 58,405 accessions of genus *Lens* are housed in gene banks around the world. These accessions include wild species (3%), landraces/old cultivars (36%), breeding lines (4%), advanced cultivars (3%), and others - the type is unknown or a mixture of two or more types (54%) from more than 70 countries (FAO 2010). Among the collected germplasm, Mediterranean germplasm has high intraspecific diversity while Asian germplasm has a narrow genetic base (Ferguson et al. 1998; Ferguson and Robertson 1999; Ferguson and Erskine 2001; Erskine et al. 2009). The International Center for Agriculture Research in the Dry Areas (ICARDA) is the principal holder of the *Lens* germplasm, having 19% of the world collection, followed by the National Bureau of Plant Genetic Resources (NBPGR), India, which holds 17% of the total collection.

2.3 Growth and Development Stages

Erskine et al. (1990) described a set of universally accepted growth and developmental stages of lentil applicable to diverse environments and cultivars. The details of their description are summarized below:

Vegetative stages

VE (seedling emergence)	Visible nodes
V1	First simple leaf has unfolded at the first node
V2	Second simple leaf has unfolded at the second node
V3	First bifoliate leaf has unfolded at the third node
V4	Second bifoliate leaf has unfolded at the fourth node
V5	First multi-foliate leaf has unfolded at the fifth node

| Vn | n^{th} multi-foliate leaf has unfolded at the nth node |

Reproductive stages

R1 – Early bloom	One open flower at any node
R2 – Full bloom	Flower open or has opened on nodes 10-13 of the basal primary branch
R3 – Early pod	Pods on nodes 10-13 of the basal primary branch visible
R4 – Flat pod	Pods on nodes 10-13 have reached full length and are mostly flat
R5 – Early seed	Seeds in any single pod on nodes 10-13 fill the pod cavity
R6 – Full seed	Seeds on nodes 10-13 fill the pod cavities
R7 – Physiological maturity	Leaves start yellowing, and 50% of the pods have turned yellow
R8 – Full maturity	90% of pods on the plant are golden-brown

Among these stages, R1 and R2 are the most important to study adaptation of genotypes.

Flowering of lentil is indeterminate, i.e., flowering starts acropetally from lower to higher nodes and from axillary buds on the main stem and then to branches. Before emergence (VE stage) there is also a pre-emergence stage, i.e., the period between seeding to first shoot appearance. Roberts et al. (1986) defined pre-emergence stage as the photoperiod insensitive stage.

2.4 Temperature and Photoperiod Effects on Days to Flower

The expansion of lentil from its origin has been governed by the selection of traits important for adaptation to new environments; including both abiotic and biotic stresses (Sarker et al. 1999; Erskine et al. 2009). DTF is an important phenological event to determine the adaptation of lentils to new environment (Erskine et al. 2009). Previous studies showed environmental factors (esp. temperature and photoperiod) have a major role to play in DTF which ultimately determines the adaptation of germplasm from one environment to another (Summerfield et al. 1985; Roberts et al. 1986; Summerfield and Roberts 1988; Erskine et al. 1994). Lentil is a quantitative long day plant and wide genetic variation in DTF due to photoperiod and temperature has been reported in the global germplasm repository at ICARDA (Erskine et al. 1990a, 2009).

For temperature effects, thermal flowering time (TFT) is a widely-accepted approach for the analysis of the effect of temperature. In this method, daily mean temperature is converted to growing degree days (GDD) by subtracting a crop specific base temperature (below which physiological activity stops) from the daily average temperature (Yin et al. 1997; Baker and Reddy 2001; Iannucci et al. 2008). Several crop specific resources are available with information about base temperature. The sum of the GDD from seeding to flowering is TFT. Most recently, the TFT approach was used in faba bean *(Vicia faba)* by Catt and Paull (2017) by considering 0°C as the base temperature.

To evaluate the effect of photoperiod on DTF in lentil, Summerfield et al. (1985), working under controlled growth conditions, proposed a linear model by considering the rate of progress towards flowering (reciprocal of days to flower). This was also tested in pea *(Pisum sativum)*, faba bean and chickpea (*Cicer arietinum)* (Summerfield and Roberts 1988). The proposed model was:

$$1/f = a + bT + cP \text{ ……………………………………………...……… [2.1]}$$

where, f is number of DTF from seeding, and 1/f is defined as rate of progress towards flowering; T and P are the mean temperatures (°C) and photoperiod (h) from seeding to flowering; a is the intercept, and b & c are temperature and photoperiod coefficients.

From the perspective of a simple understanding of physiology, plant growth and development are non-linear functions of the environmental factors, and the reciprocal of DTF provides a linear relationship. The above mentioned studies were considered to be a breakthrough in the understanding of the effects of these significant environmental factors on DTF and were found applicable to many crops. However, such methods of calculation are now considered rudimentary, especially for field grown material. The primary concern is that the calculations and values used are based on mean temperature and photoperiod during flowering. Plants continuously react with the several biotic and abiotic factors for their growth and development, which means the same amount (mean) of temperature and photoperiod may not work for field growing situations. Hence, these two factors might better be taken into consideration by accumulating the values after reaching critical, or ceiling, or base values as plants approach the full bloom stage. Furthermore, for better understanding of the photoperiod effect, equation [2.1]

needs information about photoperiod inductive and non-inductive phases as outlined by Roberts et al. (1986), which could not be achieved properly under field conditions.

Soon after the development of equation [2.1], Roberts et al. (1986) revised it to observe the photoperiod effect by using a fixed temperature value. Roberts et al. (1986) defined four different photoperiod sensitive and insensitive stages to elucidate the exact effect of photoperiod on flowering by calculating 'critical photoperiod' (below which the physiological activities stop, which is, $CP = -a/c$) and 'photoperiod sum' (light hours above the critical photoperiod, which is, $P_{sum} = 1/c$). For lentil, being a long day plant, critical photoperiod is the amount of photoperiod below which flowering either will not occur or will be delayed. The later calculation-based model was considered more convincing than just the mean temperature and photoperiod. However, in a field situation, temperature is not constant, and the case would be even more challenging when field trials are conducted at diverse locations. In the discussion section of the same study, Roberts et al. (1986) recommended an alternative method where those photoperiod (in)sensitive stages are not separately identified by deliberately defining 'nominal base photoperiod' and 'nominal light sum' instead of 'critical photoperiod' and 'photoperiod sum'. The nominal base photoperiod' and 'nominal light sum' were calculated using the same constants derived from equation [2.1]. Nominal base photoperiod and nominal light sum were different from critical photoperiod and cumulative photoperiod as the temperature was also included in the latter calculation. Moreover, nominal base photoperiods are generally shorter than daily experienced photoperiods and often have negative values. In addition, the nominal light sum calculated with this approach might not always give a real value in field situations (Roberts et al. 1986). Roberts et al. (1986) again made another change by calculating the daily contribution to the nominal light sum based on that value of the nominal based photoperiod. The general formula used to calculate the contribution each day to the nominal light sum was:

$$l_d = P_d - (a+bT_d)/c \ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \ [2.2]$$

where, l_d is daily contribution to the nominal light sum in a particular day in a period from seeding to flowering; P_d is the experienced photoperiod (sunrise to sunset) that particular day; T_d is the mean temperature; and a, b, and c are the intercept and slope derived from the multiple regression equation [2.1] which were calculated separately for each accession. This model uses temperature data while evaluating the effect of photoperiod which means the model parametric

value might fluctuate with variation of the temperature in the diverse environmental situations. Thus, it could be better to determine the effect of photoperiod alone by using the critical photoperiod derived from the equation [2.1]. One potential approach could be accumulating the daily photoperiod, after it reaches the critical photoperiod for each genotype, till DTF.

All the procedures mentioned above are separate analytical approaches used to determine the effect of photoperiod and temperature; however, it is not always appropriate to determine the effect of temperature and photoperiod separately, since doing so may overlook the potential interactive effect of both factors. Pre-defined methods exist for determining the effects of both temperature and photoperiod; the most common and widely accepted is the Photo-Thermal Time (PTT) model used by Masle et al. (1989). They found that this PTT model predicts the effect of temperature and photoperiod significantly better than a thermal time model. Recently, this model was used to predict days to flower in different plants, notably in *Arabidopsis thaliana*. Brachi et al. (2010) made a slight modification to a previously developed PTT model for *A. thaliana* which was also used by Chew et al. (2012) and Springate and Kover (2014) for the same species. The modified model is described by the equation below.

$$\text{PTT} = \sum_{j}^{ft} \lambda_j (\mu_j - \mu_b) \ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \ [2.3]$$

PTT is the photothermal time and expressed in photothermal units (PTU; in °C.hours); ft is flowering date; j spans from germination date to the flowering date, counting only the days with a mean temperature above a pre-defined base temperature; λ_j is daily photoperiod from sunrise to sunset; μ_j is daily mean temperature; and μ_b is base temperature.

Chew et al. (2012) and Springate and Kover (2014) used this model to study genetics of flowering time plasticity in Arabidopsis by considering days to flower as a linear function of temperature when the crop is grown in specific environments having a constant photoperiod throughout the growing season.

2.5 Flowering Time Genetics and Association Studies

Flowering time is a complex trait in plant growth and development which is generally governed by multiple genes (Yin et al. 1997; Shrestha et al. 2005; Weller and Ortega 2015). Molecular

pathways responsible for flowering time in response to various environmental factors have been studied in many crops, most intensively in the model species Arabidopsis and in rice (*Oryza sativa*). Over 100 genes contributing to flowering time have already been identified in these two crops (Weller and Ortega 2015).

In legumes, Weller and Ortega (2015) studied the genetic analysis of flowering time in pea and soybean (*Glycine max*) considering these as a representative of gallegoid and phaseoloid clades respectively within the larger group of papilionoid legumes. These studies were conducted from the perspective of photoperiod and vernalization and authors have made a broader understanding of the genes and/or markers associated with this. From this study, Weller and Ortega (2015) identified more than 20 loci in pea and at least 10 loci in soybean. Studies of the *Flowering Time (FT)* gene family have also been made in pea, Medicago (*M. truncatula*), soybean and lotus (*L. japonicus*) by different groups at different times (Weller et al. 2009; Kong et al. 2010; Laurie et al. 2011; Jung et al. 2012). More than five *FT-like* genes were found in these crop species; among them, *FTa, FTb*, and *FTc* are unique to the legumes (Weller et al. 2012; Weller and Ortega 2015). *FTa* and *FTb* genes are believed important for vernalization and photoperiod responses in pea and Medicago. All three genes could promote flowering at some level, but FTb matches the characteristics of the 'florigen' hormone responsible for controlling and/or triggering flowering. *FT2a* (an *FTa* type gene) and *FT5a* (sometimes also called *FTc)* gene appeared to be important genes governing photoperiod response in soybean (Kong et al. 2010). It could be summarized broadly that the FT genes rely on the interactions between light perception and the circadian clock to induce flowering. However, a clear picture of how these photoreceptor and clock inputs integrate for photoperiod specific regulation at the gene level has yet to be identified. In the case of lentil, one locus involved in controlling the flowering time was characterized when Weller and Ortega (2015) made their pea and soybean study. After research in pea, they also suggested *HIGH RESPONSE TO PHOTOPERIOD (HR)* and functionally associated clock genes, including *ELF4* and *LUX,* might also be candidates behind the genetic variation in the broader range of legumes. Lentil genotypes from the different countries of origin flower at different times when grown at the same time. A noticeable example can be seen in lentils from South Asia as they tend to flower very early compared to lentils originating from other environments. Sarker et al. (1999) suggested that the early flowering is determined by a single recessive gene *sn* in addition to some polygenes. A study conducted by Weller et al.

(2012) found that the *EARLY FLOWERING 3 (ELF3)* gene has a potential role in the reduced photoperiod response in lentil cultivars developed for short-season environments.

Until now, genetic studies for flowering time in lentil have been limited to bi-parental populations (reviewed in Table 1.2 of Rajandran 2016). Rajandran's (2016) study was an expansion of the understanding of the genetic control in lentil flowering after the *Sn* locus discovery. He reported an Arabidopsis *ELF3* orthologue as a candidate for the *Sn* locus, and Medicago orthologues *FTa1* and *FTa2* as well as Arabidopsis paralogue *PSEUDO-RESPONSE REGULATOR (PRR59c*) as other candidates for controlling early-flowering. All his studies were conducted on a bi-parental population under short days in controlled conditions. However, it remains to be determined if these will be applicable under long day field conditions and in different population sets.

Association mapping is a set of statistical methods developed to study the link between phenotypic performance and genotypes in a group of unrelated individuals. Association studies were designed to address shortcomings of linkage mapping, i.e. the methods take care of whole populations studies rather than just the segregation within a specific bi-parental family. Genome-Wide Association Studies (GWAS) are tools to assess the association between genotyped markers and phenotyped traits of interest from a large population of diverse material. It provides an opportunity to analyze the genetic architecture of complex traits (Li et al. 2012; Scherer 2017). GWAS started with the study of animal diseases, but it has now become a popular technique in plant research when working with large amounts of variation beyond just a bi-parental population (Korte and Farlow 2013). Several studies are reported on different traits using GWAS in different crops, however, GWAS in lentil is still in the novice stage.

Different software and online sources are available to conduct GWAS with built-in statistical analysis packages. TASSEL (Trait Analysis by aSSociation, Evolution and Linkage - http://www.maizegenetics.net/tassel), PLINK (Whole genome association analysis toolset - http://zzz.bwh.harvard.edu/plink/) and GAPIT (Genome Association and Prediction Integrated Tool - http://www.zzlab.net/GAPIT/index.html) are the most widely used software for GWAS at this time. Among these, GAPIT is considered better for large marker datasets. It can handle more than 10,000 individuals and 1 million SNPs with minimal computational time (Lipka et al. 2012). GAPIT also uses the state-of-the-art methods developed for statistical genetics, such as

the compressed mixed linear model (CMLM) and CMLM-based genomic prediction and selection (Lipka et al. 2012).

The use of genotypic and phenotypic covariates for GWAS has also become a standard method to increase the efficiency and to improve the statistical power of association studies by reducing residual variance (Aschard et al. 2015, 2017). A population structure (Q) matrix and a kinship (K) matrix are often used as the covariates with the aim of reducing the false discovery rate (FDR) (Yu et al. 2006; Tang et al. 2016). The inclusion of these confounding factors is mainly to take into account the bias of SNP effect estimates (Aschard et al. 2015, 2017). In a few cases, phenotypic traits related to the primary trait of interest have also been used as phenotypic covariates while running GWAS (e.g. Crowell et al. 2016). The use of the different factors as covariates helps in accounting for the actual risk factors by taking care of the residual variance which, therefore, increases the statistical power. Furthermore, the use of covariates also helps to reduce the effect of the factor in association studies; for example, for yield-related association studies, if a soil fertility factor is added as a covariate, the influence of the fertility factor in analysis is removed.

GWAS results in conjunction with a sequenced and annotated genome are useful for predicting potential candidate genes as they provide information about the position of the significant marker relative to genes on the genome. Candidate genes can be identified either from a scan of the genomic region around the marker on a genome browser with annotations or through shared synteny with other legumes where genes are already known. In the case of lentil, potential candidate genes can be identified using the genome browser available through the KnowPulse web portal (http://knowpulse.usask.ca/portal/jbrowse/Lentil).

PROLOGUE TO CHAPTER 3

The variation in DTF of 324 diverse set of genotypes tested in 10 different field locations representing three major lentil growing macro-environments will be described in chapter 3. Furthermore, this chapter will focus on the procedural requirements to evaluate the temperature and photoperiod effects on DTF variation in contrasting site-years and includes discussions related to the models to see the effects of these two factors.

CHAPTER 3. EFFECT OF TEMPERATURE AND PHOTOPERIOD ON THE DAYS TO FLOWER OF DIVERSE LENTIL (*Lens culinaris* MEDIK.) GERMPLASM IN MULTIPLE ENVIRONMENTS

3.1 Introduction

Lentil (*Lens culinaris* Medik.) is a self-pollinated, annual pulse crop grown in over 50 countries, mostly in mediterranean, sub-tropical and temperate environments around the globe (Tullu et al. 2011, FAOSTAT 2018). Although lentil is grown in different macro-environments, lentil breeding programs in the respective countries or environments are generally based on only a fraction of total available genetic diversity (Khazaei et al. 2016). This is mainly due to adaptation constraints of lentil germplasm from one environment when grown in a contrasting environment. Unadapted germplasm typically flowers at inappropriate times resulting in yield reduction. Days to flower (DTF) is one of the primary phenological stages determining crop acclimatization to diverse environments because it affects overall crop production (Marx 1979; Daba et al. 2016a). Hence, understanding the flowering time response of diverse genotypes when growing in varied environments is essential for utilizing additional genetic diversity in lentil breeding and ultimately addressing the global yield demand.

Environmental factors such as photoperiod, solar radiation, ambient temperature and vernalization affect DTF, as do agronomic factors such as soil fertility, soil moisture and seeding time (Hadley et al. 1984b; Summerfield et al. 1985; Roberts et al. 1985; Cockram et al. 2007; Springate and Kover 2014). However, temperature and photoperiod are the two most significant factors to be considered (Roberts et al. 1993; Coupland 1995), as changes to either may alter the timing and duration of flowering, which ultimately determines the ability of a species to adapt to a new region (Summerfield et al. 1985; Roberts et al. 1986; Summerfield and Roberts 1988; Erskine et al. 1994).

For a better understanding of temperature effects in a field setting, calculation of thermal flowering time (TFT) is a widely-accepted approach. In this method, daily mean temperature is converted into growing degree days (GDD) by subtracting a crop specific base temperature (below which the physiological activity is assumed to stop) from the daily average temperature (Yin et al. 1997; Baker and Reddy 2001; Iannucci et al. 2008). The sum of the GDD from the

day of seeding to DTF is the TFT. More recently, the TFT approach was used in faba bean by Catt and Paull (2017) where they considered 0 °C as the base temperature for that crop. A base temperature that helps to explain the effect of temperature on DTF in lentil has not yet been determined.

To evaluate the photoperiod effect on DTF, there is no defined approach as with temperature. Summerfield et al. (1985) proposed a linear model to estimate the effect of photoperiod and temperature together by considering the rate of progress towards flowering (reciprocal of DTF) in lentil based studies under controlled conditions. This technique was also tested in pea *(Pisum sativum)*, faba bean *(Vicia faba)* and chickpea (*Cicer arietinum)* chickpea later by the same research group (Summerfield and Roberts 1988). Besides pulses, the same model has also been applied in other legumes (Iannucci et al. 2008) and in rice *(Oryza sativa)* (Summerfield et al. 1992). However, a concern about that model is the use of mean temperature and photoperiod at flowering not the accumulation of either factors until flowering. The model requires that temperature be constant to evaluate the effect of photoperiod and vice-versa. Plants continuously react to multiple biotic and abiotic factors to regulate their growth and development, and temperature and photoperiod are not constant throughout, which suggests average temperature and photoperiod may not work under field situations. Roberts et al. (1986) suggested a model that accumulates the daily contribution of photoperiod above a critical photoperiod or nominal base photoperiod and also has temperature as an integral component. The model redescribed by Roberts et al. (1986) uses the temperature data while evaluating the effect of photoperiod, which means the amount of photoperiod estimation fluctuates with the variation of the temperature in diverse environmental situations. This indicates the need for an updated model to see the effect of photoperiod alone using the derived critical photoperiod. For this, one potential approach could be to accumulate the daily photoperiod after it reaches the critical photoperiod, until flowering. A validation of the methods to calculate the critical photoperiod is necessary to see the applicability of these models in a field situation.

In addition to the sole effect of photoperiod or temperature accumulation, there may be an interaction effect of temperature and photoperiod in determining DTF in differing environments. Different models have been proposed for studying the interactive effect of these two factors. The most widely accepted is the Photo-Thermal Time (PTT) model used by Masle et al. (1989) in

13

Arabidopsis. They found the PTT model predicts the effect of temperature and photoperiod significantly better than a thermal time model alone.

The hypothesis of this study was that temperature, photoperiod, and their interaction define the days to flower of lentil germplasm when grown in differing field environments. To test this hypothesis, this study examined at the effect of temperature and photoperiod alone, and then the combination of both temperature and photoperiod. For this, a large number of diverse accessions were grown in multiple locations with differing photoperiod and temperature profiles throughout the growing season. The objective of this study was to characterize variation in days to flower of diverse lentil genotypes grown in field experiments with contrasting temperature and photoperiod regimes. In addition, this study aimed to identify suitable models for temperature and photoperiod as well as their interaction causing the variation in DTF in diverse environments.

3.2 Materials and Methods

3.2.1   Field Experiments and Data Collection

Three hundred and twenty-four different genotypes of lentil, obtained from the gene banks of the International Center for Agricultural Research in the Dry Areas (ICARDA), United States Department of Agriculture (USDA), Plant Gene Resources of Canada (PGRC) as well as cultivars developed at the Crop Development Centre (CDC) of the University of Saskatchewan (U of S) (list of genotypes are available in Appendix 1) were evaluated in separate field trials for two seasons in each of the ten geographic locations representing three lentil growing macro-environments (Table 3.1). The experimental design followed was a randomized complete block, replicated three times at each location.

Table 3. 1. Details of the field trial locations, their macro-environments, growing season and year of the experiment, and date of seeding along with the number of seeds used with individual plot size.

Field Trial Location	Latitude, Longitude	Country	Macro-environment	Season	Year	Seeding Date	Number of seeds sown and plot size
Bhopal	23.1103, 76.8805	India	South Asia	Fall/Winter	2016/17	2016-12-04	25 / single row (1 meter long)
Jessore	23.1911, 89.1913	Bangladesh	South Asia	Fall/Winter	2016/17	2016-11-15	25 / single row (1 meter long)
Bardiya	28.2521, 81.5015	Nepal	South Asia	Fall/Winter	2016/17	2016-11-14	25 / two short rows (each 1 meter long)
Metaponto	40.3901, 16.7803	Italy	Mediterranean	Fall/Winter	2016/17	2016-11-29	25 / single row (1 meter long)
Cordoba	37.9001, -4.8017	Spain	Mediterranean	Fall/Winter	2016/17	2016-12-13	25 / single row (1 meter long)
Rabat	33.6205, -6.7201	Morocco	Mediterranean	Fall/Winter	2016/17	2016-11-21	25 / single row (1 meter long)
Sutherland	52.1677, -106.5054	Canada	Temperate	Summer	2016	2016-04-27	60 / 1 sq. meter microplots
Rosthern	52.6892, -106.2945	Canada	Temperate	Summer	2016	2016-05-06	60 / 1 sq. meter microplots
Sutherland	52.16832, -106.5108	Canada	Temperate	Summer	2017	2017-05-04	60 / 1 sq. meter microplots
Rosthern	52.6915, -106.2897	Canada	Temperate	Summer	2017	2017-05-19	60 / 1 sq. meter microplots

Days to emergence (DTE) and days to flower (DTF) were recorded on a plot basis when 10% of the plants had emerged and had one open flower, respectively. Temperature data were collected using iButtons (https://www.maximintegrated.com/en/products/digital/ibutton.html) in the SK field trials. In each SK experiments, three iButtons were installed, one in each replication, by placing the devices in a wire-mesh bag and hanging them on stakes 30 cm above ground level. All iButtons recorded air temperature at six hour intervals which were then converted into daily maximum and minimum. Temperature data were gathered from on-farm meteorological stations in all other locations. Photoperiod (number of hours between sunrise and sunset) data were extracted using the daylength function in the 'insol' package in R (Corripio 2015) by providing latitude, longitude, and specific day and time zones.

3.2.2   Data Analysis

All statistical analyses were performed in R studio version 1.1.453 (RStudio Team 2016). Normality and homogeneity of all data were visually assessed by graphing the residual distribution using a scatter plot of residuals and a Q-Q plot of residuals for each site-year prior to further analyses. Data visualization was done using the 'ggplot2' (Wickham 2016) package in R. Genotypes that did not flower or were missing in one or more replications at any location were considered as missing data in the analysis (at a particular plot in particular location) and are indicated in Appendix 1.

An Additive Main effect Multiplicative Interaction (AMMI) model (Gauch 1988) was used to investigate the contribution of either genotype or environment (site-year), or genotype by environment interaction (G × E) to overall variation in DTF. Stability analysis was also conducted to assess the consistency of genotypes for DTF in different environments by calculating the AMMI Stability Value (ASV) using the index.AMMI() function in the 'agricolae' package (Mendiburu 2017) in R.

The 'lmerTest' package (Kuznetsova et al. 2017) was used to conduct an analysis of variance (ANOVA) using a mixed model. While analyzing the data, location and year were merged by making a single term: site-year. Random effects of replication nested within site-year were included in the model. Multiple comparisons of means were performed using the 'emmeans' package (Russell et al. 2018) in R.

16

The effect of temperature was analyzed by using the mean temperature from seeding to DTF for each genotype for each site-year and after transforming the daily mean temperature to thermal flowering time (TFT). TFT was calculated for each plot using following formula:

$$\text{TFT} = \sum_{i}^{ft} \text{GDD}$$ …………………………………………………………………… [3.1]

Where, ft is number of days from seeding to flower; i spans from seeding to ft in increments of 1 day; GDD is growing degree day, which is daily mean temperature minus the base temperature. Two different base temperature values, 0°C and 5°C, were used to calculate TFT and identify the best TFT results. 0°C as the base temperature was used in similar study in fababean by Catt and Paull (2017) and 5°C is the commonly used base temperature for many cool season crops, this is also the recommended minimum soil temperature for seeding lentil in Saskatchewan (Saskatchewan Pulse Growers 2000).

Two different methods were tested to assess the effect of photoperiod on flowering. First was simple accumulation of daily photoperiod hours from emergence to flowering. The second was accumulation of photoperiod only after reaching the critical photoperiod. Critical photoperiod (in hours) for each genotype was calculated using the formula,

$$\text{Critical Photoperiod} = -a/c$$. …………………………………………………….. [3.2]

The values a and c are the intercept and photoperiod coefficient derived from the regression model, $1/f = a + b\overline{T} + cP$ from Summerfield et al. (1985), where, f is the number of DTF from emergence, and 1/f is rate of progress towards flowering. $\overline{T}$ is mean temperature (°C) from emergence to DTF, P is photoperiod (hours from sunrise to sunset) on the day of first flower for the genotype at individual plots in a given location.

After identifying critical photoperiod, photoperiod effect on DTF was analyzed in the form of cumulative photoperiod (CPP), which is the cumulative photoperiod hours after the individual genotypes crossed their critical photoperiod. CPP (in hours) was calculated for each genotype in each plot in each site-year using following formula:

$$\text{CPP} = \sum_{j}^{ft} P\wedge$$ …………………………………………………………………... [3.3]

Where, ft is the flowering day; j spans from DTE to the particular day when genotypes reached to the critical photoperiod; P^ is the daily photoperiod of an individual genotype from j^{th} day. In the case when the critical photoperiod was greater than the daily photoperiod or if the daily photoperiod did not reach the critical photoperiod before the genotype flowered, the CPP was considered to be zero.

The interactive effect of temperature and photoperiod on DTF were analyzed with a modified (considering DTE to DTF rather than from seeding to DTF) Photo-Thermal Time (PTT) model from Brachi et al. (2010). PTT for individual genotypes in every single plot was calculated using the formula,

$$PTT = \sum_{j}^{ft} \lambda_j(\mu_j - \mu_b) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots [3.4]$$

Where, PTT is the photothermal time and expressed in photothermal units (PTU); ft the flowering date; j spans from germination date to the flowering date, counting only the days with a mean temperature above a pre-defined base temperature; λ_j is the daily photoperiod from sunrise to sunset; μ_j is the daily mean temperature; and μ_b is the base temperature with both 0°C and 5°C.

All analysis scripts are available in Appendix 5.

3.3 Results

3.3.1   Variation in days to flower among genotypes and across site-years

Temperatures and day lengths were distinctly different among the 10 site years (Fig 3.1.B). The daily average temperature trended towards longer numbers of days over the growing season in the South Asian site-years, i.e., in Bhopal 2016, Jessore 2016 and Bardiya 2016. Within the South Asian site-years, the daily average temperature during flowering was highest in Bhopal 2016, followed by Jessore 2016 and then Bardiya 2016. The daily average temperature was lowest in all Mediterranean site-years, i.e., Metaponto 2016, Cordoba 2016 and Rabat 2016. The daily average temperature was in the same range at all temperate site-years, i.e., in Sutherland and Rosthern in both 2016 and 2017. The day-length at flowering reached at least 16.17 h in the

temperate site-years, whereas, the maximum daylength was 14.07 h in the Mediterranean and 12.67 h in South Asian site-years (Fig 3.1.B).

Considerable variation was observed for DTF among the 324 lentil genotypes across all site-years (Fig 3.1.A). DTF was earliest in the temperate site-years, whereas, flowering was delayed the most in Mediterranean site-years. The earliest DTF was noted on 33 days after seeding (DAS) in Rosthern 2017, SK, whereas, the latest DTF was 154 DAS in Metaponto 2016. It should be noted however, that 143 genotypes in Bhopal 2016 and 66 genotypes in Jessore 2016 did not flower in any of the three replications before dying from excessive heat. Likewise, 29 genotypes in Bardiya 2016 and ten genotypes in Cordoba 2016 did not flower at least in one replication (Refer to Appendix 1 for details).
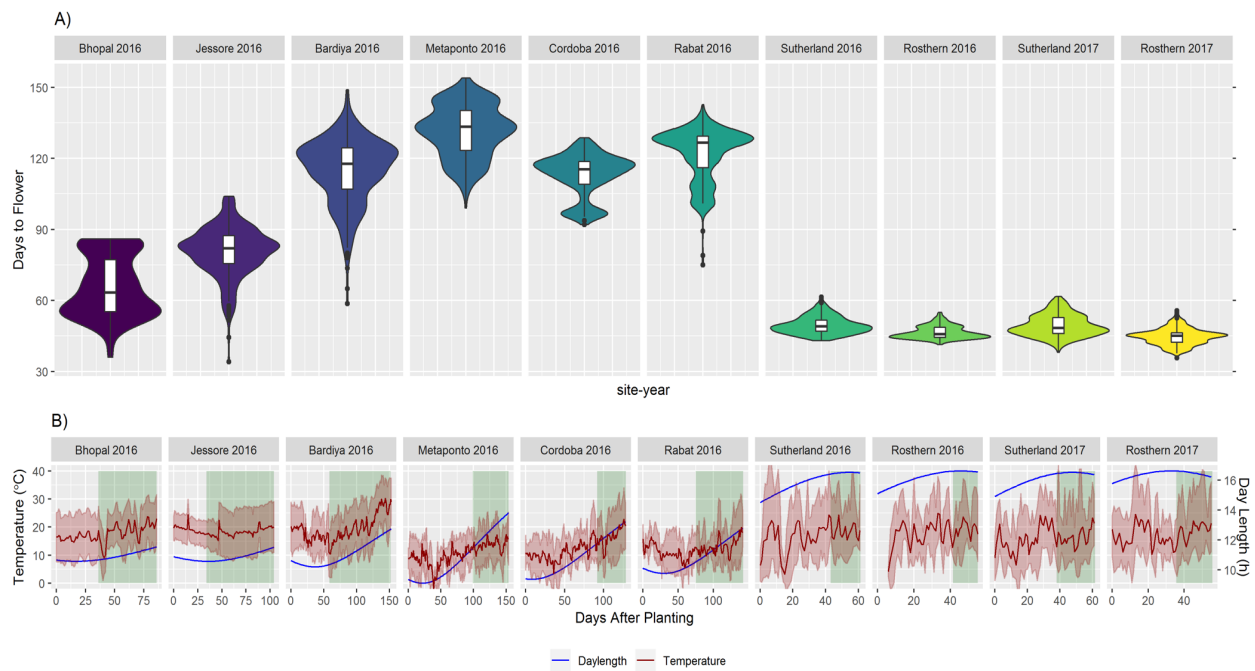
Figure 3. 1. A) Distribution of average days to flower from days after planting of 324 genotypes in ten different site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017). The width of the plots indicates the density of the distributions. The whiskers on the boxes represent 1.5 times the quantile of the data. Individuals falling outside the range of the whiskers are represented as dots. B) Variation in the daily average temperature (°C) (Red line represents the daily average temperature and pink shades represents the range of the daily temperature) and day length (h) (blue line) from seeding to flowering in the same ten site-years. Light green shades in each plot represents the flowering time window in respective site-years.

AMMI analysis showed DTF was significantly ($p<0.001$) affected by site-year (E), genotype (G), and genotype by site-year interaction (G × E) (Table 3.2). Further analysis indicated genotype governed only 3.3% of the variation and only 2.5% variation was due to G × E. The majority of the DTF variation, i.e., 93.6% of the total sum of squares, was justified by site-years alone. AMMI analysis also showed the first three interaction principal components (IPCA1, IPCA2 and IPCA3) accounted for 88.8% of the G × E interaction (Table 3.2). The AMMI biplot (Fig 3.3) indicated different macro-environments have a different contribution to the G × E interaction; however, site-years within each macro-environment contributed similarly to the variation.

Table 3. 2. Additive Main effect Multiplicative Interaction (AMMI) analysis of variance (ANOVA) for days to flower of 324 genotypes evaluated across ten site-years.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean of Squares	Explained Sum Squares (%)
genotype (G)	323	373313	1156***	3.3
site-year (E)	9	10583068	1175896***	93.6
genotype × site-year (G × E)	2697	287797	107***	2.5
IPCA$_1$	331	149902.58	452.877***	58.2
IPCA$_2$	329	49596.48	150.74***	19.3
IPCA$_3$	327	29149.62	89.14***	11.3
Error	5769	62263	11	
Total	8798	11306441	1177170	

*** indicates significance at $p < 0.001$
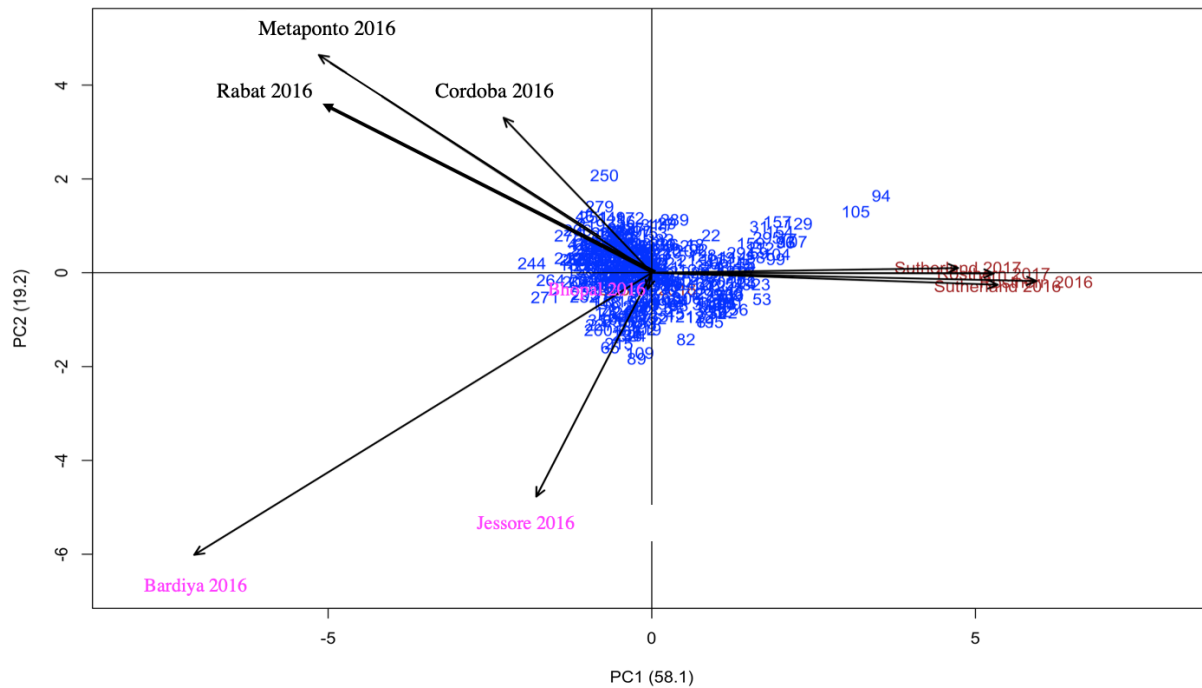
Figure 3. 2. AMMI Biplot for days to flower of 324 genotypes evaluated across ten site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017). Different colored texts along with the vectors indicate the site-years (Bhopal 2016 is hidden at the genotype cluster), similar text color represents site-years within the individual macroenvironments, and blue colored numbers represent the 324 genotypes. The distance from the origin (center) indicates the extent of interaction with that particular site-year with genotypes.

In addition, genotypes were ranked based on their consistency in DTF across different site-years using AMMI stability value (ASV). Results showed ILL 5888, ILL 7663, LIRL-22-46, ILL 7716, ILL 4605, PI 244046, PI 251248 LSP, ILL 6002, and CN 105791 consistently in the top ten early flowering genotypes, whereas, PI 163589, PI 298122 LSP, CDC Impower, PI 308614 LSP, PI 300250 LSP, PI 370481 LSP, Indianhead, PI 458503 LSP, ILL 4671 were consistently the latest flowering genotypes, across all site-years (Appendix 2).

3.3.2   Temperature effects on days to flower across site-years

By computing thermal flowering time (TFT), it should be possible to better account for unequal daily contributions to temperature accumulation for DTF across the very different site-years. Similar to simple DTF results, considerable variation for TFT requirement was observed with different site-years when considering either 0°C or 5°C base temperatures. The TFT

requirements for Mediterranean site-years shifted towards those for the temperate site-years, although they were still greater when considering 0°C base temperature (Fig 3.3.A). The variation for the TFT requirements across site-years was reduced more when 5°C was taken as the base temperature (Fig 3.3.B). Analysis of variance showed that TFT requirement (after base at 5°C) was still significantly different ($p<0.001$) among macro-environments, as well as site-years within single macro-environments. There was greater overlap among site-years, and unlike with DTF, the TFT requirement was higher in Bardiya 2016 and Jessore 2016 than in the Mediterranean site-years (Fig 3.3.B). All South Asian site-years might also have been more similar to each other if the complete set of genotypes in Bhopal 2016 and Jessore 2016 had flowered in at least one replication.

The highest TFT requirement (from the base at 5°C) in temperate site-years was 774.8 degree-days in Sutherland 2016 and the highest TFT requirement in the Mediterranean site-years was noted in Rabat 2016 which was 1014.5 degree-days, whereas, the highest TFT requirement in South Asian site-years was 2133.2 degree-days in Bardiya 2016 (in South Asia) (Fig 3.3B).
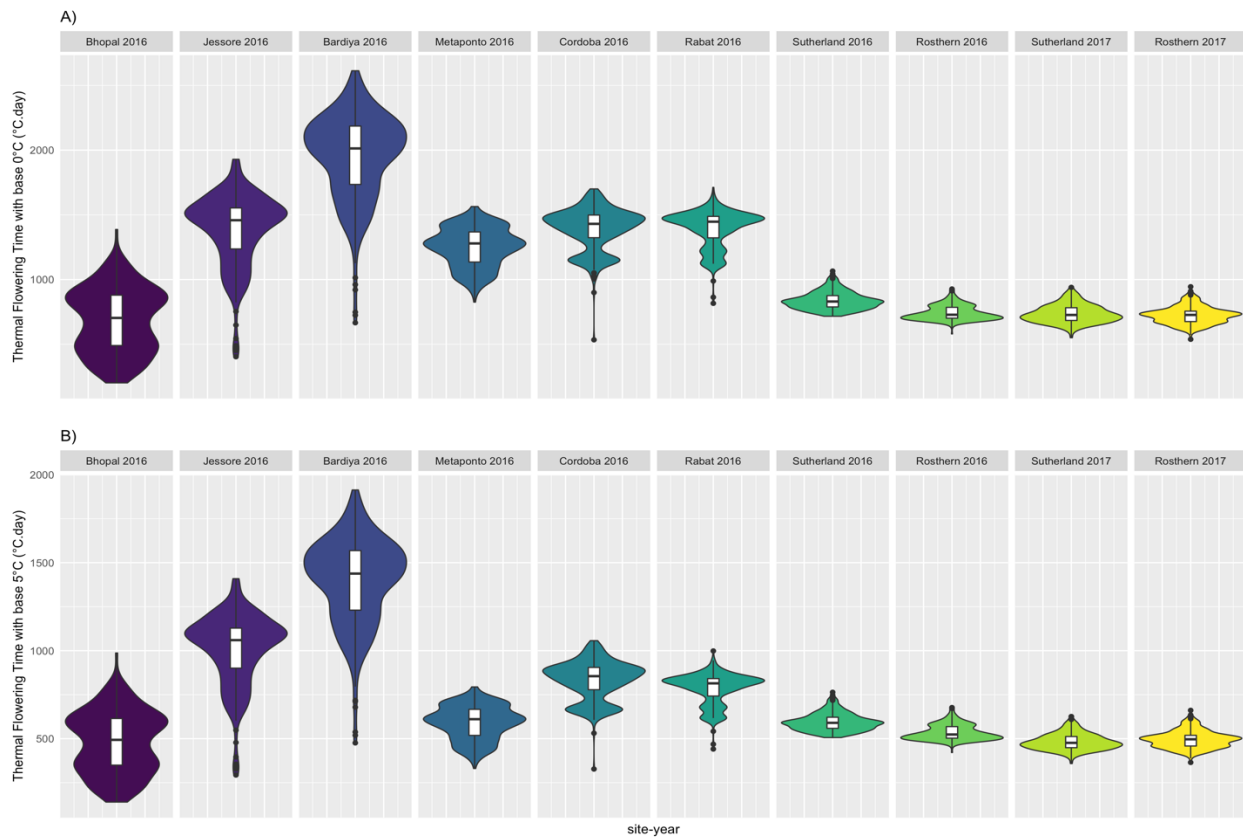
Figure 3. 3. Variation in average thermal flowering time (based on accumulated growing degree days from seeding to flowering) of 324 genotypes grown in the field at ten different site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017). The width of the plots shows the density of the distributions. The whiskers on the boxes represent 1.5 times the quantile of the data. Individuals falling outside the range of the whiskers are represented as dots. A) TFT calculated using 0°C as the base temperature and B) TFT calculated using 5°C as the base temperature.

The AMMI biplot (Fig 3.4) derived using TFT at 5°C as the base temperature indicated temperate and Mediterranean site-years are closed to each other compared to those from South Asia. South Asian site-years dispersed all around two principle components, however, Bardiya 2016 appeared near the Mediterranean site-years.
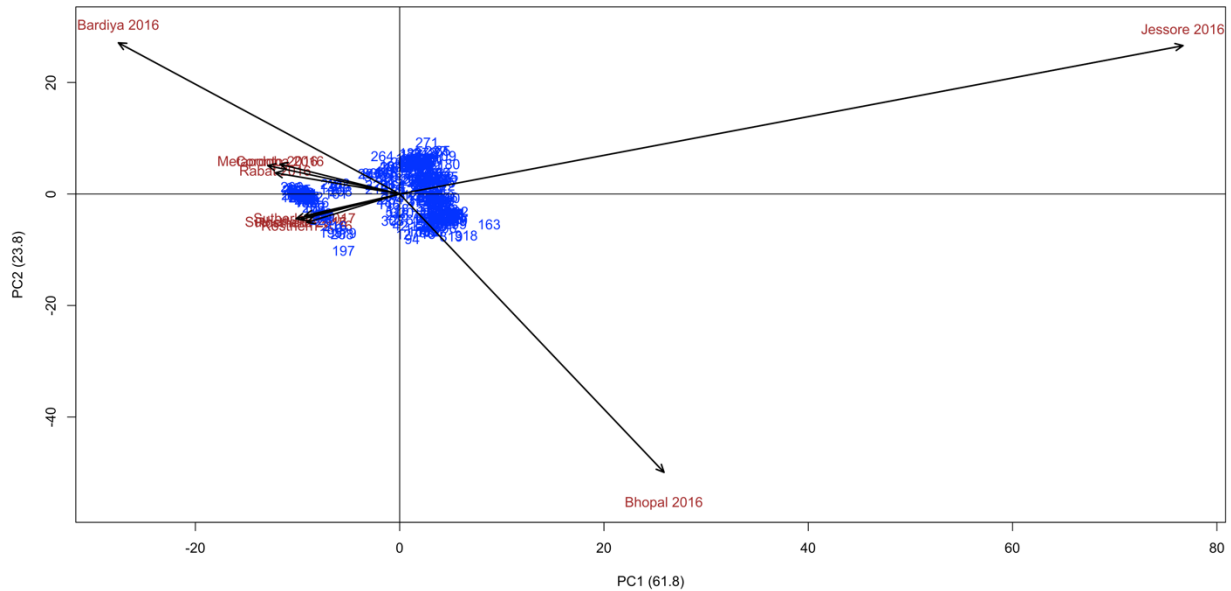
Figure 3. 4. AMMI Biplot for thermal flowering time (TFT) of 324 genotypes evaluated across ten site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017). The red colored texts along with the vectors indicate the site-years (three Mediterranean site-years are clustered together, likewise the temperate site-years), and the blue colored numbers represent the 324 genotypes. The distance from the origin (center) indicates the extent of interaction with that particular site-year with genotypes.

3.3.3 Photoperiod effects on days to flower across site-years

The raw cumulative photoperiod (summation of daily photoperiod from emergence to flowering) among the site-years presented in Fig 3.5 showed distributions are similar to those of simply DTF as shown in fig 3.3.A. There may be a critical photoperiod that is required in a model to determine the effect of photoperiod on DTF.
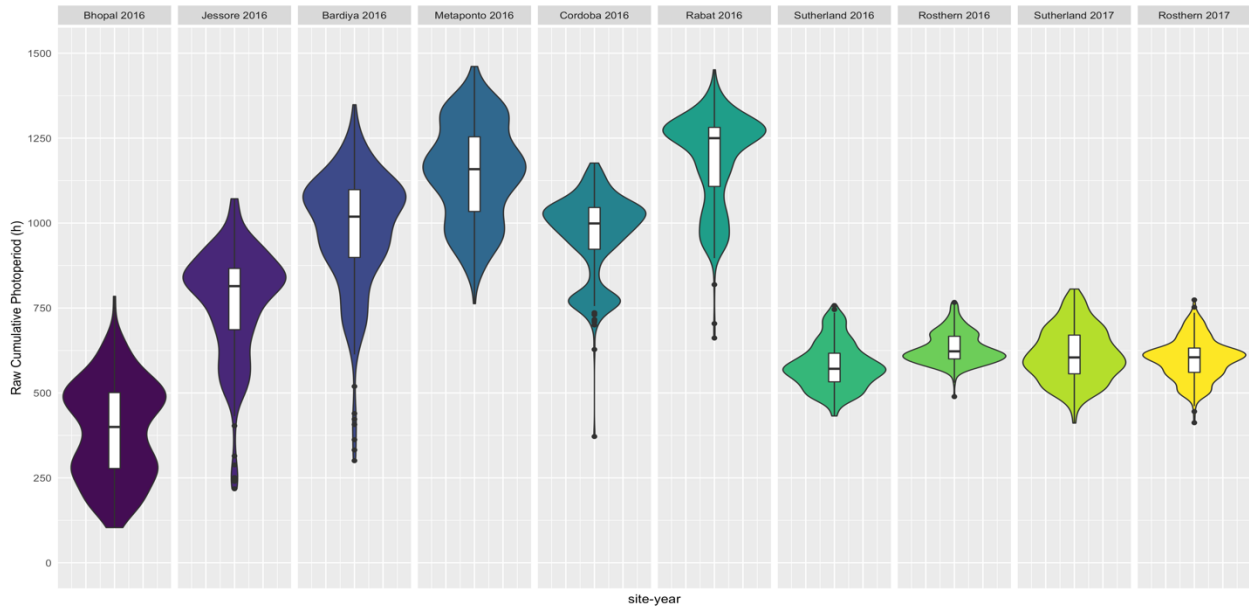
Figure 3. 5. Variation in cumulative photoperiod (h) from seeding to the days to flowering when 324 genotypes were tested in the field at ten different site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017). The width of the plots shows the density of the distributions. The whiskers on the boxes represent 1.5 times the quantile of the data. Individuals falling outside the range of the whiskers are represented as dots.

The results showed a minimum calculated critical photoperiod of 9.2 h for genotype PI 299366 LSP and a maximum of 16.12 h for genotype PI 431679 LSP, after excluding the unrealistic values of 1.97 h and 22.78 h for PI 472588 LSP and ILL 5888, respectively (Fig 3.6). The list of the genotypes with critical photoperiod along with the coefficients derived from the equation $1/f = a + b\overline{T} + cP$, which were used to calculate the critical photoperiod are available in Appendix 3.
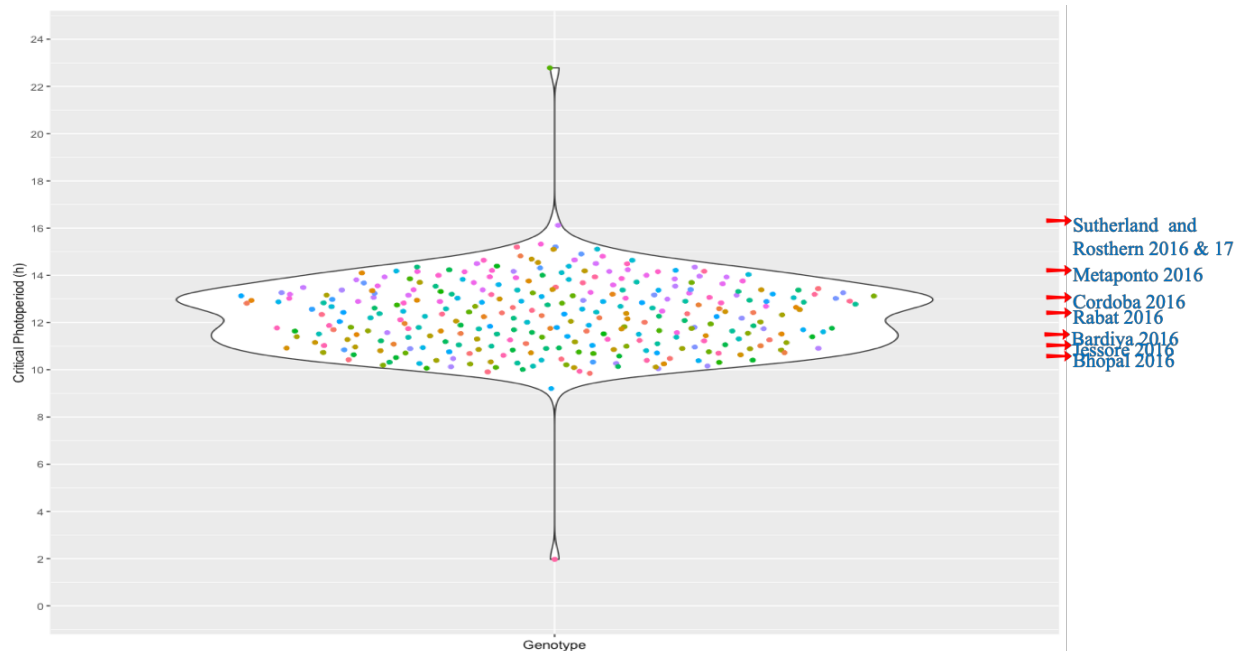
Figure 3. 6. Distribution of calculated critical photoperiod (h) values of 324 genotypes calculated using the formula, CP = -a/c. The values of a and c are the intercept and photoperiod coefficient derived from the regression model $1/f = a + b\overline{T} + cP$ (Summerfield et al. 1985) calculated using data from ten site-years. f is the number of days to flower from emergence, $\overline{T}$ is the mean temperature (°C) from emergence to first flowering, P is the photoperiod (hours from sunrise to sunset) on a first day of flowering day. The blue texts accompanying red arrows represent the maximum photoperiod hours experienced in indicated site-years.

Consideration of the critical photoperiod resulted in overall cumulative photoperiod (CPP) variation for both genotypes and site-years differently than just the simple accumulation of daily photoperiod. Many genotypes were observed to have a CPP of zero which led to a distinct skewness in distribution of CPP in the Mediterranean and South Asian site-years (Fig 3.7). The outliers for critical photoperiod stretch the distribution towards higher levels in the Mediterranean and South Asian and towards zero in temperate site-years. Several genotypes have higher critical photoperiod values than the maximum day-length hours in different site-years (Fig 3.6). This caused a zero CPP value for 213, 20 and 137 genotypes respectively in Bhopal 2016, Jessore 2016 and Bardiya 2016 site-years; 37, 64 and 110 genotypes in Metaponto 2016, Cordoba 2016 and Rabat 2016; and one genotype (ILL 5888) had a CPP of zero in temperate site-years (Fig 3.7). An additional 50 genotypes in Bhopal 2016, 68 genotypes in Jessore 2016 and 59 genotypes in Bardiya 2016; and 46 genotypes in Metaponto 2016, 60 genotypes in Cordoba 2016 and 66 genotypes in Rabat 2016, had a zero CPP because they flowered before they reached their calculated critical photoperiod.
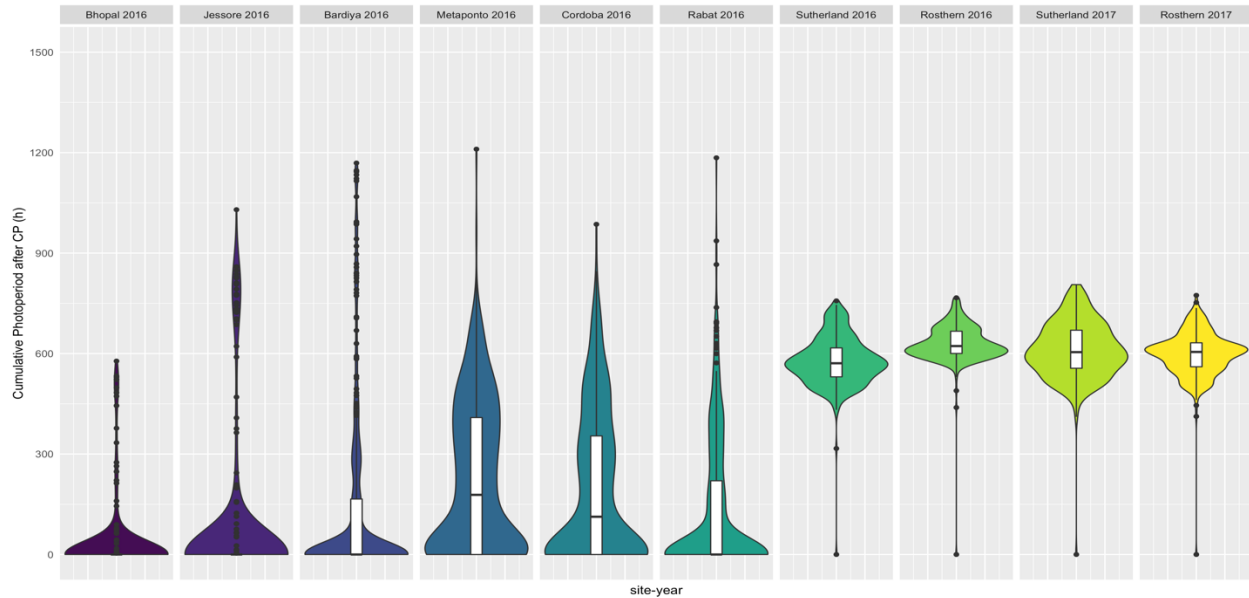
Figure 3. 7. Variation in the average cumulative photoperiod, after reaching a calculated critical photoperiod, for 324 genotypes grown in ten different site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017). The width of the plots shows the density of the distributions. The whiskers on the boxes represent 1.5 times the quantile of the data. Individuals falling outside the range of the whiskers are represented as dots.

3.3.4 Interactive effect of temperature and photoperiod on days to flower across site-years

The analysis conducted to determine the interaction effect of the temperature and photoperiod by using photothermal time (PTT) model which was based on TFT with 5°C base temperature and simple accumulation of photoperiod from days to emergence to DTF revealed a mix between TFT and raw cumulative photoperiod (Fig 3.5). The PTT requirement for the genotypes in the South Asian site-years were highest except for Bhopal 2016 where almost half of the genotypes did not flower, and the PTT requirement for the temperate site-years was lowest. The PTT requirements for DTF in the Mediterranean site-years was intermediate compared to South Asian and temperate site-years (fig 3.8).
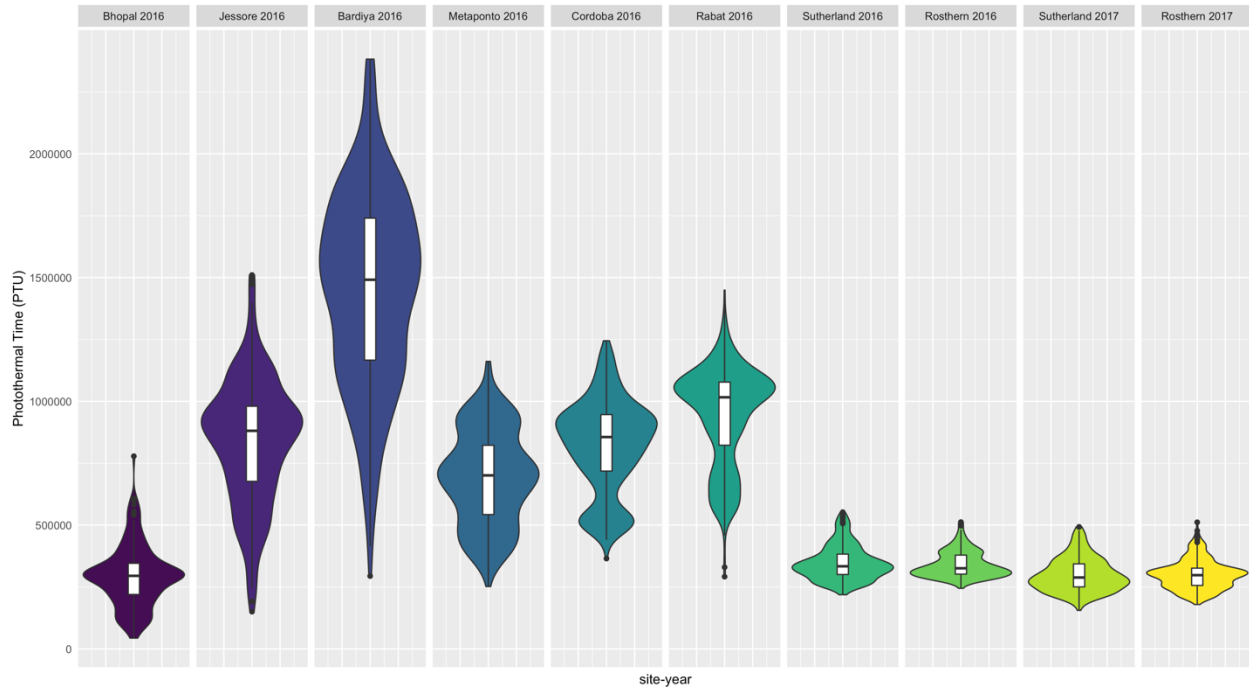
Figure 3. 8. Variation in the average photothermal time to flowering among 324 genotypes grown at ten different site-years (South Asia: Bhopal 2016, Jessore 2016 and Bardiya 2016; Mediterranean: Metaponto 2016, Cordoba 2016 and Rabat 2016; Temperate: Rosthern and Sutherland 2016 & 2017) using 5°C as base temperature. The width of the plots shows the density of the distributions. The whiskers on the boxes represent 1.5 times the quantile of the data. Individuals falling outside the range of the whiskers are represented as dots.

3.4 Discussion

This study was conducted with the objective to better understand the variability in DTF when a diverse set of lentil germplasm is grown in multiple field environments. As previous studies in different crops concluded the variation in DTF is mostly because of temperature and photoperiod (Hadley et al. 1984a; Summerfield et al. 1985; Roberts et al. 1985, 1993; Coupland 1995; Cockram et al. 2007; Springate and Kover 2014), this study focused on the effect of temperature and photoperiod on DTF and their interaction by using the variability inherent in different growing environments. Furthermore, this study was also an attempt to identify a suitable model explaining variation in DTF based on temperature and photoperiod as well as their interaction in diverse environments.

The results confirmed that site-year has the highest impact on variation in DTF when looking across diverse environments. AMMI analysis results demonstrated that DTF for site-years within individual macro-environments are similar, however, DTF was latest in the Mediterranean site-years, intermediate in South Asian (sub-tropical Savannah), and earliest in temperate site-years. This indicates environmental factors viz. temperature and photoperiod are the likely causes of variation observed when a large number of genotypes are grown in diverse site-years as these factors are distinct in different macro-environments.

The daily average temperature during the growing season in temperate site-years never went below the optimum temperature for lentil growth and development as defined by Clarke et al. (2005). In these site-years higher temperatures were always accompanied by long days. It could be generalized that the environmental factors viz. temperature and photoperiod both may not have restrictions on DTF in this environment. The early flowering and a narrow variation in DTF compared to the other macro-environments could be considered the result of a lack of restrictions from these two factors. Summerfield et al. (1985) also reported early flowering of lentils in long days and warm temperature situations. It could be concluded that the DTF variation in temperate macro-environment is the result of genotypic differences with in a large population with minimal impact from the temperature and photoperiod.

Temperature below the optimum requirements during initial days of field experiments have likely resulted in delayed flowering in the Mediterranean site-years. Summerfield et al. (1985) reported that low temperature extends the vegetative period and delays flowering in lentils. Furthermore, low temperature in Mediterranean site-years was also accompanied by short days which might have added to the effect and caused further lengthening of the vegetative period especially for the photoperiod sensitive genotypes. However, the day length in the later days started increasing which could have helped genotypes to flower despite the temperature-induced delay. Thus, it could be broadly concluded that DTF variation in Mediterranean site-years was due to the both genotypic variability and temperature as the primary environmental factor.

Daily average temperature was mostly within the optimum requirement in South Asia except for the later days of the field experiments in India and Bangladesh where temperature crossed the optimum upper limit. As discussed earlier, higher temperatures might have influenced the genotypes to flower, however, temperatures higher than the optimum resulted in flower abortion

for many genotypes. Similar findings were also reported in previous studies (Summerfield et al. 1985; Roberts et al. 1988; Kumar J et al. 2016). In addition to the higher temperature, South Asian environments always experienced shorter days during flowering compared to the Mediterranean and temperate regions. The shorter days might have restricted flowering for the photoperiod sensitive genotypes in South Asia. Studies in related crops, i.e., in pea (Berry and Aitken 1979), chickpea (Daba et al. 2016b) and faba bean (Catt and Paull 2017), also observed delayed flowering in short day and warm temperature situations. Thus, it could be broadly concluded that the DTF variation in South Asian site-years were because of both genotype and the interaction effect of temperature and photoperiod.

Temperature and photoperiod were observed to be the main factors defining DTF variation in Mediterranean and South Asian site-years, however, the average of these factors could not be used directly in comparing the DTF variation across different site-years. Accounting for temperature via TFT (summation of GDD from seeding to DTF) was a better approach for comparing the variation in DTF. More precisely, TFT using a 5°C base temperature appeared to be a better approach over TFT using a 0°C base temperature. Using TFT with a 5°C base temperature helped make site-years more comparable with each other. After conversion of temperature to TFT, variation in the Mediterranean site-years looked similar to temperate site-years. AMMI analysis result also showed temperate and Mediterranean site-years close to each other. Hence it could be concluded that TFT is a reasonable method for explaining the DTF variation under long day situations. Moreover, this result indicates that TFT method could be used for predicting DTF of lentil in long day environments; as has already used to predict flowering in many crops (Cross and Zuber 1972; Blanchard and Runkle 2006; Eshraghi-Nejad et al. 2015).

As temperature and photoperiod both are the major factors governing DTF variation in South Asian site-years, TFT alone did not account for the DTF variation and indicates a need of an interaction factor. To get a better interaction model, it is a essential to have the best method to see the photoperiod effect. Unfortunately, it was not possible to explain the model for studying effect of photoperiod due to the constraints associated with the critical photoperiod calculation. The cumulative photoperiod (CPP) after the critical photoperiod had a zero value for several genotypes in many site-years. This may be explained by the presence of insensitive genotypes as, by definition, insensitive genotypes do not have a critical photoperiod and flower under any

photoperiod. The critical photoperiod and a DTF score of early-medium-late at each site-year were used to try to differentiate the sensitive from insensitive genotypes. Those genotypes which had a lower critical photoperiod value and a low CPP should be photoperiod insensitive and should have flowered early irrespective of the site-year, however, this was not the case (data not shown here). Additionally, despite modifying the critical photoperiod calculation to only account for days post emergence and using the photoperiod on the day of flowering as used by Catt and Paull (2017) for faba bean (Catt, personal communication 2018), the calculated values still did not appear realistic for many genotypes. Hence, clearly, the $1/f = a + b\overline{T} + cP$ equation did not work for field situations even with some modifications, even though it fitted best in controlled condition experiments (Iannucci et al. 2008; Roberts et al. 1988; Summerfield et al. 1992). An alternative model suggested by Roberts et al. (1986) could also not be applied as their model needs predefined photoperiod (in)sensitive stages which were not definable from this study. Thus, development of some other model is advisable to identify critical photoperiod and then this could be applied to accumulation of daylength after crossing this point. Another variable that was not accounted for but maybe should be, is light quality. Yuan et al. (2017) reported that *Lens* species are sensitive to light quality (esp. to high R/FR) and as such disturbances such as smoke, air pollution or excessively cloudy days could have an impact on the genes controlling flowering time. Lastly, the common interaction model developed by Masle et al. (1989) and later modified by Brachi et al. (2010) could not help to make site-year comparable to each other. The result obtained from this approach looked simply like the multiplication of TFT and simple photoperiod accumulation.

3.5 Conclusion

In conclusion, significant variation among genotypes (G) with a large influence of site-years (E), and interaction between genotype and site-year (G × E) defines DTF in lentil. The variation in DTF for genotypes in differing environments was observed to be the result of both temperature and photoperiod and their interaction. The DTF variation in temperate macro-environment observed mainly because of the genotypic differences as the both temperature and photoperiod were sufficient for flowering. Whereas, DTF variation in Mediterranean site-years was observed as the result of genotype and temperature, and the variation in South Asian site-years were mostly because of genotype and the interaction between temperature and photoperiod. TFT using 5°C as a base temperature appeared to be a reasonable method for explaining DTF variation

under long day situations. However, proper methods to define DTF variation in South Asian environments could not be identified due to the constraints associated with critical photoperiod calculation. Moreover, existing equations appeared to be insufficient to account for the environmental effects that are interacting with the genes of the flowering time pathway. Thus, this study concludes with the recommendation of developing a complex model including temperature and photoperiod that have a visible effect on DTF variation as well as some other factors, e.g., solar radiation, precipitation, vernalization and light quality information as well as the genetic information of the individual germplasm.

PROLOGUE TO CHAPTER 4

Chapter 3 involved the effect of environmental factors on variation in DTF across differing environments. Chapter 4 focuses on the genetic contribution to the DTF results by associating the phenotypic data (DTF and TFT) with genotypic data for each of the 324 lines. The aim of Chapter 4 is to identify the genomic regions and/or potential candidate genes related to flowering time.

CHAPTER 4. IDENTIFICATION OF GENOMIC REGIONS ASSOCIATED WITH DAYS TO FLOWER OF DIVERSE LENTIL GERMPLASM GROWN IN MULTIPLE ENVIRONMENTS

4.1 Introduction

Successful crop introduction to new environments depends on ability to acclimatize to the local environmental conditions. Adaptation of lentil germplasm from one environment to another is a complex process and largely depends on DTF which is mainly governed by temperature, photoperiod, and their interactions (Chapter 3). While the local environment plays a large role in determining DTF, the genotype of the individual and how it interacts with the environment that determines when an individual will start to flower. Detailed information on genomic regions associated with DTF, in combination with the environmental information, will allow plant breeders to more efficiently access a wider range of germplasm. Association studies are one approach used to uncover genotypic variations associated with phenotypic outcome in different environments.

Genome-wide association studies (GWAS) have been used to identify markers, and sometimes genes, associated with flowering time in many species (e.g., *A. thaliana* – Brachi et al. 2013; *O. sativa* – Begum et al. 2015; *B. napus* – Li et al. 2015; *P. vulgaris* – Nascimento et al. 2018; *G. max* – Zhang et al. 2015; *Z. mays* – Romero Navarro et al. 2017). A GWAS assesses the statistical association between genetic markers and traits of interest and provides an opportunity to dissect the genetic architecture of complex traits (Li et al. 2012). There are several approaches to GWAS for different traits and crops; a standard approach uses a mixed linear model (MLM) (Yu et al. 2006; Zhang et al. 2010). Population structure (Q) and kinship (K) matrices are often used as covariates in MLM to reduce the error and false associations (Yu et al. 2006; Yang et al. 2014; Liu et al. 2016b, 2016a; Huang et al. 2017).

Genetic studies for flowering time in lentil have thus far been limited to bi-parental populations. The latest of this kind is by Rajandran's (2016), his study was an expansion of the understanding of the genetic control in lentil flowering after the *Sn* locus discovery. Rajandran (2016) reported Arabidopsis orthologue *ELF3*, Medicago orthologues *FTa1* and *FTa2,* and Arabidopsis paralogue *PSEUDO-RESPONSE REGULATOR (PRR59c)* as the potential candidates for early-flowering. Markers associated with temperature (esp. vernalization) and photoperiod responses

have been identified under short days, mostly in controlled conditions (Weller et al. 2012; Weller and Ortega 2015; Rajandran 2016) and very few in field situations (Kumar et al. 2018). However, it remains to be determined if these will be applicable under diverse field conditions, under varied photoperiod and temperature regimes, or in a broader set of genotypes. The phenotypic and environmental data described in Chapter 3, combined with genotypic data already generated for the 324 lines makes a powerful dataset for a GWAS.

The research hypothesis was that candidate genes controlling DTF, or markers for regions containing these genes, could be identified using a GWAS approach. In addition, it is expected that inclusion of environmental data along with the phenotypes such as thermal flowering time (TFT) could be helpful in identifying candidate genes related to adaptation of lentils in diverse environments. The objectives of this research were to associate genomic regions with DTF in lentil germplasm grown in different environments and identify candidate genes based on previous and ongoing flowering time research.

4.2 Materials and Methods

4.2.1 Phenotyping

Details about plant material, field experiments, and phenotypic as well as environmental data and statistical analysis can be found in Chapter 3. For this study, the least square means of DTF from each site-year were used. Thermal Flowering Time (TFT) using the base temperature 5°C was also used as an additional phenotypic factor.

4.2.2 Genotyping

Genotyping of all 324 accessions was done using a custom lentil exome capture array as described by Ogutcen et al. (2018). Genotypic data in the form of a 'high confidence SNP array' were accessed from http://knowpulse.usask.ca/portal/project/AGILE%3A-Application-of-Genomic-Innovation-in-the-Lentil-Economy managed by the Pulse Crop Research Group at the University of Saskatchewan (U of S). Markers with more than 5% missing data and a minor allele frequency of less than 5% were removed prior to analysis. Redundant markers that had squared pairwise correlation of one were removed. The remaining 255,714 SNPs and indels, with high coverage sites across the lentil genome (Fig 4.1), were used for GWAS.
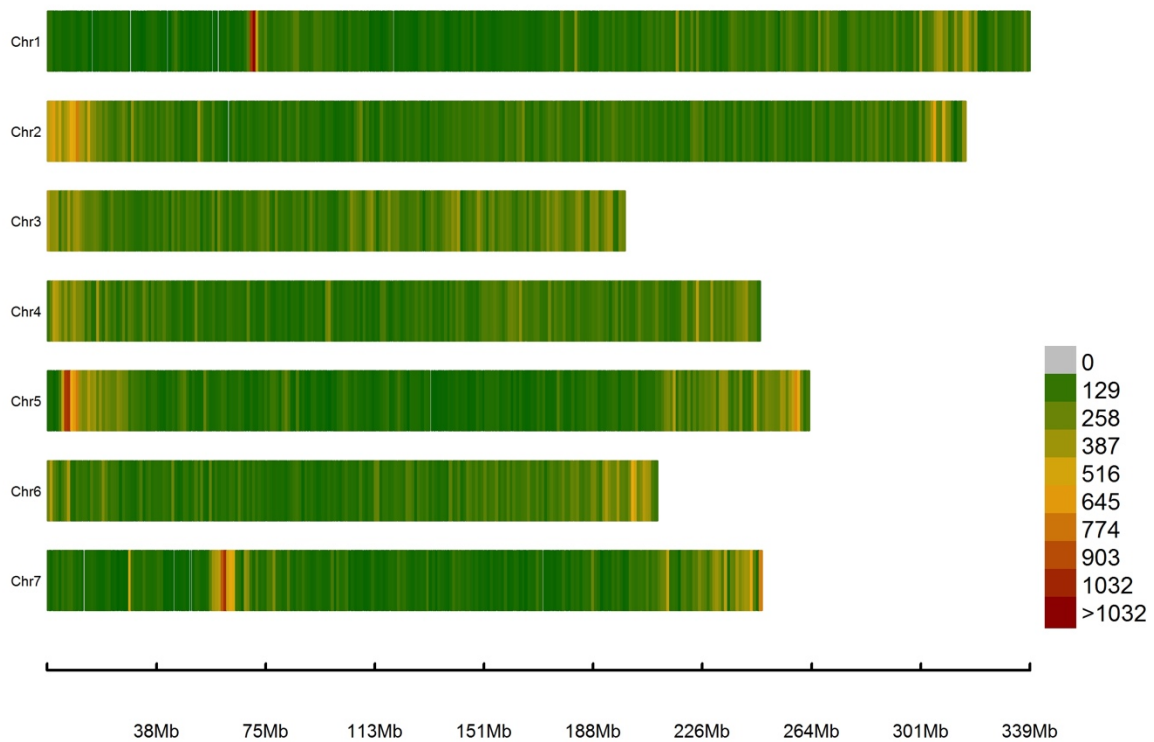
Figure 4. 1. Distribution of 255,714 single nucleotide polymorphisms (SNPs) and indels across the seven chromosomes of the lentil genome. Different colors depict marker density (the number of markers per 1Mb window).

4.2.3   Genome-Wide Association Analysis

Association analysis was performed using a Mixed Linear Model (MLM) algorithm (Zhang et al. 2010) implemented in the Genome Association and Prediction Integrated Tool (GAPIT) (Lipka et al. 2012) in R. The kinship matrix (K) and population membership coefficients (Q-matrix) were calculated using TASSEL 3.0 (Bradbury et al. 2007) and fastSTRUCTURE (Raj et al. 2014) and were provided by Drs. Teketel Haile and Ezgi Ogutcen, respectively. While running GWAS, the model used the Q-matrix as a covariate, the K-matrix was a random factor, and the fixed phenotypic factors were DTF and TFT. As shown in Chapter 3, significant variation among genotypes (G) with a large influence of site-years (E), GWAS will be presented separately for each site-year.

The threshold for significant association was set at 6.7 which is equal to -log10(0.05 / 255,714) [i.e. (P = 0.05 / no. of markers)] using the Bonferroni correction method (Holm 1979). The

Bonferroni correction was used to correct for multiple testing. Quantile-quantile (Q-Q) and Manhattan plots were constructed using the R package "qqman" (Turner 2014) to display the significance of SNP markers across the genome. In a Q-Q plot, the observed -log10(P) values of a subset of the markers are plotted against the expected -log10(P) values under the null hypothesis of no association. Observed -log10(P) should follow the expected values and deviations should only occur towards the upper end if the model was corrected for population structure and kinship. Manhattan plots were generated for traits that had significant associations after multiple testing correction.

4.2.4 Candidate Gene Identification

Regions of the genome that were significantly associated with DTF and TFT were identified by filtering the significant markers from the GWAS results. The physical positions of these markers were examined in the annotated lentil genome on JBrowse (http://knowpulse.usask.ca/portal/jbrowse/Lentil) by pointing the QTL associated to that region. Known legume flowering time genes had been annotated on the lentil genome by Dr. Raul Ortega Martinez (University of Tasmania).

After identifying candidate gene or genomic regions harbouring significant markers, the allelic proportions at the most significant marker was determined and the composition relative to the population structure data ascertained.

4.3 Results

4.3.1 GWAS using DTF and TFT as phenotypic factors

Significant peaks were identified for DTF on chromosome 2 for Jessore 2016, Bardiya 2016 and Cordoba 2016 (Fig 4.2). Significant peaks were also observed for DTF on chromosome 5 for Cordoba 2016 (Fig 4.2). No significant associations were observed for the other site-years nor when TFT was used as the phenotypic factor in any site-year (Appendix 4).
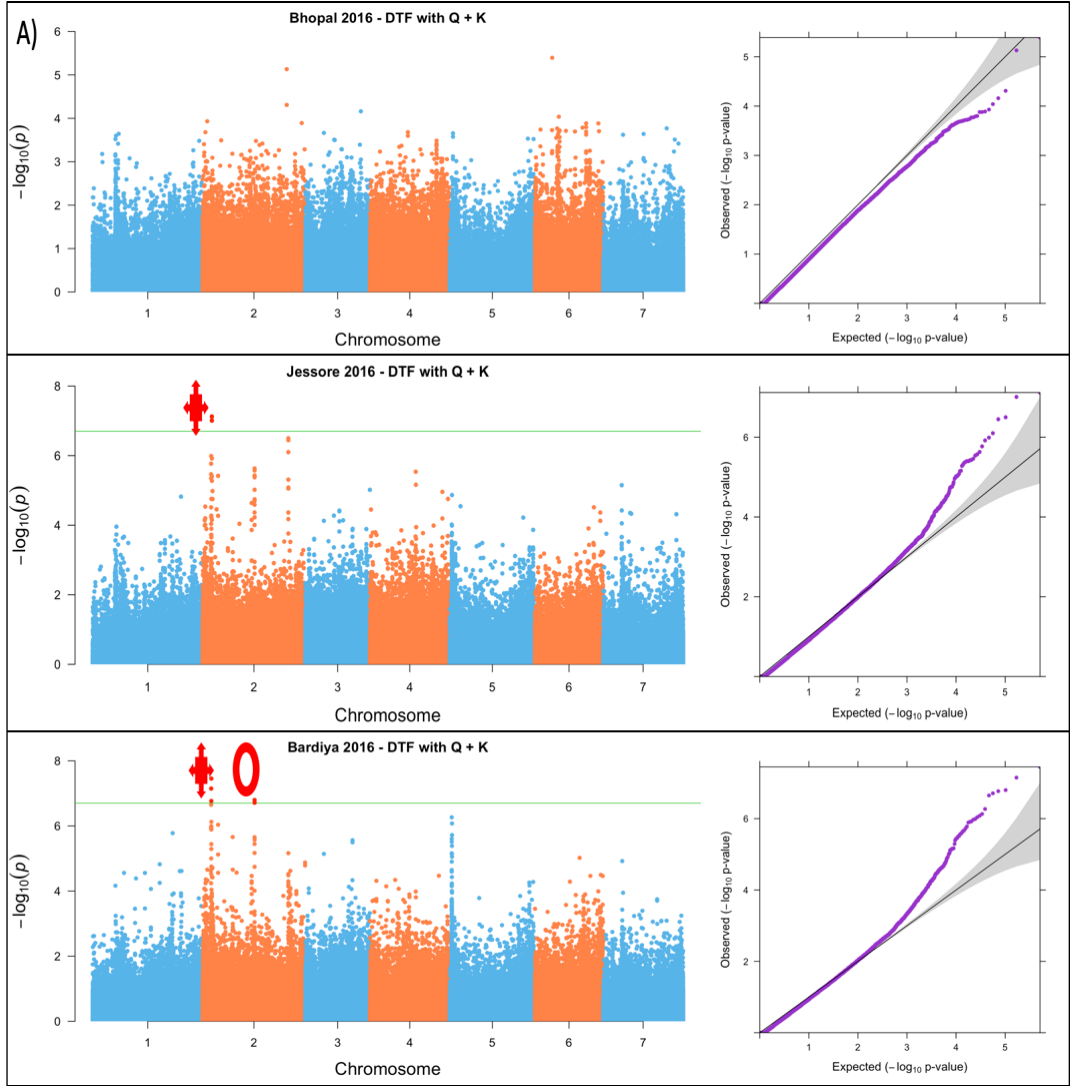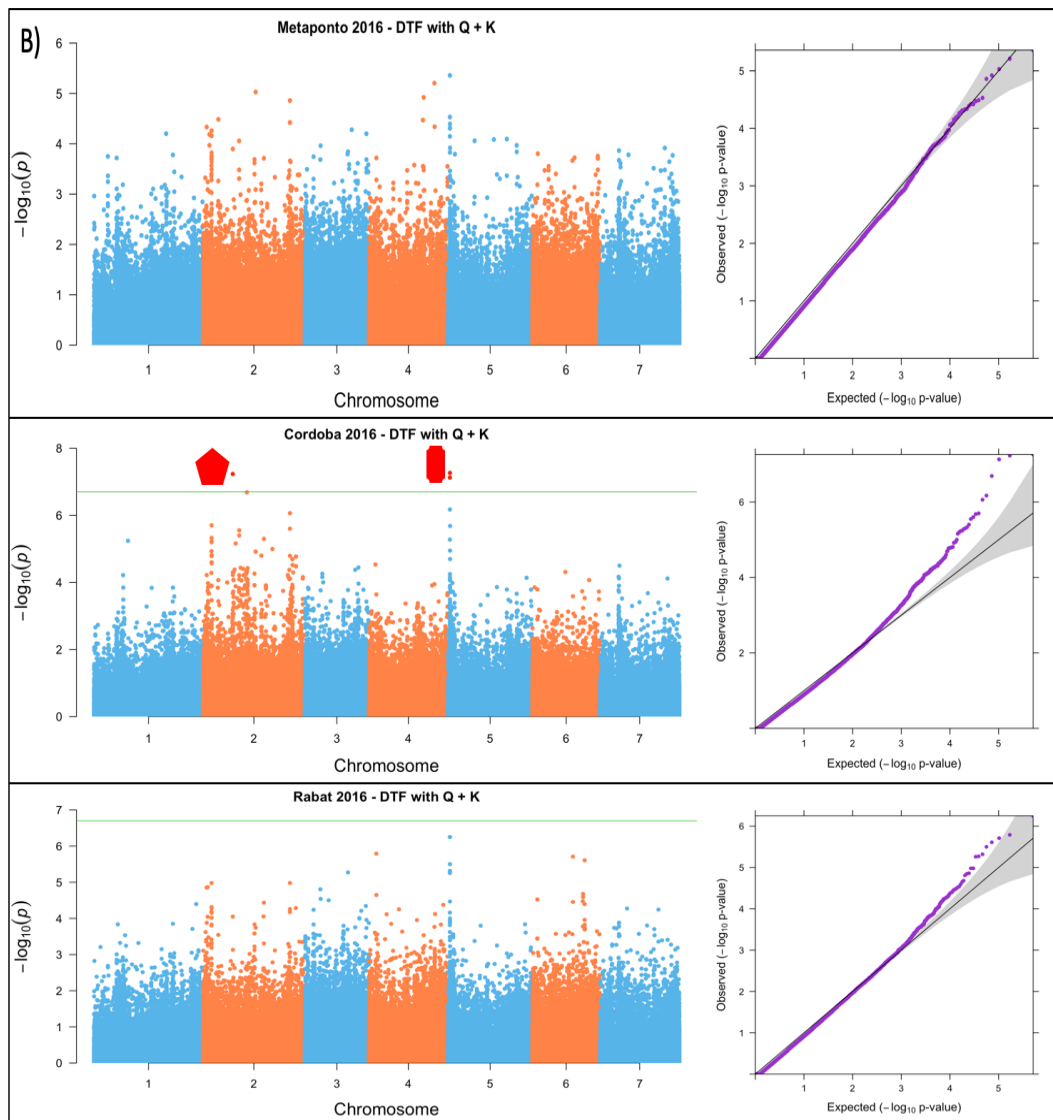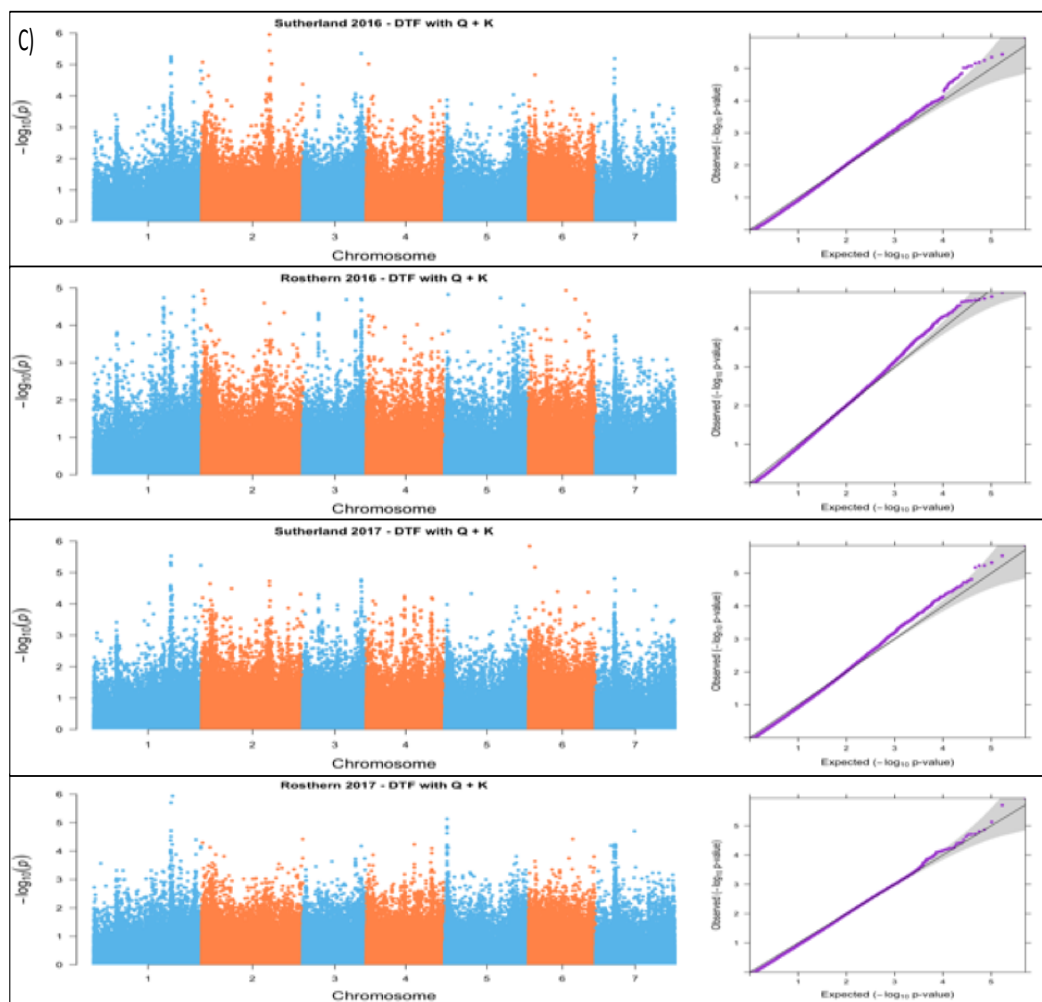
Figure 4. 2. Manhattan plots (left) and Quantile-quantile (Q-Q) plots (right) derived from association studies using days to flower (DTF) of 324 lentil lines grown in the field in – A) South Asian macro-environments - Bhopal 2016, Jessore 2016 and Bardiya 2016; B) Mediterranean macro-environments - Metaponto 2016, Cordoba 2016 and Rabat 2016; and C) Temperate macro-environments - Rosthern and Sutherland, SK 2016 & 2017. The X-axis of Manhattan plots represents lentil chromosomes, and the Y-axis is -log10 of P-values. Adjacent chromosomes are separated by color. The green line on the Manhattan plots indicates the significance threshold [-log10(*P*) > 6.7]. Different red colored shapes on the Manhattan plots indicate different QTL representing significant markers. The X-axis on the Q-Q plots is the expected -log10 of P-values and the Y-axis is the observed -log10 of P-values.

4.3.2   Examination of significant genomic regions to predict candidate genes

A total of 95 markers were significantly associated with DTF which are clustered in four different loci (Table 4.1). On chromosome 2, 53 markers were identified as the significant at both Jessore 2016 and Bardiya 2016 site-years which were clustered in QTL qDTF.2-1; single marker was significant at Cordoba 2016 which was on QTL qDTF.2-2; and three markers were

significant at Jessore 2016 which were clustered in QTL qDTF.2-3. Likewise, 58 markers were significant at chromosome 5 at Cordoba 2016 and clustered in QTL qDTF.5-1. One unique flowering time related candidate gene *LcELF4c* was identified in vicinity of most significant marker LcChr2p28456076 which was clustered with other markers in QTL qDTF.2-1 (Table 4.1).

Table 4. 1. Genomic regions associated with days to flower and flowering time related candidate genes extracted from Lentil JBrowse at KnowPulse along with the information about the number of significant markers within the loci which were used to identify candidate genes. Different shaped red colored boxes represent QTL as indicated on Manhattan plots.

QTL	Chr	site-year	Physical position (BP)		Markers #	MS Marker	P-value	FDR Adjusted P-values	Candidate Gene
			BP Start	BP End					
qDTF.2-1	2	Jessore 2016 Bardiya 2016	26091732	28674286	7 26	LcChr2p28456076	7.54E-08	0.012534378	*ELF4c*
qDTF.2-2	2	Cordoba 2016	92674515	92845580	1	LcChr2p92674515	5.86E-08	0.006388081	
qDTF.2-3	2	Bardiya 2016	160151349	160173160	3	LcChr2p160171626	1.60E-07	0.009539251	
qDTF.5-1	5	Cordoba 2016	6256891	6800429	58	LcChr5p6364346	5.45E-08	0.006388081	

QTL, Quantitative Trait Loci; Chr, Chromosome; BP, Base Pair Position, Marker #, Number of significant markers within a particular QTL; MS Marker, Most Significant Marker; FDR, False Discovery Rate

Classification of genotypes according to allelic composition at LcChr2p28456076 (qDTF.2-1) resulted in a group consisting of most members of the [1]AFRICA 2 group and many admixed individuals that have the alternate allele, while the rest have the reference allele (fig 4.4.A). Similar results as for LcChr2p28456076 (qDTF.2-1) were observed for LcChr2p160171626 (qDTF.2-3) (fig 4.4.A). For LcChr2p92674515 (qDTF.2-2) none of the alleles had a governing role for DTF variation as there was no dominating structure group in either of the alleles (fig 4.4A). While for LcChr5p6364346 (qDTF.5-1), the majority of genotypes observed to be governed by reference allele where mostly from the SOUTH ASIA1 group (fig 4.4A).

---

[1] These are the population structure group derived from fastSTRUCTURE analysis (and provided for this study by Dr. Ezgi Ogutcen. Population structure analysis identified ten different related groups for the 324 genotypes. They are – AFRICA 1, AFRICA 2, AMERICA 1, AMERICA 2, ASIA 1, ASIA 2, EUROPE 1, EUROPE 2, SOUTH ASIA1, and OTHER (admix).
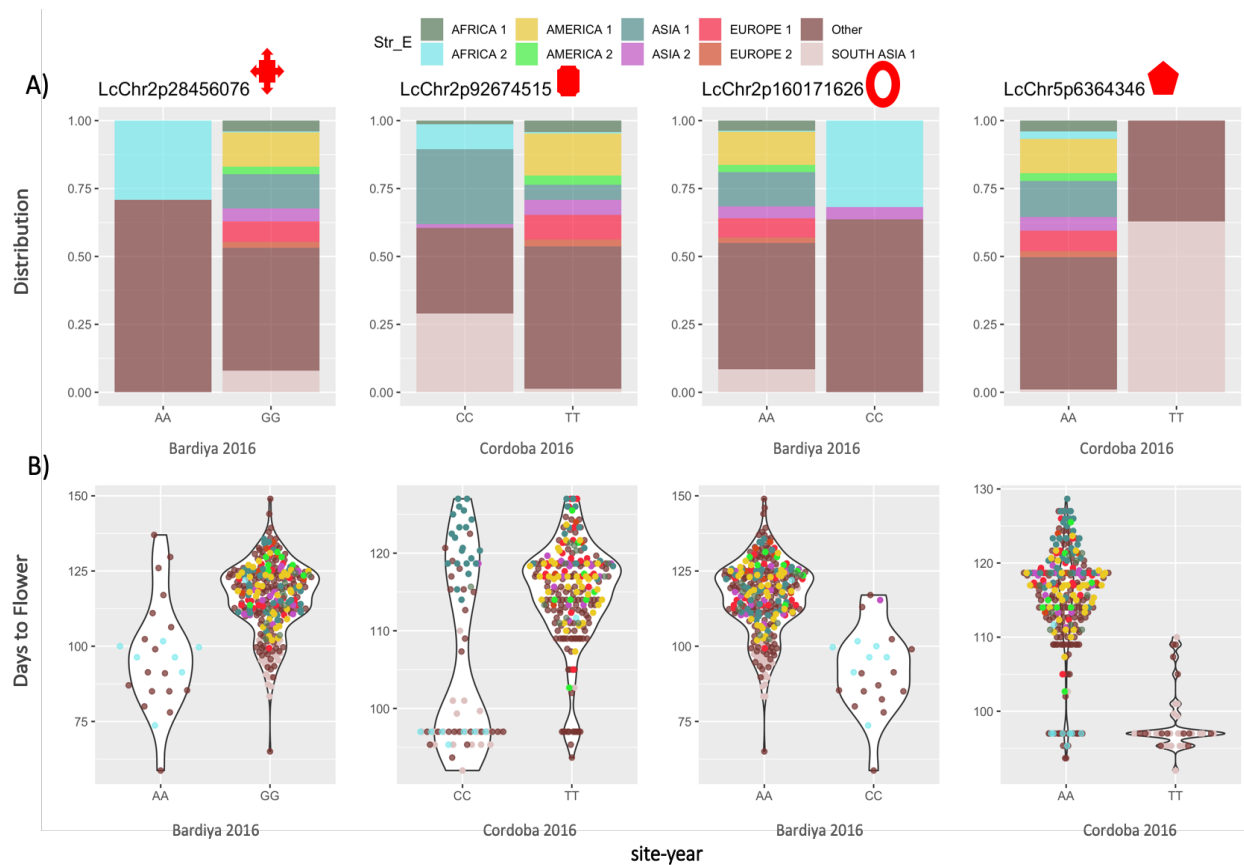
Figure 4. 3. A) Allele distribution across the sub-populations derived from fastSTRUCTURE analysis for four most significant markers resided in different QTL as indicated by the different red colored shapes corresponds to Table 4.1 and fig 4.2. B) Distribution of days to flower for individual significant markers separated by allele state within site-year. Different colors in all figures represent different groups of the genotypes based on the population structure (as shown in the legend as Str_E).

4.4. Discussion

The objectives of this research were to identify genomic regions associated with DTF in lentil germplasm grown in different environments and detect potential flowering time related candidate genes. Significant associations between DTF as the phenotypic data and genotypic markers were noticed for two South Asian site-years (Bardiya 2016 and Jessore 2016) and one Mediterranean site-year (Cordoba 2016). The significantly associated markers were clustered in four QTL among which three were resided at chromosome 2 and one at chromosome 5.

A flowering time related candidate gene *LcELF4c (EARLY FLOWERING 4)* was detected at QTL qDTF.2-1 on chromosome 2 where significant association was detected in both Jessore 2016 and Bardiya 2016 site-years. This gene is believed to be involved in photoperiod perception and circadian regulation. Orthologs of *ELF4* have been observed to promote the clock accuracy and is essential for sustained circadian rhythms in *Arabidopsis* (Doyle et al. 2002; Putterill et al. 2004; Kikis et al. 2005)*. Doyle et al. (2002) also reported that mutations in this gene result in early flowering in short day situations. *DIE NEUTRALIS (DNE)*, an ortholog of *Arabidopsis ELF4*, elevates the expression of a *FLOWERING LOCUS T (FT)* homolog under short-day conditions in garden pea (Liew et al. 2009). In general, *FT* is believed to be crucial for accelerated flowering in response to long days (Turck et al. 2008). As *ELF4* was identified in the genomic regions observed to be associated with DTF in South Asia, introgression of this gene in the genotypes from long day environment could be helpful in adopting temperate genotypes in South Asia or other short-day environments. However, additional research such as identifying the biological function of *ELF4* in lentil should be done to confirm this conclusion.

Further analysis with allelic proportion at all significant SNPs gave ideas about the role of specific allele combinations for early flowering in selected site-years. However, a full picture of their role was not clear, which could be because DTF is governed by multiple genes and putative candidate genes are not the only genes governing DTF in lentil at those site-years. For marker LcChr5p6364346 in qDTF.5-1, early flowering for the majority of genotypes was associated with the reference allele and is prevalent in most South Asian genotypes. This marker is found within the LD for the genomic region of another flowering time related time candidate gene *SUPPRESSOR OF PHYA-105 (SPA1)*. *SPA1* gene is considered to have a significant role in flowering in a short-day situation (Laubinger et al. 2006; Ishikawa et al. 2006; Chen et al. 2010). Thus, introgression of this marker in the genotypes from long day environments could also help them to flower at on optimum time in short day situations esp. in Mediterranean locations.

Association studies results provided a positive sign in identifying potential solution for the adaptation of germplasm in differing environments. The result would have been better if there were a greater number of markers identified as significant. However, only a small number (N=95) of markers were significantly associated with DTF although a large number of markers (N= 255,714) were used for the association studies. Likewise, some peaks fall just shy of the threshold limit for some site-years (fig 4.2). If the significance threshold was lowered, more

markers would be declared significant. Part of the problem arises from the large number of SNPs in the denominator of the equation to determine the significant threshold (P-values divided by the total number of polymorphic markers) by using the Bonferroni correction approach (Holm 1979). This approach is believed as the suitable for controlling type-I error (Yi et al. 2014); however, many GWAS related discussions have concluded this approach is too stringent and might produce false negatives or type-II errors. The alternate method over this is using false discovery rate (FDR) suggested P-value using the Benjamini and Hochberg (1995) approach. However, no differences in results were observed between these two approaches when considering significant threshold level for FDR value ≤ 0.05. Another potential option could be to decrease the number of SNP markers used by strategically selecting a subset across the genome, this could be done by using haplotypes. Furthermore, this study followed the mixed linear model (MLM) procedure, a widely accepted approach for GWAS because of its capacity to better control for the confounding effects of both population structure and cryptic relatedness among individuals (Miao et al. 2018). However, the statistical power of this method rapidly decreases with the increase in the number of variants controlling the variation in the given trait (Miao et al. 2018). Therefore, testing of more complex approaches which requires more statistical power (e.g., Multi-locus mixed-model (MLMM) and models using a Bayesian approach) are recommended for future analyses, which could yield more significant markers.

In addition, this study is expected to be important in identifying the genomic regions and potential candidate genes for TFT which was derived from seeing the effect of temperature on DTF variation in contrasting environments. However, none of the markers were observed to be significantly associated with this trait in none of the ten site-years. As Chapter 3 of this thesis concluded, TFT is a better approach to deal with the DTF variation in long day environments, but non-occurrence of significant markers for TFT from both temperate and Mediterranean site-years could also be because of the high significant threshold. Thus, future study in line with the previous recommendations for DTF is also advisable for this trait.

4.5. Conclusion

Association studies identified four QTLs and one candidate gene associated with DTF in lentil using a MLM GWAS approach including both population structure and kinship matrix. The candidate gene and markers could be used by lentil breeding programs in short day environments

to overcome the flowering time related problems for the lentil genotypes from tong day situations. It could be concluded from this study that the GWAS is a promising tool to dissect complex traits by identifying genomic regions and potential candidate genes, however, it was not possible to dissect the genetic components as thoroughly as hoped. Additionally, use of some other statistical models, SNPs filtering options, as well as multiple testing correction procedures are recommended for future studies.

CHAPTER 5. GENERAL DISCUSSION, CONCLUSIONS, AND FUTURE RESEARCH

5.1 General Discussion of Results

Lentil is currently grown in more than four dozen countries representing mostly temperate, Mediterranean and South Asian environments (Tullu et al. 2011; Khazaei et al. 2016). However, most of the lentil breeding programs are based on a narrow genetic base (Khazaei et al. 2016) leading to the risk of development of the new lentil varieties as per the grower's demand. The dependency of lentil breeding programs on narrow genetic resources is mainly because of the constraints of adaptation. Adaptation of lentil genotypes to differing environments mostly relies on days to flower (DTF) as a result of genotypes interacting with environmental factors, especially temperature and photoperiod (Summerfield et al. 1985; Roberts et al. 1986; Summerfield and Roberts 1988; Erskine et al. 1990a, 1994). In this context, this research was conducted with the objective of understanding the variation of DTF governed mainly by temperature and photoperiod and their interactions as well as to identify genomic regions and candidate genes or markers associated with DTF in specific environment. To accomplish this, a large set of genotypes was grown in ten different field environments across four continents and the consequence of the interaction with temperature and photoperiod with respect to DTF was investigated at the phenotypic and genotypic level.

The first study was focused on studying the variation in DTF of 324 diverse set of genotypes tested in ten different field locations representing three major lentil growing macro-environments. There was enormous variation in DTF among site-years representing different macro-environments. This indicates environmental factors viz. temperature and photoperiod are the likely causes of variation observed when a large number of genotypes are grown in diverse site-years as these factors are distinct in different macro-environments. DTF was earliest in temperate site-years, intermediate in South Asian site-years and latest in the Mediterranean site-years. The DTF variation with in temperate site-years was mostly because of the genotypic factors as the daily average temperature during the growing season in those site-years were always above the minimum requirements as defined by Clarke et al. (2005). In addition, the higher temperature was always accompanied by the long days in those site-years. Previous studies by Summerfield et al. (1985) also reported early flowering of lentils in long days and warm temperature situations. In Mediterranean site-years, DTF variation was governed by the

genotypes as well as temperature as the environmental factor. Lower temperature than the optimum during initial days have resulted delayed flowering in those site-years. Summerfield et al. (1985) also reported delaying flowering due to the low temperature in lentils. The lower temperature during the first few weeks of the crop season was also accompanied by the short day in Mediterranean site-years, however, the day length in the later days in the field started increasing. In South Asian site-years, higher temperature above the optimum resulted in flower abortion for late flowering genotypes. In addition, short days accompanying those high temperatures resulted in large variations in flowering date. Accounting for temperature via TFT (summation of GDD from seeding to DTF) using a 5°C base was a better approach for comparing the DTF variation in Mediterranean site-years and made them comparable to temperate site-years. However, model parameters to account for temperature and photoperiod interaction to neutralize the DTF variation in south Asia could not be identified due to the constraints associated with the critical photoperiod calculation. Besides, previous models designed to determine the effect of photoperiod and the interaction between temperature and photoperiod failed to define the DTF variation in diverse field experiments conducted with large number of genotypes.

Due to the significant genotype by environment effect on DTF noted in Chapter 3, the genotypic effect on the DTF was tested for individual site-years in Chapter 4. Genomic regions and potential candidate genes related to DTF were identified at three site-years representing South Asian and Mediterranean macro-environments. DTF variation in South Asian site-years was because of an interaction with both temperature and photoperiod (Chapter 3) so the significant associations at Bardiya and Jessore 2016 site-years could be the result of specific genes interacting with these two factors. Three QTLs on chromosome 2 and a single locus on chromosome 5 contained genes or markers that are causing variation in DTF in one or more site-years. A flowering time related candidate gene *ELF4* was noticed at q.DTF.2-1 on chromosome 2. This gene interacts with photoperiod at flowering and is part of the circadian pathway (Doyle et al. 2002; Putterill et al. 2004; Kikis et al. 2005; Turck et al. 2008; Liew et al. 2009). DTF was highly variable across South Asian site-years as compared to the long day temperate site-years, and some genotypes (esp. from the temperate origin) did not flower in South Asia. If this gene is introgressed in genotypes from long day environments, those may flower at a more appropriate time.

A key flowering time related candidate gene and few QTLs were identified through this study even with the high significant threshold (-log10(P) > 6.7), if the significance threshold was lowered, more markers from other site-years would be declared significant. There are some methods other than the GWAS MLM used in this thesis which could be tested to obtain more significant markers but getting into these other methods was beyond the scope of a MSc thesis project.

5.2 Conclusions

Chapter 3 of this study demonstrated significant variation in DTF with genotypes (G), site-years (E), and genotype and environment interaction (G × E). However, the contribution of the site-year in DTF variation was very high. DTF variation within a macroenvironment was due to the genotypes alone, whereas among macroenvironments it was the result of genotype by site-year interaction. In temperate site-years, the DTF variation occurred mainly because of the genotypic variability. Temperature was the major factor defining DTF variation in Mediterranean site-years, whereas, the interaction between temperature and photoperiod was the determining factor in South Asian site-years. This showed that the temperature and photoperiod as well their interaction define the DTF variation in diverse field situation, hence the first hypothesis of this research study, i.e., temperature and photoperiod define the variation in DTF across differing environments, is accepted. Furthermore, this study also made an attempt to identify the appropriate models to identify the effects of temperature and photoperiod and their interaction, while doing so, TFT by considering 5°C as the base temperature was observed as the best approach to deal with the temperature in long day environments. However, this study was not able to confirm the models to account effect of photoperiod and the interaction between temperature and photoperiod.

Association studies conducted using MLM GWAS approach identified four QTLs for DTF on chromosome 2 and chromosome 5. A flowering time related candidate genes *ELF4* was identified at QTL qDTF.2-1 from Bardiya 2016 and Jessore 2016 site-years. This *ELF4* gene may serve as a promising target for flowering time related studies in lentil and may assist the adaptation of lentil germplasm from a long day to short day situations or from temperate to South Asian locations. This finding showed that association study is one of the reliable approaches in identifying key genes defining DTF to a specific environment.

5.3 Future Research

This study confirms the interaction of temperature and photoperiod with genetics to determine DTF among a large set of genotypes grown in diverse environmental conditions in the field. This study was able to identify some of the components of a model to study the effect of temperature, however, could not provide a complete model to explain the photoperiod and interactive effects. GWAS was used as to dissect complex traits by identifying genomic regions and potential candidate genes, however, it was not possible to dissect the genetic components as thoroughly as hoped. Therefore, the following recommendations for future research are made:

- The development of a model to address the effect of photoperiod and the interactive effect of temperature and photoperiod on DTF remains elusive. This study relied on a regression model $1/f = a + b\overline{T} + cP$ by Summerfield et al. (1985) which has several shortcomings. The primary issue is that they used indoor studies in which they could manipulate temperature and daylength independently for this method. This limits its use for field studies as it doesn't take into consideration the interaction of these factors, and other environmental conditions will also be confounded. Although this model has been applied in many crops, most were indoor experiments (Summerfield and Roberts 1988; Summerfield et al. 1992; Iannucci et al. 2008; Catt and Paull 2017), and it seems they cannot be directly applied in contrasting field situations. This may have been even more complicated by the use of such a diverse set of germplasm. Thus, a complex model, which includes other environmental factors like solar radiation, soil moisture, precipitation, vernalization and light quality, is recommended for future studies.
- This study relied on the MLM approach of GWAS to identify significant markers based on the information available at analysis time. However, other complex approaches like Multi-locus mixed-model (MLMM) and models using Bayesian approaches which have higher statistical power are recommended for future studies. Although a candidate gene related to DTF was identified from two site-years, confirmatory candidate gene analysis that includes sequencing of contrasting haplotypes, expression analysis, mutant experiments and physiological characterization of the gene are recommended.
- Both chapters of this thesis relied on only the first year of data from the Mediterranean and South Asian environments, which might have misled the results, hence, inclusion of an additional year of data is recommended to confirm the findings.

REFERENCES

Alo, F., Furman, B.J., Akhunov, E., Dvorak, J., and Gepts, P. 2011. Leveraging Genomic Resources of Model Species for the Assessment of Diversity and Phylogeny in Wild and Domesticated Lentil. J. Hered. **102**: 315–329. Oxford University Press. doi:10.1093/jhered/esr015.

Arumuganathan, K., and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Report. **9**: 208–218. doi:10.1007/BF02672069.

Aschard, H., Guillemot, V., Vilhjalmsson, B., Patel, C.J., Skurnik, D., Ye, C.J., Wolpin, B., Kraft, P., and Zaitlen, N. 2017. Covariate selection for association screening in multiphenotype genetic studies. Nat. Genet. **49**: 1789–1795. Nature Publishing Group. doi:10.1038/ng.3975.

Aschard, H., Vilhjálmsson, B.J., Joshi, A.D., Price, A.L., and Kraft, P. 2015. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. Am. J. Hum. Genet. **96**: 329–39. Elsevier. doi:10.1016/j.ajhg.2014.12.021.

Baker, J.T., and Reddy, V.R. 2001. Temperature Effects on Phenological Development and Yield of Muskmelon. Ann. Bot. **87**: 605–613. No longer published by Elsevier. doi:10.1006/ANBO.2001.1381.

Barulina, O.H. 1930. Lentil of USSR and of other countries. Bull. Appl. Bot. Genet. Pl. Breed. **40**: 1–319. [Online] Available: https://ci.nii.ac.jp/naid/10006227016/ [2018 Nov. 8].

Begum, H., Spindel, J.E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Virk, P., Collard, B., and McCouch, S.R. 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (Oryza sativa). PLoS One **10**: e0119873. Public Library of Science. doi:10.1371/journal.pone.0119873.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B **57**: 289–300. [Online] Available: http://www.jstor.org/stable/2346101 [2019 Jan. 9].

Berry, G.J., and Aitken, Y. 1979. Effect of Photoperiod and Temperature on Flowering in Pea (Pisum sativum L.). Aust. J. Plant Physiol. [Online] Available: http://www.publish.csiro.au/fp/pdf/pp9790573 [2019 Jan. 9].

Bhatty, R.S. 1988. Composition and Quality of Lentil (Lens culinaris Medik): A Review. Can. Inst. Food Sci. Technol. J. **21**: 144–160. Elsevier. doi:10.1016/S0315-5463(88)70770-1.

Blanchard, M.G., and Runkle, E.S. 2006. Temperature during the day, but not during the night, controls flowering of Phalaenopsis orchids. J. Exp. Bot. **57**: 4043–4049. Oxford University Press. doi:10.1093/jxb/erl176.

Brachi, B., Faure, N., Bergelson, J., Cuguen, J., and Roux, F. 2013. Genome-wide association mapping of flowering time in Arabidopsis thaliana in nature: genetics for underlying components and reaction norms across two successive years. Acta Bot. Gallica **160**: 205–219. NIH Public Access. doi:10.1080/12538078.2013.807302.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., and Roux, F. 2010. Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. PLoS Genet. **6**: e1000940. Public Library of Science. doi:10.1371/journal.pgen.1000940.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23**: 2633–2635. doi:10.1093/bioinformatics/btm308.

Catt, S., and Paull, J. 2017. Effects of ambient temperature and photoperiod on fl owering time in faba bean (Vicia faba L .). Crop Pasture Sci. **68**: 893–901. doi:https://doi.org/10.1071/CP17187.

Chen, H., Huang, X., Gusmaroli, G., Terzaghi, W., Lau, O.S., Yanagawa, Y., Zhang, Y., Li, J., Lee, J.-H., Zhu, D., and Deng, X.W. 2010. Arabidopsis CULLIN4-Damaged DNA Binding Protein 1 Interacts with CONSTITUTIVELY PHOTOMORPHOGENIC1-SUPPRESSOR OF PHYA Complexes to Regulate Photomorphogenesis and Flowering Time. Plant Cell **22**: 108–123. doi:10.1105/tpc.109.065490.

Chew, Y.H., Wilczek, A.M., Williams, M., Welch, S.M., Schmitt, J., and Halliday, K.J. 2012. An augmented Arabidopsis phenology model reveals seasonal temperature control of flowering time. New Phytol. **194**: 654–665. John Wiley & Sons, Ltd (10.1111). doi:10.1111/j.1469-8137.2012.04069.x.

Clarke, H., Khan, T., Croser, J., White, P., Singh, S.P., Lulsdrof, M., Hunbury, C., and Ryan, M. 2005. Temperature tolerance in food legumes. Page 21 *in* M.C. Kharkwal, ed. Abstracts of

Fourth lnternational Food Legume Research Conference (IFLRC IV), 18-22 October.

Cockram, J., Jones, H., Leigh, F.J., O'Sullivan, D., Powell, W., Laurie, D.A., and Greenland, A.J. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J. Exp. Bot. **58**: 1231–1244. Oxford University Press. doi:10.1093/jxb/erm042.

Corripio, J.G. 2015. Package ' insol .' [Online] Available: http://www.meteoexploration.com/R/insol/index.html.

Coupland, G. 1995. Genetic and environmental control of flowering time in Arabidopsis. Trends Genet. **11**: 393–7. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/7482765 [2018 Nov. 8].

Cross, H.Z., and Zuber, M.S. 1972. Prediction of Flowering Dates in Maize Based on Different Methods of Estimating Thermal Units. Agron. J. **64**: 351. American Society of Agronomy. doi:10.2134/agronj1972.00021962006400030029x.

Crowell, S., Korniliev, P., Falcão, A., Ismail, A., Gregorio, G., Mezey, J., and McCouch, S. 2016. Genome-wide association and high-resolution phenotyping link Oryza sativa panicle traits to numerous trait-specific QTL clusters. Nat. Commun. **7**: 10527. Nature Publishing Group. doi:10.1038/ncomms10527.

Daba, K., Tar'an, B., Bueckert, R., and Warkentin, T.D. 2016a. Effect of Temperature and Photoperiod on Time to Flowering in Chickpea. Crop Sci. **56**: 200–208. The Crop Science Society of America, Inc. doi:10.2135/cropsci2015.07.0445.

Daba, K., Tar'an, B., and Warkentin, T.D. 2016b. Flowering response of diverse chickpea (Cicer arietinum L.) accessions to photoperiod. Genet Resour Crop Evol **63**: 1161–1172. doi:10.1007/s10722-015-0308-5.

Doyle, M.R., Davis, S.J., Bastow, R.M., McWatters, H.G., Kozma-Bognár, L., Nagy, F., Millar, A.J., and Amasino, R.M. 2002. The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. Nature **419**: 74–77. Nature Publishing Group. doi:10.1038/nature00954.

Erskine, W. 1996. Seed-size effects on lentil (Lens culinaris) yield potential and adaptation to temperature and rainfall in West Asia. J. Agric. Sci. [Online] Available: https://www.cambridge.org/core/journals/journal-of-agricultural-science/article/seedsize-

effects-on-lentil-lens-culinaris-yield-potential-and-adaptation-to-temperature-and-rainfall-in-west-asia/BB45F7B720DAB0F5F6DBFC12D2C6C07D [2018 May 16].

Erskine, W., Ellis, R.H.H., Summerfield, R.J.J., Roberts, E.H.H., and Hussain, A. 1990a. Characterization of responses to temperature and photoperiod for time to flowering in a world lentil collection. Theor. Appl. Genet. **80**: 193–199. doi:10.1007/BF00224386.

Erskine, W., Hussain, A., Tahir, M., Bahksh, A., Ellis, R.H.R., Summerfield, R.R.J., and Roberts, E.E.H. 1994. Field evaluation of a model of photothermal flowering responses in a world lentil collection. Theor. Appl. Genet. **88**: 423–428. doi:10.1007/BF00223655.

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

Erskine, W., Muehlbauer, F.J., and Short, R.W. 1990b. Stages of Development in Lentil. Exp. Agric. **26**: 297–302. doi:10.1017/S0014479700018457.

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

Eshraghi-Nejad, M., Bakhshandeh, A., Gharineh, M.H., and Soltani, A. 2015. Prediction of Spring Barley Flowering Time Based on Multiplicative Approach of Temperature × Photoperiod. Int. J. Agric. Biosci. **4**: 21–26. [Online] Available: www.ijagbio.com [2018 Nov. 6].

FAO 2010. The State of The World's Plant Genetic Resources for Food and Agriculture. FAO, Rome. [Online] Available: http://www.fao.org/docrep/013/i1500e/i1500e00.htm.

FAOSTAT 2018. FAOSTAT Statistics Database. [Online] Available: http://www.fao.org/faostat/en/?#data/QC [2018 Apr. 18].

Ferguson, M., and Erskine, W. 2001. Lentils (Lens L.). Pages 125–133 *in*. Springer, Dordrecht. doi:10.1007/978-94-015-9823-1_7.

Ferguson, M.E., and Robertson, L.D. 1999. Morphological and phenological variation in the wild relatives of lentil. Genetic Resources and Crop Evolution. Kluwer Academic Publishers. doi:10.1023/A:1008645029658.

Ferguson, M.E., Robertson, L.D., Ford-Lloyd, B. V., and Newbury, H.J. 1998. Contrasting

Genetic Variation amongst LentilLlandraces from Different Geographical Origins.

Gauch, H.G. 1988. Model Selection and Validation for Yield Trials with Interaction. Biometrics **44**: 705. doi:10.2307/2531585.

Hadley, P., Robert, E.H., Summerfield, R.J., and Minchin, F.R. 1984a. Effects of temperature and photoperiod on flowering in soya bean (Glycine max (L.) Merill) : a quantitative model. Ann. Bot. **53**: 669–681.

Hadley, P., Roberts, E.: H., Summerfield, R.J., and Minchinf, A.R. 1984b. Effects of Temperature and Photoperiod on Flowering in Soya bean [Glycine max (L.) Merrill] a Quantitative Model. Annals of Botany. [Online] Available: https://academic.oup.com/aob/article-abstract/53/5/669/157561 [2019 Jan. 9].

Harlan, J.R., and Harlan, J.R. 1992a. Views on Agricultural Origins. Pages 29–60 *in* Crops & Man. American Society of Agronomy, Crop Science Society of America. doi:10.2135/1992.cropsandman.c2.

Harlan, J.R., and Harlan, J.R. 1992b. The Dynamics of Domestication. Pages 115–133 *in* Crops & Man. American Society of Agronomy, Crop Science Society of America. doi:10.2135/1992.cropsandman.c6.

Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. Scand. J. Stat. **6**: 65–70. [Online] Available: https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf [2018 Oct. 23].

Huang, C., Nie, X., Shen, C., You, C., Li, W., Zhao, W., Zhang, X., and Lin, Z. 2017. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. Plant Biotechnol. J. **15**: 1374–1386. Wiley-Blackwell. doi:10.1111/pbi.12722.

Iannucci, A., Terribile, M.R.R., and Martiniello, P. 2008. Effects of temperature and photoperiod on flowering time of forage legumes in a Mediterranean environment. F. Crop. Res. **106**: 156–162. doi:10.1016/j.fcr.2007.11.005.

Ishikawa, M., Kiba, T., and Chua, N.-H. 2006. The Arabidopsis *SPA1* gene is required for circadian clock function and photoperiodic flowering. Plant J. **46**: 736–746. Wiley/Blackwell (10.1111). doi:10.1111/j.1365-313X.2006.02737.x.

Jung, C.-H., Wong, C.E., Singh, M.B., and Bhalla, P.L. 2012. Comparative Genomic Analysis of

Soybean Flowering Genes. PLoS One **7**: e38250. Public Library of Science. doi:10.1371/journal.pone.0038250.

Khazaei, H., Caron, C.T., Fedoruk, M., Diapari, M., Vandenberg, A., Coyne, C.J., McGee, R., and Bett, K.E. 2016. Genetic Diversity of Cultivated Lentil (Lens culinaris Medik.) and Its Relation to the World's Agro-ecological Zones. Front. Plant Sci. **7**: 1093. Frontiers. doi:10.3389/fpls.2016.01093.

Kikis, E.A., Khanna, R., and Quail, P.H. 2005. ELF4 is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY. Plant J. **44**: 300–313. Wiley/Blackwell (10.1111). doi:10.1111/j.1365-313X.2005.02531.x.

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B.M., Watanabe, S., Yamada, T., Tabata, S., Kanazawa, A., Harada, K., and Abe, J. 2010. Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. Plant Physiol. **154**: 1220–31. American Society of Plant Biologists. doi:10.1104/pp.110.160796.

Korte, A., and Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods **9**: 1. Plant Methods. doi:10.1186/1746-4811-9-29.

Kumar, J., Gupta, S., Biradar, R.S., Gupta, P., Dubey, S., and Singh, N.P. 2018. Association of functional markers with flowering time in lentil. J. Appl. Genet. **59**: 9–21. Springer Berlin Heidelberg. doi:10.1007/s13353-017-0419-0.

Kumar J, Kant R, Kumar S, Basu PS, Sarker A, and Singh NP 2016. Heat Tolerance in Lentil under Field Conditions. Legume Genomics Genet. **7**: 1–11. doi:10.5376/lgg.2016.07.0001.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. **82**: 1–26. doi:10.18637/jss.v082.i13.

Ladizinsky, G. 1979. The origin of lentil and its wild genepool. Euphytica **28**: 179–187. Kluwer Academic Publishers. doi:10.1007/BF00029189.

Laubinger, S., Marchal, V., Le Gourrierec, J., Wenkel, S., Adrian, J., Jang, S., Kulajta, C., Braun, H., Coupland, G., Hoecker, U., and Gentilhomme, J. 2006. Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Dev. Dev. Dev. **133**: 3213–3222. doi:10.1242/dev.02661.

Laurie, R.E., Diwadkar, P., Jaudal, M., Zhang, L., Hecht, V., Wen, J., Tadege, M., Mysore, K.S., Putterill, J., Weller, J.L., and Macknight, R.C. 2011. The Medicago FLOWERING LOCUS

T Homolog, MtFTa1, Is a Key Regulator of Flowering Time. PLANT Physiol. **156**: 2207–2224. doi:10.1104/pp.111.180182.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., Liu, J., Warburton, M.L., Cheng, Y., Hao, X., Zhang, P., Zhao, J., Liu, Y., Wang, G., Li, J., and Yan, J. 2012. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. **45**: 43. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. [Online] Available: http://dx.doi.org/10.1038/ng.2484.

Li, L., Long, Y., Zhang, L., Dalton-Morgan, J., Batley, J., Yu, L., Meng, J., and Li, M. 2015. Genome Wide Analysis of Flowering Time Trait in Multiple Environments via High-Throughput Genotyping Technique in Brassica napus L. PLoS One **10**: e0119425. Public Library of Science. doi:10.1371/journal.pone.0119425.

Liew, L.C., Hecht, V., Laurie, R.E., Knowles, C.L., Vander Schoor, J.K., Macknight, R.C., and Weller, J.L. 2009. DIE NEUTRALIS and LATE BLOOMER 1 contribute to regulation of the pea circadian clock. Plant Cell **21**: 3198–211. American Society of Plant Biologists. doi:10.1105/tpc.109.067223.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics **28**: 2397–2399. Oxford University Press. doi:10.1093/bioinformatics/bts444.

Liu, S., Fan, C., Li, J., Cai, G., Yang, Q., Wu, J., Yi, X., Zhang, C., and Zhou, Y. 2016a. A genome-wide association study reveals novel elite allelic variations in seed oil content of Brassica napus. Theor. Appl. Genet. **129**: 1203–1215. Springer Berlin Heidelberg. doi:10.1007/s00122-016-2697-z.

Liu, X., Huang, M., Fan, B., Buckler, E.S., and Zhang, Z. 2016b. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genet. **12**: e1005767. Public Library of Science. doi:10.1371/journal.pgen.1005767.

Marx, G.A. 1979. Some Photo-Dependent Responses in Pisum. I. Physiological Behavior. Physiol. Behav.: 273–276. [Online] Available: https://dl.sciencesocieties.org/publications/cs/abstracts/9/3/CS0090030273 [2018 Apr. 21].

Masle, J., Doussinault, G., Farquhar, G.D., and Sun, B. 1989. Foliar stage in wheat correlates better to photothermal time than to thermal time. Plant, Cell Environ. **12**: 235–247. John Wiley & Sons, Ltd (10.1111). doi:10.1111/j.1365-3040.1989.tb01938.x.

Mendiburu, F. De 2017. Package ' agricolae .' [Online] Available: http://tarwi.lamolina.edu.pe/~fmendiburu.

Miao, C., Yang, J., and Schnable, J.C. 2018. Optimizing the identification of causal variants across varying genetic architectures in crops. Plant Biotechnol. J. Wiley/Blackwell (10.1111). doi:10.1111/pbi.13023.

Nascimento, M., Nascimento, A.C.C., Silva, F.F. e, Barili, L.D., Vale, N.M. do, Carneiro, J.E., Cruz, C.D., Carneiro, P.C.S., and Serão, N.V.L. 2018. Quantile regression for genome-wide association study of flowering time-related traits in common bean. PLoS One **13**: e0190303. Public Library of Science. doi:10.1371/journal.pone.0190303.

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

Putterill, J., Laurie, R., and Macknight, R. 2004. It's time to flower: the genetic control of flowering time. BioEssays **26**: 363–373. Wiley-Blackwell. doi:10.1002/bies.20021.

Raj, A., Stephens, M., and Pritchard, J.K. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics **197**: 573–89. Genetics. doi:10.1534/genetics.114.164350.

Rajandran, V. 2016. Genetic control of flowering time in lentil. University of Tasmania. [Online] Available: https://eprints.utas.edu.au/23513/1/Rajandran_whole_thesis.pdf [2018 Nov. 19].

Roberts, E., Summerfield, R., Ellis, R., and Qi, A. 1993. Adaptation of Flowering in Crops to Climate. OutloOk Agric. **22**: 105–110. [Online] Available: https://journals.sagepub.com/doi/pdf/10.1177/003072709302200207 [2018 Nov. 8].

Roberts, E., Summerfield, R., Muehlbauer, F., and Short, R. 1986. Flowering in Lentil (Lens culinaris Medic.): The Duration of the Photoperiodic Inductive Phase as a Function of Accumulated Daylength above the Critical Photoperiod. Ann. Bot. **58**: 235–248. Oxford University Press. doi:10.1093/oxfordjournals.aob.a087201.

Roberts, E.H., Hadley, P., and Summerfield, R.J. 1985. Effects of Temperature and Photoperiod

on Flowering in Chickpeas (Cicer arietinum L.). Ann. Bot. **55**: 881–892. Oxford University Press. doi:10.1093/oxfordjournals.aob.a086969.

Roberts, E.H., Summerfield, R.J., Ellis, R.H., and Stewart, K.A. 1988. Photothermal Time for Flowering in Lentils (Lens culinaris) and the Analysis of Potential Vernalization Responses. Ann. Bot. **61**: 29–39. Oxford University Press. doi:10.1093/oxfordjournals.aob.a087525.

Romero Navarro, J.A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., Preciado, E., Terron, A., Delgado, H.V., Vidal, V., Ortega, A., Banda, A.E., Montiel, N.O.G., Ortiz-Monasterio, I., Vicente, F.S., Espinoza, A.G., Atlin, G., Wenzl, P., Hearne, S., and Buckler, E.S. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat. Genet. **49**: 476–480. Nature Publishing Group. doi:10.1038/ng.3784.

RStudio Team 2016. RStudio: Integrated Development for R. [Online] Available: https://www.rstudio.com/ [2018 Jul. 1].

Russell, L., Henrik, S., Love, J., Buerkner, P., and Herve, M. 2018. emmeans: Estimated Marginal Means, aka Least-Squares Means. [Online] Available: https://cran.r-project.org/web/packages/emmeans/index.html [2018 Jul. 1].

Salunkhe, D., and Kadam, S. 1989. Crc handbook of world food legumes: nutritional chemistry, processing technology, and utilization. [Online] Available: http://agris.fao.org/agris-search/search.do?recordID=XF2016047208 [2018 Apr. 22].

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

Sarker, A., Erskine, W., Sharma, B., and Tyagi, M.C. 1999. Inheritance and linkage relationship of days to flower and morphological loci in lentil (Lens culinaris Medikus subsp. culinaris). J. Hered. **90**: 270–275. doi:10.1093/jhered/90.2.270.

Saskatchewan Pulse Growers 2000. Pulse Production Manual : 7. Lentil. [Online] Available: https://www.ndsu.edu/pubweb/pulse-info/resources-pdf/Lentil production guide from Saskatchewan.pdf [2018 Dec. 13].

Scherer, A. 2017. GWAS. Third. Golden Helix, Inc. [Online] Available: http://goldenhelix.com/resources/ebooks/GWAS.html.

Shrestha, R., Siddique, K.H.M., Turner, N.C., Turner, D.W., and Berger, J.D. 2005. Growth and seed yield of lentil (Lens culinaris Medikus) genotypes of West Asian and South Asian

origin and crossbreds between the two under rainfed conditions in Nepal. Aust. J. Agric. Res. **56**: 971–981. doi:10.1071/AR05050.

Sonnante, G., Hammer, K., and Pignone, D. 2009. From the cradle of agriculture a handful of lentils: History of domestication. Rend. LINCEI **20**: 21–37. Springer Milan. doi:10.1007/s12210-009-0002-7.

Springate, D.A., and Kover, P.X. 2014. Plant responses to elevated temperatures: A field study on phenological sensitivity and fitness responses to simulated climate warming. Glob. Chang. Biol. **20**: 456–465. doi:10.1111/gcb.12430.

Summerfield, R.J., Collinson, S.T., Ellis, R.H., Roberts, E.H., and Vries, F.W.T.P.D.E. 1992. Photothermal Responses of Flowering in Rice (Oryza sativa). Ann. Bot. **69**: 101–112.

Summerfield, R.J., and Roberts, E.H. 1988. Photo-thermal regulation of flowering in pea, lentil, faba bean and chickpea. Pages 911–922 *in*. Springer, Dordrecht. doi:10.1007/978-94-009-2764-3_72.

Summerfield, R.J., Roberts, E.H., Erskine, W., and Ellis, R.H. 1985. Effects of Temperature and Photoperiod on Flowering in Lentils (Lens culinaris Medic.). Ann. Bot. **56**: 659–671. doi:10.1093/oxfordjournals.aob.a087055.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Su, Z., Pan, Y., Liu, D., Lipka, A.E., Buckler, E.S., and Zhang, Z. 2016. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. Plant Genome **9**: 0. doi:10.3835/plantgenome2015.11.0120.

Tullu, A., Diederichsen, A., Suvorova, G., and Vandenberg, A. 2011. Genetic and genomic resources of lentil: status, use and prospects. Plant Genet. Resour. **9**: 19–29. Cambridge University Press. doi:10.1017/S1479262110000353.

Turck, F., Fornara, F., and Coupland, G. 2008. Regulation and Identity of Florigen: FLOWERING LOCUS T Moves Center Stage. Annu. Rev. Plant Biol. **59**: 573–594. Annual Reviews . doi:10.1146/annurev.arplant.59.032607.092755.

Turner, S.D. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv: 005165. Cold Spring Harbor Laboratory. doi:10.1101/005165.

USDA 2016. USDA National Nutrient Database for Standard Reference. [Online] Available: https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-

research-center/nutrient-data-laboratory/docs/usda-national-nutrient-database-for-standard-reference/ [2016 Mar. 10].

Weller, J.L., Hecht, V., Liew, L.C., Sussmilch, F.C., Wenden, B., Knowles, C.L., and Vander Schoor, J.K. 2009. Update on the genetic control of flowering in garden pea. J. Exp. Bot. **60**: 2493–2499. Oxford University Press. doi:10.1093/jxb/erp120.

Weller, J.L., Liew, L.C., Hecht, V.F.G., Rajandran, V., Laurie, R.E., Ridge, S., Wenden, B., Schoor, J.K. Vander, Jaminon, O., Blassiau, C., Dalmais, M., Rameau, C., Bendahmane, A., Macknight, R.C., and Lejeune-Hénaut, I. 2012. A conserved molecular basis for photoperiod adaptation in two temperate legumes. Proc. Natl. Acad. Sci. U. S. A. **109**: 21158–63. doi:10.1073/pnas.1207943110.

Weller, J.L., and Ortega, R. 2015. Genetic control of flowering time in legumes. Front. Plant Sci. **6**: 1–13. doi: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 [2018 Jul. 1].

Wong, M.M.L., Gujaria-Verma, N., Ramsay, L., Yuan, H.Y., Caron, C., Diapari, M., Vandenberg, A., and Bett, K.E. 2015. Classification and characterization of species within the genus lens using genotyping-by-sequencing (GBS). PLoS One **10**: 1–16. Public Library of Science. doi:10.1371/journal.pone.0122025.

Yamashino, T., Yamawaki, S., Hagui, E., Ueoka-Nakanishi, H., Nakamichi, N., Ito, S., and Mizuno, T. 2013. Clock-Controlled and FLOWERING LOCUS T (FT)-Dependent Photoperiodic Pathway in Lotus japonicus I: Verification of the Flowering-Associated Function of an FT Homolog. Biosci. Biotechnol. Biochem. **77**: 747–753. doi:10.1271/bbb.120871.

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W., Liu, J., Li, J., and Yan, J. 2014. Genome Wide Association Studies Using a New Nonparametric Model Reveal the Genetic Architecture of 17 Agronomic Traits in an Enlarged Maize Association Panel. PLoS Genet. **10**: e1004573. Public Library of Science. doi:10.1371/journal.pgen.1004573.

Yi, N., Xu, S., Lou, X.-Y., and Mallick, H. 2014. Multiple comparisons in genetic association studies: a hierarchical modeling approach. Stat. Appl. Genet. Mol. Biol. **13**: 35–48. NIH Public Access. doi:10.1515/sagmb-2012-0040.

Yin, X., Kropff, M.J., Horie, T., Nakagawa, H., Centeno, H.G.S., Zhu, D., and Goudriaan, J. 1997. A model for photothermal responses of flowering in rice I . Model description and parameterization. **51**: 189–200.

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., and Buckler, E.S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**: 203–8. doi:10.1038/ng1702.

Yuan, H.Y., Saha, S., Vandenberg, A., and Bett, K.E. 2017. Flowering and Growth Responses of Cultivated Lentil and Wild Lens Germplasm toward the Differences in Red to Far-Red Ratio and Photosynthetically Active Radiation. **8**: 1–10. doi:10.3389/fpls.2017.00386.

Zhang, J., Song, Q., Cregan, P.B., Nelson, R.L., Wang, X., Wu, J., and Jiang, G.-L. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. BMC Genomics **16**: 217. BioMed Central. doi:10.1186/s12864-015-1441-4.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **42**: 355–360. Nature Publishing Group. doi:10.1038/ng.546.

A.1. Chapter 3 Supplementary Tables and Figures

Table A.1.1 List[2] of the genotypes used for the field experiments along with the information about the missing DTF data in certain site-years.

Red colored cells represent specific replications (indicated as Rep) in respective site-years which did not flower in Bhopal 2016, Jessore 2016, Bardiya 2016 and Cordoba 2016. Yellow colored cells represent specific replications in Bardiya 2016 and Rosthern 2016 which were missed to seeding or damaged before flowering due to some other reasons. Other site-years, namely Metaponto 2016, Rabat 2016, Rosthern 2017, Sutherland 2016 & 2017, where all genotypes flowered are not included in the table. In total 901 observations had missing DTF data out of a total of 9720 plots.

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
3156-11 AGL	■	■	■												
CDC Asterix	■	■	■	■	■	■									
CDC Cherie	■	■	■												
CDC Glamis	■	■	■												
CDC Gold	■	■	■	■	■	■	■								
CDC Greenstar	■	■	■												
CDC Imax		■	■	■	■	■									
CDC Impower	■	■	■												
CDC KR-1															
CDC LeMay			■												
CDC Maxim		■	■												
CDC QG-1		■	■				■								

---

[2]The list of genotypes used in the field experiments, along with their origin, is also available at
http://knowpulse.usask.ca/portal/project/AGILE%3A-Application-of-Genomic-Innovation-in-the-Lentil-Economy?pane=germplasm&page=0

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
CDC Red Rider	█	█	█												
CDC Redcoat	█	█	█	█	█	█	█								
CDC Redwing	█	█	█	█	█	█									
CDC Robin	█		█				█								
CDC Rosebud															
CDC Rosetown	█	█	█		█										
CDC Rosie			█												
CDC Rouleau	█		█												
CDC Royale		█	█												
CDC Ruby															
CDC SB-1	█		█												
CDC Sedley			█												
CDC Vantage	█		█												
CN 105604															
CN 105605															
CN 105715	█														
CN 105732	█		█												
CN 105767			█												
CN 105777	█														
CN 105789															
CN 105791															
CN 105862	█	█	█	█	█	█									
CN 105863	█	█	█	█	█	█									
CN 105864	█	█	█	█	█	█									
CN 105865	█	█	█	█	█	█									
CN 105866	█	█	█	█	█	█				█					
CN 105895															
CN 106265	█	█	█												
CN 108369	█	█	█												

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
CN 108370	■	■	■												
Crimson					■										
DPL 62															
Eston			■												
Gudo	■		■												
IG 1046	■	■	■												
IG 1706	■	■													
IG 1959			■		■										
IG 4258															
IG 4781	■	■	■	■	■	■	■		■		■				
IG 858			■		■										
ILL 10657	■														
ILL 10748															
ILL 11547															
ILL 11548															
ILL 11553		■													
ILL 11555															
ILL 11557															
ILL 11558	■														
ILL 1220	■	■	■												
ILL 1553		■	■												
ILL 1762	■	■	■	■	■	■									
ILL 1983			■												
ILL 213	■	■	■					■							
ILL 2194	■	■	■												
ILL 2507															
ILL 2580	■	■	■												
ILL 28	■	■	■	■	■	■									
ILL 3025			■												

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
ILL 313	■	■	■												
ILL 3347			■												
ILL 358	■	■													
ILL 3597			■												
ILL 4164	■	■	■	■	■	■									
ILL 4400	■	■	■	■	■	■									
ILL 4605			■												
ILL 4609	■	■	■	■	■	■									
ILL 4665	■		■												
ILL 4671	■	■	■	■	■	■			■						
ILL 4768															
ILL 4782															
ILL 4783	■	■	■		■										
ILL 4804		■	■												
ILL 4875	■	■	■	■	■	■									
ILL 4956		■	■												
ILL 5058	■	■	■	■	■	■					■	■			
ILL 5151	■	■	■	■	■	■									
ILL 5209		■	■												
ILL 5480		■	■												
ILL 5639															
ILL 5722			■												
ILL 5883	■		■												
ILL 5888															
ILL 5945															
ILL 6002															
ILL 618	■		■												
ILL 6182	■	■	■		■										
ILL 6211															

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
ILL 624			■				■								
ILL 6821	■	■	■	■	■	■									
ILL 6853	■	■	■	■	■										
ILL 7089	■	■	■			■	■								
ILL 7558															
ILL 7663															
ILL 7668															
ILL 7716															
ILL 7747	■	■	■	■	■	■									
ILL 7946	■	■	■												
ILL 7978															
ILL 7979															
ILL 8007	■														
ILL 8009															
ILL 8010					■										
ILL 8072	■	■	■	■	■	■									
ILL 8174			■		■										
ILL 8595															
ILL 9		■			■										
ILL 975	■	■	■		■										
ILL 9888	■		■			■									
ILL 9932															
ILL 9945		■	■												
ILL 9977															
ILL 9997															
ILWL 118															
Indianhead	■	■	■	■	■	■									
IPL 220						■									
Laird	■		■												

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
LIRL-22-46															
PI 163589	■	■	■	■	■	■	■								
PI 169534 LSP	■	■	■												
PI 175754 LSP	■	■	■												
PI 177430 LSP	■	■	■	■	■	■									
PI 178939 LSP	■	■	■												
PI 178952		■	■												
PI 178971 LSP	■	■	■												
PI 179324 LSP	■	■	■	■	■	■			■						
PI 179330	■		■												
PI 181771 LSP	■	■	■												
PI 181886 LSP	■	■	■	■	■	■									
PI 182217 LSP	■		■		■										
PI 193546	■	■	■	■	■	■									
PI 193547 LSP	■	■	■		■										
PI 193548 LSP															
PI 193550															
PI 207492 LSP	■	■	■		■										
PI 209858 LSP	■	■		■		■									
PI 211052 LSP	■	■	■	■	■	■									
PI 212100 LSP	■	■	■	■	■	■									
PI 212610 LSP	■	■	■												
PI 217949 LSP		■	■												
PI 238758 LSP			■												
PI 244046															
PI 250156 LSP	■	■	■			■									
PI 250158 LSP															
PI 251248 LSP															
PI 273664 LSP	■	■	■	■	■	■									

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
PI 280686			R												
PI 283604 LSP	R	R	R	R	R	R			R						
PI 289066	R	R	R	R		R			R						
PI 289073 LSP		R	R												
PI 289079 LSP	R		R				O	O	O						
PI 290716 LSP	R		R		R										
PI 297285 LSP		R	R												
PI 297287	R	R	R			R									
PI 297754 LSP	R	R	R												
PI 297767	R	R	R	R	R	R									
PI 297772 LSP	R	R	R	R	R	R									
PI 297783		R													
PI 297787 LSP		R	R												
PI 298023 LSP		R	R												
PI 298121 LSP			R												
PI 298122 LSP	R	R	R												
PI 298357 LSP															
PI 298631 LSP	R		R									R			
PI 298644 LSP	R	R	R												
PI 298645		R	R												
PI 298921	R	R	R	R	R	R	R		R						
PI 298922 LSP	R	R	R						R			R			
PI 298923 LSP	R	R	R									R			
PI 299116 LSP		R	R						R						
PI 299120 LSP															
PI 299121 LSP	R														
PI 299126 LSP	R	R	R		R										
PI 299163 LSP	R	R	R												
PI 299164 LSP			R												

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
PI 299165	■		■	■	■	■									
PI 299177 LSP	■	■	■	■	■	■									
PI 299237 LSP			■												
PI 299289			■												
PI 299312	■		■												
PI 299345	■	■	■	■	■	■									
PI 299351 LSP			■												
PI 299366 LSP	■			■	■	■									
PI 299375 LSP	■	■	■	■	■	■									
PI 300250 LSP	■			■	■	■			■						
PI 302398 LSP															
PI 308614 LSP	■	■	■	■	■	■		■							
PI 311107 LSP	■	■	■	■	■	■									
PI 312175 LSP	■	■	■	■	■	■									
PI 320935 LSP	■	■	■												
PI 320936 LSP															
PI 320937 LSP		■	■												
PI 320941 LSP	■	■	■	■	■	■			■						
PI 320945 LSP	■	■	■	■	■	■									
PI 320946 LSP	■	■	■	■		■									
PI 320952 LSP	■	■	■												
PI 320953 LSP	■	■	■	■	■	■									
PI 320954 LSP	■	■	■												
PI 329157 LSP			■												
PI 329167 LSP	■		■												
PI 339266 LSP		■	■												
PI 339283 LSP	■	■	■	■	■	■									
PI 339285	■	■	■	■	■	■		■	■						
PI 339289 LSP	■		■		■										

Note: Red-filled cells are marked with ■; empty cells are blank.

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
PI 339292 LSP	■	■													
PI 339296			■												
PI 343026 LSP		■	■												
PI 345627 LSP			■												
PI 358602 LSP	■	■	■												
PI 368647 LSP	■		■												
PI 368651 LSP		■	■												
PI 370481 LSP	■	■	■	■	■	■	■								
PI 374116 LSP	■	■	■												
PI 374117 LSP	■	■	■												
PI 374118		■	■	■	■	■									
PI 374120 LSP	■	■	■		■										
PI 374121			■												
PI 379368 LSP			■												
PI 383682 LSP	■	■	■	■	■	■									
PI 420818	■	■	■	■	■	■									
PI 420924 LSP	■	■	■	■	■	■									
PI 420925 LSP	■	■	■	■	■	■									
PI 420929 LSP	■	■	■	■	■	■									
PI 426202 LSP	■	■	■	■		■									
PI 426778 LSP	■	■													
PI 426784 LSP															
PI 426797 LSP															
PI 426807 LSP	■	■	■												
PI 431618 LSP			■												
PI 431622 LSP	■	■	■												
PI 431630 LSP		■													
PI 431633 LSP			■		■										
PI 431636 LSP		■	■												

89

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
PI 431643 LSP	■	■	■												
PI 431662 LSP	■	■	■												
PI 431663 LSP	■	■	■												
PI 431679 LSP	■	■	■												
PI 431684 LSP	■	■	■	■	■	■									
PI 431705 LSP	■	■	■												
PI 431710 LSP	■	■	■												
PI 431714 LSP	■		■												
PI 431717 LSP	■	■	■												
PI 431728 LSP		■	■												
PI 431731 LSP		■	■												
PI 431739 LSP															
PI 431753 LSP			■												
PI 431756 LSP		■	■									■			
PI 431809 LSP	■	■	■		■										
PI 431824 LSP	■	■	■												
PI 431863 LSP	■	■	■												
PI 431873 LSP	■	■	■		■										
PI 431884 LSP	■	■	■						■						
PI 431888 LSP	■	■	■	■	■	■									
PI 431893 LSP															
PI 431923 LSP	■		■		■										
PI 431980 LSP	■	■	■												
PI 432000 LSP	■	■	■	■	■	■						■			
PI 432001 LSP	■	■	■		■										
PI 432002 LSP	■	■	■												
PI 432005 LSP	■	■	■												
PI 432028 LSP	■										■				
PI 432033 LSP	■	■	■						■						

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
PI 432085 LSP	■	■	■					■							
PI 432106 LSP			■		■										
PI 432124 LSP	■	■	■												
PI 432145 LSP	■	■	■												
PI 432147 LSP		■	■		■										
PI 432184 LSP		■	■												
PI 432188 LSP			■												
PI 432190 LSP			■												
PI 432201 LSP			■												
PI 432212 LSP		■	■		■										
PI 432236 LSP	■	■	■						■						
PI 432245 LSP		■	■												
PI 432271 LSP	■	■	■												
PI 432286 LSP	■	■	■	■	■	■	■	■		■					
PI 451763 LSP															
PI 458503 LSP			■												
PI 468900 LSP	■	■	■		■										
PI 468901	■	■	■	■	■	■							▨		
PI 468902 LSP		■	■												
PI 472136 LSP															
PI 472205 LSP															
PI 472213 LSP			■												
PI 472327 LSP															
PI 472380 LSP															
PI 472416 LSP															
PI 472488 LSP															
PI 472559 LSP	■	■	■												
PI 472561 LSP															
PI 472569 LSP															

Genotype	Bhopal 2016			Jessore 2016			Bardiya 2016			Cordoba 2016			Rosthern 2016		
	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII	RepI	RepII	RepIII
PI 472588 LSP	■	■	■												
PI 472590 LSP	■	■	■	■		■									
PI 472615 LSP			■												
PI 472629 LSP			■												
PI 477921 LSP					■										
PI 490288 LSP			■												
PI 490289 LSP	■	■	■												
PI 508090			■												
PI 518731 LSP	■	■	■		■										
PI 518733 LSP	■	■	■		■										
PI 518734 LSP	■		■												
PI 533688 LSP	■	■	■	■	■	■									
PI 533693 LSP	■	■	■												
PI 612875	■	■	■	■	■	■									
PI 643451															
PI 643452															
Shasta		■					■		■						
W6 27754 LSP		■	■		■				■						
W6 27760 LSP			■												
W6 27763 LSP	■	■	■	■	■	■						■			
W6 27766 LSP															
W6 27767 LSP	■	■	■	■	■	■									

Table A.1.2. AMMI Stability (ASV) Rank of the genotypes with average DTF when the analysis was conducted using all site-years data.

Genotype	AMMI Stability Rank	Average DTF (from all site-year)	Genotype	AMMI Stability Rank	Average DTF (from all site-year)
ILL 5888	324	56	PI 374120 LSP	161	83
ILL 7663	323	60	CDC Vantage	160	81
LIRL-22-46	322	65	PI 179324 LSP	159	86
ILL 7716	321	69	PI 177430 LSP	158	87
ILL 4605	320	69	PI 432106 LSP	157	84
CN 105791	319	66	PI 169534 LSP	156	85
ILL 6002	318	68	ILL 3597	155	75
PI 251248 LSP	317	64	CN 105732	154	82
PI 244046	316	65	PI 339292 LSP	153	83
ILL 7558	315	71	PI 432271 LSP	152	89
ILL 6211	314	71	PI 299345	151	85
PI 431630 LSP	313	91	PI 472615 LSP	150	76
PI 472205 LSP	312	68	PI 320937 LSP	149	86
CN 105777	311	67	PI 299177 LSP	148	85
ILL 10748	310	71	ILWL 118	147	79
ILL 8007	309	67	PI 426807 LSP	146	82
PI 432002 LSP	308	91	PI 472629 LSP	145	81
ILL 11558	307	68	PI 298357 LSP	144	82
ILL 11557	306	69	Gudo	143	78
W6 27766 LSP	305	71	CDC Greenstar	142	84
PI 280686	304	70	PI 431636 LSP	141	83
PI 431873 LSP	303	89	PI 431714 LSP	140	85
ILL 11548	302	69	PI 175754 LSP	139	84
PI 431679 LSP	301	89	CDC QG-1	138	90
PI 472136 LSP	300	70	PI 468902 LSP	137	84
PI 320936 LSP	299	76	ILL 8595	136	78
IPL 220	298	69	PI 368647 LSP	135	87
ILL 11555	297	70	CN 105605	134	80
ILL 9997	296	71	ILL 213	133	82
PI 432145 LSP	295	91	PI 508090	132	79
PI 193550	294	69	ILL 8072	131	82

Genotype	AMMI Stability Rank	Average DTF (from all site-year)	Genotype	AMMI Stability Rank	Average DTF (from all site-year)
PI 250158 LSP	293	71	PI 432001 LSP	130	83
CN 105895	292	72	ILL 6182	129	78
PI 339289 LSP	291	85	IG 4781	128	89
ILL 11547	290	71	ILL 7089	127	82
ILL 8010	289	71	PI 533688 LSP	126	87
PI 320941 LSP	288	88	PI 290716 LSP	125	87
IG 1046	287	84	PI 432201 LSP	124	85
PI 472569 LSP	286	73	ILL 4164	123	79
ILL 2507	285	72	PI 468900 LSP	122	87
ILL 5945	284	71	PI 518731 LSP	121	85
ILL 7978	283	71	CDC Royale	120	84
PI 431888 LSP	282	87	ILL 358	119	81
PI 298922 LSP	281	92	PI 217949 LSP	118	79
PI 426784 LSP	280	73	ILL 4956	117	86
PI 472561 LSP	279	73	PI 368651 LSP	116	88
IG 1706	278	86	ILL 4804	115	81
PI 297783	277	69	PI 420925 LSP	114	82
ILL 10657	276	72	ILL 28	113	81
PI 320945 LSP	275	89	ILL 1553	112	81
ILL 7979	274	74	CDC Rouleau	111	88
ILL 11553	273	72	ILL 618	110	81
PI 432000 LSP	272	86	CDC Sedley	109	84
DPL 62	271	71	CDC Cherie	108	83
PI 432147 LSP	270	87	PI 432188 LSP	107	81
ILL 1762	269	88	PI 320953 LSP	106	92
CN 105789	268	71	PI 238758 LSP	105	82
PI 431684 LSP	267	86	ILL 4783	104	88
ILL 2194	266	84	PI 178939 LSP	103	86
PI 426202 LSP	265	88	CDC Red Rider	102	87
PI 193548 LSP	264	71	CDC KR-1	101	84
PI 339296	263	83	PI 374118	100	80
PI 273664 LSP	262	89	PI 299121 LSP	99	77
PI 432124 LSP	261	85	PI 431643 LSP	98	82

Genotype	AMMI Stability Rank	Average DTF (from all site-year)	Genotype	AMMI Stability Rank	Average DTF (from all site-year)
PI 298923 LSP	260	85	ILL 5639	97	83
PI 383682 LSP	259	86	PI 612875	96	77
PI 311107 LSP	258	88	CDC Asterix	95	88
ILL 5209	257	81	PI 178952	94	82
ILL 8009	256	71	ILL 2580	93	77
PI 299366 LSP	255	77	ILL 9	92	75
CN 105866	254	91	PI 298644 LSP	91	90
PI 432184 LSP	253	84	PI 289066	90	92
PI 643451	252	75	CDC Rosie	89	87
ILL 4768	251	72	PI 432190 LSP	88	84
PI 431717 LSP	250	87	CN 105864	87	85
PI 374117 LSP	249	85	PI 431731 LSP	86	84
PI 472488 LSP	248	73	PI 343026 LSP	85	82
CDC Ruby	247	78	PI 358602 LSP	84	87
CN 105863	246	87	PI 302398 LSP	83	79
ILL 3347	245	74	CN 105767	82	76
PI 432236 LSP	244	88	PI 207492 LSP	81	85
PI 339283 LSP	243	83	CN 108369	80	84
ILL 3025	242	72	CN 105715	79	79
ILL 1220	241	84	PI 432245 LSP	78	78
ILL 7946	240	81	ILL 5151	77	81
PI 339266 LSP	239	82	PI 312175 LSP	76	88
PI 472559 LSP	238	73	PI 299312	75	87
PI 432028 LSP	237	88	PI 533693 LSP	74	86
PI 472380 LSP	236	72	Shasta	73	82
ILL 7668	235	76	CDC Redwing	72	86
ILL 4782	234	74	PI 289079 LSP	71	89
PI 431923 LSP	233	88	CDC Rosetown	70	87
PI 431663 LSP	232	87	CDC Imax	69	85
Crimson	231	79	PI 490288 LSP	68	84
PI 431622 LSP	230	88	ILL 5480	67	87
PI 426797 LSP	229	73	PI 431893 LSP	66	82
ILL 5058	228	92	CDC SB-1	65	81

Genotype	AMMI Stability Rank	Average DTF (from all site-year)	Genotype	AMMI Stability Rank	Average DTF (from all site-year)
PI 431756 LSP	227	87	PI 297287	64	84
PI 643452	226	77	CDC Redcoat	63	86
IG 1959	225	85	ILL 9888	62	77
PI 431809 LSP	224	84	CDC Glamis	61	87
PI 211052 LSP	223	86	PI 379368 LSP	60	83
PI 299163 LSP	222	88	PI 178971 LSP	59	85
ILL 9945	221	73	PI 432085 LSP	58	80
PI 472416 LSP	220	76	PI 339285	57	84
IG 4258	219	74	Eston	56	81
PI 431753 LSP	218	85	PI 431739 LSP	55	82
PI 472213 LSP	217	84	PI 345627 LSP	54	88
W6 27760 LSP	216	79	PI 182217 LSP	53	82
CN 105604	215	76	PI 297285 LSP	52	84
PI 431980 LSP	214	83	PI 431618 LSP	51	82
PI 420924 LSP	213	82	PI 477921 LSP	50	80
ILL 4875	212	92	3156-11	49	90
IG 858	211	85	PI 298645	48	86
PI 431710 LSP	210	85	PI 299164 LSP	47	83
PI 420929 LSP	209	83	CDC Gold	46	87
ILL 9977	208	72	PI 299126 LSP	45	77
PI 472590 LSP	207	84	PI 374116 LSP	44	86
ILL 313	206	85	CN 106265	43	88
ILL 1983	205	76	ILL 624	42	85
W6 27767 LSP	204	82	PI 518734 LSP	41	88
PI 297772 LSP	203	86	PI 299120 LSP	40	79
PI 320935 LSP	202	83	PI 298921	39	81
PI 193546	201	84	PI 329157 LSP	38	81
ILL 6821	200	81	PI 298121 LSP	37	81
PI 472327 LSP	199	73	PI 212610 LSP	36	86
PI 209858 LSP	198	93	PI 320946 LSP	35	80
Laird	197	84	PI 250156 LSP	34	81
PI 420818	196	86	ILL 4665	33	84
CN 105862	195	90	PI 472588 LSP	32	83

Genotype	AMMI Stability Rank	Average DTF (from all site-year)	Genotype	AMMI Stability Rank	Average DTF (from all site-year)
PI 432033 LSP	194	86	PI 320954 LSP	31	85
PI 431728 LSP	193	85	CN 108370	30	89
PI 297754 LSP	192	86	PI 431633 LSP	29	81
CN 105865	191	88	CDC Maxim	28	84
PI 431863 LSP	190	87	PI 518733 LSP	27	88
ILL 5722	189	74	ILL 4400	26	83
PI 299165	188	85	PI 329167 LSP	25	82
ILL 4609	187	84	PI 193547 LSP	24	86
PI 431884 LSP	186	81	PI 299289	23	85
PI 298023 LSP	185	81	PI 432286 LSP	22	86
CDC Rosebud	184	82	PI 468901	21	90
PI 179330	183	86	CDC LeMay	20	85
ILL 5883	182	81	PI 181771 LSP	19	84
PI 451763 LSP	181	77	ILL 8174	18	82
PI 432005 LSP	180	85	W6 27754 LSP	17	88
PI 212100 LSP	179	87	PI 299116 LSP	16	81
PI 297767	178	89	PI 320952 LSP	15	91
PI 299375 LSP	177	85	PI 289073 LSP	14	86
PI 431662 LSP	176	85	CDC Robin	13	83
PI 431705 LSP	175	86	PI 299237 LSP	12	82
ILL 6853	174	83	PI 283604 LSP	11	85
PI 299351 LSP	173	85	PI 374121	10	83
PI 490289 LSP	172	85	ILL 4671	9	83
ILL 975	171	79	PI 458503 LSP	8	83
ILL 7747	170	83	Indianhead	7	92
W6 27763 LSP	169	81	PI 370481 LSP	6	85
PI 298631 LSP	168	91	PI 300250 LSP	5	82
PI 432212 LSP	167	84	PI 308614 LSP	4	84
PI 297787 LSP	166	84	CDC Impower	3	86
PI 181886 LSP	165	83	PI 298122 LSP	2	85
ILL 9932	164	76	PI 163589	1	84
PI 426778 LSP	163	85			
PI 431824 LSP	162	83			

Table A.1.3. List of genotypes along with the intercept (a), temperature (b) and photoperiod (c) coefficients, the critical photoperiod (-a/c) derived using the equation $1/f = a + b\overline{T} + cP$ using DTF and mean temperature from seeding to the days to flower, and photoperiod at a flowering day from all ten site-years.

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
3156-11	-0.027899215	0.000644778	0.002352056	11.86162654
CDC Asterix	-0.040590146	0.000671822	0.003271236	12.40819894
CDC Cherie	-0.03238888	0.00061319	0.002874926	11.26598672
CDC Glamis	-0.027655994	0.000655235	0.002364297	11.69734563
CDC Gold	-0.042541251	0.000550437	0.003542277	12.00957759
CDC Greenstar	-0.026085912	0.000622266	0.002347915	11.11024572
CDC Imax	-0.039415026	0.00054823	0.003387211	11.63642315
CDC Impower	-0.032827226	0.00085735	0.002561223	12.81701343
CDC KR-1	-0.023370825	0.000922154	0.001881609	12.4206605
CDC LeMay	-0.031794238	0.000804102	0.002586067	12.29443639
CDC Maxim	-0.03303667	0.000804502	0.002708008	12.19961944
CDC QG-1	-0.034169367	0.001126745	0.002306009	14.81753115
CDC Red Rider	-0.031536014	0.000601588	0.002734066	11.53447493
CDC Redcoat	-0.031693415	0.000368023	0.002955585	10.723229
CDC Redwing	-0.037736479	0.000357407	0.003425761	11.01550341
CDC Robin	-0.03521334	0.00085592	0.002798002	12.58517544
CDC Rosebud	-0.026461367	0.000984698	0.002076182	12.74520296
CDC Rosetown	-0.029447451	0.000490673	0.002656566	11.08478074
CDC Rosie	-0.026496131	0.000743747	0.002215825	11.95768214
CDC Rouleau	-0.033780482	0.000613367	0.002874242	11.75283025
CDC Royale	-0.027568674	0.0008565	0.002251373	12.2452743
CDC Ruby	-0.024713436	0.001147473	0.001872138	13.20064699
CDC SB-1	-0.043057547	0.001012492	0.003328257	12.93696636
CDC Sedley	-0.024988002	0.000738818	0.002175004	11.48871515
CDC Vantage	-0.028882055	0.000974503	0.002283997	12.64540173
CN 105604	-0.03526221	0.000884902	0.003062371	11.5146746
CN 105605	-0.048587086	0.001373962	0.003445333	14.10229036
CN 105715	-0.050827449	0.001082156	0.003926387	12.94509368
CN 105732	-0.035532161	0.001108247	0.002582536	13.75863177

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
CN 105767	-0.043462427	0.001272868	0.003255476	13.35056178
CN 105777	-0.032869454	0.001097342	0.002823663	11.64071389
CN 105789	-0.032454469	0.000730391	0.003165776	10.25166399
CN 105791	-0.025618414	0.001222632	0.00223854	11.44424886
CN 105862	-0.042416604	0.000550256	0.003489121	12.15681538
CN 105863	-0.043246874	0.000147308	0.004034674	10.71880165
CN 105864	-0.037722332	0.000347199	0.003455326	10.91715637
CN 105865	-0.043393636	0.000654898	0.003503536	12.38566745
CN 105866	-0.043218792	0.000387284	0.00368442	11.73014663
CN 105895	-0.02526432	0.000899488	0.002375925	10.63346741
CN 106265	-0.035971686	0.000744003	0.00286684	12.54750458
CN 108369	-0.032223395	0.00085467	0.002585603	12.46262129
CN 108370	-0.033139729	0.000845056	0.002549804	12.99697187
Crimson	-0.022542304	0.001166459	0.001682564	13.39759195
DPL 62	-0.029943363	0.000982274	0.002683739	11.15733002
Eston	-0.038311362	0.001092109	0.002861785	13.38722574
Gudo	-0.036671337	0.00060123	0.003393443	10.80652846
IG 1046	-0.051305565	0.001425692	0.003397665	15.1002445
IG 1706	-0.053245072	0.001310802	0.00366114	14.54330463
IG 1959	-0.043993974	0.001351025	0.002995552	14.68643426
IG 4258	-0.036324543	0.000761628	0.003346218	10.85540286
IG 4781	-0.045445866	0.000459517	0.003768304	12.06003297
IG 858	-0.045750735	0.001206954	0.003224496	14.18849216
ILL 10657	-0.030153504	0.000707101	0.002981579	10.11326715
ILL 10748	-0.0159786	0.000995058	0.001583377	10.09147152
ILL 11547	-0.027929801	0.000975316	0.002547455	10.96380334
ILL 11548	-0.028628511	0.001073623	0.002537955	11.28015108
ILL 11553	-0.029870017	0.000753932	0.002891033	10.33195345
ILL 11555	-0.026874435	0.001128386	0.002350371	11.43412434
ILL 11557	-0.028767838	0.001065126	0.002545149	11.303008
ILL 11558	-0.030526399	0.000947004	0.002811235	10.85871517
ILL 1220	-0.048189059	0.001086146	0.003517065	13.70149965
ILL 1553	-0.045558823	0.001051327	0.003462018	13.15961443
ILL 1762	-0.052720296	0.000428003	0.004374592	12.05147711

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
ILL 1983	-0.033837534	0.000856736	0.002968446	11.39907218
ILL 213	-0.051265545	0.000766488	0.00415931	12.32549267
ILL 2194	-0.044496257	0.000429995	0.003979607	11.18106746
ILL 2507	-0.030603624	0.000771746	0.00294348	10.3970879
ILL 2580	-0.04203401	0.000559087	0.003860557	10.88806986
ILL 28	-0.041891923	0.000324148	0.003902769	10.73389795
ILL 3025	-0.035535345	0.000799861	0.003259837	10.90095866
ILL 313	-0.045589031	0.000637658	0.003791615	12.02364389
ILL 3347	-0.033260509	0.00078108	0.003069566	10.83557317
ILL 358	-0.041420957	0.000691438	0.003555702	11.64916537
ILL 3597	-0.036538948	0.000828023	0.003276939	11.15032912
ILL 4164	-0.042274731	0.000212964	0.004126615	10.24440992
ILL 4400	-0.039962012	0.00060561	0.003400454	11.75196285
ILL 4605	-0.018447801	0.001240404	0.001560359	11.82279228
ILL 4609	-0.053049259	0.000401619	0.004594684	11.54578968
ILL 4665	-0.034428085	0.000889932	0.002714264	12.684132
ILL 4671	-0.038067234	0.000395864	0.003461816	10.9963187
ILL 4768	-0.034208521	0.000658753	0.003355749	10.1940041
ILL 4782	-0.035338458	0.000635147	0.003461613	10.20866874
ILL 4783	-0.037309627	0.000847934	0.002843328	13.12181386
ILL 4804	-0.042790558	0.000712113	0.003626153	11.80053995
ILL 4875	-0.045453706	0.000765684	0.003429002	13.25566794
ILL 4956	-0.039047015	0.00102148	0.002851369	13.69413003
ILL 5058	-0.043200787	0.000440246	0.003618375	11.93927886
ILL 5151	-0.041671865	0.00037583	0.003872254	10.7616547
ILL 5209	-0.038829732	0.000455515	0.003625771	10.70937315
ILL 5480	-0.036817705	0.000895873	0.00280305	13.13487148
ILL 5639	-0.037298181	0.001018515	0.002909793	12.81815689
ILL 5722	-0.03871257	0.001122484	0.003110379	12.44625414
ILL 5883	-0.039966132	0.000628105	0.003511595	11.38119191
ILL 5888	-0.013814209	0.002231389	0.000606265	22.78575321
ILL 5945	-0.030204865	0.000718247	0.002991331	10.09746524
ILL 6002	-0.020309087	0.001188954	0.001742142	11.65753737
ILL 618	-0.045860141	0.001217758	0.003310542	13.85275831

79

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
ILL 6182	-0.038738517	0.000485086	0.003624915	10.68673815
ILL 6211	-0.021045073	0.000966008	0.001957752	10.74961214
ILL 624	-0.038228828	0.000876188	0.003000414	12.74118438
ILL 6821	-0.043466621	4.33E-05	0.004318526	10.06515109
ILL 6853	-0.0357477	0.000314713	0.003466816	10.31139193
ILL 7089	-0.046854865	0.001069563	0.003519188	13.31411147
ILL 7558	-0.019676029	0.001132317	0.001708039	11.51966098
ILL 7663	-0.014786585	0.001739655	0.001027494	14.39091711
ILL 7668	-0.033396924	0.000911339	0.002870391	11.63497534
ILL 7716	-0.018147837	0.001163782	0.001592406	11.39649108
ILL 7747	-0.044761636	0.000129667	0.00430025	10.40907681
ILL 7946	-0.042558682	0.00032811	0.004048083	10.51329227
ILL 7978	-0.029131705	0.000888497	0.002753209	10.58099944
ILL 7979	-0.029457408	0.000807446	0.00276979	10.6352494
ILL 8007	-0.034241369	0.000992927	0.003160154	10.83534866
ILL 8009	-0.034989016	0.001092586	0.002975446	11.75925142
ILL 8010	-0.032006571	0.001121856	0.002691628	11.89115561
ILL 8072	-0.044129033	0.000378933	0.003991064	11.05695891
ILL 8174	-0.033088005	0.001249244	0.002363023	14.00240679
ILL 8595	-0.037441272	0.001345801	0.002630213	14.23507053
ILL 9	-0.038055095	0.000922241	0.003285346	11.58328445
ILL 975	-0.042991269	0.000121328	0.004293917	10.01213433
ILL 9888	-0.039198581	0.000605856	0.003588349	10.92384867
ILL 9932	-0.037126586	0.00080665	0.003317447	11.19131108
ILL 9945	-0.036138505	0.000553546	0.003564798	10.13760266
ILL 9977	-0.036622391	0.001047566	0.003132097	11.69260968
ILL 9997	-0.027039398	0.000847003	0.002620363	10.31895062
ILWL 118	-0.031116024	0.001064144	0.002467225	12.61175119
Indianhead	-0.032460293	0.000585642	0.002686136	12.08438102
IPL 220	-0.031320058	0.001065465	0.002790435	11.22407923
Laird	-0.023830375	0.000849929	0.001942585	12.26735266
LIRL-22-46	-0.018265236	0.001495442	0.001365378	13.37741583
PI 163589	-0.03755962	0.000372679	0.003445186	10.90205851
PI 169534 LSP	-0.040387536	0.001183642	0.002813866	14.35303999

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
PI 175754 LSP	-0.045006529	0.000855236	0.003530687	12.74724302
PI 177430 LSP	-0.043658115	0.000634316	0.003583795	12.18209214
PI 178939 LSP	-0.040615485	0.00100382	0.003000068	13.53818688
PI 178952	-0.040146346	0.001087687	0.003015103	13.31508313
PI 178971 LSP	-0.041140542	0.000939631	0.003115797	13.20385919
PI 179324 LSP	-0.054017972	0.000240936	0.004692838	11.51072548
PI 179330	-0.043326803	0.000878638	0.003346879	12.9454328
PI 181771 LSP	-0.041070472	0.000883453	0.003193943	12.85886096
PI 181886 LSP	-0.043754497	0.000374121	0.003958995	11.05191965
PI 182217 LSP	-0.042914174	0.000922476	0.003358752	12.77682077
PI 193546	-0.04643029	0.000283606	0.004220999	10.99983415
PI 193547 LSP	-0.038747143	0.000550966	0.003328429	11.64127176
PI 193548 LSP	-0.033614303	0.000706119	0.003269104	10.28242069
PI 193550	-0.033542966	0.000986364	0.003020951	11.1034446
PI 207492 LSP	-0.043033372	0.000847632	0.003353149	12.8337201
PI 209858 LSP	-0.028109223	0.000811132	0.002178662	12.90205878
PI 211052 LSP	-0.05129573	0.000910727	0.003833962	13.37930136
PI 212100 LSP	-0.048087774	0.000606209	0.003885317	12.37679603
PI 212610 LSP	-0.041864209	0.000797895	0.00331892	12.61380356
PI 217949 LSP	-0.047481333	0.000805206	0.00404751	11.7309981
PI 238758 LSP	-0.035266885	0.001140679	0.002571277	13.71570607
PI 244046	-0.032116616	0.001178775	0.002823884	11.37320618
PI 250156 LSP	-0.046646913	0.000658978	0.003955884	11.79178001
PI 250158 LSP	-0.030146516	0.000797082	0.002968901	10.15410013
PI 251248 LSP	-0.034058114	0.001278476	0.002957719	11.51499349
PI 273664 LSP	-0.052245201	0.000633504	0.004153138	12.57969417
PI 280686	-0.030001664	0.000981228	0.002652711	11.30981144
PI 283604 LSP	-0.039215517	0.000444883	0.003445855	11.38048901
PI 289066	-0.035398752	0.000592734	0.002901794	12.19891842
PI 289073 LSP	-0.034012062	0.000947606	0.002606396	13.04945872
PI 289079 LSP	-0.033465201	0.000948511	0.002501733	13.37680785
PI 290716 LSP	-0.040281223	0.000793073	0.003175727	12.68409646
PI 297285 LSP	-0.029260409	0.000865706	0.002386202	12.26233399
PI 297287	-0.041081591	0.000777252	0.003302692	12.43881859

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
PI 297754 LSP	-0.044397149	0.00111373	0.003145541	14.11431042
PI 297767	-0.040472124	0.000547148	0.003369676	12.0106858
PI 297772 LSP	-0.04339717	9.03E-05	0.004169277	10.4088002
PI 297783	-0.034193446	0.000872557	0.003162784	10.81118731
PI 297787 LSP	-0.041534532	0.001428705	0.002748857	15.10974921
PI 298023 LSP	-0.045897975	0.00139574	0.0031359	14.6363016
PI 298121 LSP	-0.033475165	0.001221162	0.002421556	13.82382473
PI 298122 LSP	-0.036174865	0.000842208	0.002854373	12.67348891
PI 298357 LSP	-0.042117962	0.001276367	0.003000945	14.03490197
PI 298631 LSP	-0.034139301	0.001072909	0.002372928	14.38699373
PI 298644 LSP	-0.026968205	0.000603633	0.002344629	11.50212179
PI 298645	-0.035201632	0.000824107	0.002821625	12.47566057
PI 298921	-0.038779176	0.000194002	0.003777225	10.26657829
PI 298922 LSP	-0.050373974	0.00094528	0.003642909	13.82795255
PI 298923 LSP	-0.04953494	0.001175298	0.003492856	14.18178527
PI 299116 LSP	-0.040832507	0.00106054	0.003133597	13.0305554
PI 299120 LSP	-0.031784217	0.001062177	0.002529004	12.5678783
PI 299121 LSP	-0.033595187	0.000891938	0.002894259	11.60752638
PI 299126 LSP	-0.043427028	0.000509749	0.003970195	10.93826039
PI 299163 LSP	-0.042100946	0.00081999	0.003262268	12.90542266
PI 299164 LSP	-0.033285483	0.001021315	0.002585395	12.87443097
PI 299165	-0.030639008	0.001143674	0.002252052	13.60493107
PI 299177 LSP	-0.047210109	0.000541056	0.003971714	11.88658444
PI 299237 LSP	-0.037790717	0.001025122	0.00288457	13.10098785
PI 299289	-0.030353633	0.000952995	0.002355321	12.88726075
PI 299312	-0.033297223	0.00095276	0.002490135	13.37165557
PI 299345	-0.034610744	0.00024779	0.003333063	10.38406422
PI 299351 LSP	-0.0438931	0.001203814	0.003088483	14.21186598
PI 299366 LSP	-0.023114436	0.000482112	0.002511941	9.201821338
PI 299375 LSP	-0.04470819	0.000502977	0.003823434	11.69320248
PI 300250 LSP	-0.039718447	0.000968697	0.003090051	12.85365286
PI 302398 LSP	-0.036438	0.000822776	0.003192206	11.41467622
PI 308614 LSP	-0.043057045	0.000320565	0.003902863	11.03217038
PI 311107 LSP	-0.045067477	0.000135433	0.004188276	10.76038875

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
PI 312175 LSP	-0.039260384	0.00050487	0.0033299	11.79026054
PI 320935 LSP	-0.052919557	0.001007243	0.004006827	13.20734626
PI 320936 LSP	-0.020856173	0.000845339	0.001946996	10.7119776
PI 320937 LSP	-0.022154105	0.000959217	0.00168747	13.1285936
PI 320941 LSP	-0.060730869	0.000628582	0.00472942	12.8410827
PI 320945 LSP	-0.050054132	0.000570151	0.004050268	12.35822745
PI 320946 LSP	-0.041353199	0.000747892	0.003492827	11.83946306
PI 320952 LSP	-0.029955694	0.000675318	0.002444987	12.25188428
PI 320953 LSP	-0.03986347	0.000578362	0.003207256	12.42915103
PI 320954 LSP	-0.032720069	0.000724979	0.002717441	12.04076485
PI 329157 LSP	-0.040175362	0.001025946	0.003116672	12.89047034
PI 329167 LSP	-0.041196909	0.00090087	0.003279544	12.56178009
PI 339266 LSP	-0.048718818	0.000969893	0.003753553	12.97938553
PI 339283 LSP	-0.045041757	9.26E-05	0.00436186	10.32627248
PI 339285	-0.051297983	6.61E-05	0.004730038	10.8451514
PI 339289 LSP	-0.053902258	0.001280023	0.003766588	14.31063101
PI 339292 LSP	-0.047192093	0.001048534	0.003523076	13.39513845
PI 339296	-0.04930556	0.001413748	0.003308254	14.90380105
PI 343026 LSP	-0.042490983	0.001171603	0.003106658	13.67739346
PI 345627 LSP	-0.016118964	0.001171315	0.001060503	15.19935741
PI 358602 LSP	-0.037999886	0.000909477	0.002871291	13.23442669
PI 368647 LSP	-0.040332389	0.000881695	0.003097707	13.02007922
PI 368651 LSP	-0.038747791	0.000887955	0.002942324	13.16911217
PI 370481 LSP	-0.038016781	0.000304399	0.003468117	10.96179337
PI 374116 LSP	-0.039401148	0.000929662	0.002971223	13.26091836
PI 374117 LSP	-0.046968364	0.000587829	0.003981211	11.797507
PI 374118	-0.040917	0.00043005	0.003755893	10.89407898
PI 374120 LSP	-0.028524784	0.000839711	0.0023413	12.18330906
PI 374121	-0.036385856	0.001057153	0.002742941	13.26526866
PI 379368 LSP	-0.039445406	0.000950928	0.003048472	12.93940347
PI 383682 LSP	-0.053079939	0.000387922	0.004522058	11.73800466
PI 420818	-0.039702668	6.35E-05	0.003908144	10.15895817
PI 420924 LSP	-0.04046725	8.78E-05	0.004029778	10.04205383
PI 420925 LSP	-0.037772331	0.000246572	0.003678138	10.26941743

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
PI 420929 LSP	-0.040010871	0.000102774	0.003950791	10.12730637
PI 426202 LSP	-0.050570037	0.000927107	0.003780107	13.37793745
PI 426778 LSP	-0.047230979	0.001189806	0.003334058	14.16621322
PI 426784 LSP	-0.032521294	0.000744186	0.003105362	10.47262597
PI 426797 LSP	-0.034606726	0.000823843	0.003172498	10.90835127
PI 426807 LSP	-0.047437405	0.000583641	0.004099261	11.57218562
PI 431618 LSP	-0.037794979	0.001139489	0.002790904	13.54220077
PI 431622 LSP	-0.047892602	0.00088419	0.003639717	13.15833266
PI 431630 LSP	-0.055272993	0.001129598	0.003852537	14.34716788
PI 431633 LSP	-0.038475172	0.001057428	0.002945573	13.06203155
PI 431636 LSP	-0.046557977	0.001087027	0.00345259	13.4849426
PI 431643 LSP	-0.049329874	0.000980478	0.00379754	12.98995588
PI 431662 LSP	-0.047142725	0.001132254	0.003376248	13.96304963
PI 431663 LSP	-0.051117859	0.001049581	0.003701572	13.80977017
PI 431679 LSP	-0.052211798	0.001487967	0.003238756	16.120942
PI 431684 LSP	-0.056607294	0.001064124	0.004064696	13.92657622
PI 431705 LSP	-0.045066821	0.001072774	0.003254757	13.84644513
PI 431710 LSP	-0.049123943	0.001244023	0.003388024	14.49929108
PI 431714 LSP	-0.045693521	0.001107294	0.003280198	13.93011098
PI 431717 LSP	-0.052081572	0.001131593	0.003677179	14.16345769
PI 431728 LSP	-0.045602928	0.001240481	0.003145857	14.4961879
PI 431731 LSP	-0.04267224	0.000962442	0.00324357	13.15594913
PI 431739 LSP	-0.03842917	0.001102823	0.002871273	13.38401798
PI 431753 LSP	-0.047251561	0.001298607	0.003225292	14.65032154
PI 431756 LSP	-0.051771568	0.001155361	0.003636224	14.23772786
PI 431809 LSP	-0.048476809	0.0009796	0.003673528	13.19625591
PI 431824 LSP	-0.050544506	0.000918606	0.00392064	12.89190192
PI 431863 LSP	-0.042522255	0.001005899	0.003116863	13.64264458
PI 431873 LSP	-0.057135244	0.000822765	0.004313003	13.24720586
PI 431884 LSP	-0.0482717	0.001129643	0.003551639	13.59138683
PI 431888 LSP	-0.055271991	0.000603959	0.004427363	12.48417921
PI 431893 LSP	-0.035501826	0.00121572	0.002535698	14.00080958
PI 431923 LSP	-0.045988473	0.000958009	0.003393467	13.55206328
PI 431980 LSP	-0.046690942	0.00110871	0.00341037	13.69087421

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
PI 432000 LSP	-0.056636293	0.000649297	0.004464398	12.68620973
PI 432001 LSP	-0.046453545	0.000971636	0.003498725	13.27727989
PI 432002 LSP	-0.053751778	0.000853127	0.004014635	13.38895855
PI 432005 LSP	-0.047184016	0.000880853	0.003661028	12.88818697
PI 432028 LSP	-0.04933934	0.001038758	0.003550717	13.89559806
PI 432033 LSP	-0.047629523	0.001152694	0.003363709	14.15982156
PI 432085 LSP	-0.045296012	0.000614722	0.003914195	11.57224139
PI 432106 LSP	-0.044204644	0.001101791	0.003212606	13.75974659
PI 432124 LSP	-0.055081995	0.001217603	0.003893134	14.14849653
PI 432145 LSP	-0.054857457	0.001220339	0.003705196	14.80554712
PI 432147 LSP	-0.048724895	0.001394782	0.003179856	15.32298768
PI 432184 LSP	-0.050851882	0.000933703	0.003903195	13.02827091
PI 432188 LSP	-0.044804596	0.001239747	0.003231883	13.86330909
PI 432190 LSP	-0.035149697	0.000923355	0.002698998	13.02323923
PI 432201 LSP	-0.040262185	0.001145562	0.002889279	13.93502735
PI 432212 LSP	-0.048571578	0.001207026	0.003419248	14.20533865
PI 432236 LSP	-0.052208329	0.001089641	0.003729655	13.99816721
PI 432245 LSP	-0.039161335	0.001063946	0.003053111	12.82669704
PI 432271 LSP	-0.037470573	0.001049099	0.002647298	14.15427114
PI 432286 LSP	-0.041893682	0.000350152	0.003710374	11.29095869
PI 451763 LSP	-0.035638084	0.001443698	0.002434692	14.6376161
PI 458503 LSP	-0.03878624	0.000980339	0.003022478	12.83259671
PI 468900 LSP	-0.042810624	0.000611177	0.003533528	12.11554767
PI 468901	-0.034058863	0.000409175	0.003024149	11.262297
PI 468902 LSP	-0.039449071	0.001000947	0.003019918	13.06295926
PI 472136 LSP	-0.029300293	0.001040067	0.002610911	11.22224972
PI 472205 LSP	-0.026257965	0.001248577	0.002231002	11.76958577
PI 472213 LSP	-0.046559849	0.0013026	0.003213907	14.48699185
PI 472327 LSP	-0.036652992	0.001065857	0.003087193	11.87259374
PI 472380 LSP	-0.035171522	0.001051139	0.002995336	11.74209684
PI 472416 LSP	-0.034344784	0.000781481	0.003114567	11.02714622
PI 472488 LSP	-0.032671579	0.000797516	0.003079248	10.61024447
PI 472559 LSP	-0.036645003	0.000434956	0.003688417	9.93515841
PI 472561 LSP	-0.029692068	0.000782129	0.002855489	10.39824176

Genotype	Intercept Coefficient (a)	Temperature Coefficient (b)	Photoperiod Coefficient (c)	Critical Photoperiod (h)
PI 472569 LSP	-0.029515124	0.000871738	0.002758312	10.70043118
PI 472588 LSP	-0.048805331	-0.016386994	0.024766703	1.970602667
PI 472590 LSP	-0.048395594	0.000978949	0.003670115	13.18639639
PI 472615 LSP	-0.034974614	0.0008505	0.003110964	11.24237061
PI 472629 LSP	-0.044905399	0.001239497	0.003223328	13.93137623
PI 477921 LSP	-0.035446594	0.001320128	0.00250143	14.17053416
PI 490288 LSP	-0.035788538	0.000962468	0.002712442	13.19421282
PI 490289 LSP	-0.047010259	0.001070308	0.003435178	13.68495546
PI 508090	-0.03511715	0.001496567	0.002310805	15.19693034
PI 518731 LSP	-0.042306741	0.000904441	0.003278115	12.90581501
PI 518733 LSP	-0.030841674	0.000737952	0.002495453	12.35914925
PI 518734 LSP	-0.030099277	0.000790307	0.002404828	12.51618735
PI 533688 LSP	-0.03990605	0.000677925	0.003233407	12.34179505
PI 533693 LSP	-0.038680476	0.000963711	0.002865823	13.49716134
PI 612875	-0.04653146	4.83E-05	0.004697142	9.906333506
PI 643451	-0.031404544	0.000716787	0.00300476	10.45159856
PI 643452	-0.036593743	0.000576467	0.003509587	10.42679366
Shasta	-0.030463919	0.001127005	0.002267648	13.43414528
W6 27754 LSP	-0.030121055	0.00076068	0.002434568	12.37224048
W6 27760 LSP	-0.02390061	0.001104109	0.001891272	12.63732358
W6 27763 LSP	-0.046233491	0.000247075	0.004326153	10.68697535
W6 27766 LSP	-0.022746646	0.000928789	0.002171203	10.47651919
W6 27767 LSP	-0.040297386	4.17E-05	0.004091203	9.849765066

A.2. Chapter 4 Supplementary Tables and Figures

Figure A.2.1. Manhattan and Q-Q plots for Thermal Flowering Time (TFT).

B

Metaponto 2016 - TFT with Q + K

Cordoba 2016 - TFT with Q + K
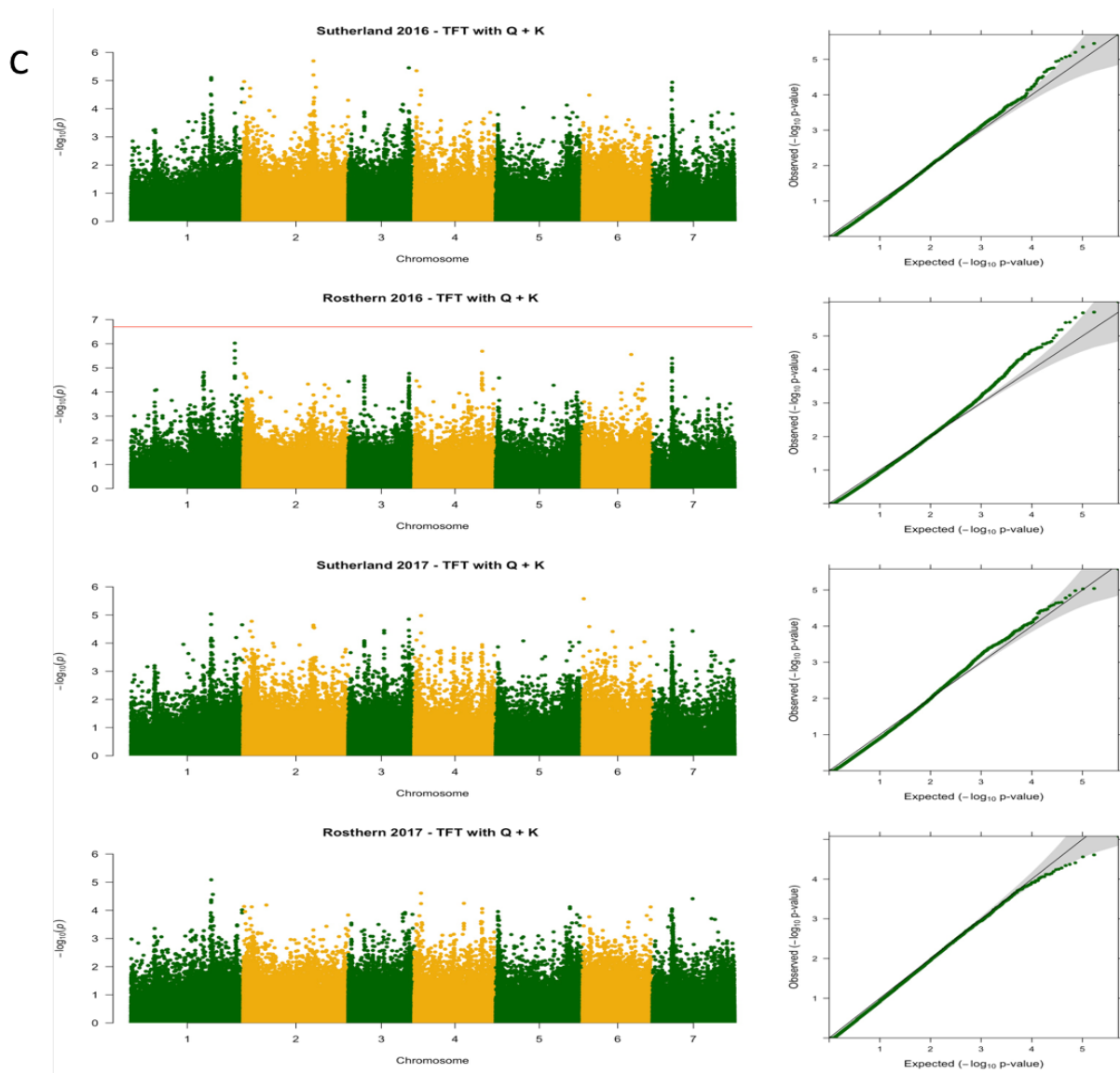

Rabat 2016 - TFT with Q + K
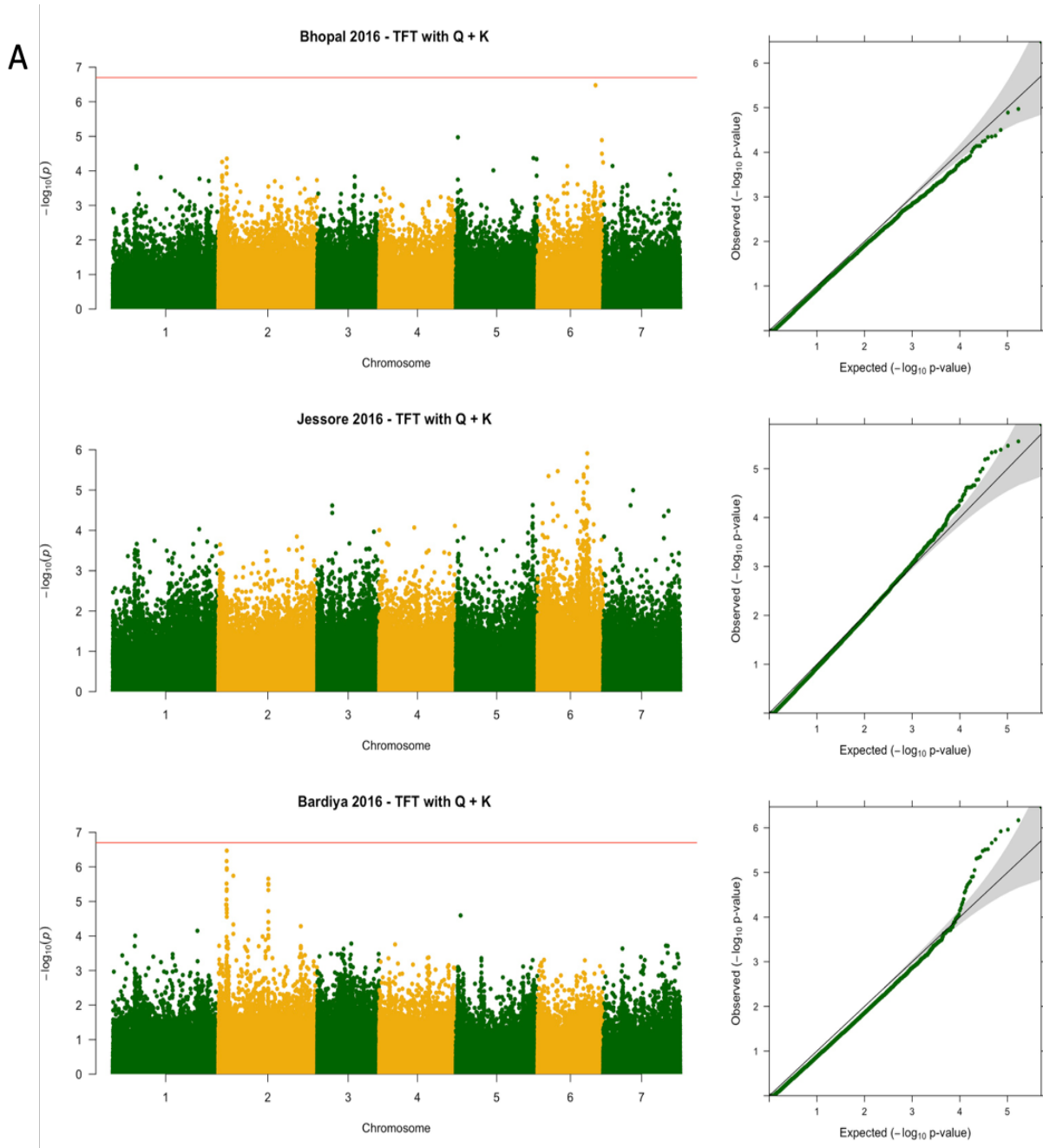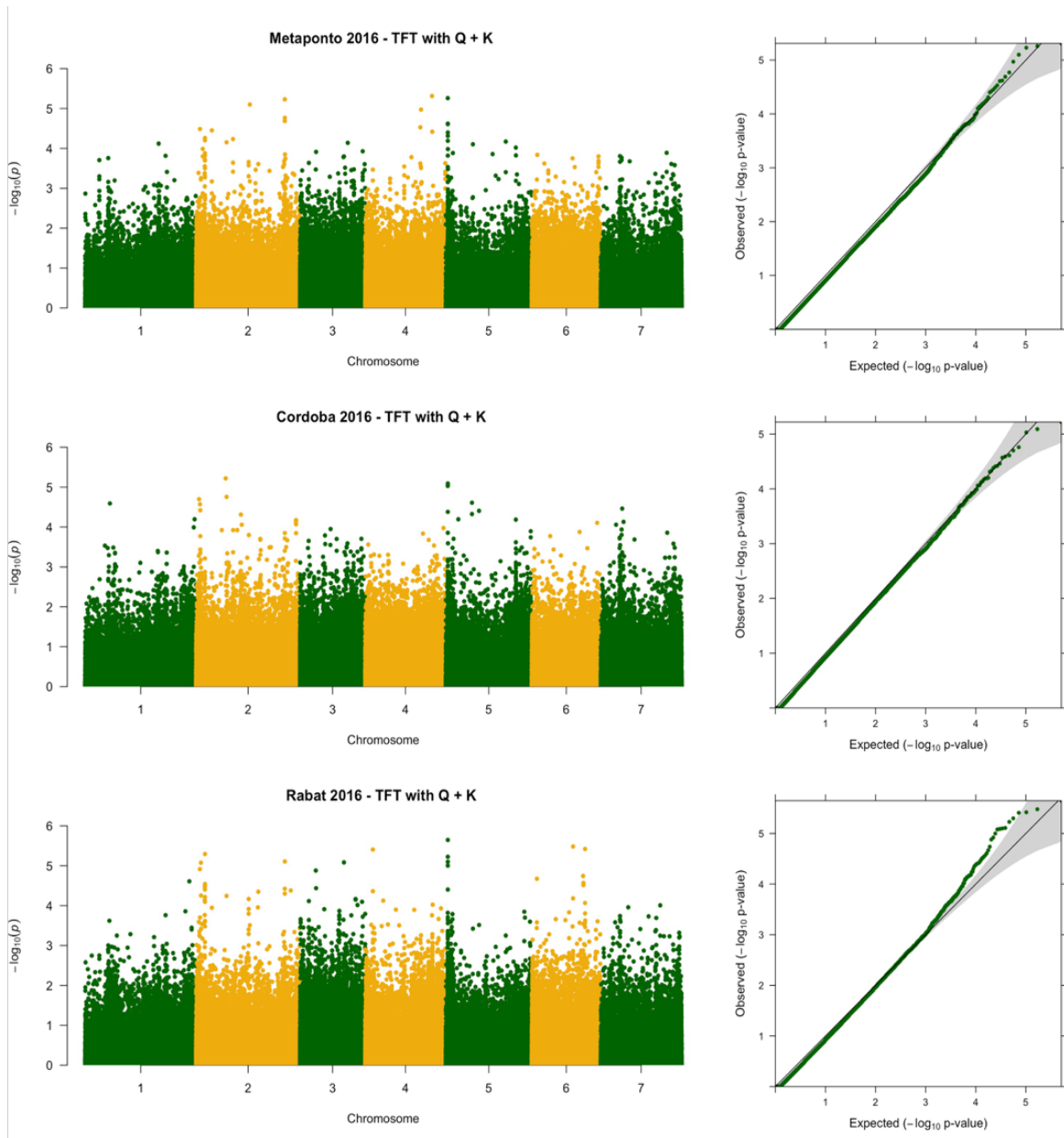
Figure A.1. Manhattan plots (left) and Quantile-quantile (Q-Q) plots (right) derived from the association studies using thermal flowering time (TFT) as the phenotypic factors from field trials at - A. South Asian macro-environment - Bhopal 2016, Jessore 2016 and Bardiya 2016; B. Mediterranean macro-environment - Metaponto 2016, Cordoba 2016 and Rabat 2016; and C. Temperate macro-environment - Rosthern and Sutherland 2016 & 2017. The X-axis of Manhattan plots represents the genomic position of SNPs in lentil genome, and the Y-axis is the -log10 of P-values. Adjacent chromosomes are colored by green and yellow. The red line on Manhattan plots indicates the significant threshold at [-log10(P) > 6.7]. The X-axis on Q-Q plot indicates expected -log10 of P-values and the Y-axis indicates observed -log10 of P-values.

Appendix. R-scripts used for all data analysis.

```
##############################################################################
##CHAPTER 3 ####################

library(openxlsx)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(viridis)

AGILE_RawData<- read.xlsx("2018-10-05_dataforSandesh.xlsx", "all") %>% # getting raw
phenotyped data from file
  mutate(Planting.Date..date. = as.Date(Planting.Date..date., origin = "1899-12-30")) ##changing
  planting date as date

names(AGILE_RawData)

xx <- AGILE_RawData %>%
  select(Plot, Entry, Name, Rep, Expt,
      PlantingDate = Planting.Date..date., DTE = Days.to.Emergence..days.,
      DTF = Days.till.10..of.Plants.have.One.Open.Flower..R1..days.,
      DTM = Days.till.10..of.Plants.have.1.2.Pods.Mature..R7..days.)


xx <- xx %>%
  mutate(RDTF = 1 / DTF,
      ETF = DTF - DTE,
      RETF = 1 / ETF) ## To get rate of pogress towards flowering, emergence to flower and 1/f
from etof



xx<- xx %>% filter(Expt %in% c("Rosthern 2016", "Sutherland 2016", "Spain 2016",    "India
2016",    "Italy 2016" ,
                    "Morocco 2016",  "Bangladesh 2016", "Bardiya 2016",   "Rosthern 2017",
                    "Sutherland 2017" )) # filtering required experiments


names(xx)[names(xx)=="Expt"] <- "siteyear" #chaning Expt name as Location
names(xx)[names(xx)=="Name"] <- "genotype" #changing name to accession
xx$siteyear[xx$siteyear == "Nepal 2016"] <- "Bardiya 2016" #changing Nepal2016 to just
Bardiya 2016
xx$siteyear[xx$siteyear == "Italy 2016"] <- "Metaponto 2016" #changing Nepal2016 to just
Metaponto 2016
xx$siteyear[xx$siteyear == "Bangladesh 2016"] <- "Jessore 2016" #changing Bangladesh 2016
to just Isurdi 2016
```

```
xx$siteyear[xx$siteyear == "Morocco 2016"] <- "Rabat 2016" #changing Morocco 2016 to just
Rabat 2016
xx$siteyear[xx$siteyear == "India 2016"] <- "Bhopal 2016" #
xx$siteyear[xx$siteyear == "Spain 2016"] <- "Cordoba 2016" #


####Data distribution check DTE and DTF
#DTE
dte.plot<- ggplot(xx, aes(genotype, DTE)) +
  geom_point(aes(color= as.factor(Rep)), alpha=0.5) +
  facet_grid(siteyear ~.) +
  ylim(0,160) +
  theme(axis.text.x=element_blank()) +
  xlab("genotype") + ylab("Days to Emergence") +
  #ggtitle("ETF Distribution") +
  theme(legend.position="top")
dte.plot
#DTF

dtf.plot<- ggplot(xx, aes(genotype, DTE)) +
  geom_point(aes(color= as.factor(Rep)), alpha=0.5) +
  facet_grid(siteyear ~.) +
  ylim(0,160) +
  theme(axis.text.x=element_blank()) +
  xlab("genotype") + ylab("Days to Flower") +
  theme(legend.position="top")
dtf.plot

#DTF site-yearwise
dtf.plot1 <- ggplot(data=xx.avg, aes(x=siteyear, y=DTF)) +
  geom_violin(aes(color= siteyear, fill=siteyear), scale = "count")  +
  geom_boxplot(width = 0.08) +
 scale_y_continuous(breaks=seq(20, 160, 20)) +
  #ylim(0, 160) +
  #scale_color_brewer(c("#E31B14", "#631311", "#60CC4F")) +
  xlab("site-year") + ylab("Days to Flower")
  #theme(legend.position="top")
dtf.plot1

######Getting ENvironmental Data
ee<- read.xlsx("2018-10-05_AGILE_EnvData_From_Knowpulse.xlsx", "upto DTF") # getting
environment data from file
str(ee)
ee<- ee %>% mutate(Date = as.Date(Date, origin = "1899-12-30")) #making date in proper YY-
MM-DD format

names(ee)[names(ee)=="Expt"] <- "siteyear" #chaning Expt name as Location
names(ee)[names(ee)=="Name"] <- "genotype" #changing name to accession
```

```
ee$siteyear[ee$siteyear == "Nepal 2016"] <- "Bardiya 2016" #changing Nepal2016 to Bardiya
2016
ee$siteyear[ee$siteyear == "Italy 2016"] <- "Metaponto 2016" #changing Nepal2016 to
Metaponto 2016
ee$siteyear[ee$siteyear == "Bangladesh 2016"] <- "Jessore 2016" #changing Bangladesh 2016
tojesssore 2016
ee$siteyear[ee$siteyear == "Morocco 2016"] <- "Rabat 2016" #c6
ee$siteyear[ee$siteyear == "India 2016"] <- "Bhopal 2016" #
ee$siteyear[ee$siteyear == "Spain 2016"] <- "Cordoba 2016" #

ee$siteyear <- factor(ee$siteyear, levels=c("Sutherland 2016", "Rosthern 2016", "Sutherland
2017", "Rosthern 2017",
                        "Metaponto 2016", "Cordoba 2016", "Rabat 2016",
                        "Bhopal 2016", "Jessore 2016", "Bardiya 2016")) #ordering the site-years
for plots



############### filtering and making min and max for DTF siteyearwise
xx1 <- xx %>% filter(DTF>0)

ll <- xx1 %>% group_by(siteyear) %>%
  summarise(MinDTF = min(DTF), MaxDTF = max(DTF)) %>% ungroup()

ll$siteyear <- factor(ll$siteyear, levels=c("Sutherland 2016", "Rosthern 2016", "Sutherland
2017", "Rosthern 2017",
                        "Metaponto 2016", "Cordoba 2016", "Rabat 2016",
                        "Bhopal 2016", "Jessore 2016", "Bardiya 2016")) #ordering the site-
years for plots



######################### scaling Daylength
ee <- ee%>%
  mutate(DayLen1 = scales::rescale(daylen, to = c(0, 40)))

mp2 <- ggplot(ee) +
  geom_rect(data = ll, aes(xmin = MinDTF, xmax = MaxDTF), ymin = 0, ymax = 40, fill =
"darkgreen", alpha = 0.2) +
  geom_line(aes(x = DAS, y = DayLen1, color = "Blue")) +
  geom_line(aes(x = DAS, y = meanT, color = "darkred") ) +
  geom_ribbon(aes(x = DAS, ymin = minT, ymax = maxT),
        fill = alpha("darkred", 0.25),
        color = alpha("darkred", 0.25)) +
  facet_grid(.~siteyear, scales = "free_x") +
  scale_y_continuous(sec.axis = sec_axis(~ (16.62 - 9.11) * . / (40 - 0) + 9.11,
                        breaks = c(10, 12, 14, 16), name = "Daylength(h)")) +
  scale_color_manual(values = c("Blue", "darkred"), labels = c("Daylength","Temperature"),
name = NULL) +
  coord_cartesian(ylim=c(0, 40)) +
```

```
  #theme_bw() +
  theme(strip.placement = "outside", legend.position = "bottom") +
  labs(title = "B.",
      x = "Days After Planting", y = "Temperature (°C)")

mp2

#####################################
###plotiing DTF
xx22.avg<- xx %>%
  group_by(Entry, genotype, siteyear) %>%
  summarise_all(funs(mean), na.rm=TRUE) #generating mean with genotype and siteyear
```

xx22.avg$siteyear <- factor(xx22.avg$siteyear, levels=c("Sutherland 2016", "Rosthern 2016", "Sutherland 2017", "Rosthern 2017",
                                    "Metaponto 2016", "Cordoba 2016", "Rabat 2016",
                                    "Bhopal 2016", "Jessore 2016", "Bardiya 2016")) #ordering
the site-years for plots

```
dtf.1<- ggplot(xx22.avg, aes(x = 1, y = DTF, fill = siteyear)) +
  geom_violin() + geom_boxplot(width = 0.1, fill = "white") + #scale = "count"
  facet_grid(.~siteyear, scales = "free_x") +
  #scale_y_continuous(sec.axis = dup_axis(name = "")) +
  scale_fill_viridis(discrete = T) +
  #theme_bw() +
  theme(legend.position = "none", strip.placement = "outside",
      axis.text.x = element_blank(), axis.ticks.x = element_blank(),
      axis.text.y.right = element_blank(), axis.ticks.y = element_blank()) +
  #scale_y_continuous(sec.axis = sec_axis(limits=c(0,150), breaks=seq(30,150, by = 30), name =
"")) +
  labs(title = "A.",
      x = "", y = "Days to Flower")

dtf.1

##arranging dtf and env plot
dtfe<- ggarrange(dtf.1, mp2, nrow=2, ncol = 1)

##############################################################
##ANOVA and lsmeans  using mixed model approach
library(lmerTest)
library(emmeans)

ldp.m<- lmer(DTF~ genotype*siteyear+ (1|siteyea/Rep), data=ldp) #mixed model- genotype and
site-years as fixed and rep nested to site-year as random
```

```
anova(ldp.m)

difflsmeans(ldp.m, test.effs = "siteyear")
m2<- emmeans(ldp.m, list(pairwise ~ siteyear), adjust = "tukey")
m3<- emmeans(ldp.m, list(pairwise ~ genotype), adjust = "tukey")
#same approach was followed for individual macro-environment

##############################################################################
##AMMI analysis for DTF (same process was applied for TFT)

library(agricolae)
library(stability)
model.dtf<- with(xx,AMMI(siteyear, genotype, Rep, DTF, console=FALSE)) #for better biplot
in change genotype to entry
dtfanova<- model.dtf$ANOVA #Analysis of Variance Table
par(cex=0.4,mar=c(4,4,1,2))
plot(model.dtf,type=1,las=1,xlim=c(-8,8)) #AMMI biplot
##############################################################################
######### EFFECT of TEMPERATURE #########
###Calculating TFT at 0 , 5 and 7 oC base temp

# creaing fucntion to caluclate GDD, Average Temp from DTE to DTF and Avergae Daylenth
for the same
i<-163
envFunct1 <- function(xx, ee) {
  for (i in 1:nrow(xx)) {
    envD <- ee %>%
      filter(siteyear == xx$siteyear[i], Date >= xx$PlantingDate[i] + xx$DTE[i], Date <=
xx$PlantingDate[i] + xx$DTF[i])

    xx$meanTetf[i] <- mean(envD$meanT)
    #xx$meanPetf[i] <- mean(envD$daylen)
    xx$PatDTF[i] <- ifelse(is.na(xx$DTF[i]), NA, envD$daylen[envD$Date ==
(xx$PlantingDate[i] + xx$DTF[i])])
    xx$TatDTF[i] <- ifelse(is.na(xx$DTF[i]), NA, envD$meanT[envD$Date ==
(xx$PlantingDate[i] + xx$DTF[i])])
  }
  xx
}

#applying the fucntion to xx
xx11 <- envFunct1(xx, ee)

##Calculate TFT (thermal flowering from seeding)
i<-94
envFunct11 <- function(xx11, ee) {
  for (i in 1:nrow(xx11)) {
    envD11 <- ee %>%
```

```
    filter(siteyear == xx11$siteyear[i], Date >= xx11$PlantingDate[i], Date <=
xx11$PlantingDate[i] + xx11$DTF[i])
    xx11$TFTat0[i] <- sum(envD11$meanT - 0)
    xx11$TFTat2[i] <- sum(envD11$meanT - 2)
    xx11$TFTat5[i] <- sum(envD11$meanT - 5)
    xx11$TFTat8[i] <- sum(envD11$meanT - 8)
    xx11$TFTat10[i] <- sum(envD11$meanT - 10)
    xx11$meanTdtf[i] <- mean(envD11$meanT)
  }
  xx11
}


##applying values in xx file

xx22<- envFunct11(xx11, ee)
##plot TFT at 5, process will be similar for TFT at 0, 8 and 10
tft5 <- ggplot(xx22.avg1, aes(siteyear, TFTat5)) +
  geom_violin()  + geom_violin(scale = "count")  +
  geom_violin(aes(color= siteyear, fill=siteyear)) +
  geom_boxplot(width = 0.08) +
  #theme(axis.text.x=element_blank())+
  #scale_y_continuous(breaks=seq(0, 3000, 500)) +
  scale_y_continuous(limits=c(0,2500), breaks=seq(0,2500, by = 500)) +
  xlab("") + ylab("Thermal Flowering Time with base 5°C (°C.day)") +
  theme(legend.position="none")
tft5


#################################################################################
###### EFFECT of PHOTOPERIOD #######

# Calculating a,b and c for critical and Cumulative photoperiod as well as nominal base
photoperiod for each entry
abc11 <- data.frame(Entry = unique(xx22$Entry),
            genotype = xx22$genotype[match(unique(xx22$Entry),xx22$Entry)],
            a = NA, b= NA, c = NA) #creating a dataframe for a and b and c
#creating fucntion to get a,b,c
i<-120
for (i in unique(xx22$Entry)) {
  x1 <- xx22 %>% filter(Entry==i)
  model1<- lm(RETF~meanTetf+PatDTF, data=x1) #using model by Summerfield et al. 1985,
1/f~a+bt+cp
  #anova(model1)
  #coef(model1)
  #summary(model1)
  abc11[match(i, abc11$Entry), c("a","b", "c")] <- coef(model1) ##applying in file name abc
}
abc11 <- abc11 %>% mutate(critical_photo=-a/c) #calculating critical photoperiod and
overwiritng on previous data frame
```

abc11<- left_join(abc11, mm %>% select(Entry, Origin), by = "Entry") #adjoing origin wrt entry and making a single file
xx33 <- left_join(xx22, abc11 %>% select(Entry, critical_photo), by = c("Entry")) #renaming as xx33 and adding the critical photoperiod info on xx22


```
# creaing fucntion to caluclate simple cumulative photoperiod and photoperiod after critical photoperiod
psFunct <- function(xx33, ee, abc11) {
  for (i in 1:nrow(xx33)) {
    yi <- ee %>%
      filter(siteyear == xx33$siteyear[i], Date >= xx33$PlantingDate[i] + xx33$DTE[i], Date <= xx33$PlantingDate[i] + xx33$DTF[i]) %>%
      mutate(DPPwithCP = daylen-((-abc11$a[match(xx33$Entry[i], abc11$Entry)])/(abc11$c[match(xx$Entry[i], abc11$Entry)])),
          DPP = daylen,
          DPPmorethanCP = ifelse (daylen <((-abc11$a[match(xx33$Entry[i], abc11$Entry)])/(abc11$c[match(xx33$Entry[i], abc11$Entry)])), 0, daylen)) %>%
      mutate(dpp1 = ifelse(DPPwithCP< 0, 0, DPPwithCP))
    xx33$sumDPP[i] <- sum(yi$DPP)
    xx33$CPPwithCP[i] <- sum(yi$DPPwithCP)
    xx33$CPPafterCP[i] <- sum(yi$DPPmorethanCP)
  }
  xx33
}
xx44 <- psFunct(xx33, ee, abc11) #getting the caluclated info into the X file using the created function

############
###plotiing critical photoperiod and DPP and CPP
xx44.avg<- xx44 %>%
  group_by(Entry, genotype, siteyear) %>%
  summarise_all(funs(mean), na.rm=TRUE) #generating mean with genotype and siteyear
xx44.avg$siteyear <- factor(xx44.avg$siteyear, levels=expts) #ordering the site-years for plots
xx44.avg1 <- xx44.avg %>% filter (sumDPP >0)
dpp<- ggplot(xx44.avg1, aes(siteyear, sumDPP)) +
  geom_violin()  + geom_violin(scale = "count")  +
  geom_violin(aes(color= siteyear, fill=siteyear)) +
  geom_boxplot(width = 0.08) +
  #theme(axis.text.x=element_blank())+
  scale_y_continuous(limits=c(0,1500), breaks=seq(0,1500, by = 200)) +
  xlab("site-year") + ylab("Cumulative Photoperiod (h)") +
  theme(legend.position="none")
dpp

##critical photoperiod
cp <- ggplot(abc11, aes(genotype, y= critical_photo)) +
  geom_point (aes(color=genotype, alpha=0.5)) +
```

```
  scale_y_continuous(limits=c(0,27), breaks=seq(0,27, by = 2)) +
  geom_hline(yintercept = 12.12, color="grey") +
  #theme(axis.text.x=element_blank())+
  theme(axis.text.x = element_text(angle = 90, size= 3, hjust = 0, vjust=0.5)) +
  xlab("Genotype") + ylab("Critical Photoperiod (h)") +
  theme(legend.position="none")
cp

##cumulative photoperiod after CP
cpp<- ggplot(xx44.avg1, aes(siteyear, CPPafterCP)) +
  geom_violin()  + geom_violin(scale = "count")  +
  geom_violin(aes(color= siteyear, fill=siteyear)) +
  geom_boxplot(width = 0.08) +
  #theme(axis.text.x=element_blank())+
  scale_y_continuous(limits=c(0,1250), breaks=seq(0,1250, by = 200)) +
  xlab("") + ylab("Cumulative Photoperiod after CP (h)") +
  theme(legend.position="none")
cpp

################################################################################
############ INTERACTION EFFECT ########

xx55 <- xx44 %>%   mutate(PTT0 =TFTat0*sumDPP, PTT5 =TFTat5*sumDPP) #calculate
interactive effect

###plotting PTT
xx55.avg<- xx55 %>%
  group_by(Entry, genotype, siteyear) %>%
  summarise_all(funs(mean), na.rm=TRUE) #generating mean with genotype and siteyear
ptt5<- ggplot(xx55.avg1, aes(siteyear, PTT5)) +
  geom_violin()  + geom_violin(scale = "count")  +
  geom_violin(aes(color= siteyear, fill=siteyear)) +
  geom_boxplot(width = 0.08) +
  #theme(axis.text.x=element_blank())+
  scale_y_continuous(limits=c(0,3000000), breaks=seq(0,3000000, by = 500000)) +
  xlab("site-year") + ylab("Photothermal Time (PTU)") +
  theme(legend.position="none")
ptt5

################################################################################
################################################################################
## CHAPTER 4 ####################
#Installing packages and library needed for GAPIT GWAS

source("https://bioconductor.org/biocLite.R")
#biocLite("BiocUpgrade")

biocLite("multtest")
```

```
biocLite("chopsticks")

install.packages("gplots")
install.packages("LDheatmap")
install.packages("genetics")
install.packages("ape")
install.packages("EMMREML")
install.packages("scatterplot3d")

library(multtest)
library(gplots)
library(LDheatmap)
library(genetics)
library(ape)
library(EMMREML)
library(compiler)
library(scatterplot3d)
library(Matrix)

source("http://zzlab.net/GAPIT/gapit_functions.txt") # to install GAPIT Package
source("http://zzlab.net/GAPIT/emma.txt") #to get emma library


##############################################################################
######DTF only

#####Nepal 2016

myGAPIT <- GAPIT(
  Y=myY.np, #DTF data from Nepal 2016
  G=myG, #Genotyped data
  KI=myKI, #kinship Matrix
  CV=myCV # Q matrix
) #followed same process for all site-years


################################# Creating customized Manhattan and Q-Q plots

GWAS_append <- function(folder, package = "MVP", colors = c("Dark Green",
"darkgoldenrod2"), threshold = 6.7) {
  library(magick)
  ####################################################
  # MVP
  ####################################################
  if(package == "MVP") {
    phefiles  <- list.files(folder)[grep("MVP.Phe", list.files(folder))]
    manfiles1 <- list.files(folder)[grep("Multraits.Rectangular", list.files(folder))]
    manfiles2 <- list.files(folder)[grep("Multracks.Rectangular", list.files(folder))]
```

98

```r
  qqfiles   <- list.files(folder)[grep("Multraits.QQplot", list.files(folder))]
  for (i in 1:length(manfiles1)) {
    im1 <- image_read(paste(folder, manfiles1[i], sep = "/"))#image_convert(image_scale(,
"x500"), "jpg")
    im2 <- image_read(paste(folder, manfiles2[i], sep = "/"))
    im3 <- image_read(paste(folder, phefiles[i],  sep = "/"))
    im4 <- image_read(paste(folder, qqfiles[i],   sep = "/"))
    im5 <- image_append(c(im1, im2), stack = T)
    im6 <- image_append(c(im3, im4), stack = T)
    im7 <- image_append(image_scale(c(im5, im6)))
    image_write(im7, path = paste0(folder, "/Appended.", substr(phefiles[i], 22,
nchar(phefiles[i])-4), ".jpg"))
  }
}


  ########################################################
  GAPIT_Man <- function(folder, colors = c("Dark Green", "darkgoldenrod2")) {
  library(qqman)
  library(dplyr)
  resultfiles <- list.files(folder)[grep("GWAS.Results.csv", list.files(folder))]
  for (i in 1:length(resultfiles)) {
    x <- read.csv(paste(folder, resultfiles[i], sep = "/")) %>%
      rename(CHR = Chromosome, BP = Position, P = P.value)
    png(paste0(folder, "/", substr(resultfiles[i], 1, nchar(resultfiles[i])-16), ".GenomeWide.png"),
      width = 2000, height = 1000, res = 200)
    manhattan(x, suggestiveline = F, main = "Sutherland 2017 - DTF with Q + K ", col = colors,
genomewideline = threshold)
    dev.off()
  }
}
  GAPIT_QQ <- function(folder) {#col.points = "Dark Green", col.conf = "lightgray"
  library(lattice)
  library(dplyr)
  #library(qqman)
  #library(snpStats)
  #library(ggpubr)
  qqunif.plot<-function(pvalues,
              should.thin=T, thin.obs.places=2, thin.exp.places=2,
              xlab=expression(paste("Expected (",-log[10], " p-value)")),
              ylab=expression(paste("Observed (",-log[10], " p-value)")),
              draw.conf=TRUE, conf.points=1000, conf.col="lightgray", conf.alpha=.05,
              already.transformed=FALSE, pch=20, aspect="iso", prepanel=prepanel.qqunif,
              par.settings=list(superpose.symbol=list(pch=pch)),  ...) { #col.points=col.points,
    #error checking
    if (length(pvalues)==0) stop("pvalue vector is empty, can't draw plot")
    if(!(class(pvalues)=="numeric" ||
      (class(pvalues)=="list" && all(sapply(pvalues, class)=="numeric"))))
```

```r
    stop("pvalue vector is not numeric, can't draw plot")
  if (any(is.na(unlist(pvalues)))) stop("pvalue vector contains NA values, can't draw plot")
  if (already.transformed==FALSE) {
    if (any(unlist(pvalues)==0)) stop("pvalue vector contains zeros, can't draw plot")
  } else {
    if (any(unlist(pvalues)<0)) stop("-log10 pvalue vector contains negative values, can't draw
plot")
  }
  #
  grp<-NULL
  n<-1
  exp.x<-c()
  if(is.list(pvalues)) {
    nn<-sapply(pvalues, length)
    rs<-cumsum(nn)
    re<-rs-nn+1
    n<-min(nn)
    if (!is.null(names(pvalues))) {
      grp=factor(rep(names(pvalues), nn), levels=names(pvalues))
      names(pvalues)<-NULL
    } else {
      grp=factor(rep(1:length(pvalues), nn))
    }
    pvo<-pvalues
    pvalues<-numeric(sum(nn))
    exp.x<-numeric(sum(nn))
    for(i in 1:length(pvo)) {
      if (!already.transformed) {
        pvalues[rs[i]:re[i]] <- -log10(pvo[[i]])
        exp.x[rs[i]:re[i]] <- -log10((rank(pvo[[i]], ties.method="first")-.5)/nn[i])
      } else {
        pvalues[rs[i]:re[i]] <- pvo[[i]]
        exp.x[rs[i]:re[i]] <- -log10((nn[i]+1-rank(pvo[[i]], ties.method="first")-.5)/(nn[i]+1))
      }
    }
  } else {
    n <- length(pvalues)+1
    if (!already.transformed) {
      exp.x <- -log10((rank(pvalues, ties.method="first")-.5)/n)
      pvalues <- -log10(pvalues)
    } else {
      exp.x <- -log10((n-rank(pvalues, ties.method="first")-.5)/n)
    }
  }
  #this is a helper function to draw the confidence interval
  panel.qqconf<-function(n, conf.points=1000, conf.col="gray", conf.alpha=.05, ...) {
    require(grid)
    conf.points = min(conf.points, n-1);
```

```r
      mpts<-matrix(nrow=conf.points*2, ncol=2)
      for(i in seq(from=1, to=conf.points)) {
        mpts[i,1]<- -log10((i-.5)/n)
        mpts[i,2]<- -log10(qbeta(1-conf.alpha/2, i, n-i))
        mpts[conf.points*2+1-i,1]<- -log10((i-.5)/n)
        mpts[conf.points*2+1-i,2]<- -log10(qbeta(conf.alpha/2, i, n-i))
      }
      grid.polygon(x=mpts[,1],y=mpts[,2], gp=gpar(fill=conf.col, lty=0), default.units="native")
    }
    #reduce number of points to plot
    if (should.thin==T) {
      if (!is.null(grp)) {
        thin <- unique(data.frame(pvalues = round(pvalues, thin.obs.places),
                          exp.x = round(exp.x, thin.exp.places),
                          grp=grp))
        grp = thin$grp
      } else {
        thin <- unique(data.frame(pvalues = round(pvalues, thin.obs.places),
                          exp.x = round(exp.x, thin.exp.places)))
      }
      pvalues <- thin$pvalues
      exp.x <- thin$exp.x
    }
    gc()
    #
    prepanel.qqunif= function(x,y,...) {
      A = list()
      A$xlim = range(x, y)*1.02
      A$xlim[1]=0
      A$ylim = A$xlim
      return(A)
    }
    #draw the plot
    xyplot(pvalues~exp.x, groups=grp, xlab=xlab, ylab=ylab, xlim=range(exp.x),
ylim=range(pvalues), #aspect=aspect,
        prepanel=prepanel, scales=list(axs="i"), pch=pch,
        panel = function(x, y, ...) { #col.points=col.points,
          if (draw.conf) {
            panel.qqconf(n, conf.points=conf.points,
                   conf.col=conf.col, conf.alpha=conf.alpha)
          };
          panel.xyplot(x,y, col = "Dark Green", ...);
          panel.abline(0,1);
        }#, par.settings=par.settings, ...
    )
  }
  #
  resultfiles <- list.files(folder)[grep("GWAS.Results.csv", list.files(folder))]
```

```r
   for (i in 1:length(resultfiles)) {
     x <- read.csv(paste(folder, resultfiles[i], sep = "/")) %>%
       rename(CHR = Chromosome, BP = Position, P = P.value)
     png(paste0(folder, "/", substr(resultfiles[i], 1, nchar(resultfiles[i])-16), ".QQ.png"),
         width =1000, height = 1000, res = 200)
     print(qqunif.plot(x$P))
     dev.off()
   }
 }

 #
 if(package == "GAPIT") {
   GAPIT_Man(folder = folder, colors = colors)
   GAPIT_QQ(folder = folder)
   resultfiles <- list.files(folder)[grep("GWAS.Results.csv", list.files(folder))]
   manfiles    <- list.files(folder)[grep("GenomeWide.png", list.files(folder))]
   qqfiles     <- list.files(folder)[grep("QQ.png", list.files(folder))]
   for (i in 1:length(resultfiles)) {
     im1 <- image_read(paste(folder, manfiles[i], sep = "/"))#image_convert(image_scale(,
"x500"), "jpg")
     im2 <- image_read(paste(folder, qqfiles[i], sep = "/"))#image_convert(image_scale(, "x500"),
"jpg")
     im3 <- image_append(c(im1, im2))
     image_write(im3, path = paste0(folder, "/", substr(resultfiles[i], 1, nchar(resultfiles[i])-16),
".ManQQ.png"))
   }
 }
}

#GWAS_append(folder =
"/Users/owner/Documents/LDP_Final_Analysis/GWAS/180929_all_final/180930_GAPIT_Resu
lts_DTF_Sutherland 17", package = "GAPIT", colors = c("Dark Green", "darkgoldenrod2"))
#followed same process for TFT

#############################################################################

GWAS_PeakTable <- function(myG, myResultsFolder, myGenes, outfile = "peaktable",
peakthreshold = 6.7, windowthreshold = 6.7, g.range, package = "GAPIT") {
  oo <- NULL
  gg <- read.csv(myG)
  #
  if(package == "GAPIT") {
    fnames <- list.files(myResultsFolder)[grepl(".GWAS.Results.csv",
list.files(myResultsFolder))]
    #k <- fnames[1]
    for(k in fnames) {
      gw <- read.csv(paste(myResultsFolder,k, sep="/"))
      gw$logP <- -log10(gw$P.value)
```

```
    gw <- gw[gw$logP > windowthreshold,]
    gp <- gw[gw$logP > peakthreshold,]
    if(nrow(gp) > 0) {
      gp$Start <- NA
      gp$End <- NA
      gp$allM <- NA
      #i<-1
      for(i in 1:nrow(gp)) {
        gi <- gw[gw$Chromosome == gp$Chromosome[i] & gw$Position > gp$Position[i] -
g.range & gw$Position < gp$Position[i] + g.range,]
        gp$Start[i] <- min(gi$Position)
        gp$End[i]   <- max(gi$Position)
        gp$allM[i] <- paste(gi$SNP[gi$logP > peakthreshold], collapse = ";")
      }
      gp <- gp[rev(order(gp$logP)),]
      gp <- gp[!duplicated(gp$Start),]
      #
      gp$Trait <- substr(k, 7, gregexpr(".GWAS.Results.csv",k)[[1]][1]-1)
      oo <- rbind(oo, gp)
    }  }  }
  #
  # if(package == "MVP) {}
  #
  oo$Num <- NA
  oo$Genes <- NA
  oo$Description <- NA
  gl <- read.csv(myGenes)
  #i<-123
  for(i in 1:nrow(oo)) {
    gi <- gl[gl$Chromosome == oo$Chromosome[i] & gl$End > oo$Start[i] & gl$Start <
oo$End[i],]
    oo$Num[i] <- nrow(gi)
    oo$Genes[i] <- paste(gi$ID, collapse = ";")
    oo$Descriptions[i] <- paste(gi$Description, collapse = ";")
  }
  write.csv(oo, "~/Documents/LDP_Final_Analysis/GWAS/180929_all_final/181218_Singificant
Peaks.csv" , row.names = F) #outfile
}

GWAS_PeakTable(myG =
"/Users/owner/Documents/LDP_Final_Analysis/GWAS/180929_all_final/324_LDP_common._I
D_changed_LD.Filtered_hapmap.csv",
        myResultsFolder
="/Users/owner/Documents/LDP_Final_Analysis/GWAS/180929_all_final/181013_GWAS_Res
ults",
        #myResultsFolder
="/Users/owner/Documents/LDP_Final_Analysis/GWAS/180929_all_final/",
```

```
        myGenes =
"/Users/owner/Documents/LDP_Final_Analysis/GWAS/180929_all_final/genelist.csv",
        outfile =
"~/Documents/LDP_Final_Analysis/GWAS/180929_all_final/181218_Singificant Peaks.csv",
        peakthreshold =  6.7,
        windowthreshold =  5,
        g.range =  1000000,
        package =  "GAPIT")


#####################
##Allelic proportion plot
#######Qgroup and DTF
sandPlot <- function(x, marker, phenodata) {
 x <- x %>%
   filter(rs==marker) %>%
   select(12:ncol(.)) %>% t() %>% as.data.frame() %>%
   rownames_to_column() %>% rename(Name = rowname) %>%
   left_join(select(mm, Name, grpQ , Str_E), by = "Name") %>%
   left_join(phenodata, by = c("Name"="genotype"))
 x <- x %>% filter (V1 %in% c("AA", "GG", "CC", "TT")) #selecting only homozygous
 x <- x %>% filter (grpQ !="NA")
 x <- x %>% filter (siteyear !="NA")
 ##plotting q group and str_E
 #
 qg.1<-ggplot(x, aes(x = V1)) + geom_bar(aes(fill = grpQ), alpha=0.7, position = "fill") +
   scale_fill_manual(values = c("#517553A9", "#70E9EDA9", "#EBCA23", "#2BF032",
"#498C8B", "#C251C4", "#F72338"))+
   labs(title = marker, x = "", y= "")
 #
 dtf.qg.1 <- ggplot(x, aes(x = V1, y = DTF)) +
   geom_violin() +
   geom_quasirandom(aes(color= grpQ, fill=grpQ)) +
   scale_color_manual(values = c("#517553A9", "#70E9EDA9", "#EBCA23", "#2BF032",
"#498C8B", "#C251C4", "#F72338")) +
   labs(title = "", x = "", y= "")
 #
 dtf.1 <- ggplot(x, aes(x = siteyear, y = DTF)) +
   geom_violin() +
   geom_quasirandom(aes(color= grpQ, fill=grpQ)) +
   scale_color_manual(values = c("#517553A9", "#70E9EDA9", "#EBCA23", "#2BF032",
"#498C8B", "#C251C4", "#F72338")) +
   labs(title = "", x = "", y= "")


 #
 qg1 <- list(x1 = qg.1, x2 = dtf.qg.1, x3=dtf.1)

 qg1
```

```
}

mpl1 <- sandPlot(myG, as.character(yy$SNP[1]), dtf.np)
mpl2 <- sandPlot(myG, as.character(yy$SNP[2]), dtf.np)
mpl3 <- sandPlot(myG, as.character(yy$SNP[3]), dtf.spn)
mpl4 <- sandPlot(myG, as.character(yy$SNP[4]), dtf.spn)
mpl5 <- ggarrange(mpl1[[1]], mpl2[[1]], mpl3[[1]], mpl4[[1]],
          mpl1[[2]], mpl2[[2]], mpl3[[2]], mpl4[[2]],
          mpl1[[3]], mpl2[[3]], mpl3[[3]], mpl4[[3]],
          ncol = 4, nrow=3, common.legend = T), labels = c("a","b","c","d"))
mpl5
```

##
##