# *A Study on Adaptive Strategies of Wild and Cultivated Linum Populations Across Western Europe and Their Implications for Linum Trait Development and Ecology.*

HORASMAN FEBRICO HABEAHAN

**How to cite:**

HABEAHAN, HORASMAN FEBRICO (2023) A Study on Adaptive Strategies of Wild and Cultivated Linum Populations Across Western Europe and Their Implications for Linum Trait Development and Ecology. Doctoral thesis, Durham University.

# A Study on Adaptive Strategies of Wild and Cultivated *Linum* Populations Across Western Europe and Their Implications for *Linum* Trait Development and Ecology.

*Collections of studies on local adaptation, and effects of local climates on genotype and phenotype of wild and cultivated Linum populations found across Western Europe with implications for ecology and development agricultural traits.*

*by*

Horasman Febrico Habeahan (Rico)

Department of Biosciences

Durham University

Supervised by: Dr. A.C. Brennan, Prof. A.R. Hoelzel, Dr. C. Kidner

This thesis is submitted in candidature of the degree 'Doctor of Philosophy' (Ph.D) at the Department of Biosciences, Durham University (2018-2022).

# Table of Contents:

# DECLARATION

Materials contained within this thesis have not previously been summitted, published or reviewed for a degree at Durham University and/or other institutions. The research, methods, and analyses in this thesis were conducted by the author, unless otherwise stated in the thesis. References are made where the author have used, in part or in full, materials which were already published in the literatures.

# DEDICATION

I dedicate this work to my mother, father and brother, my mentors/supervisors both past and present and my friends who have helped me and are the foundations to the person I am today.

“In the sciences, the authority of thousands of opinions is not worth as much as one tiny spark of reason in an individual man” – Galileo Galilei

# ACKNOWLEDGEMENTS

Works and studies conducted during this thesis would not have been possible without the support of many people and institutions within my time as a PhD student at the Department of Biosciences in Durham University. I would like to first thank all the institutions which have contributed to the completion of this thesis both technically and financially. I would like to thank the Department of Biosciences at Durham University for providing the correct space and technical equipment needed for the completion of this thesis. I would like to thank the UK Research and Innovation (UKRI) and especially the Biotechnology and Biological Sciences Research Council (BBSRC) for financial provision during my time at Durham University both in forms of laboratory research grants and stipend support towards my living costs while being a student at Durham University. I would like to also extend my thanks to the Royal Botanic Garden Edinburgh (RBGE) as our BBSRC CASE Partner and the financial provision which I have received as part of my research grant from the RBGE.

I would like to thank and acknowledge my current supervisor, Dr. Adrian Brennan for his excellent patience and his professional guidance throughout the work of this thesis. I hope to one day be as much of a guide and an excellent supervisor as Dr. Adrian Brennan to me during my time at Durham University, you have made my PhD a richer experience with your extensive knowledge and guidance, despite restrictions caused by the Covid-19 pandemic. I would like to extend this for my other supervisors, Prof. Rus Hoelzel of Durham University and Dr. Catherine Kidner of the Royal Botanic Garden Edinburgh. I would like to further acknowledge Dr. Pilar Suarez-Montes (CONACYT fellow) for her kind help and insights during her time as a Postdoctoral research assistant at Durham University and for her contributions during the chapter two and chapter five of this thesis. I would like to thank all the good friends I have in Laboratory 19 (past and present) of the Department of Biosciences for all the support you have shown me.

Finally, I want to acknowledge Dr. Rocio Perez-Barrales and Beatrice Landoni (Portsmouth University) as our excellent collaborators during my time at Durham University. I want to thank you for sharing wild flax materials along with your knowledge for the enrichment of this thesis. I hope the outcome of this thesis will have further implications within your own future research and collaborations.

# LIST OF ABBREVIATIONS

AP1 = APETALA 1 (Gene)

CO = CONSTANS (Gene)

FLC = FLOWERING LOCUS C (Gene)

FLM = FLOWERING LOCUS M (Gene)

FRI = FRIGIDA (Gene)

FT = FLOWERING LOCUS T (Gene)

FUL = FRUITFUL (Gene)

GAPDH = Glyceraldehyde-3-phosphate dehydrogenase (Gene)

GI = GIGANTEA (Gene)

GOI = Gene of interest

HKGs = House Keeping Genes

LFY = LEAFY (Gene)

NGS = Next Generation Sequencing

NRT = No Reverse Transcriptase

NTC = No Template Controls

PCR = Polymerase Chain Reaction

RADSeq = Restriction-site Associated DNA Sequencing

RGE = Relative Gene Expression

RT-qPCR = Real-time quantitative polymerase chain reactions

S0 = Selfing 0 (Wild selfing generation)

S1 = Selfing 1 (First selfing Generation)

SD = Standard Deviation

SEP3 = Sepallata 3

SNPs = Single Nucleotide Polymorphisms

SOC = Suppressor of Overexpression of Constans

SVP = SHORT VEGETATIVE PHASE (Gene)

UBI = Ubiquitin (Gene)

VRN = Vernalization

# ABSTRACT

Local adaptation plays a major part in plant survival and reproduction. *Linum* represents a genus of potential study models which provide insights for both applied and evolutionary biology. As a flowering plant, they may have evolved adaptations to achieve optimal flowering time. Most flowering plants have developed their own strategies to flower in specific habitats. There are interests from both evolutionary, and agricultural points of view regarding flowering development. In evolutionary terms, flowering time may affect offspring and population fitness. In agriculture, faster flowering time is a desirable trait for production. For this reason, *Linum* is a versatile model to study. *Linum usitatissimum* (cultivated flax) is useful in several industries. Their wild predecessors, the wild flax (*Linum bienne*) is relatively less studied in comparison to their cultivated relative. As their predecessor, we suggest that implications of local adaptation in the wild flax, may aid the development of their cultivar relatives. We examined both wild and cultivar type to determine local adaptation strategies, particularly in terms of flowering.

In chapter 2 of this thesis, we examined three flowering time genes and two duplicate genes. *Linum* is a temperate plant and as such requires vernalization. Vernalization is the process in which plants require colder temperatures to induce the flowering process. Plants that require vernalization often flower earlier when in colder temperatures than those that have not experienced cold induction. We studied expression of five flowering time genes for implication of local adaptation after treatment to vernalization. *L. bienne* appeared to express genes differently in comparison to the cultivars. The expression of *Linum* FLOWERING LOCUS T (*LuFT*) revealed a positive correlation with number of days to flower. This potentially identifies FLOWERING LOCUS T as one of the important genes regulating vernalization in *Linum*. Our result revealed variation in relative flowering time gene expressions. Wild and cultivated *Linum* demonstrate different relationships between flowering time and environmental variables.

In chapter 3, we quantified the viability of pollen, an important part in the transfer of the male gamete in flowering plants, under different temperature treatments. *Linum* is an established temperate plant. Sensitivity to temperature changes maybe more predominant in temperate plants as seasonal changes would reflect a challenge to flower in temperate environments. In this chapter, *Linum* pollen was treated under different temperatures to observe their ability to germinate. This is important to determine whether temperature plays a major part in affecting the viability of pollen, which in turn plays a major role in the formation of seeds. This chapter revealed a reduction in the number of pollen tubes formed under different treatments and across the two *Linum* species. In

addition to this, correlations to local climates were also observed, with variation in trends across the temperature treatments.

In chapter 4 of this thesis, we examined population genetics of wild *L. bienne* samples originating from different latitudes across western Europe. We examined Wild flax *(L. bienne)* populations across western Europe to provide insights into their genetic structure and diversity.  This population analysis will develop our understanding of adaptation in wild *Linum* in response to their environment. A double-digest RAD sequencing (DDRadSeq) protocol was utilized to look at variation in SNPs across different populations. Both *L. bienne* and *L. usitatissimum* samples were sequenced and aligned to an *L. usitatissimum* whole genome. Genetic structuring of our *Linum* samples were revealed across Western Europe. Cultivars in our collection revealed to be more genetically related to the Northern accessions of our wild samples.

In chapter 5 we summarized plant architectures. In this chapter traits and their relationship to the latitude were summarized in relation to the requirement for vernalization. Four traits were measured: overall height of the plants after first flowering, the number of stems of the plants after first flowering, and the number of flower buds on the plants at first flowering. In addition to this, seed size was examined in terms of its area. We also examined correlation between traits and environmental variables. We found that there was a relationship between traits when no-vernalization occurred. With vernalization, these relationships became less significant and, in some cases, not significant. This illustrates that vernalization influences the relationship of traits beyond flowering initiation. There were also suggestions that the traits measured correlate with latitude under no-vernalization treatments. For vernalized individuals, the correlation for the traits measured and latitude was not significant. Seed sizes were strongly correlated with both latitude and climatic variables in all cases of the treatments. The findings suggest there are effect of environmental variables in these measured traits, which suggests differentiation within wild *Linum* species.

---

# CHAPTER 1: INTRODUCTION

For living organisms, adaptation to the local environment is important for species survival and the success of their offspring. This adaptation to the environment is often referred to as “local adaptation” (Brandon R N, 2014). Local adaptation often results in a higher mean fitness of a population in their native environment. Large numbers of reciprocal studies and garden experiments have illustrated this in plants (Lascoux M. et al., 2016). Local adaptation usually arises from biotic and abiotic factors which represent a selection pressure that organisms must adapt to (Rúa M. et al., 2016). Studies also suggests that local adaptation could be climate-driven, with suggestion that locally adapted tree population managing to track climate change due to local adaptation (Moran E., 2020). It is well established that local adaptation arises from selection by the local environment which favours specific phenotypes. This can translate into a genotype-by-environment interaction in response to biotic and abiotic factors (Sork V., 2018).

More generally, local adaptation is understood to be pervasive and is a key evolutionary process that contributes towards the success of a population (Sork V., 2018; Meek M. et al., 2023). It is generally a result of divergent selection on traits which can lead to reproductive isolation and even speciation (White N. and Butlin R., 2021). One outcome of local adaptation is the maintenance of ecologically important genes (Whitlock M., 2015). Adaptation to local environment may provide insight into improving survival rate for populations not on the brink of extinction, to maximize their likelihood of long-term persistence (Bay R. et al., 2018).

Meta-analysis of plant population suggested that locally adapted populations of plants tend to perform better than foreign plants when at the site of their origin (Leimu M. and Fischer M., 2008). In addition to this, they also suggested that population size have a positive role on the ability of a population to locally adapt (Leimu M. and Fischer M., 2008). In plants, responses to selection pressure in their environment may lead to differences in phenotypic traits, such as the size and shapes of their leaves (Dudley S., 1996). Local adaptation of species to the environment have also been previously observed in germination behaviour of plants (Donohue K. et al., 2010). Variation in local adaptation can affect biodiversity of a given environment (Atkins K.E. et al., 2010) (Adams, J., 2009).

There are various ways in which plants can adapt to their local environment, this is often illustrated not only in phenotypic, but also genotypic changes of a plant’s biological make-up. A study in various alpine landscapes found that diversity in the grass species *Poa alpina*, was affected by their localities. This along with variation in phenotypes was found as a result of local adaptations

(Stöcklin J. et al., 2009). Local adaptation along with variation in breeding strategies may suggest significance in an ever-changing environment. This is of interest for different species under different environments. Evolutionary processes such as genetic selection and drift may provide opportunities for further study to the importance of ecological genetics to plant adaptations, whereby genetic mechanisms can be outlined for a specific species to understand their evolutionary processes and ecological implications (Anderson J. et a*l*., 2011). Differences in each gene pool can be used to detect processes like genetic drift and speciation. In turn, this can influence performance of different populations. Local adaptation studies in aquatic plants, reveal variation in performance between Northern and Southern populations which alludes to different genetic performance due to genetic drift and inbreeding (L. SantamarÍa *et al*., 2003). Population genetic studies of different *Linum* population could potentially reveal genetic factors which could contribute to performance of different populations. This can then be utilised to underpin genetic loci to manipulate *Linum* under different environments. This is particularly of interests in agriculture.

In the flowering plants, flowering strategies often form an important set of adaptations to the local environment. For example, in *Mimulus guttatus* populations, selection to flower at the correct time causes them to exhibit local adaptation to the environment (Hall M. et al., 2007). One environmental factor influencing flowering is the sensitivity of an individual’s flowering time to the availability of sunlight in their local area (photoperiod sensitivity). Environmental selection on flowering times was observed in *Campanulastrum americanum* to influence reproduction under different natural light conditions (Galloway F., 2012). Responses to selection in flowering time is of interest as flowering affects reproductive mechanisms of plants in the wild and affects offspring health and survival. In the perennial plants, a single genomic region mediates local adaptation. This region contains the floral integrators *FT* (Wang J. et al., 2018). Genetically, early/late flowering populations is suggested to show differences in their expression of flowering time genes (Reeves P.H. et al., 2000). This in turn reveals genetic differences in flowering time strategies for different populations. In the model *Arabidopsis thaliana*, flowering time gene network and expression under different treatments reveals several pathways, with photoperiod and vernalization pathways being the most well established (Putterill J. et al., 2004). Studies reveal CONSTANS (*CO*) and FLOWERING LOCUS C (*FLC*) genes are key to photoperiod and vernalization perception and regulates FLOWERING LOCUS T (*FT*) (Kinmonth-Schultz *et al*., 2021). With the suggestion that the floral integrator *FT* is in the genomic region that could potentially mediate local adaptation, it is of interest to observe variation in flowering-time genes and phenotypes.

For agriculture, the importance of local adaptation may reveal conditions for optimal yield of a cultivar type. This is of interest to stakeholders in which diverse agroecosystems can be optimized to prioritize climate adaptation responses in an ever-changing environment (Lee et al., 2014). This could be important in sustainable agriculture, which is of interest to stakeholders in the agricultural sector (Öhlund et al., 2015). In cultivated Maize (*Zea mays*) for example, genetic variants were found to be associated with flowering time, an important adaptation mechanism in temperate plants (Li *et al*., 2016). In cultivated apple trees (*Malus domestica*), genetic mapping of different cultivar types suggested a cold perception mechanism which was linked to the flowering time network of genes (Allard et al., 2016). It is of interest then to observe whether local adaptations to environmental variables in cultivated lines is affecting the performance of such lines in terms of their sustainable production in certain environments and to find optimal treatments for important cultivars to yield products optimally.

As a crop, factors such as plant height, number of stems and as well as seed sizes are of interest. It is of interest to look at whether seed size contributes to differences in reproduction as well as differences in physical formation of seeds based on environmental conditions. It has been suggested that seed size is regulated, and that viability of the embryo is enhanced when the seed contains a substantial amount of starch and protein for the seedlings (Sundaresan V., 2005). This suggests an advantage to having bigger seeds. In the model *Arabidopsis*, it has been shown that environment affected seed size and that an observation was made whereby variance in progenies' morphology were affected by parental environment (Brown et al., 2019). This potentially extends to development across generations. When looking at plant heredity and performing genetic studies, seed sizes are one factor of interest in relation to parental growth conditions. Looking at wild *Linum bienne* grown in different environment may reveal the parental effect regarding morphology structure such as seed sizes. Seed sizes may also be affected by the requirement of seed to "sense cold", otherwise known as vernalization. In addition, population genetics study may reveal structuring and local adaptation mechanisms, previously untknown in *Linum*.

## 1.2 INTRODUCTION TO STUDY SPECIES

In this thesis, two species of the flowering plant in the genus *Linum* were studied. *Linum* is a genus of flowering plants which contains more than 200 identified species, distributed in the subtropics and temperate zones (Öhlund et al., 2015; Muravenko et al., 2010). *Linum* is found in temperate into the subtropical regions with species such as the blue flaxes *L. lewisii* and the *L. narbonense* often cultivated as garden ornamentals (Ionkova et al., 2013; Addicott, 1977). In horticulture, prolific flowering makes them of interest in ornamental uses (Tork D. et al., 2022. Geographically, *Linum* is distributed across Northern Americas, and the European continents. Another cultivated species which is important industrially includes the *L. usitatissimum* (Jhala A. et al., 2010). The wild ancestor of this cultivar is thought to be the pale flax *L. bienne* (Uysal et al., 2010). Allaby *et al* (2005), reveals that there is a single domestication event for *Linum,* and they suggest that the *sad2* locus of *Linum* was subject to artificial selection in the cultivars. The wild relative is significant because it represents a primary genepool for development of the cultivar. Previous study reveals that there is a genetic differentiation between the two species and suggests that 48% of *L. bienne* alleles were unique (Soto-Cerda et al*.,* 2014). With this suggestion, it is also possible to map *L. bienne* sequences to the cultivars to identify alleles favourable to the cultivars (Soto-Cerda *et al.,* 2014). In this thesis, both wild (*L. bienne*) and cultivar (*L. usitatissimum*) individuals will be examined.

Morphologically, *L. usitatissimum* tends to have taller stems when compared to *L. bienne*. They also tend to have larger seeds, capsules, and flowers. The pale flax (*L. bienne* Mill) was previously shown to vary in flowering phenology (Uysal et al., 2011) and the cultivated flax, *L. usitatissimum* has been extensively studied in terms of its phenology with studies suggesting development of certain phenotypes to optimise yield (Rehman et al., 2014). Wild and cultivated flax differ in their stem structure seeds (Figure 1) .

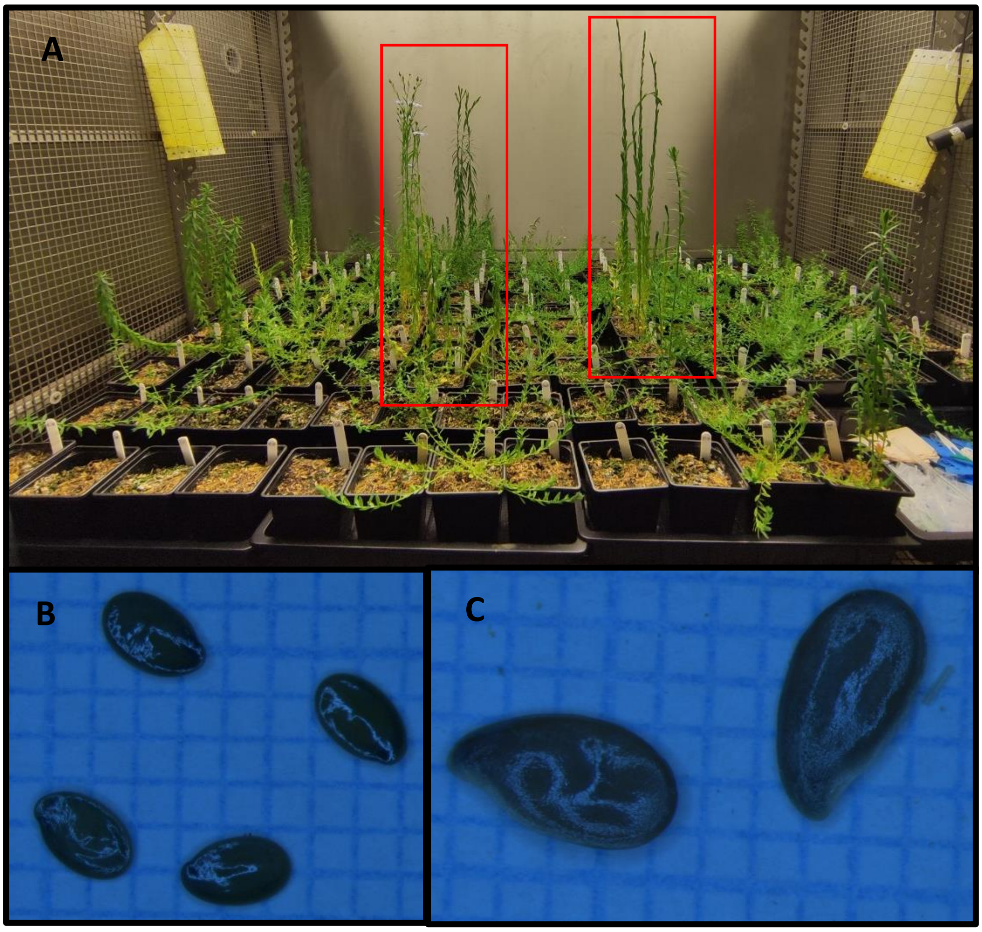


Figure 1. Collection of wild and cultivated S1 (selfing) *L. usitatissimum* and *L. bienne* grown for >3 months period. Cultivars (in red boxes) are growing taller, sturdier stems than wild relatives. There is variation in plant heights and number of stems in our *Linum* samples (1A). Wild (*L. bienne*) and cultivar (*L. usitatissimum*) relatives also have varying seed sizes. Wild relatives (1B) tend to be smaller in length and width than the cultivars (1C). Each square is 1mm in area for Figures 1B and 1C.

The availability of *L. usitatissimum* whole genome sequence (Wang et al., 2012) means that studies looking at implications in the cultivars due to domestication processes were possible and thus the ability to infer some genetic consequences from domestication were revealed (Fu, 2012; Fu, 2011). Although the availability of a whole genome is true for *L. usitatissimum,* there is currently no published whole genome for wild species of *Linum.*

### 1.2.1 *Linum usitatissimum:* Cultivated Flax

*Linum usitatissimum,* otherwise known as the "common cultivated flax" is an agriculturally important plant that serves multiple purposes (oil and fibre morphotypes). Therefore, *Linum* is of economic significance in agriculture. In terms of their morphology, You F. et al., 2017 found that fibre morphotypes have greater straw weight, plant height and protein content than the oil morphotypes while the oil morphtypes have greater seed weight, seed oil content, and branching capability (You F. et al., 2017). Measures of oil and stem fibre content of several oil/fibre morphotypes also shows variability, suggesting variation within this species for industrial use (Diederichsen and Ulrich., 2009; Rozhmina T. et al., 2021).

Historically, *Linum usitatissimum* has been cultivated for its fibres and seeds, with evidence of domestication in the Middle East, revealed by the diversity of the *sad2* gene locus (Allaby et al., 2005). Plant domestication is an important process which enables human civilizations to utilize plant materials for advancements (Diamond, 2002). In terms of agriculture, traits for larger seed (size and weight) and fibre properties have been selected for, when considering flax as cultivars with development for said traits still of interest to this day (Rahimi M. et al., 2011; Yan L. et al., 2014; Guo D. et al., 2020). As a result of domestication, genes for seed sizes, flowering time, and capsule dehisence are revealed to be artificially selected (Zhang J. et al., 2020). More recently, qualities of cell wall, stem strength and fibre properties are studied for using flax fibres as an environmentally friendly composite alternative (Goudenhooft C. et al., 2019). Selective breeding in a trait may imply better selection for that trait, however in long term, the consequences to those selected lines in terms of survival and loss of functions in other genes have not been tested yet. It has been shown that the temperate region of the world started to cultivate flax thousands of years ago. Archaeological records illustrate that flax were cultivated during the Viking and Medieval age (Ejstrud, et al., 2011) (Nag, et al., 2015). This suggests a process of long-term artificial selection in *L. usitatissimum* with implications in several field of interests for agriculture. We are interested in flowering time for this study. We suggest that selection for faster flowering in the cultivars may result in loss of signalling in flowering time genes.

*Linum* seeds produces oleic compounds composed of triglycerides, and these are particularly rich in α-linolenic acid, which gives linseed oil its ability to polymerize into solids. This is useful in productions of industrial resins and solvents (Vereshchagin, & Novitskaya, 1965; McCullough, *et al*., 2011), suggesting significance in the industry sector. In addition to this, recent research has found that linseed oil has a positive effect on reducing the level of cholesterol in rabbits, illustrating potential for linseed in the health food sector (Króliczewska, et al., 2018). Molecularly, it is suggested that *Linum usitatissimum* seeds contains several biologically active components that are useful both medically and industrially. This includes the presence of lignin in linseed, useful in resin production (Del Rio *et al*., 2011; Touré & Xueming, 2010). These lignans have shown to be medically beneficial in relation to their antioxidant properties (Hosseinian, et al., 2006). Fatty acid profiling has also revealed that linseeds provide significant nutritional values (El-Beltagi, et al., 2007), suggesting values in food security. Other potential uses of cultivated *Linum* species is their biogeographic trait variation. This could be useful for ornamental purposes. In recent study of the ornamental *L. lewisii*, traits such as flowering indeterminacy, seed mass and stem numbers suggest its potential use in field restoration and agriculture (Innes, et al., 2022). These findings reveal that not only the common cultivated flax was cultivated early on for their fibres, but also research into flax biproducts and genetics is still ongoing with a strong potential in multiple fields such as industry, agriculture, the health food sectors and horticulture. With these in mind, it is then of interest to optimise *Linum* crop production. Examples of research in this area includes seed priming which influences crop growth and development (Rehman H. *et al*., 2014), and experiments to employ plant growth regulators (giberellic acid) for optimal yield and growth in *L. usitatissimum* (Rastogi A. *et al*., 2013). We propose that, information into local adaptations and genetic implications of *Linum* wild crop relatives could aid these efforts to further improve *Linum* as a crop plant. This may be done by exploring genetic and phenotypic responses to environment changes to benefit agriculture. To this day, research for implementation of wild relatives towards genetic engineering cultivated flax is relatively scarce. For studies in this thesis, 18 varieties of *Linum usitatissimum* seeds were received from IPK World Collection (https://www.ipk-gatersleben.de/en/). A list of cultivars and their morphotypes can be found in appendix 1.

### 1.2.2 *Linum bienne:* Wild Flax

*Linum bienne* is a biennial plant and a wild relative of *Linum usitatissimum* (Gill KS, 1966). Its common name is the "pale flax", considered the wild forebear of the cultivated *Linum usitatissimum*. A study by Allaby *et al*, reveals this by exploring *sad2* loci from 30 accessions of wild and cultivated flax. They found phylogenetic evidence that the wild type *L. bienne* was first domesticated for oil and that there is an artificial selection of the loci, indicative of cultivation of the wild types *L. bienne* (Allaby, et al., 2005). Further studies also revealed that *sad2* locus is a candidate domestication locus associated with increased unsaturated fatty acid production in cultivated flax (Fu Y.B. et al., 2012). This suggests implication for seed selection when *Linum* was first domesticated as opposed to fibre. In the literature, there are suggestions that *L. bienne* are widely distributed around Western Europe, the Mediterannean basins, North Africa and into Iran and the Caucasus. This represents its wide distribution in sub-temperate into temperate zones (Zohary and Hopf, 1993). Just like the cultivar *L. usitatissimum*, this wild progenitor has the same number of chromosome (*n=15*) and they're both homostylous, in contrast to some distylous relatives such as the *Linum Tenue* (Gutiérrez-Valencia *et al*., 2022).

Previously the wild *L. bienne* were considered as sub-species of *L. usitatissimum* (Uysal et al., 2011), but now they are considered as the closest relative to *L. usitatissimum*. Phylogenetic studies have illustrated that *L. bienne* is the closest relative of L. usitatissimum (Uysal, et al., 2010). Further phylogenies from studies of *Linum* orthologs also suggests that *L. bienne* is the closest relatives of *L. usitatissimum* (McDill J. et al., 2009; Sveinsson S. et al., 2011). This species represents a wider gene pool, which can be of aid for *Linum usitatissimum* development in agriculture.

*L. bienne* is native to the Mediterranean and parts of Western Europe and are spread as far north as the Scandinavian countries. Naturally, growing in different environments compared to *Linum usitatissimum*, the *bienne* represent a wider gene pool, which could be of interest in the enhancement of *Linum usitatissimum*. Differences between wild populations could shed light on the evolutionary trends of this species as well as their cultivar relatives. In the literature, population genetic study of wild relative of flax is relatively scarce. Studies conducted in this thesis will suggest some implications about the importance of wild relatives for future studies looking at the genetics of *Linum*. Across Western Europe wild *L. bienne* are distributed across a range of latitude. For studies in this thesis, collection of seeds was essential. Seeds from the wild were first collected in 2016 by Dr. Adrian Brennan of Durham university and Dr. Rocío Pérez-Barrales of University of Granada. The plants collected throughout western Europe. Different populations with different number of individuals were used in each aspect of studies in this thesis. We used this collection in the corresponding chapters.

### 1.2.3 Production Values of *Linum* Biproducts:

With the above put into consideration, the current observation is that flax output is economically significant. One of the biproducts of *Linum* is its fibre. Flax fibre production in 2020 stands at 976,113 tonnes (FAOSTAT data, 2022). In 2016 alone flax fibre output reached a production value of 535 million USD. This value increased more in 2020, with flax world trade value estimated at 726 million USD (Flax Fibers | OEC, 2022). The world's largest flax fibre exporter in 2020 is France, with an estimate of 51.4% in total world export value. The export share in the UK only stands at 0.12% with an export value of 870 thousand USD. The world's top importer of flax fibres is China with 47.8% total import value which stands at 347 million USD. The UK imports 0.14% of total flax fibre import value, with an import value of 1.03 million USD (Flax Fibers | OEC, 2022). The export value growth from the year 2000 – 2020 in the UK is -76.7%, with export values, massively reduced from 3.74 million USD at the start of this period. This reveals, that at least in the UK, current export value in flax fibre is decreasing. This being the case, in developing countries demands for flax fibre are increasing. In India alone, the import value increased from 5.93 million USD in the year 2000 to 47.7 million USD in 2020, an increase of 703%. Other countries such as China reveals an import value growth of 254% between the year 2000 and 2020 while Vietnam shows a >1000% increase (Flax Fibers | OEC, 2022). This new trend in demand of flax fibre in developing countries could provide a reason for temperate countries such as the UK to increase export of flax fibres.

Other major biproduct of *Linum* cultivars is linseed. In 2020, the total world production of linseed is 3,367,331 tonnes, with Kazakhstan producing the most at 105,8247 tonnes (FAOSTAT data, 2022). In the same year the trade value of linseed worldwide is 980 million USD. Major exporters of linseed include Canada (25.8%), Russia (23.5%) and Kazakhstan (21.7%). Major importers of linseed include China (26.8%), Belgium (22.7%) and Germany (8.75%). In the UK, export of linseed grew 57.7% between the years 2000 – 2020. The total export value of linseed in the UK in 2020 is 7.28 million USD. The biggest increase is seen in Kazakhstan between the years 2000 – 2020 (Flax Fibers | OEC, 2022). The increase in export of linseed illustrates the economical values linseed represents worldwide in terms of industry and agriculture. Potential to improve cultivar types will be of huge interest to the economic values represented in *Linum* biproducts derived from flax fibres and linseeds.

---

## 1.3 LITERATURE REVIEW OF *LINUM* GENETIC DIVERSITY

*Linum* underwent events of genome duplication and domestication. With *Linum* domestication, it is worth considering the consequences that may have come with this process. While vital to human development, domestication of wild plants often leads to a genetic disadvantage because of divergence from the wild that could be caused by selective breeding, associated with the process of domestication. This is often due to selection for “preferred traits” (Purugganan & Fuller, 2009). Selected traits, either directly or indirectly selected, may in turn have genetic implications within the domesticated populations. Fu Y. B., (2012) revealed that cultivated flax have 27% reduction of nucleotide diversity when compared to the wild pale flax (*L. Bienne*), perhaps because of domestication. Genetic effects such as the pleiotropic effects (Conner, 2002) and linkage disequilibrium (Falconer, D.S., 1996) may also have occurred because of selective breeding in domestication. This could lead to further genetic consequences such as reduced genetic diversity, and increased chances of genetic drift (Rauf, et al., 2010).

Genetic diversity analyses and studies with different *Linum usitatissimum* cultivars are widely available (Nag, et al., 2015; Diederichsen, 2001). These resources can be used for breeding programmes in *L. usitatissimum*. Past studies have identified morphological traits-based diversity in *L. usitatissimum* seed variation (seed colour, seed weight, and seed oil contents). They suggest indirect selection for seed weight and seed colours (Diederichsen and Raney, 2006). Seven linseed genotypes have previously found to have quantitative traits differences between seed traits (Nôžková J, *et al.*, 2014), suggesting variation in *L. usitatissimum* genotypes. Zhang *et al*, (2020) conducted a genome wide association study for *L. usitatissimum* and suggested that during flax domestication, genes relevant to flowering, dehiscence, oil production, and plant architecture were artificially selected and that selection in these genes may shape their morphology . There is a suggestion here of artificial selection in *L. usitatissimum*. In addition, molecular markers are also widely available for *L. usitatissimum* genetic studies. Cloutier *et al,* (2009) have defined 83 Single-Sequence Repeat (SSR) motifs for 23 *L. usitatissimum* accessions. In addition to this, another study defined 28 SSR markers for the study of genetic fingerprinting in *L. usitatissimum*, which is useful in assessing genetic purity (Pali V *et al.*, 2014). A consensus of genetic and physical maps of *L. usitatissimum* is also available under high resolution (74% of the estimated flax genome) (Cloutier S. *et al.*, 2012). These genetic resources don’t consider wild relatives, which may be of further interest in this area, and when looking at breedable traits. Initial studies looking at the wild relatives *Linum bienne,* illustrates that they represent a wider genetic diversity from which the cultivar is derived (Uysal et al., 2010). Soto-Cerda et al also suggested genetic association mapping is possible using wild relatives of *Linum* to identify favourable alleles within the wild (Soto-Cerda *et al.,* 2014).

### 1.3.1 Methods for studying population genomics in *Linum*

Within this study, we compare genetic diversity of wild *Linum bienne* samples, relative to the cultivars. With regards to genetic analysis, it is worth noting Next-Gen sequencing (NGS) techniques. NGS makes it possible to study local adaptation at a population level without a whole genome sequence. This is of interest when looking at population genetics of wild relatives of agricultural plants (Park & Kim, 2016). In relation to our study model, NGS is valuable in terms of gaining insights into the genetic diversity of *L. bienne* and *L. usitatissimum.* Population analysis of wild relatives may also be informative in the development of crop relatives. Several NGS techniques are available, and the sequence data can be mapped de-novo or to a reference genome. This is relatively faster and cheaper than producing a whole genome for the wild relatives.

With next generation sequencing technologies available, several *Linum* molecular markers (microsattelites) have been developed (Fu & Peterson, 2010). It was shown that molecular markers are transferable in many species of *Linum* (Soto-Cerda, et al., 2011). These transferable molecular markers, often referred to as simple sequence repeats (SSRs), have been useful for data enhancement for genetic and evolutionary studies. Because of this transferability, genetic and evolutionary studies can compare between wild and cultivars of flax. Numerous species in the *Linum* genus represent an interest to study their evolution, while the cultivar (*L. usitatissimum*) represents an interest in the development of *Linum* as a crop through study of their genetic makeups and local adaptation mechanisms such as self-incompatibility mechanisms (SI) and inbreeding depression which their wild relatives may be able to reveal.

For whole genome sequencing, Illumina’s Hi-C sequencing has become a mainstream sequencing technique. It is preceded by the 3C method (Downes et al., 2021). The method is widely used to examine organisation and conformation of chromosomes and secondary genetic structures into chromosomal level organisation based on whole genome sequencing (Belton et al., 2012; Lieberman-Aiden et al., 2009). The method involves the use of isoschizomer restriction enzymes such as DpnII to digest cell nuclei isolates, recognising and cutting amino acid sites to generate an overhang. These overhangs can then be used to to enable paired-end sequencing using adaptor ligation (Belaghzal et al., 2017). However, to gain whole genome sequences often requires laborious procedures that are relatively more expensive than reduced representation sequencing techniques such as RAD Sequencing. This sequencing technique can help with identification of markers such as single-nucleotide polymorphisms (SNP) that can be used to identify genetic signals such as structuring between population and even identify novel loci for quantitative trait locus analyses.

The genome of *L. Usitatissimum* is estimated at 370Mb in size (Wang et al., 2012). Whole genome sequences are publicly available down to chromosomal levels for the cultivated *L. usitatissimum* having first been assembled using shotgun sequencing technologies in 2012 (Wang Z. et al., 2012). The contig assembly contained 302 Mb of non-redundant sequence, representing an estimated 81% genome coverage (Wang Z. et al., 2012). The initial genome is now referred to as CDC Bethune v1 (Sa et al., 2021). The most widely used *L. usitatissimum* whole genome is the CDC Bethune v2 (Sa et al., 2021; You et al., 2018). For this whole genome contigs contains a total N50 of 6.64Mb summing up 316 Mb of reads and a 97% coverage of annotated genes, considerably higher than the previous CDC Bethune v1 (You et al., 2018). Another genome was assembled using long reads for the Atlant variety of *L. usitatissimum*. Using a combined Oxford Nanopore and Illumina sequencing technique, whole genome with a total length of 361.7 Mb, N50 of 350 kb, and 97.40% completeness was recently achieved for the Atlant variant of the *L. usitatissimum* (Dmitriev et al., 2021). The availability of this whole genome also suggests that continuity of these assemblies is relatively poor as revealed by the relatively small contig N50. The largest contig N50 was found to be 365Kb (Sa et al., 2021; Dmitriev et al., 2021). They suggest that events such as recent whole-genome duplication may result to the collapse of homologs and repeat sequences during the assembly process, both under short and long reads sequencing (Sa et al., 2021; Dmitriev et al., 2021).

For the wild *L. bienne*, whole genome sequences are under development and not publicly available. However, past studies using inter simple sequence repeats (ISSR) and RADSeq markers have shown that the wild *L. bienne* illustrates ancestral relationship with the cultivated *L. usitatissimum* and there were indications that suggests the contribution of *L. bienne* to *L. usitatissimum* genome through gene flow (Fu, 2012; Gutaker RM., 2014). Molecular and cytogenic studies of both cultivated and wild *Linum* were also undergone using high-throughput ribosomal RNA. The studies suggests that 5S rDNA and ITS phylogeny is closely related between *L. usitatissimum* and *L. bienne*. Both species have identical karyotypes and distribution of rDNA sites. In both species, the coding sequences of 5S rDNA genes were similar (Bolsheva N. et al., 2017). Comparison of ITS1 to 5.8S rDNA-ITS2 sequences showed that rDNA sequences were conservative in all studied flax specimens (Bolsheva N. et al., 2017). In addition to this, Assessment of the number of polymorphic sites in a genetic marker (*LuTFL* homologs) suggests mixed ancestry of the locus and different copy of the locus suggests ancestry to the wild pale flax *L. bienne* (Gutaker RM., 2014). More Northern populations are suggested to have high similarity to the *L. usitatissimum* (Gutaker RM., 2014). This high similarity illustrates that whole genome for the cultivar is acceptable for use in the wild *L. bienne* before the availability for its whole genome.

Other genetic events have also occurred as *Linum* species have undergone polyploidy events. In plants, polyploidy is an event whereby whole genome duplication occurred increasing the number of chromosomes. An example of this has also occurred in sugarcanes (Vilela et al., 2017). Genetically, the genus *Linum* has experienced multiple polyploidy events. The last polyploidy event occurring around 5-9 million years ago (mya) and a previous polyploidy event around 20-40 mya. This reveals that every gene in the flax genome is potentially duplicated and multiplied through polyploidy (Sveinsson et al., 2013). This needs to be accounted for when looking at their genetic make-ups and for studying gene expression (Sveinsson et al., 2013). Polyploidy is important, as it is a widely accepted to be a pervasive mechanism of plants and is often consequently followed by the selective silencing of genes (Wendel & Adams, 2005). Studies have previously inferred that DNA alterations in flax could induce changes that are heritable. These changes were also linked to environmental factors (Schneeberger & Cullis, 1991). Polyploidy events can consequently rid of genes, which were previously present in predecessors of cultivated flax. With the aid of cultivation, breeding strategies of the genus, conditions may have changed and thus the genetic make-up of cultivars is heritably different to those of the wild progenitor, representing a wider gene pool to be studied.

### 1.3.2 Self-Incompatibility in *Linum*

In plants, control of outcrossing is often due to self-incompatibility (SI) genes at the S locus (Newbigin E. et al., 1993). This locus is responsible for the prevention of self-fertilization and therefore selfing in plants. Inactivation of the S-locus receptor kinase (SRK) and the S-locus cysteine-rich protein (SCR) within this locus has led to the loss of self-incompatibility in some plants in the *Brassicaceae* family (Sherman-Broyles, et al., 2007) and *Arabidopsis* (Suzuki, et al., 1999). Self-Incompatibility genes is not yet fully understood in *Linum*, research focuses mainly on the formation of hetero/homo styly. Differences in the S-genes expressed in the male and female reproductive structure trigger self-rejection, controlling outcrossing, and leading to high heterozygosity (Thompson & Kirch, 1992). Heterostyly in *Linum* means that there are morphological differences between styles and anther filaments of *Linum* and it has been found as an outcrossing mechanism in *Linum* (McDill, et al., 2009). Therefore, in *Linum,* polymorphisms of breeding organs contribute to heterozygosity  (Ebert, et al., 1989). In *Linum grandiflorum*, the S-locus controls the flower morphology through regulations of transcriptional S-locus products (Ushijima, et al., 2011). This reveals the connection of the S-locus regulation with the ability for *Linum* to promote heterozygosity and therefore maintain genetic variation. The mating system of wild flax (*L. bienne*) is not yet fully understood, although related species show distyly to ensure outcrossing, research regarding polymorphisms in the wild relatives are still scarce to date. This thesis will explore aspects of *Linum* reproduction by looking at pollen morphology, especially pollen viabilities measures such as the availability of pollen to germinate under environmental stress, This will provide trait data to study genetic mechanisms which may affect such morphology.

### 1.3.3 Population Genetics Analysis

In modern genetics, there is an interest in determining genotypes which represent underlying phenotypes. These are often crucial to a study model with potentially many populations, representing a wide gene pool. To do this, it is crucial to gain insight into genetic sequences of many individuals relatively quickly and cheaply. For the purposes of sequencing, high capacity and relatively low costing technology for sequencing were first achieved using microarrays-based sequencing technologies (Schena et al., 1995). Later, widely used genetic markers were introduced as “microsatellites”, also known as Single Sequence Repeats (SSRs) across the whole genome (Vieira et al., 2016). This genotyping technology was a pre-cursor of Next-Generation Sequencing (NGS) technologies such as RAD Sequencing. NGS provides sequence data for nucleic acids, DNA and RNA, that can be analysed as biomarkers specific to regions of a genome. NGS also consists of several techniques used today (Lemuth & Rupp, 2015). NGS genotyping technologies can often be an alternative to the vastly more expensive and time consuming, whole genome strategies to look at a study species. This is especially useful in studies involving species without publicly available genome such as the wild relatives of *Linum* (Hu et al., 2021).

Peterson *et al* (2012) described one type of NGS genotyping technique as double-digest RAD Sequencing or ddRADSeq. The protocol is a variation of the RAD sequencing protocol (Davey *et al*, 2010). DdRADSeq in comparison to RAD sequencing, uses a second restrictive digestion step to improve the size selection step of the protocol (Peterson B. et al., 2012). This involves the use of a second restriction enzyme and a second indexing step which allows for combinatorial indexing, more specific to every individual (Peterson B. et al., 2012). When compared to a traditional RAD-Sequencing technique, a double-digest RAD Sequencing can exclude regions which are flanked by a very close or very distant recognition site. This then recovers libraries which consists of fragments close to the specific target sequencing size (Peterson B. et al., 2012). The use of multiple enzymes also contributes to a more diverse size selection availability in comparison to the traditional RAD Sequencing technique (Peterson B. et al., 2012). In the study by Iguchi *et al* (2020), they were able to identify diversity and selection mechanisms which leads to genetic variation in deep-sea amphipods using a RAD sequencing protocol. This illustrates the significance of RAD sequencing for the use of population diversity analysis, which is of interest when looking at our set of wild *L. bienne* samples.

ddRADSeq identifies Single Nucleotide Polymorphisms (SNPs) as genetic markers. SNPs markers are simpler to genotype than microsatellites and are widely used in fine-scale population genetic studies (Liu et al., 2019). The use of ddRADSeq have also been implied in herbarium samples which have a relatively lower quality of preserved DNA, this protocol proves useful for obtatining genetic datasets from a study species which is relatively fast to degrade such as plant leaf materials (Jordon-Thaden et al., 2020). ddRADSeq have also been effective for genotyping oragnisms with a larger genome size. A study using a non-model orchid species (genome size of 31.6Gbp) were viable for ddRADSeq protocol, with inference of genetic diversity and differentiation (Gargiulo and Fay., 2020). They also showed the requirement for a higher standard of quality (DNA extraction to library preparation) when working with non-model organism of which their whole genome is not available for reference (Gargiulo and Fay., 2020). This will prove as a challenge under the study model of this thesis, as there is currently no whole genome publically available for the species *L. bienne*. This will mean that any *de-novo* mapping of the extracted short sequences, will have to be of exceptionally high quality. However, the use of *L. usitatissimum* whole genome is viable for reference to the later mapping steps.

DdRADSeq uses indexing adaptors to “barcode” certain regions of a genome that have been digested using two specific restriction enzymes. Subsequent bioinformatics analysis is then able to highlight SNPs. The technique can be more cost effective than the previously developed SNP Chip methods (Vieira et al., 2016). For any sequencing purposes, the first challenge is to extract a sufficient amount of good quality DNA materials. There are multiple DNA extraction protocols available to use. Nowadays, there are DNA extraction kits available to purchase such as Qiagen’s Dneasy kits. For ddRADSeq purposes, the most used protocol is a CTAB DNA extraction protocol (Jordon-Thaden et al., 2020). CTAB DNA extractions can also be modified to target for high polysaccharides, which may be the case in plant material extractions (Porebski et al., 1997). For ddRADSeq purposes, other studies have used between 0.1 – 0.8 µg of total DNA per sample (Jordon-Thaden, et al., 2020). This relatively high amount of DNA may represent a challenge in the extraction of *Linum* DNA materials. During this study, optimisation of ddRADSeq protocols for *Linum* will be performed and DNA extraction protocols will be explored in the interest of extracting relatively high-quality DNA required to build ddRADSeq libraries. To build ddRADSeq libraries, one of the most important step is to select restriction enzymes which are specified for digestion at associated length (250-500bp for this study). This would enable short reads over many individuals which are subsequently barcoded. For the barcoding process, a ligation process can be applied to the digested

DNA materials. Adapters (single indexing) can be ligated to the DNA materials at both ends of the restrictedly cut DNA materials. Libraries can then be pooled and sequenced.

When successful, ddRADSeq outputs forward and reverse reads of the library. Reads can then be mapped either *de-novo* or to a reference genome before further analysis could take place (Peterson B. et al., 2012). In the interest of illustrating population structure and diversification within our wild and cultivated samples, population genetic analysis is essential. Population genetic analysis is often useful for genotyping studies involving samples from a wide range of localities. Population genetic studies often utilise mathematical models to evaluate how a population varies in terms of their genetics (Servedio et al., 2014). As part of an evolutionary biology study, protocols into population genetics may infer adaptation, speciation, and structuring across populations which may infer biological conclusion such as local adaptation, inbreeding depression, and potential transfer of genetic material from one population to another in the sample set (Hoban et al., 2016).

Often local adaptation can lead to population genetic structures, with population forming specific niche to the local environments (Leimu & Fischer, 2008). In plants this can be an important mechanism in which biotic interactions can often lead to responses to environmental changes such as elevational changes (Grassein et al., 2014). For *Linum,* observations of population genetics for the wild relatives are scarce compared to wild relatives of other species such as the wild crop relatives of cabbage, *Brassica oleracea*. Population genetics studies in *Brassica oleracea* have revealed population structuring along Western Europe for their wild relatives (Mittell et al., 2020). The structuring can suggest diversity in wild relatives of cultivars, which represent a wider genepool for the improvement of cultivar relatives. While this is explored in the *Brassicaceae*, it hasn't been explored in the wild relatives of flax, *L. bienne*. In another example, population genetic analyses suggests gene flow and hybridization in 12 plant species suggesting extinction due to intogression of cultivar types to the wild (Ellstrand N. et al., 2003). This adds to the relevance of population genetic analysis, especially in illustrating genetic relation in wild and cultivar relatives of plants. In this study, initial population structure analysis will look to answer the question of "Can genetic distance infer population structuring within the Western wild samples of *L. bienne* and its cultivar relatives?" Other relationships in population genetics will also be investigated for evidence of genetic diversity within our samples.

### 1.3.4 Genetic Expression in Flowering Time:

Within the model *Arabidopsis,* various environmental cues are known to affect the floral regulatory network genes with light and temperature being the most significant (Li et al., 2015).

Timing of flowering in plants is often determined by environmental and endogeneous factors. Environmental factors that have been shown to influence flowering time includes availability of nutrients, ambient temperatures, drought, salinity, and the presence of pathogenic microbes (Cho L. et al., 2017). Further, genetic expression of certain genes in the network affects other genes. As an example, expression in genes such as *SOC1* have been shown to affect further genes in the pathway such as *AP1* and *LFY*. There is an importance in quantitative input to each gene in the network for expression in flowering times (Leal Valentim et al., 2015).

To look at specific gene expression within plants, specific genes can be targeted using primers and real-time qPCR studies (Higuchi et al., 1992). Primers are relatively short stretches of nucleic acid sequences which can uniquely identify a specific region of the whole genome (Cox and Doudna, 2015). They complement and amplify specific nucleic acids using apolymerase enzyme. Therefore, these primers are useful for studies looking at specific gene expressions.

With additional information, and access to sequencing technologies; this thesis will aim to better understand flowering responses in *Linum* by generating genetic expression data. In the studies performed in this thesis, both *L. usitatissimum* and *L. bienne* genetic materials were collected, tested, and analysed regarding differences in species and population localities. It is of interest to find out whether there are differences in genetic expression of flowering time related genes in response to environmental factors such as temperature. How would these changes in genetic expression measure up against population localities, and climate within those localities? Within this thesis, we hypothesize that differences in relative gene expression are present in our *Linum* collection and that this is correlated to population localities and climatic variables, reflective of local adaptation.

### 1.3.5 Observation of Pollen Viabilities

In addition to observing variation in flowering time due to response to the environment, we wanted to measure other adaptation in flowering for *Linum*. Since most studies already looked at hetero/homo styly, we approached this using the viability or *Linum* pollen. In addition to flowering time, we investigated if temperature affects the number of pollen, pollen tube formation and its ability to germinate the ovule. Pollens are an important part of a plant's breeding strategy and act as a male counterpart to the ovule. Plants often show differences in pollen response related to differences in breeding systems of said plants (Hanley et al., 2008). In addition, climate has been observed to affect pollen performance in plants (Jan Ejsmond et al., 2011; Iovane M. et al., 2022).

Temperature acts as a major climatic factor which may limit the germination of pollen in plants. It is deemed that temperature changes affects distribution of angiosperm species as the reproductive stage is proven to be vulnerable to temperature. In the cultivated peach, it is observed that increases of temperature to 20°C from their original 5.7°C in the field results in reduction in receptivity of stigmas to pollens (Hedhly A. et al., 2005). They suggested that effects of temperature on male and female organs of a flowering plants may be species-specific and could provide plants with a level of plasticity to withstand environmental effects, such as ambient temperature, to make sure they achieve fertilization (Hedhly A. et al., 2005). An older study in *Trifolium repense*, suggests that growth of pollen tubes is affected by temperature and that pollen tube penetration into the ovary is also observed to be negatively affected by colder and warmer temperatures (Chen and Gibson, 1973). This effect in pollen viability due to temperatures could even be observed in plants originating in warmer climates. In a species of groundnut (*Arachis hypogea*) it is observed that warmer air temperatures (>35°C) results in failure of the setting of pods and pegs due to its lower pollen viability. They found that for pollen germination and tube growth rate were less predictive in discriminating genotypes for higher temperature than other parameters such as pollen tube length and pollen germination rates (Kakani V. et al., 2002). This suggests plasticity in pollen viability parameters for discriminating genotypes against temperatures in plants. In tomatoes (*Lycopersicon esculentum*), it is suggested that temperatures affect *in-vitro* pollen germination and hybridization of populations originating from different altitudes (Zamir D. et al., 1981; Maisonneuve and Den Nijs, 1984). Studies regarding pollen germination sensitivity to temperature stimulus can be achieved *in-vivo* or *in-vitro*, with *in-vitro* procedures proving to be more elusive in plants (Boavida and McCormick, 2007).

In several *Linum* species, it has been recently suggested that *in-vitro* germination is possible to achieve up to 50% pollen germination rates (Lyakh and Soroka, 2021). However, they didn't include the species *L. usitatissimum* or *L. bienne* within their study. Using the various sample populations across western Europe that were in our collection, it is with interest to look at whether pollen viability as pollen performance measure were affected under different treatments and whether populations correlate with geographic and climatic variables, such as latitude, temperature, and precipitation. We hypothesize that geographic and climatic variables affect pollen performance and viability in our *Linum* collection and therefore its potentially affecting plasticity within population fertilization strategy.

# CHAPTER 2: FLOWERING TIME IN LINUM DUE TO VERNALIZATION AND EXPRESSION OF FIVE FLOWERING TIME GENES.

In plants, it is widely accepted that environmental and climatic variables affect gene expression changes which may in turn affect their performance as a response to changes in the environment (Bigot et al., 2018; Elfving et al., 2011). Flowering initiation is of interest in population survival as flowering is vital for reproduction and fitness in flowering plants. Flowering initiation also have a major influence in seed dispersal and maternal effects (Galloway and Burgess, 2009; Giménez-Benavides, Escudero and Iriondo, 2007). In species with agricultural significance, early flowering types are actively selected for advantageous benefits such as late season drought avoidance and better fruit development (Lotz L., 1990; Shavrukov et al., 2017; Ibrahim et al., 2018). In the model *Arabidopsis thaliana*, environmental changes in photoperiod and temperature affects induction of the FLOWERING LOCUS T (*FT*) gene in their leaves, affecting activation of downstream floral meristem genes (Song et al., 2013). Expression of *FT* has been linked to internal and environmental factors such as plant age, secretion of phytohormones (gibberellic acid) and ambient temperature (Song et al., 2013). Furthermore, Lee et al suggests that the alteration of *FT* activities, under different temperatures, regulate expression of downstream the floral meristem gene SEPALLATA 3 (*SEP3*) (Lee et al., 2012).

Several studies have identified key flowering time genes that act as primary determinants of the flowering-time network in the model *Arabidopsis thaliana* (Ballerini and Kramer., 2011; Welch S. et al., 2004; Figure 1). Amongst these genes, the most rigorously studied are, the floral pathway integrators such as FLOWERING LOCUS T (*FT*) and the SUPPRESSOR OF OVEREXPRESSION OF CONSTANS (*SOC*), with FLOWERING LOCUS C (*FLC*) also studied (Sasaki *et al.*, 2017; Li *et al.*, 2015; Valentim *et al.*, 2015). In the model *Arabidopsis,* it has been illustrated that, *FT*, and *SOC* loci expression activates downstream flowering meristem identity genes such as LEAFY(*LFY)*, APETALA1(*AP1*), SEPALLATA3 (*SEP3*) and FRUITFULL(*FUL*). This is because both *FT* and *SOC* are considered transcription factor loci, and in the *Arabidopsis* flowering network, they are considered as floral integrators (Welch S. et al., 2004; Lee and Lee, 2010). Furthermore, Ballerini et al (2011) suggested that there is an element of conservation of *FT* homologs within flowering plants. Although this is the case, they suggest that there is complexity in their regulation and evolution that is still of interest in different angiosperm species to this day (Ballerini and Kramer., 2011). This suggests the importance of *FT* as a floral regulator and the potential expression variation it may have in different species, potentially due to environmental stimulus. In addition, there are two pathways that are worth looking at in response to light and temperature regulation of *FT*. These are the Photoperiod and Vernalization pathway. The expression of photoperiod genes CONSTANS (*CO*) and GIGANTEA (*GI*) can activate *FT* due to photoperiodicity (Kurokura et al., 2017; Song Y. et al., 2014) and the vernalization gene FRIGIDA (*FRI*) is responsible in determining variation of flowering time due to vernalization requirements, defined as the requirement for seeds to sense cold to flower faster (Shindo et al., 2005). Finally, the downstream meristem genes function as an initiator for floral organ developments. Upregulation in these flowering meristem genes signals groups of meristem cells to develop into flowering cells, instead of shoot cells (Teper-Bamnolker and Samach, 2005), thus resulting in floral intitation after complex regulations of several genes in response to the environment.

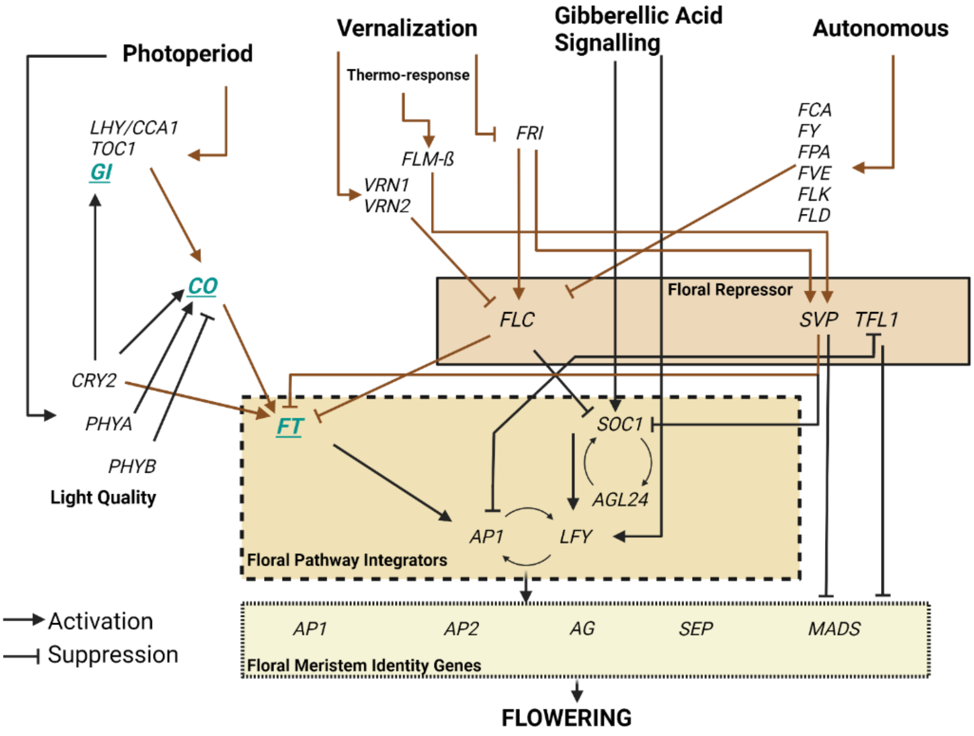


Figure 1. An overview of the flowering time network in *Arabidopsis thaliana* adapted from the literature (Sharma N. et al., 2020; Leitjen W. et al., 2018; Chen et al., 2018; Ballerini and Kramer., 2011; Welch S. et al., 2004) Underlined are the genes tested. We also highlighted pathways that affects *FT* (pathways in brown).

Comparative genomics have also been used to model flowering time pathways in other species such as the temperate grasses (*Brachypodium distachyon*) (Higgins et al., 2010). They suggest that flowering time pathways in *Brachypodium distachyon* are highly like the model *Arabidopsis*. Reviews from Leitjen et al (2018) suggests that despite the overall conservation of these flowering time network genes, there is evidence of divergence of flowering time regulation in both the model *Arabidopsis thaliana* and several crop species due to environmental factors such as temperature. With this, it can be assumed that different plant populations express varied responses to flowering due to the environment. To what extent and which direction each gene is affecting each plant population is still of research interest to this day, especially in non-model, cultivated plants. As an addition, this variation due to flowering responses doesn't just occur in the model *Arabidopsis*. Flowering initiation in non-model organisms such as the common sunflower (*Helianthus annuus)*, have been suggested to be "exploited" by natural selection through genetic expression and tissue-specific expressions of the *SOC1* genes affected by environmental factors. One of these factors is photoperiodicity. Sunflower (*Helianthus annuus*) has shown diversity in photoperiodic responses, due to changes in genetic expression modulated through natural selection (Blackman, et al., 2011). Blackman et al, 2011, also suggests the occurrence of convergent evolution of photoperiod responses in cultivated and wild types of sunflowers, which suggests phenotypic plasticity is weakly constrained by genetic regulation (Blackman et al., 2011). In Japanese wild radish (*Raphanus sativus*), recent research suggest that northern accessions were more sensitive to exposure to prolonged cold conditions, in the sense that they require colder conditions to flower. This was suggested by FLC expression, which is part of the vernalization pathway (see below). The Southern wild radish population seem to be more sensitive to photoperiodicity, suggested by the expression of photoperiodic genes (Han Q. et al., 2021). This is observed to correlate with multiple flowering time gene expressions (Han Q, et al., 2021).

Vernalization is an adaptation, often found in temperate plants, which sense cold to allow optimal timing of flowering initiation in spring (Kim *et al.*, 2009). Vernalization is the requirement for seeds to be cooled from germination for individuals to flower earlier. In the model *Arabidopsis* vernalization is widely studied. The vernalization network has been shown to progress around the MADS-domain protein complex comprising of several flowering time genes forming complexes with the MADS-protein SVP. One of these genes is the *FLC* (Flowering Locus C) which acts as a repressor of *SOC1* (Chen et al., 2018). The closely related *FLM* (Flowering Locus M) also have been shown to form complexes with the MADS-protein *SVP*. These proteins have shown to mediate response to changes in ambient temperature by forming a complex with more specifically the β form of *FLM* at lower temperatures. This in turn represses flowering at colder temperatures (Lee J. et al., 2013; Chen et al., 2018; Posé et al., 2013). At relatively higher temperature, the MADS-protein *SVP* degrades and the complex with *FLM-β* is not formed, thus producing reduced repressive complex, allowing flowering initiation to proceed. Several vernalization related loci play an important part in vernalization sensitivity and interact with other pathways leading to flowering initiation (Blázquez, Koornneef and Putterill, 2001) (Chandler, Wilson and Dean, 1996). In this chapter, we will investigate natural variation in flowering time responses for the temperate plant genus *Linum*. Both wild (*L. bienne*) and cultivar types (*L. usitatissimum*) will be observed in terms of their phenotypic and genetic responses to flowering initiation due to temperature changes in relation to vernalization. This will imply potential adaptation in vernalization responses for both wild and cultivar types. The main environmental variable tested will be temperature, as this corresponds more towards vernalization requirements. However, expression of flowering time genes that may regulate photoperiodicity were also observed.

Studies of *Arabidopsis thaliana* have also established that the latitude differences in populations is associated with co-variation in growth and flowering time (Debieu et al., 2013a). In respect of the wide distribution of *L. bienne* (as discussed in chapter 1), it is of interest to look at differences in genetic expression from a flowering time perspective, based on phenotypic and genotypic results. In *Linum*, vernalization requirement has been little studied, with studies mainly focused on the cultivar, *L. usitatissimum* (Darapuneni *et al.*, 2014a). The wild relative (*L. bienne*) is of interest, especially for questions regarding climates, as the wild population are likely to be locally adapted to their environments (Landoni et al., 2022). This adaptation can be observed as clines of flowering initiation times within the wild species. Several genes are of interest when looking at the development of *Linum* against latitude which are part of the flowering genes network. *Gutaker et al.* have previously identified the alteration and expression of the flowering gene *LuTFL1,* which could reflect latitudinal adaptation and crop selection for fibre production in northern populations (Gutaker *et al.*, 2019). However, other flowering time genes were not explored in this case. Other flowering time genes could involve genes that are in the photoperiodicity network. Several flowering time genes in cultivated *Linum* such as *LuCO* (CONSTANS) and *LuGI* (GIGANTEA) were previously explored by Sun et al (2019) where they found variation in the expression of these genes under several experiments with long days (22 °C/16 °C (12 h/12 h) for a 16 h/8 h photoperiod at a light intensity of 300 µmol photons m− 2 s− 1) and short days (Similar to long days treatment but with 10 h/14 h photoperiod instead) which revolves around photoperiodicity (Sun *et al.*, 2019).

Under temperate wild *Arabidopsis* accessions, it has been recently observed that different expression of genes of interest in this thesis correlates with the flowering time of the temperate accessions (Kinmonth-Schultz *et al.*, 2021). The main aim for this chapter is to find out whether expression of important flowering time genes, such as Flowering Locus C (FLC) and Constans (CO) influence the flowering initiation of different, vernalized, and non-vernalized *Linum* populations reflecting varying environmental conditions across the range. The hypothesis was that flowering initiation requirements based on vernalization differ within the *Linum* collection dependent on source location and type (wild or cultivated); and that different flowering behaviour is supported by differences in genetic expression of flowering time genes. During this experiment, vernalized and non-vernalized individuals were examined to explore phenotypic and genotypic differences both between species and within species. We asked: “How do flowering time responses to vernalization differ across wild populations in western Europe and are there differences between wild and cultivars?”.

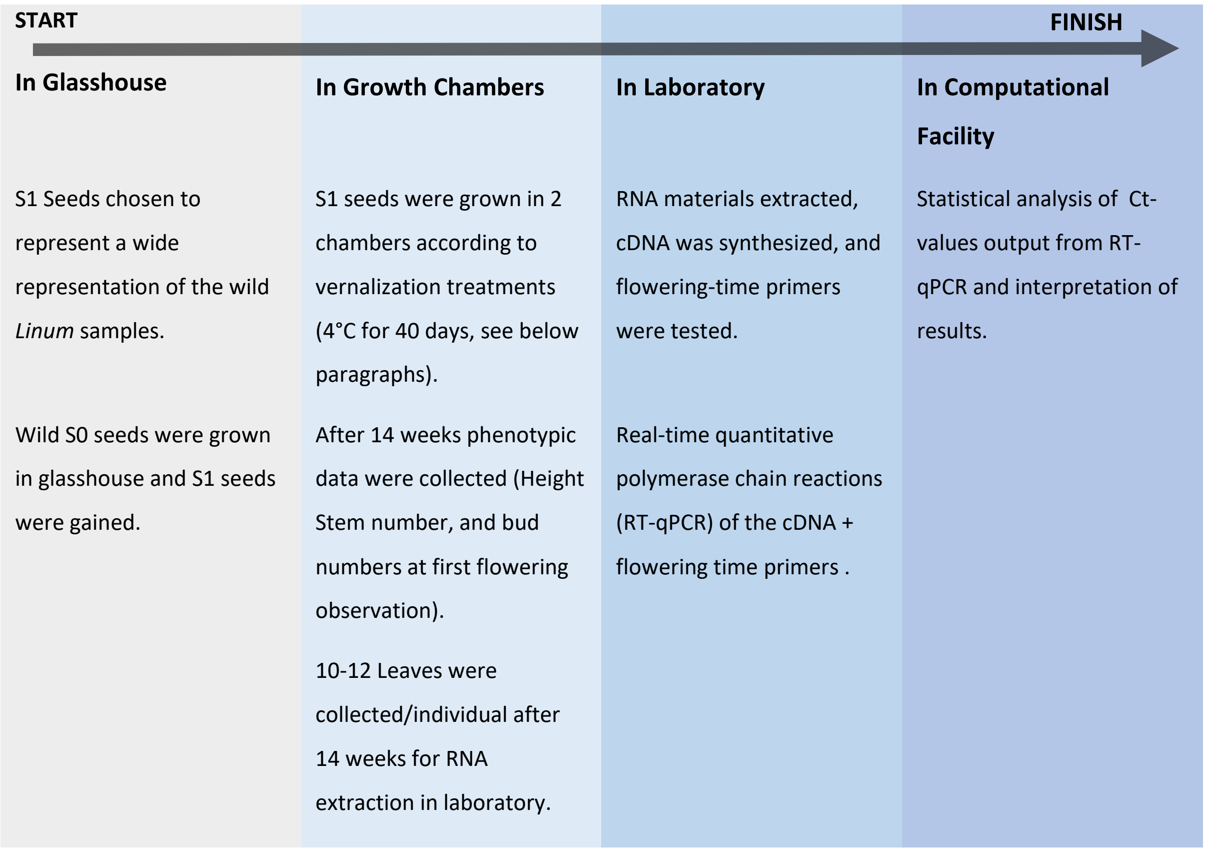
Differences in vernalization requirement within different plant species have been observed to subsequently affect their initial flowering time, this can be observed in species that requires both long or short days to flower (Adhikari, Buirchell and Sweetingham, 2012) (Ream et al., 2014). These flowering time interactions have been related to source population latitude, climate, and species. To look for climatic variables, The R package BioClim available at, https://rdrr.io/github/jjvanderwal/climates/man/bioclim2.html. Various climatic parameters can be used to summarize climate using PCA (principal component analysis) to extract major climatic factors that may contribute to any differences in vernalization requirement observed in this experiment involving relative gene expression. In the wild, it is important that flowering initiation responds appropriately to local climate. Regarding vernalization, the expectation is that the more northern a plant population is localized, the more dependent it is on vernalization cues such as colder temperatures and wind speed. Thus, a colder climate of origin will be associated with greater relative gene expression differences when grown at different temperatures.

There is an expectation that selection on flowering time genes is stronger in the wild, leading to different vernalization responses in different latitudinal range. This implies interests in genetically quantifying variation in vernalization requirements both between and within the two *Linum* species. The expectation is for there to be genetic differences between the flowering time genes tested here when comparing cultivars with the wild samples. These differences are then expected to be attributed to environmental variables such as local climate.

## 2.2 STUDY AND METHODS

### 2.2.1 Samples and Experiment

We set out to investigate relationship between vernalization and genetic expression of flowering time genes using the following workflow:



| **In Glasshouse** | **In Growth Chambers** | **In Laboratory** | **In Computational Facility** |
|---|---|---|---|
| S1 Seeds chosen to represent a wide representation of the wild *Linum* samples. | S1 seeds were grown in 2 chambers according to vernalization treatments (4°C for 40 days, see below paragraphs). | RNA materials extracted, cDNA was synthesized, and flowering-time primers were tested. | Statistical analysis of Ct-values output from RT-qPCR and interpretation of results. |
| Wild S0 seeds were grown in glasshouse and S1 seeds were gained. | After 14 weeks phenotypic data were collected (Height Stem number, and bud numbers at first flowering observation). | Real-time quantitative polymerase chain reactions (RT-qPCR) of the cDNA + flowering time primers . | |
| | 10-12 Leaves were collected/individual after 14 weeks for RNA extraction in laboratory. | | |

Figure 2. A workflow summary of the study methods in this chapter from the glasshouse to analyses in computing facilities.

To measure phenotypic and genetic differences between representative *Linum* populations; controlled growth of plant materials was conducted. Samples for this experiment were collected as described in sections “1.2.1 Linum usitatissimum: Cultivated flax” and 1.2.2 “Linum bienne: Wild Flax” of this thesis. Samples were collected from across Portugal, Spain, France, and the UK (Fig 3). A full list of the individuals used is available under appendix 2.

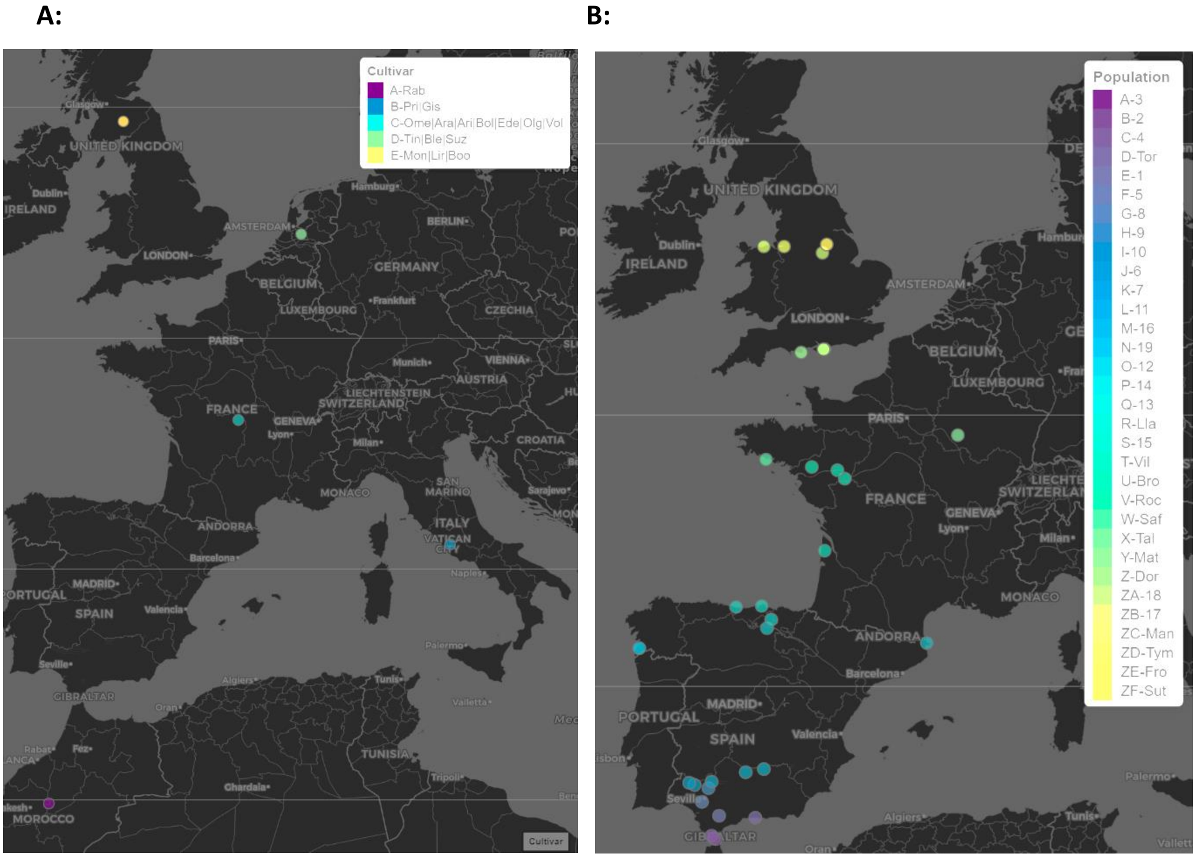


Figure 3. A map of the origin of collection of cultivars (A) and wild (B) *Linum*. Each population is alphabetically ranked by their latitude, with the most Southern population ranked A. For cultivars, since latitude was only recorded by country level, numerous populations belonging to the same latitude are depicted per latitude level which are alphabetically ranked.

Out of the total of 47 populations, 28 were wild types (*L. bienne*) and 18 were cultivars (*L. usitatissimum*). Wild seeds were initially grown in glasshouses available at the Department of Biosciences in Durham University. The controlled glasshouse had 16:8 hours daylight to no light ratio and the minimum temperature was measured at 13°C during winter seasons, with a maximum temperature measured at 28°C during summer seasons. S1 Seeds from these were utilized for vernalization experiments. A collection of 473 individuals over three experimental designs were studied. The vernalization experiments were conducted twice in the duration of this study (2018-2019 and 2020-2021). During both time frames, two controlled conditions (vernalized and non-vernalized) were specified using controlled growth chambers. Out of 473 individuals, 62 individuals were vernalized in 2020-2021; 257 individuals were vernalized in 2018-2019; and 154 individuals were non-vernalized in 2018-2019. The experiments in 2018-2019 were replicated (2 vernalized and 2 non-vernalized chambers). In comparison, there was only one vernalization chamber and one non-vernalized chamber under the 2020-2021 experiment, albeit the same conditions were applied between the two experimental time frames. For the 2020-2021 experiments, samples were gathered from a collection of S1 and S2 seeds previously grown under experimental conditions based on no vernalization or under controlled glasshouses available in the department of Bioscience of Durham University. In total, there were 47 germinating populations from both vernalized and non-vernalized experiments. After 40 days of treatment, plants were morphologically examined by their flowering time, overall height, and number of stems when first flowering. Morphological measures between these traits will be examined in chapter 5 of this thesis. The vernalization conditions (see below) were setup in two Weiss Gallenkamp growth chambers model numbers A3655 and A3658. The conditions were kept the same throughout the experiment. The number of days to flower (flower initiation) was a point of phenotypic interest and genetic expression (RNA materials) as a point of interest for gene expression studies. Therefore, collection of RNA materials also took place at the corresponding first flowering of every individual.

Preliminary measures for the S0 population grown in the controlled glasshouse were made before sowing of S1 generation for measures regarding vernalization which are plant height, stem numbers and bud numbers. The plants were exposed to vernalization conditions as follows; after sowing, *4°C (0-hour lights) for 72 hours, 22°C for 10 days with 16:8 h light ratio, and 4°C for 40 days with 16:8 h light ratio*. After vernalization plants were kept at 24°C to 16:8 h light ratio indefinitely. The non-vernalized conditions were as follows; after sowing, *4°C (0-hour lights) for 72 hours, 22°C for 10 days with 16:8 light ratio, 24°C for 40 days with 16:8 light ratio indefinitely*. Note the difference between the vernalization and the non-vernalized being the *4°C for 40 days cold treatment*. The lightings consist of numerous fluorescent tube lighting units, which were “Philips Master TL5 HE”. These light tubes are 14 Watts in power requirements for each tube and emits 4000 Kelvin light temperature in colour by specifications. Plants from these S1 generations under vernalization treatments were also collected for RNA materials, for the genetic analysis purposes of this study. The phenotypic traits measured were divided into two categories: vegetative (stem number and plant heights) and reproductive (first day to achieve first flowering (flowering initation) and seed size), with particular interest in reproductive measures, due to the interest in flowering initiation. The phenotypic measures were measured at the time of first flowering of every individual (±1 day).

Leaf materials were collected 14 weeks after sowing, and no more than three days after vernalization treatments have occurred. Leaf samples were collected from each treatment during a 2-day window period from 11am to 1pm to avoid expression differences due to time of collection. Ten to twelve leaves from the top 2cm of the longest shoot on each plant were collected inside a 1.5ml Eppendorf tube for each sample and were each labelled, and flash frozen in liquid $N^2$ before storing in the -80°C freezer. Two vernalization experiments were conducted during this study under different years (2019/20 and 2020/21). However due to restrictions on caused by the Covid-19 pandemic, the 2020/21 vernalization experiment was cut short. Due to this, differences between the two experiments were unavoidable. These differences were quantified by analysing comparative analysis of the two experiments. This can be found under the "Results" part of this chapter.

When considering locality of a wild population, its environment is the climate of the local area in terms of longitude, latitude, and altitude of the places where these wild populations were collected. Local climate was analysed based on different climate variables such as precipitation, temperatures, and sunlight availability that were extracted from over a 30-year period. The data was retrieved from WorldClim database. Data were available for precipitation (mm), solar radiation (kJ/day*m2), average temperature (°C), minimum and maximum temperature (°C), vapour pressure (kPa), and wind speed (m/s). Climate variables for principle component analysis (PCA) were used at 30arcsec resolution over seasonal months (June July August for summer, September October November for autumn, December January February for winter and March April May for spring). To include all 6 dimensions of climatic variables, they were loaded into a principle component analysis with a principle component one (PC1) level explaining 63.8% of climatic variables (data available in appendix 3). Loaded value also suggests that lower PC1 values are associated with colder temperature associated with more Northern latitude. These data were used in a linked study of flowering time study (Landoni B. et al., 2022). All the climate variables were loaded in a collective PCA, whereby climatic PC1 values for each population were processed as the final representative values for all climatic variables. This would make further downstream analysis, regarding climate variables easier to process.

When plotting PC1 and latitude of origin for each of our population, there is a visible relationship between the two variables, suggesting that lower PC1 values reflects higher latitudes. This can be seen in the plot below:

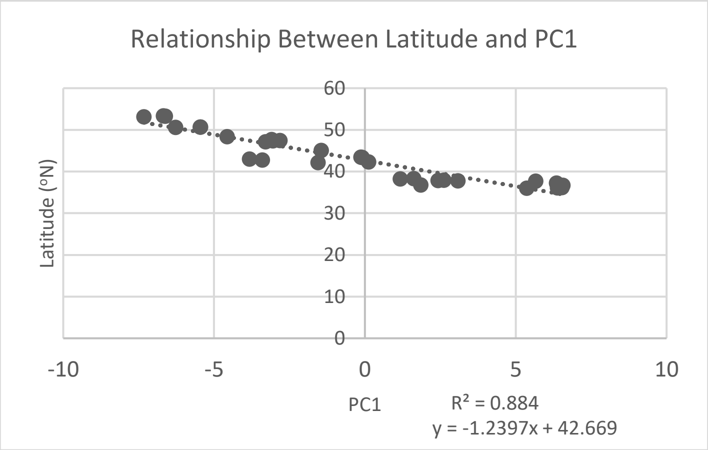


Figure 4. Scatterplot to show negative relationship between climate variable representative (PC1) and latitude. Lower PC1 values represent lower temperatures (°C) and solar radiation (kJ/day*m2).

## 2.2.2 Relative Gene Expression: RNA Extractions

Real-time quantitative PCR (rt-qPCR) is well suited for relative gene expression quantification (Livak and Schmittgen, 2001). For this, RNA is extracted and then converted back into copy DNA (cDNA) by using reverse transcriptase. This was done to an individual level, with no bulking over population. cDNA products were then subjected to polymerase chain reactions by designing primers specifically for loci of interest. The amount of cDNA target is quantified using dye markers that fluoresce when they bind DNA PCR product and comparing real-time amplicons with those of housekeeping genes from the same sample. The use of real-time polymerase chain reaction is a routine tool in molecular biology for the study of gene expression.

Leaf tissues from vernalization experiments (as mentioned under section 2.2.1 “Samples and Experiment”) were first ground using a homogeniser. Promega’s ReliaPrep RNA Tissue Miniprep System was used to extract RNA and solutions were prepared using manufacturer’s instructions. After solutions were prepared, LBA+TG buffer were added into previously ground samples and were mixed using micro-pestles and tissue homogeniser until thoroughly mixed. Materials were lysated and processed following manufacturer’s protocol. After the centrifugation steps, RNA was washed according to manufacturer’s protocol with an additional second wash. Further centrifugation steps took place at 14,000g for 2 minutes after the second wash.

After collection, samples were ready for elution. Samples were eluted according to the manufacturer’s instructions. Centrifugation steps were repeated using the same elutant to make sure all the available RNA were eluted out of the membrane. 1µL of elutant from each sample tube were used for RNA measurement under a Nanodrop ND1000. When satisfied with measurements, samples were stored in -80$^{o}$C freezer to preserve RNA integrity. Satisfactory measures were decided at >20ng/µL of RNA with 260:230 ratio of >1.20 and 260:280 ratio of >1.60.

## 2.2.3 Relative Gene Expression: cDNA synthesis:

For cDNA synthesis, Applied Biosystem’s High-Capacity cDNA Reverse Transcription Kit was used. For control, no template control (NTC) and no reverse transcriptase (NRT) samples were added. Each sample reaction was made to the same amount of RNA template in a total of 10µL volume. The amount of RNA template needed for the cDNA synthesis was calculated by standardizing all samples to the sample with the lowest amount of detectable RNA. This was done by calculating how much RNA could be extracted for a 10µL eluted RNA additive (to the mastermix), based on measures of RNA by spectrometry in ng/µL for specific samples. The maximum amount of RNA for the cDNA synthesis protocol was 2µg in 10µL final volume. Samples which had more detectable RNA were diluted down to the same concentration as the lowest concentrated sample. This dilution was done using nuclease-free water accordingly. For the no template control, RNA was replaced with nuclease free water. 10µL was added per NTC. For the NRT, 5µL of random RNA sample with 6µL of nuclease free water were added.

For the rest of the reactants, a master-mix were prepared in an Eppendorf tube per manufacturer's instructions. Master-mixes were vortexed for 10 seconds. After the addition of the master-mix ,1µL of Multiscribe Reverse Transcriptase were added into each sample and the NTC control. Samples were then centrifuged briefly and vortexed subsequently. After the addition of both master mix and Transcriptase, a thermocycling step can take place. The thermocycling steps for cDNA synthesis were as follows:

| | **Step 1** | **Step 2** | **Step 3** | **Step 4** |
|---|---|---|---|---|
| **Temperature** | 25°C | 37°C | 85°C | 4°C |
| **Time Elapsed** | 10 minutes | 120 minutes | 5 seconds | ∞ |

Table 1. Thermocycling steps for cDNA synthesis.

After the thermocycling steps were finished, samples were stored in a 4°C fridge for short term (<72 Hours) and a -21°C freezer for long term storage(>72 Hours). The final Ct-value from the later real time qPCR protocol would serve as quality control. If quality of RNA from the extractions are satisfactory there was no need to measure cDNA output (Schmitz and Amasino, 2007).

### 2.2.4 Expression of Genes of interest and House Keeping genes:

Housekeeping genes (HKGs) are required for maintenance of cells (Butte, Dzau and Glueck, 2001). These genes are usually expressed relatively constantly under different conditions, which makes them suitable for comparison against the expression of other genes. Generally, HKGs constitute an important component for rt-qPCR procedures. However, a study suggested that expression of HKGs may change with cellular density of samples and that they could be affected by experimental conditions (Greer *et al.*, 2010a). Therefore, use of multiple HKGs within the same experiment can guard against these effects. *GAPDH* (Glyceraldehyde-3-phosphate dehydrogenase) and *UBI* (Ubiquitin) represent a couple of the commonly used HKGs in plants and therefore used for assessing different abiotic stress (Liang *et al.*, 2018a) (Carmona *et al.*, 2017). For *Linum,* the HKGs which will be used for this experiment are *LuGAPDH* and *LuUBI2.* This reference genes for *Linum* were developed by *Huis et al.* and are specifically useful for real time PCR (qRT-PCR) protocols (Huis, Hawkins and Neutelings, 2010). Our study will be incorporating these HKGs, as were also found on the study investigating photoperiodicity in *Linum* by *Sun et al* (Sun *et al.*, 2019).

We hypothesized that there would be variation in gene expression for flowering time genes according to latitude and due to vernalization. We tested flowering time genes already studied in *Linum usitatissimum* by Sun et al (Sun *et al.*, 2019). In our study however, we will be testing the gene expression based on vernalization stimulus and will be controlling for other environmental stimulus such as light availability. Whilst the study by Sun *et al* (2019) had found significant differences in expression of flowering time genes for *L. usitatissimum* under different light-length treatments, they have not tested vernalization as an environmental factor. In addition to this, they have not tested the wild relatives. It would be of interest to look at the expression of floral integrator gene, *LuFT,* along with the other photoperiodicity genes when differences in vernalization treatments is applied. This is because *Linum* is a temperate plant found across the sub-temperate into the temperate regions, thus different population may have adapted differently to vernalization as a stimulus to flowering initiation.

In addition, the lineage leading to *L. usitatissimum and L. bienne* as a genus have been understood to have undergone a polyploidy event 20-40*mya* (Sveinsson *et al.*, 2014a). This has repercussions as there potentially are multiple copies (paralogues) of the same genes within *Linum*. This would mean there may be a need to test multiple copies of the same gene to see an effect which could be from one or multiple set of these copies. Wild and cultivars were tested for seven paralogue-specific flowering time genes using a modified real-time qPCR protocol. These genes were *LuFT1*(*Flowering Locus T 1*), *LuFT2*(*Flowering Locus T 2*), *LuGI1.1* (*GIGANTEA 1.1*), *LuGI1.2* (*GIGANTEA 1.2*), *LuGI2* (*GIGANTEA 2*), *LuCO1*(*CONSTANS 1*), and *LuCO2* (*CONSTANS2*). A full list of nucleobases codes used for each primer can be found in appendix 4 (Sun *et al.*, 2019).

### 2.2.5 Primer Testing

The primers for the genes of interest were initially tested by performing PCR and analysing the products on an agarose gel (Jarman, Ward, and Elliott, 2002). Primers was tested using Promega's Go-Taq green master-mix; 2.5µl of each Forward and Reverse primers along with 2.5µl of *Linum* cDNA template for a 25µl total reaction volume. This would mean that the concentration of each primer is 1µM with < 250ng of cDNA template. We added 12.5µl of the Go-Taq green mastermix for a 1x concentration solution. The rest of the solution is made up with nuclease-free water up to 25µl. PCR mixes were treated in a thermal cycler using the following programme: Denaturation at 95°C for 3 minutes; Annealing with 35 cycles of 95°C for 45 seconds, 55°C for 30 seconds, 72°C for 1 minute; and finally, an extension stage at 73°C for 5 minutes. Products were examined on a 2% TAE gel for PCR products. Products at <200bps with minimal amount of smearing observed were deemed acceptable. A decision was made to not use LuGI1.2 and LuGI2 primers based on inconsistencies with the gel such as smearing, and product sizes observed as > 200bps, leaving 5 primers (*LuFT1, LuFT2, LuCO1, LuCO2*, and *LuGI1.1* with the 2 housekeeping genes (*LuGAPDH* and *LuUBI2*).

## 2.2.6 Quantitative Real-Time Procedure

For real-time quantitative PCR reactions, an Applied Biosystem model 7300 real-time thermal-cycler was used with StarLab's 96-Well PCR Plates (96-Well PCR Plate, Skirted, Low Profile, White – STARLAB, 2022).

Diluted cDNA was prepared for each sample. Primers were ordered from Integrated DNA Technologies (IDT). Each primer was aliquoted and diluted to 10µM working solution. To perform a real time qPCR, GoTaq qPCR SYBR Green were used (Promega Corporation). A master mix of the reaction components were then prepared as follows:

| Component | Volume / reaction (µl) | **Final Concentration** |
|---|---|---|
| **SYBR Green mix** | 7.5 | 75% |
| **Forward-Primer** | 0.9 | 9% |
| **Reverse-Primer** | 0.9 | 9% |
| **Nuclease free water** | 0.7 | 7% |

*Table 2. The components required for the real-time qPCR reaction.*

SYBR Green Mix (included in the Promega qPCR GoTaq RT-qPCR system), and Nuclease free water were mixed by considering three repetitions per sample and 2 controls (No reverse transcriptase (NRT) and No template control (NTC)). For 14 samples we had (3×(14+2)) = 48 × 2 = 96. As there are possibilities for pipetting errors, 100 samples were considered for the SYBR and the water. SYBR and water were mixed in a 1.5ml Eppendorf tube and vortexed at 2000RPM for 30 seconds.

The small volume of primers was often easier to be dispensed onto the sides of each well, this way loaded wells can be marked by the presence of the primer on the side to avoid contamination by other primers. After the primers were fully loaded, the plate was centrifuged down for 1 minute at 3000RPM at room temperature using the F2096 rotor in an Allegra X22 refrigerated bench centrifuge. Making sure all the liquid has reached the bottom of the well, 8.2µL of the SYBR Green mix were loaded in with 5µL of each respective diluted cDNA samples following. Plates were sealed using an appropriate 96-well plate plastic sealer (we used Starlab's Self-Adhesive sealing films), making sure that each well was tightly secured with the sealer by pressing on each well after sealing. Another centrifugation step for 1 minute at 3000RPM were applied to the plate and samples were ready to be thermal cycled as follows:

| Stage | Temperature (°C) | Time (s) | Number of repeats |
|---|---|---|---|
| **Holding** | 50 | 120 | 1 |
| **Enzyme activation** | 95 | 600 | 1 |
| **Cycling I** | 95 | 15 | 40 |
| **Cycling II** | 60 | 60 | 40 |

*Table 3. The cycling steps for the real-time qPCR.*

SYBR green (Promega) was used to quantitatively assess amplified PCR product. SYBR green dye fluoresces at 497-520nm blue to green light when binding to double stranded DNA (ds-DNA). The SYBR dye fluorescence intensity can be used to quantify how much of each gene were amplified using specific primers (Zipper *et al.*, 2004). The ds-DNA will increase by each thermal cycle thus increasing their binding to SYBR until the amount of DNA material reaches the cycle-threshold (Ct) values which is the number of cycles it takes for the dye to be distinguishable to the background as it binds to the DNA material (Zipper *et al.*, 2004).

The Ct-value reads were collected at end of stage cycling II, whereby the real-time machine measures the cycling threshold values against amplifications of the targeted cDNA expressed in the sample. After this, Ct-values were further analysed as described under following section 2.2.10 “Relative Gene Expression: Quantification”.

### 2.2.7 Primer Efficiency

A standard curve calculation was done for the Ct-values of the different dilutions to define efficiency of primers whereby the primer efficiency is tested by calculating the slope against the concentration of tested primers (Pfaffl, 2001). This test is essential, because calculation of relative gene expression is based on the delta-delta (difference) of the Ct-values between the HKGs and the genes of interest. The difference in Ct-values is influenced by how efficient a PCR product can react with the SYBR dye.

Primer efficiency was tested by calculating the average Ct-values for runs of a sample that has been serially diluted and calculating the coefficient of Determination ($R^2$) values across sample dilutions. In theory, the more dilute the samples are, the more slowly the primers are going to amplify, and this will form a standard curve for fitting a model and performing $R^2$ calculation (Glantz, Slinker and Neilands, n.d.) using:

$$R - squared = \frac{Sum\ of\ squares\ due\ to\ regression}{Total\ sum\ of\ squares}$$

In this case the dilutions of primer tested were as follows; undiluted, 1/20, 1/200, 1/2000. After applying *Log* to the sample quantity, the coefficient of determination (slope) was calculated. The efficiency (%) was then calculated using the following equation described by Ginzinger (Ginzinger D, 2002):

$$Efficiency\ (\%) = \left(10^{\frac{-1}{Slope}} - 1\right) \times 100$$

Efficiency values from the above equation are represented in percentage. However, for the relative gene expression calculations, the percentage values were converted into decimal values whereby a value of 2.00 will indicate 100% efficiency and 1.00 will indicate 0% efficiency. This conversion is essential for the final efficiency value input calculation in relation to ΔCt of each gene further downstream. Efficiency values were converted from % using the following equation (Ginzinger D, 2002):

$$Efficiency\ Value = \frac{Effiency\ \%}{100} + 1$$

Three primer efficiency tests were performed using real-time qPCR procedures for each respective primer but only the third test values used further for down-stream calculations as these seemed the best results from several tests, based on the efficiency (%) values closer to 90-110%. The results from the primer tests are noted in appendix 5.

These primer efficiency values represent values which were inputted as a correction when calculating relative gene expression (RGE) of each gene tested. The efficiency value was calculated for each of the tested primers as well as the tested HKGs according to the separate ΔCt (avg Ct values – actual) for each gene and for each separate real time run (i.e repeats). The final Primer efficiency correction were calculated using this formula:

$$Final\ efficiency\ input = Efficiency\ value\text{\^{}}\Delta\text{Ct for each gene}$$

The measured efficiencies values were not optimal as efficiencies of some primers are <95% as recommended (Miranda and Steward, 2017). Ideally these primers would be re-designed and re-ordered, but with the RNA materials in some cases already extracted and time limit considerations, a decision was made to lower the threshold efficiency to 85% to include all the tested primers were above the threshold.

## 2.2.8 Real time qPCR of Vernalised and Non-vernalised Samples

Regarding the real time qPCR experiment described above, a set of samples that had experienced both vernalization treatments were chosen to compare flowering time gene expression differences. The final set of samples included 14 wild and 14 cultivar pairs each with samples from the non-vernalization and vernalization experiment. Some additional samples could not be utilized due to low quality RNA. Subsequently, cDNA from 2 samples at a time was added to 96 well plates designed to test the 5 Flowering time genes along with the 2 HKGs.

Primer-sample combinations tested totalled up to 14 per plate with 3 repetitions each. The primer sequences are available under appendix 4. These, with the addition of No Reverse Transcriptase (NRT) and No Template Control (NTC) made up a 96-well plate.

## 2.2.9 Clean-ups and quality controls:

As a pre-cursor to calculating relative gene expression (RGE), the Ct-values of the Housekeeping (HK) genes were checked for quality control. This was of importance because the observed primer-efficiency values were not ideal (87.273% for *LuGAPDH* and 51.464% for *LuUBI2*). Individual Ct values were removed if either; Ct-values were either too high (>29 cycle-threshold values) or too far away from the other values (i.e >1 Ct variation between the triplicates) for LuGADHP; 51/144 Ct-values were either too far away from other Ct-values or too high. These steps addressed potential experimental errors such as poor qualities and low quantities of cDNA and pipetting errors, which will impact the real-time qPCR quantification process.

When applying the delta-delta Ct ($2^{-\Delta\Delta Ct}$) method to the final relative gene expression values, all initial samples were included, except for one wild sample (Saf_10) for both Flowering Time genes. This sample was excluded because of the much bigger relative gene expression values that make it an extreme outlier. This could be due to experimental errors such as the quality of RNA.

## 2.2.10 Relative Gene Expression: Quantification

RGE data were processed in Microsoft Excel. For quantification purposes, a template was built for primer efficiency using formula stated under section 2.2.7 “Primer Efficiency” part of this chapter. In addition to these, further calculations are required to calculate relative gene expression. The Ct results for each gene of interest (GOI) was added to a Microsoft Excel template and are specified to the same House Keeping Genes (*GAPDH*) Ct-values for each respective samples. To calculate relative gene expression, a difference (delta) of Ct values were calculated between a reference house-keeping gene and the gene of interest based on previous method described by *Livak et al* (Livak and Schmittgen, 2001a). There are two methods which are available to explore using one HKG or using two HKGs. Both are described below.

The default formula for calculating relative gene expression differences is the delta-delta Ct ($2^{-\Delta\Delta Ct}$) method. This method uses only one HKG. This is where, for each gene and sample, average Ct-values for the gene of interest (GOI) and the housekeeping gene (HKG) are calculated. Differences (delta-Ct) between the gene of interest and the housekeeping gene Ct-values were calculated and then the delta-delta Ct (ΔΔCt) were calculated using a calibrator value (i.e., *delta-Ct – calibrator value*). The choice calibrator value differs depending on experimental design but in this experiment, the average value of all the non-vernalized treatment ΔCt for the gene of interest was used as the calibrator value for that specific gene. This is an important factor to the quantification of RGE, as the choice of calibrator will impact the final gene-fold value (Schmittgen and Livak, 2008).

ΔCt for a specific gene, using one HKG was calculated using (Rao X *et al*, 2013):

$$\Delta Ct = Ct(a\ Gene\ of\ interest\ (GOI)) - Ct(a\ reference\ gene\ (HKG))$$

ΔΔCt are then able to be calculated using (Rao X *et al*, 2013):

$$\Delta\Delta Ct = \Delta Ct(a\ Gene\ of\ interest\ (GOI)) - \Delta Ct(a\ reference\ gene\ (HKG))$$

The formula above can be applied to a specific GOI with a specific HKG which are constant for the entire experiment (i.e GAPDH only). Using this method, a fold-change of the target gene (either controlled (non-vernalized) or treated (vernalized) can be calculated, normalized, and related to a HKG using calibrator values as described above. The method was fully described and updated in a data analysis by Rao *et al* (Rao X et al., 2013).

Although the delta-delta Ct method is the most used, the use of multiple house-keeping genes is regarded as more reliable in giving a background HKG expression against which to calculate relative gene expression (Riedel *et al.*, 2014a) (Manoli *et al.*, 2012). The variation that may occur between the multiple different house-keeping genes is resolved by normalization steps (Huggett *et al.*, 2005). According to *Vandesompele et al,* 2002*,* the normalization is done by geometric averaging of the multiple house-keeping genes involved as controls (Vandesompele et al., 2002; Riedel et al., 2014). The equation for using multiple reference genes is as follows:

$$Relative\ Gene\ Expression = \frac{(E_{GOI})^{\Delta CtGOI}}{GeoMean[(E_{REF})^{\Delta CtREF}]}$$

The delta-Ct (ΔCt) values for each of the tested samples were determined using the following equation:

$$\Delta Ct = Control\ Ct - Treated\ Ct$$

The treated Ct values in this case will be the Ct values from the vernalized samples and the control from the non-vernalized. The PCR performed consisted of 2 reference genes (*LuGAPDH* and *LuUBI2*) with a gene of interest from one of our 5 chosen loci. Using the equations above and the primer efficiency values gained previously as formulated using methods described under section 2.2.5 "Primer Testing", it is possible to compare relative genetic expression between treated (vernalized) and control (non-vernalized) individuals in our *Linum* sample set.

---

## 2.3 STATISTICAL ANALYSES FOR RELATIVE GENE EXPRESSION

Relative Gene Expression (RGE) is calculated using the $2^{-\Delta\Delta Ct}$ formula as described in section 2.2.10 "Relative Gene Expression: Quantification" above. For follow-up analysis, various packages were used in R (R Core Team, 2022) ran under the R graphical interface of Rstudio (Rstudio Team, 2020). ΔΔCt can be calculated by creating a template using the formula to calculate ΔCt (see section 2.2.10) for each of the genes that were tested, including the HKG. The template was created in Microsoft Excel. the raw data were further analysed in R.

For comparison between the different experiments that took place during different years (2018 and 2021), an F-test (Box G, 1953) was applied in R. For genetic expression analyses, templates for the calculation of Ct-values (see "Relative Gene Expression: Quantification") were made and bar charts can be plotted in Microsoft Excel. After values for Relative Gene Expresisons (RGE) were gathered for each respective locus, species and treatment, data could be modelled against environmental variables.

RGE differences of vernalized and non-vernalized treatments, and wild or cultivar combinations was tested for each gene using a 2-sample, 2-tailed unequal variance t-test. This test compared fold-gene expression of the gene of interest between the two treatments (vernalized and non-vernalized). The tests were performed in Microsoft Excel. The formula used for the t-test is *"=TTEST(array1,array2,2,3)"*. RGE differences were visualised using R package ggplot including the command: *"+stat_compare_means(label = "p.signif", method = "t.test",ref.group = ".all.")"* (appendix 6).

The differences of flowering initation between vernalized and non-vernalized individuals were correlated against latitude of origin ($^o$N) using Pearson's coefficient of determination ($r^2$) in R. This analysis measured how correlated flowering initiation is to latitude in each data sets (wild and cultivated; Steel R. *et al.,* 1960). P-values and $r^2$ values explained by the models were visualised within scatter plots of the data, showing lines of regression using the R-package "ggplot2" (Wickham H.*,* 2016). The full command used for drawing correlation coefficients for this analysis can be found under appendix 6. To further analyse these relationships using modelling, a Linear model analysis was able to run using R. The command can also be found under appendix 7. For modelling against environmental variables such as "latitude" and "climate" (described under the "Study and Methods" part of this chapter), a GLM (General Linear Modelling) approach was used. Various commands used for the GLM analysis in R can be found under appendix 7.

## 2.4 RESULTS: FLOWERING TIME BETWEEN EXPERIMENTS FROM DIFFERENT YEARS

The vernalization results were divided into three different categories: “non-Vernalized 2018”, “Vernalized 2018” and “Vernalized 2021”. Due to COVID-19 pandemic, vernalization experiment in 2021 had to be terminated early in March 2020. Consistencies between the different years couldn’t be kept the same and a bias towards earlier flowering plants can be observed in the 2021 experiment as a result.

Differences between the 2018 and the 2021 vernalization experiments was observed when looking at variation in number of days to flower. This was due to bias towards earlier flowering plants in the 2020-2021 replicate as the experiment was stopped prematurely (due to Covid-19 restrictions), resulting on limited data collection for late flowering plants. Statistical comparison of the number of days to flower (flowering initiation) between the two vernalization experiments, under the different years found significant differences between “vernalized 2018” and “vernalized 2021” experiments (*F-test, F=2.523, df=196, p=<0.005)* (Figure 5A). When data were filtered to the same individuals present in both experiments, the difference between the two datasets remained, with biases for earlier flowering individuals expressed *(2-sample, 2 tailed t-test p=<0.001)* (Figure 5B). The significant variation between flowering initiation under the 2018 and 2021 vernalization experiment suggests that at least for phenotypic measures, data from the 2021 vernalization experiment should be excluded to avoid biases towards earlier flowering plants.

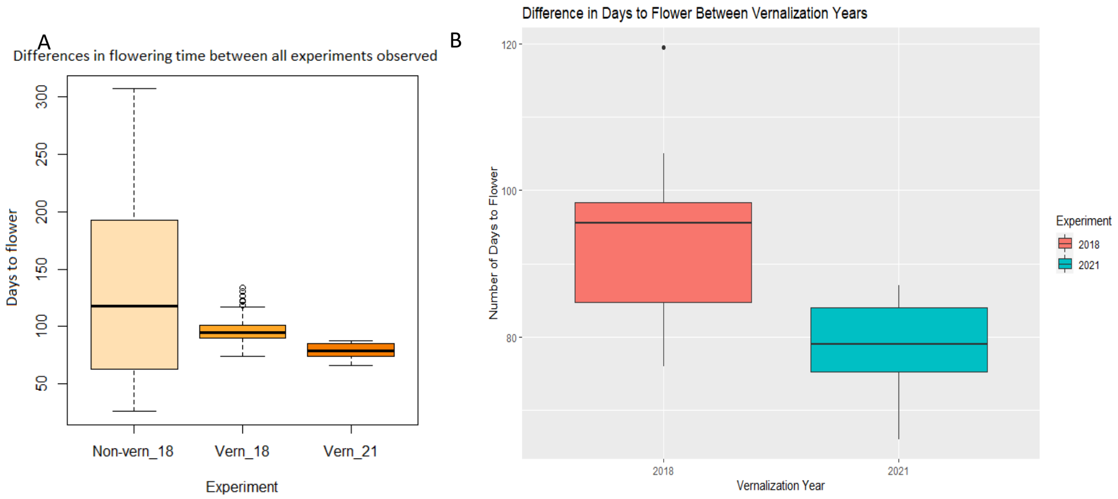


Figure 5A. A boxplot showing flowering time differences between all the individuals tested across the 2018 non-vernalization, the 2018 vernalization and the 2021 vernalization experiment (F-test, F=2.523, =196, p=<0.005). Figure 5B. A boxplot representing flowering time of a subset of all the mutual individuals found in the two vernalization experiments, showing more significant variations (2-sample, 2 tailed t-test p=<0.001). Boxplots suggests a bias for earlier flowering plants in the 2021 vernalization experiment, due to ending prematurely.

### 2.4.1 Results: Vernalization and flowering time

Data were processed without flowering initiation data from the 2021 vernalization experiment. The number of days it took for individuals to flower were tested against latitude for wild and cultivars (Figure 6). Under the 2018 vernalization experiment, dramatic reduction of number of days to flowering was observed in the wild *L. bienne* that were treated by vernalization. A positive correlation between the latitude of origin and the number of days to flower in the wild species were also observed. The cultivars showed a much smaller difference between treatments and a negative latitudinal correlation with number of days to flowering. This suggests that requirement of vernalization differs between the two species. *L. bienne* (wild) conveys a dramatic change in flowering initiation. This was observed on the more northern populations. In the most northern populations, vernalization reduced the number of days to first flowering by more than 100 days. The Northern wild species were more sensitive to vernalization, in terms of reducing the number of days required to flower. In contrast for *L. usitatissimum* (cultivar), vernalization only slightly increases the number of days it takes for individuals to flower, in comparison with non-vernalized individuals. In addition, linear modelling showed that for wild samples in non-vernalization, latitude has a significant influence on days to flowering *(Non-vernalized R=0.46, p=<0.01; Vernalized R=0.48 p=<0.01)* (Figure 6). In contrast for cultivars, the requirement to vernalize is dramatically reduced, with correlation becoming negative between the number of days a plant takes to flower and latitude, however this was not statistically significant.

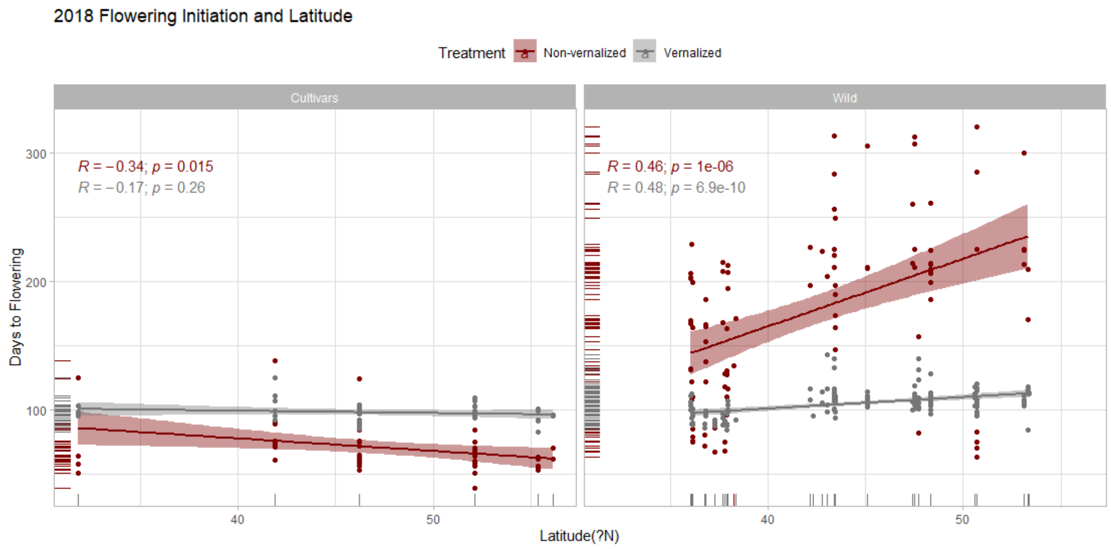


Figure 6. Scatterplots with regression line and linear models for flowering initiation to latitude. Note the differences because of latitude and vernalization in the number of days to flowering between wild and cultivated and samples. Reduction of more than 100 days to flowering initiation is observed in more Northern individuals of wild *Linum* whilst this was not observed in the cultivars. In contrast, the cultivars reveal an addition to the number of flowering due to the vernalization stimulus. In the wild number of days to flowering is positively correlated to45ignifice in both vernalization treatments (*p*=<0.05). Shaded areas represent a 95% confidence interval in the dataset respectively.

## 2.4.2 Results: Primer Efficiency Values

For primer efficiency, the desirable value would be 90-110% of efficiency. When we look at the efficiency percentage of the five flowering time loci however, only *LuFT1* and *LuCO1* fell into this category. Efficiency values for *LuUBI2* were very low. However, *LuGAPDH* showed close to ideal values at 87% efficiency. Sub-optimal primer efficiency may cause the representation of false fold-change; thus, this may affect the representation of the data in this study. Even with these sub-optimal primer efficiency value, time was of the essence during this study. It was deemed that any relationship which may be inferred in this study is taken with precautions to this sub-optimal primer efficiency, as efficiency values is significant to quantification of relative gene expression (Sreedharan S. et al., 2018).

| ***Primer*** | ***Type*** | ***R^2*** | ***Slope*** | ***Efficiency (%)*** | ***Converted value*** |
|---|---|---|---|---|---|
| ***LuGAPDH*** | *HKG* | *0.995* | *-3.670* | *87.273* | *1.873* |
| ***LuUBI2*** | *HKG* | *0.903* | *-5.546* | *51.464* | *1.515* |
| ***LuGI1.1*** | *GOI* | *0.333* | *-0.559* | *6034.644* | *61.346* |
| ***LuCO1*** | *GOI* | *0.552* | *-3.564* | *90.818* | *1.908* |
| ***LuCO2*** | *GOI* | *0.962* | *-3.614* | *89.114* | *1.891* |
| ***LuFT1*** | *GOI* | *0.799* | *-3.085* | *110.949* | *2.109* |
| ***LuFT2*** | *GOI* | *0.980* | *-2.380* | *163.090* | *2.631* |

Table 4. Primer efficiency values for genes of interest (GOI) and housekeeping genes (HKG).

## 2.4.3 Results: Comparison of RGE using one and multiple HKGs.

With the variation observed in the primer efficiency for this study, it is of interest to compare RGE results using one or multiple HKGs. Treatments are sorted into four groups: non-vernalized cultivars (NV-Cultivars), vernalized cultivars (V-Cultivars), non-vernalized wild (NV-Wild), and vernalized wild (V-Wild). A summary boxplot is shown below for the relative gene expression of the floral integrator *LuFT*1.

Relative gene expression varies when using one or two HKGs. Using one HKG, the vernalized, cultivar treatment showed a significantly increased *LuFT*1 expression (Figure 7) to the rest of the treatments. This reveals that using different methods and respective calculations could affect further analysis. Based on the primer efficiency test result that *LuUBI2* (one of the HKGs) have a very low primer efficiency (51.46%), it is thought that using *LuUBI2* may have skewed relative gene expression results. Therefore, for further analysis, only one HKG (*LuGAPDH*) was considered and the use of the delta-delta ct ($2^{-\Delta\Delta Ct}$) method would be further applied for the rest of the analysis regarding relative gene expression.

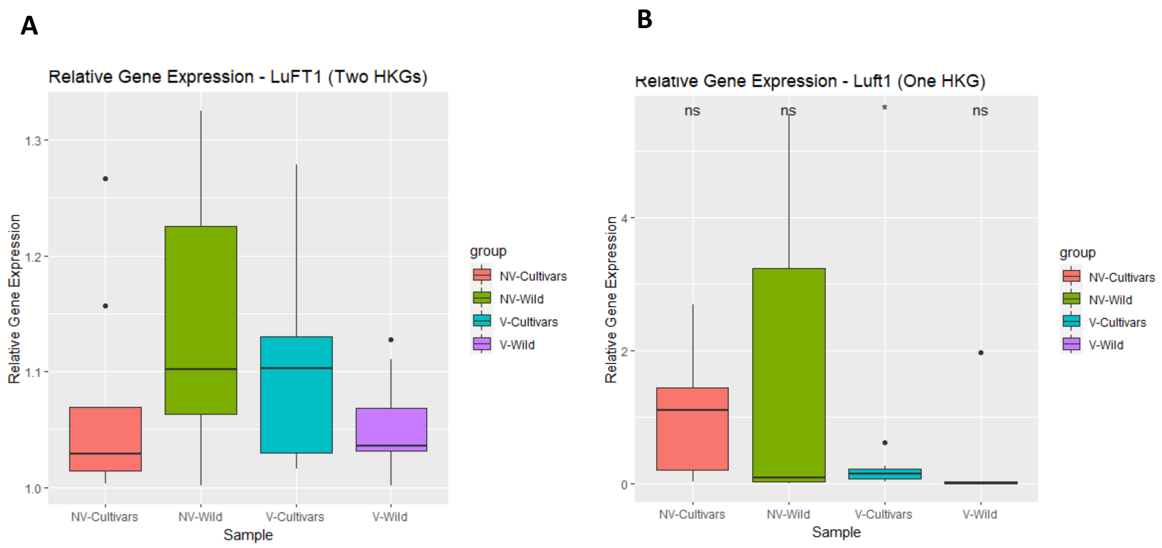


Figure 7. Boxplots comparing Relative Gene Expression for *LuFT1* using two (Figure 6A) or just one (Figure 6B) house-keeping genes. In Figure 5B the relative gene expression of *LuFT1* is found to be significantly different (marked by *) to the other treatments (p = <0.05, multiple pairwide t-test).

## 2.4.4 Results: Relative Gene Expression in Response to Vernalization

Relative gene expression results were initially sorted into groups reflecting treatments and species (non-vernalized-cultivated, vernalized-cultivated, non-vernalized-wild, and vernalized-wild). Relative gene expression ($2^{-\Delta\Delta Ct}$) was compared across each of the treatments to see whether any species/treatments show significant differences in relative gene expression respective of each gene for the treatment. Boxplots can be illustrated with a Bonferroni-corrected multiple pairwise t-test to reveal significance between treatments (Figure 8).

Variation in expression was mostly seen in cultivars with a potential down-regulation in all the tested genes when cultivars were vernalized (Figures 7 A-E). Less strong RGE differences were seen in the wild samples. In the case of the comparisons between treatments, most of the differences in the data were observed when treating cultivars to vernalization. This is interesting as phenotypically; this difference wasn't to be expected. This is because flowering initiation in the cultivars did not show faster flowering initiation in response to vernalization and instead showed slightly slower flowering initiation (Figure 4). The multiple pairwise t-test however, was comparing differences of relative gene expression when vernalized, against all wild and cultivar individuals which are perhaps behaving very differently in terms of expressing each of these genes and maybe independent of each other. Therefore, variation among individuals was explored in more detail. Comparisons between expression responses (relative gene expression) against individuals with different treatment were undergone using a paired t-test as in the next part of this chapter.

**A**

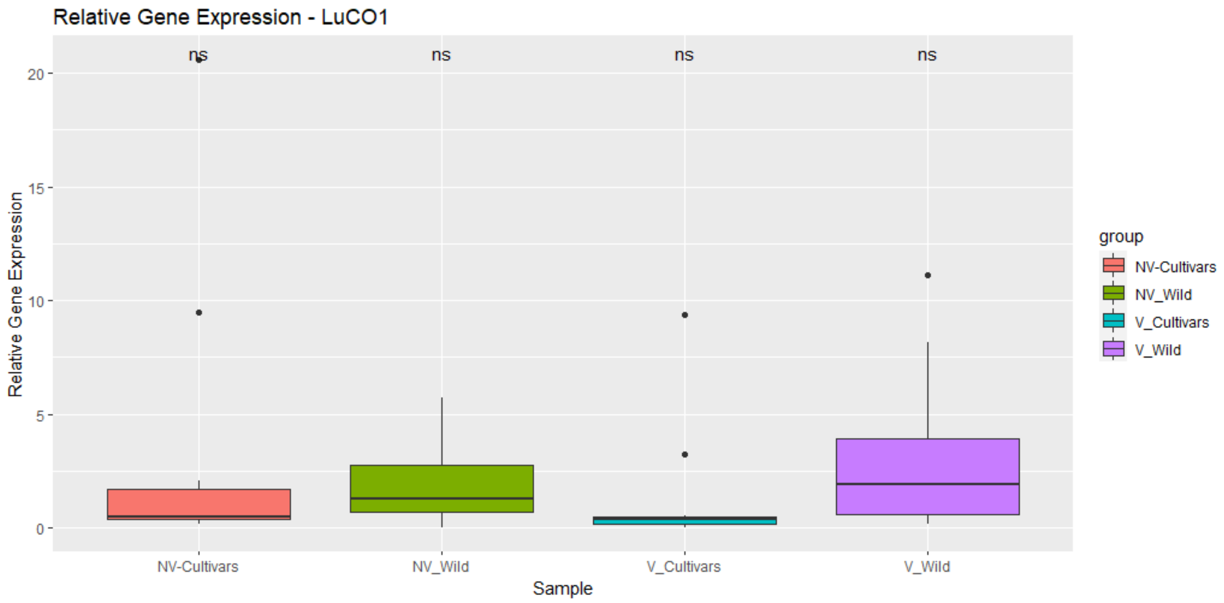


Figure 8A. Relative Gene Expression (RGE) of samples and treatments for the locus *LuCO1*. The significance of multiple pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot. The labels "ns" means no-significant differences were observed between 4 observed treatments.

**B**

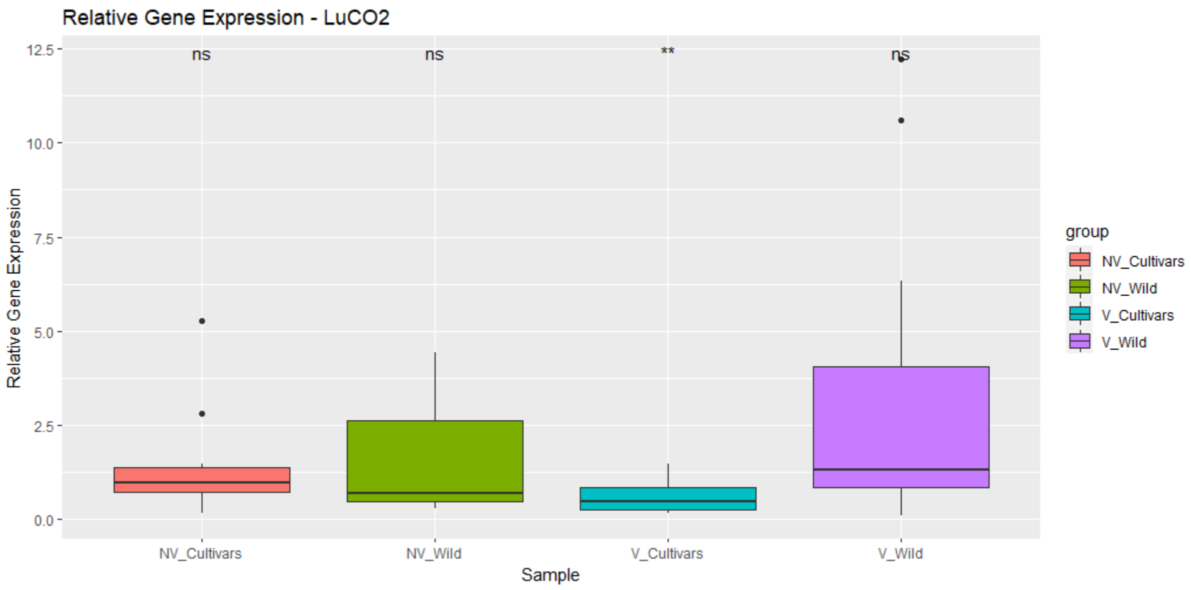


Figure 8B. Relative Gene Expression (RGE) of samples and treatments for the locus *LuCO2*. The significance of multiple pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot. “Vernalized cultivated” (“NV_Cultivars”) show a significant RGE variation when compared to other treatments (“**” means p= <0.01).

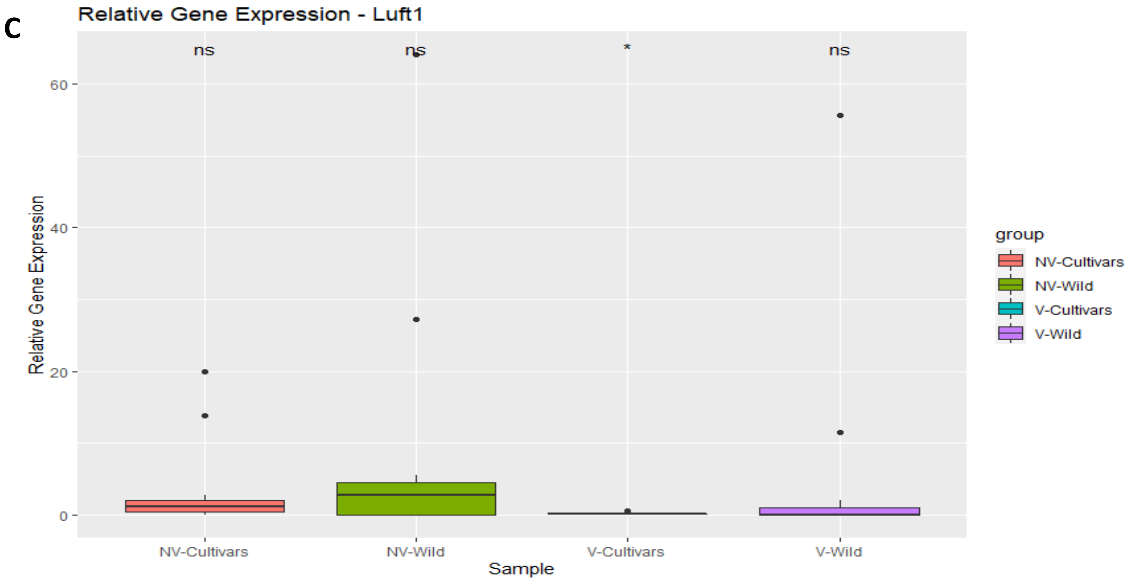


Figure 8C. Relative Gene Expression (RGE) of samples and treatments for the locus *LuFT1*. The significance of multiple pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot . “Vernalized cultivated” (V-Cultivars) shows a significant RGE reduction (“*” means p= <0.05).

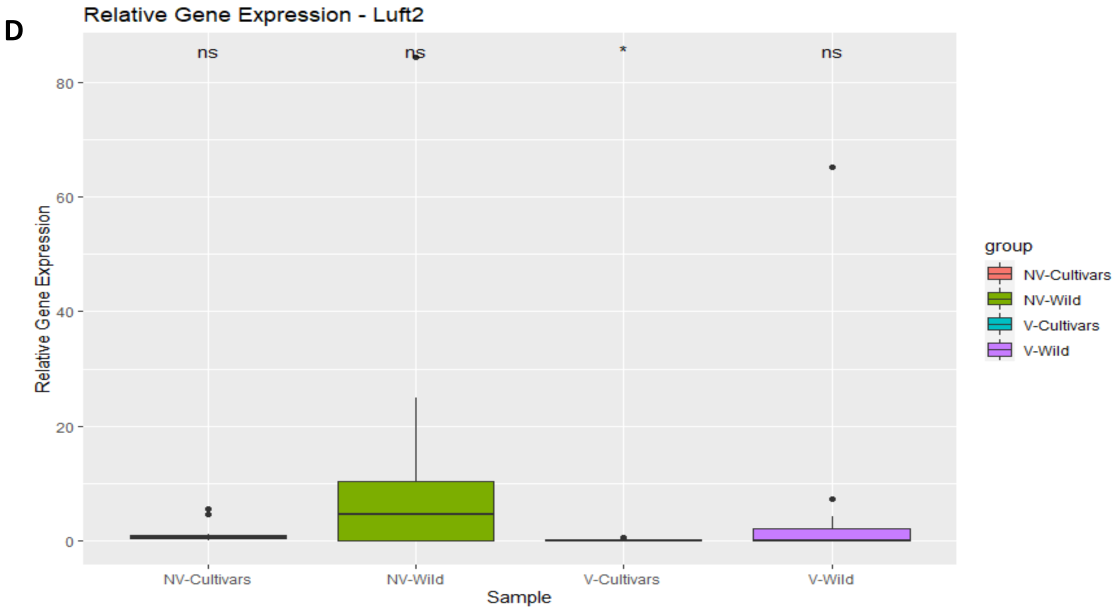


Figure 8D. Relative Gene Expression (RGE) of samples and treatments for the locus *LuFT2*. The significance of multiple pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot . “Vernalized cultivated” (V-Cultivars) shows a significant RGE reduction (“*” means p= <0.05).

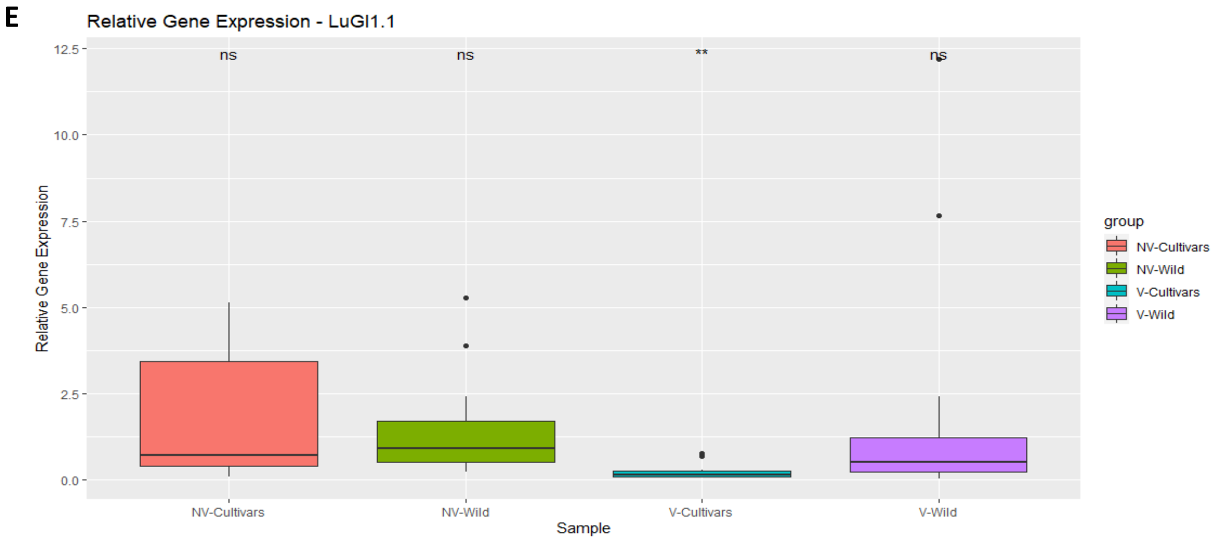


Figure 8E. Relative Gene Expression (RGE) of samples and treatments for the locus *LuGI1.1*. The significance of multiple pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot . "Vernalized cultivated" (V-Cultivars) shows a significant RGE reduction ("**" means p= <0.01).

Variation in expression was mostly seen in cultivars with a potential downregulation in all the tested genes when cultivars were vernalized. Less strong RGE differences were seen in the wild samples. In the case of the comparisons between treatments, most of the differences in the data were observed when treating cultivars to vernalization. This is interesting as phenotypically; this difference wasn't to be expected. This is because flowering initiation in the cultivars did not show faster flowering initiation in response to vernalization and instead showed slightly slower flowering initiation (Figure 6). The multiple pairwise t-test, however, was comparing differences of relative gene expression when vernalized, against all wild and cultivar individuals which are perhaps behaving very differently in terms of expressing each of these genes and maybe independent of each other. Therefore, variation among individuals was explored in more detail. Comparisons between expression responses (relative gene expression) against individuals with different treatment were undergone using a paired t-test as in the next part of this chapter.

### 2.4.5 *LuGI1.1 (GIGANTEA)* Expression Response

We tested fold-change differences (delta-delta ct) as a relative gene expression (RGE) measure to investigate whether differences in gene expression under vernalization treatment can be observed. To be more conservative, since we have two treatments compared with 5 genes, our alpha needs to be adjusted using Bonferroni's correction to 0.005 (0.05/10). Overall differences in *LuGI1.1* expression in cultivars was the closest to approaching significance, however, are not significant when considering Bonferroni's correction *(paired, 2-tailed t-test, p = 0.012 (p > α), t-stat = 3.043, t-Critical two-tail= 2.228)*. The greatest differences were observed in cultivars EDE, BLE, PRI, GIS, and MON (Figure 9).

There was no significant difference expression of *LuGI1.1* between treatments in wild samples (*paired, 2-tailed t-test, p = 0.407, t-stat = -0.864, t-Critical two-tail= 2.228*) (Figure 10).

Expression varied more among samples with observations of both up and down-expression of *LuGI1.1*. Samples Lla_A and Tor_4 increased *LuGI1.1* expression the most and vernalized sample Tal_4 decreased expression the most.

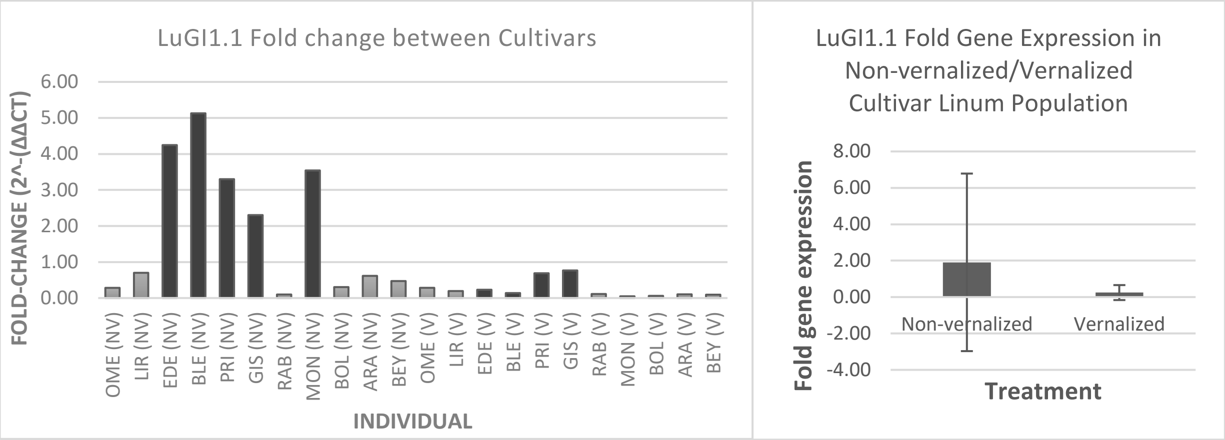


Figure 9. Bar charts showing *LuGI1.1* fold change between cultivars under non-vernalized and vernalized treatments (*paired, 2-tailed t-test, p = 0.012 (p*=>0.005*)*). The cultivars showing the largest downregulation differences are shaded in darker grey.

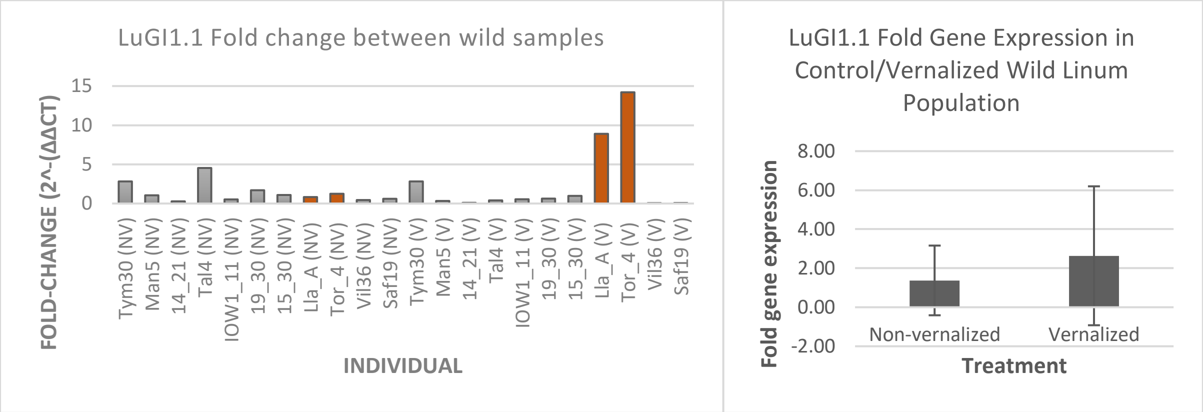


Figure 10. Bar charts showing *LuGI1.1* fold change between wild samples under non-vernalized and vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression of *LuGI1.1* when vernalized, respectively.

### 2.4.6 *LuCO1* and *LuCO2 (CONSTANS)* Expression Response

The two genes *LuCO1* and *LuCO2* did not show a significant expression difference between vernalization treatments for either wild or cultivated samples (Figure 11). However, the variation within the data could still observed.

There were no significant differences in *LuCO1* expression change in cultivars (*paired, 2-tailed t-test, p = 0.229, t-stat = 1.281, t-Critical two-tail= 2.228*). Most of the variation observed occurred in sample LIR with decreased expression when vernalized. Expression of *LuCO1* was neither significantly different between non-vernalized nor vernalized treatments of the wild samples (*paired, 2-tailed t-test, p = 0.310, t-stat = -1.070, t-Critical two-tail= 2.228*). Different samples showed increased and decreased expression of *LuCO1*. As in the case of *LuGI1.1*, samples lLa_A and Tor_4, and 19_30, increased expression after vernalization. Sample Tal_4, as well as Tym_30 and Saf_10 decreased expression after vernalization.

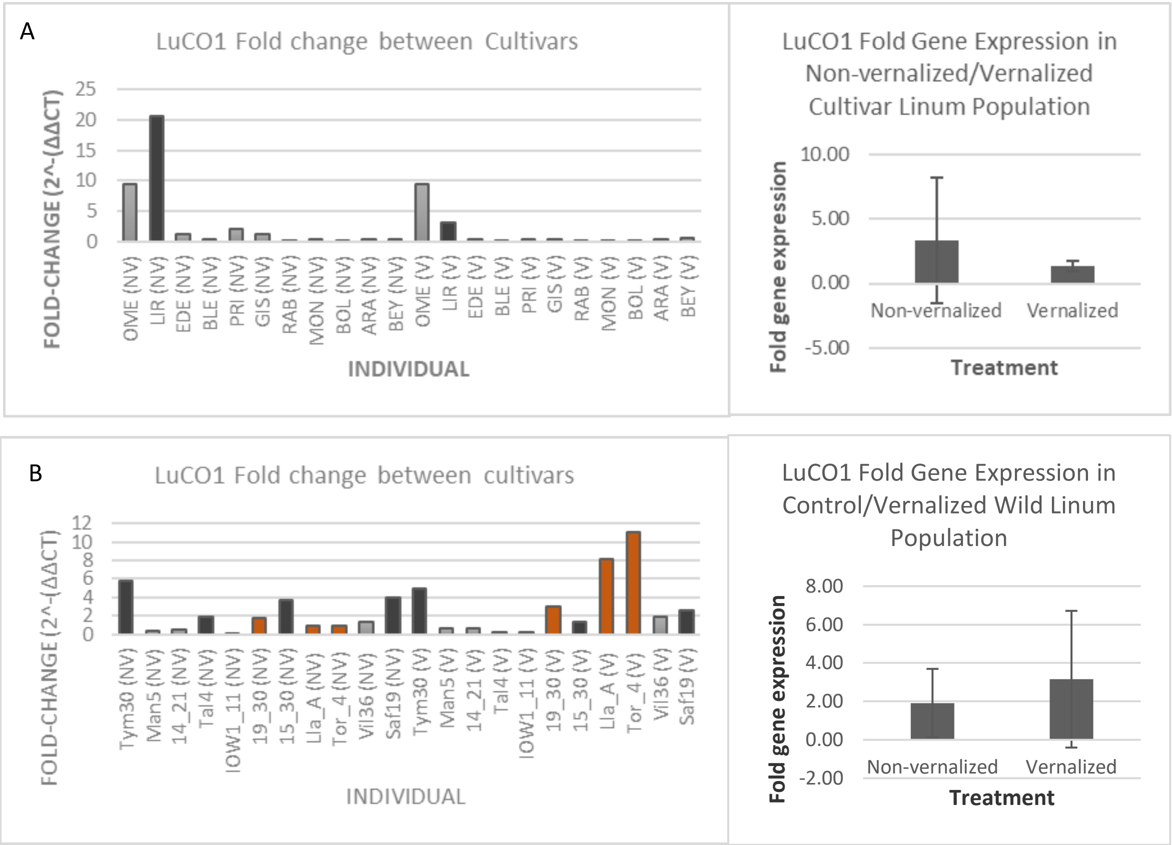


Figure 11. Bar charts showing LuCO1 fold change between cultivars (A) and wild types (B) under non-vernalized and vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression when vernalized, respectively.

There were non-significant differences in *LuCO2* expression between the non-vernalized and the vernalized treatments for cultivars *(paired, 2-tailed t-test, p = 0.132, t-stat = 1.640, t-Critical two-tail= 2.228)*, and wild samples *(paired, 2-tailed t-test, p = 0.222, t-stat = -1.301, t-Critical two-tail= 2.228)* (Figure 12). The trend was for *LuCO2* expression to decrease in cultivars during vernalization, while the trend for wild samples was to increase *LuCO2* expression. This revealed that although there were overall significant differences when comparing vernalized cultivated samples gene expression for *LuCO2*, when testing between vernalized and non-vernalized cultivated samples only, the expression of *LuCO2* shows no significant difference and results to vernalization having no significant effect in the relative gene expression of *LuCO2*, for both cultivar and wild samples.

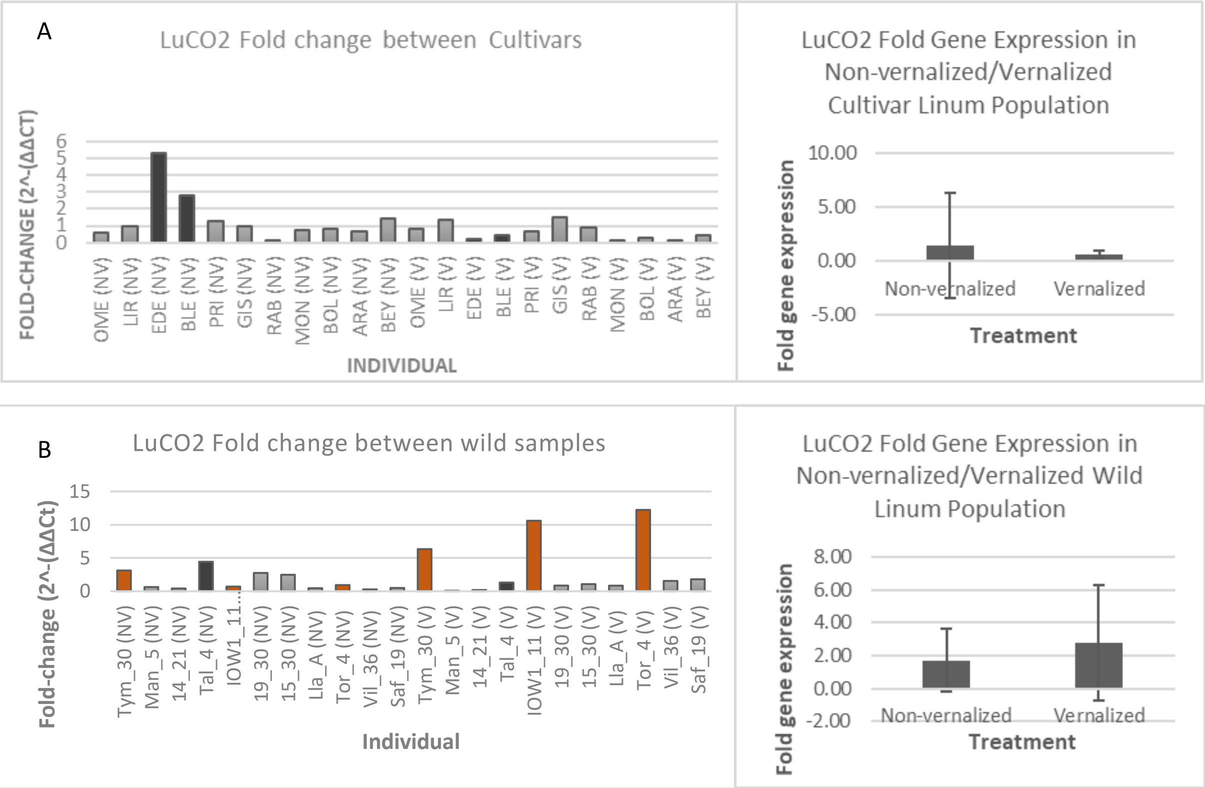


Figure 12. Bar charts showing *LuCO2* fold change between cultivars (A) and wild types (B) under non-vernalized and vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression when vernalized, respectively.

### 2.4.7 *LuFT1* and *LuFT2 (FLOWERING LOCUS T)* Expression Response

The cultivars showed a trend of decreased expression of *LuFT1* when vernalized. This was seen more in cultivar samples EDE and BLE, alike the *LuCO2* results. The overall change in gene expression was not significant *(paired, 2-tailed t-test, p = 0.093, t-stat = 1.856, t-Critical two-tail= 2.228)* (Figure 12A). Wild samples showed a mixed pattern with one potential anomaly for sample Saf_10 (not shown) with much greater *LuFT1* expression than other wild samples across both treatments. There was no significant expression change in *LuFT1* expression when wild samples were vernalized *(paired, 2-tailed t-test, p = 0.669, t-stat = 0.440, t-Critical two-tail= 2.228)* (Figure 13B).

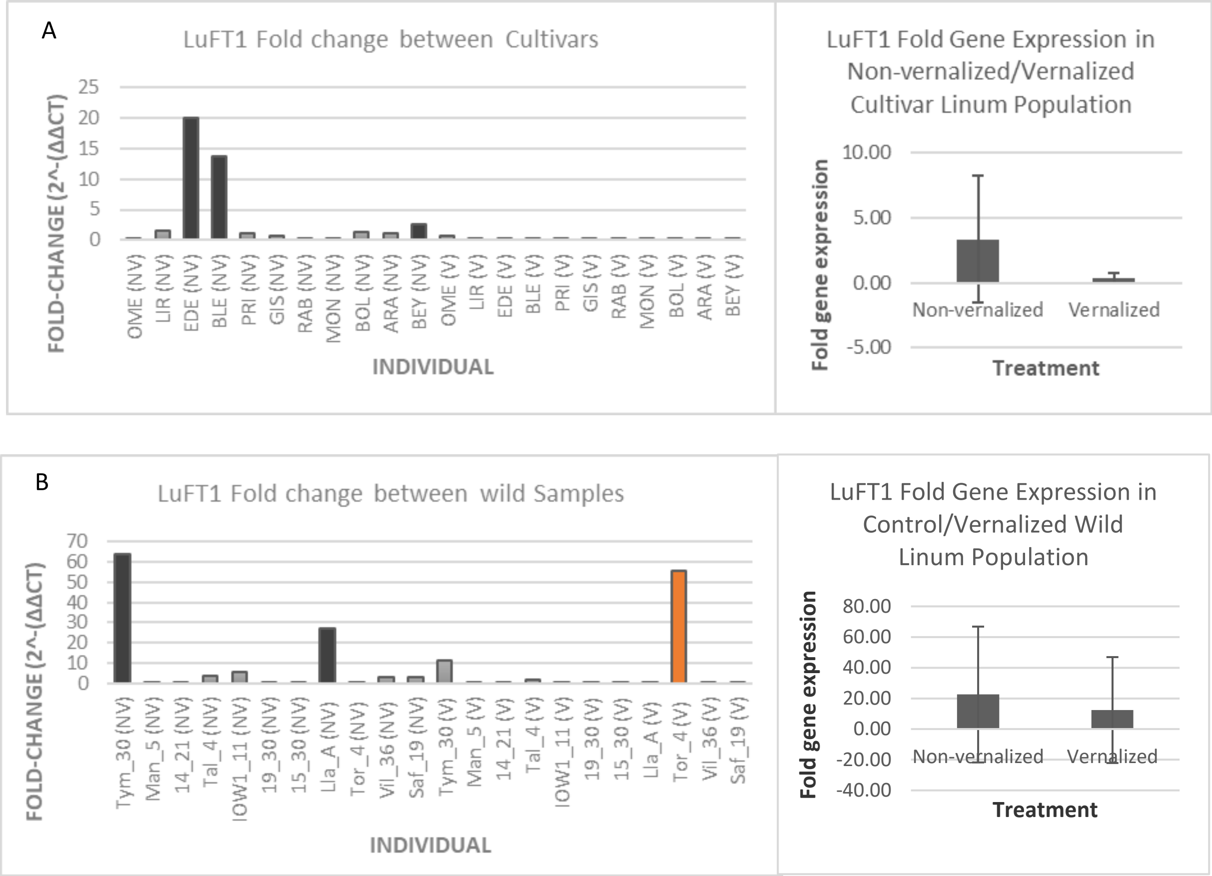


Figure 13. Bar charts showing *LuFT1* fold change between cultivars (A) and wild types (B) under non-vernalized and vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression when vernalized, respectively.

As was observed for *LuFT1*, cultivar samples EDE and BLE showed the most reduced *LuFT2* expression after vernalization with non-significance on the fold change between the two treatments for the entire cultivar sample set (*paired, 2-tailed t-test, p = 0.087, t-stat = 1.916, t-Critical two-tail= 2.262)* (Figure 14A). Among the wild samples, an exceptionally high downregulation in expression of *LuFT2* (>100 fold difference, not shown) was observed sample Saf_10. This sample was treated as an anomaly. Nearly the same trends as *LuFT1* were seen, with sample Tor_4 showing the most increase in expression of *LuFT2*. However, the results between the two treatments for the wild sample set were still not statistically significant altogether (*paired, 2-tailed t-test, p = 0.575, t-stat = 0.579, t-*

*Critical two-tail= 2.228)* (Figure 14B). Relative gene expression was also tested between species in only the vernalized samples using a paired, 2-tailed t-test. None of the tested flowering time genes have shown significant expression difference between species when vernalized (α = >0.05).

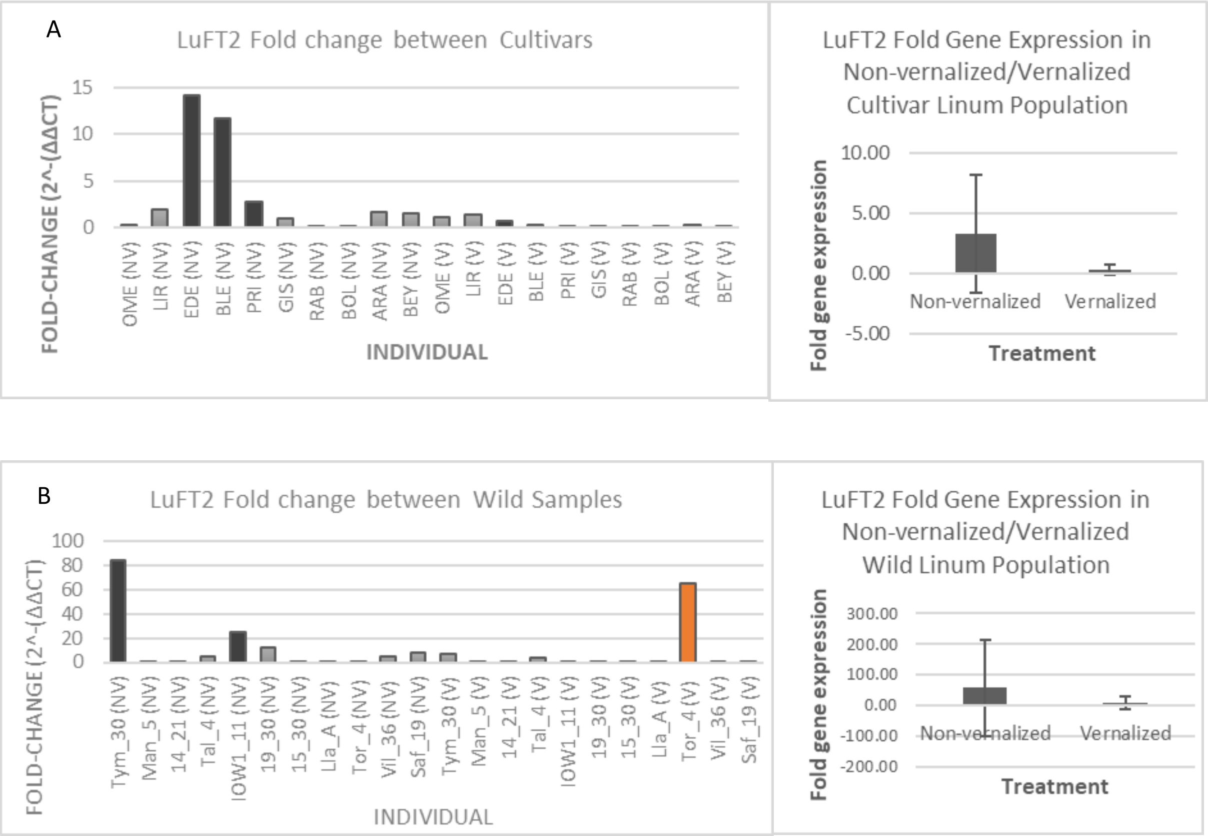


Figure 14. Bar charts showing *LuFT2* fold change between cultivars (A) and wild types (B) under non-vernalized and vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression when vernalized, respectively.

## 2.4.8 Gene Expression and Flowering Time

General linear models (GLM) were used to test whether the difference in RGE between vernalized and non-vernalized wild and cultivated *Linum* were associated with the flowering time initiation (Figure 15). In wild samples, the relative gene expression (RGE) differences of *LuFT1* and *LuFT2* were significantly positively associated with days to flower after vernalization (*LuFT1 t=6.359 p=<0.001; LuFT2 t=5.128 p=<0.001*). The other tested loci, *LuCO1*, *LuCO2*, and *LuGI1.1* showed no significance when tested for association with flowering time (*LuCO1 t=1.510 p=0.162; LuCO2 t=1.324 p=0.215; LuGI1.1 t=1.766 p=0.108*).

For the cultivars, it was found that *LuFT1* and *LuFT2* expression differences were not significantly associated with flowering initiation (*LuFT1 t=1.226 p=0.237; LuFT2 t=1.732 p=0.122*). Neither did *LuCO1* and *LuCO2* expression differences show significant associations with flowering initiation (*LuCO1 t=-1.575 p=0.150; LuCO2 t=1.287 p=0.230*). Interestingly, *LuGI1.1* expression differences showed a significant positive association with flowering initiation (*LuGI1.1 t=2.932 p=0.016*).

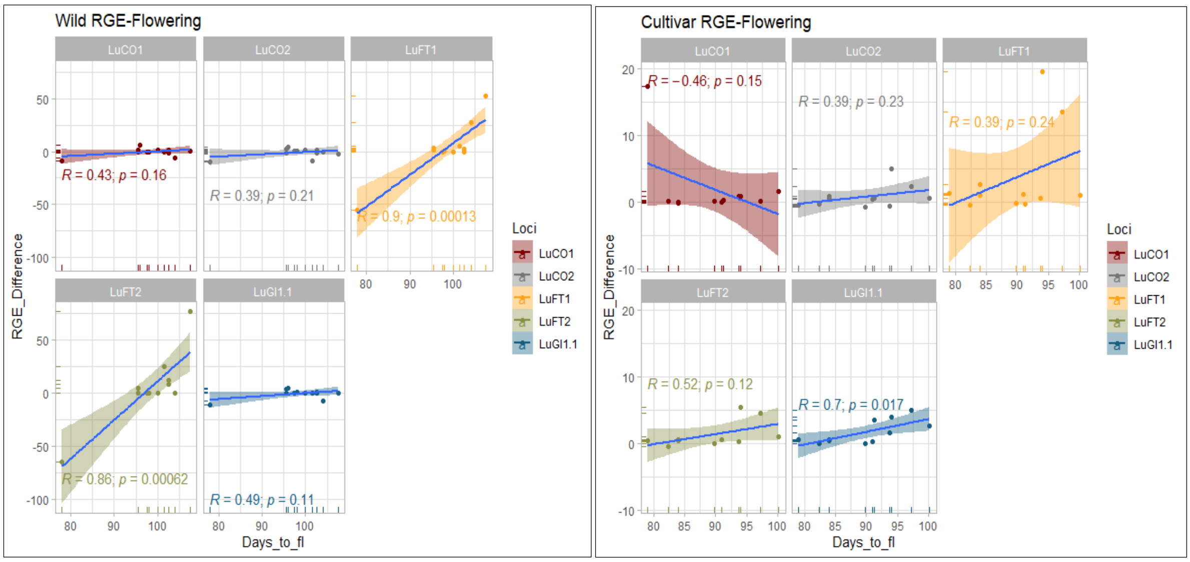


Figure 15. Scatterplots showing number of days to flower after vernalisation labelled as “Days_to_fl” in relation to relative gene expression difference in wild samples and cultivars separately.

### 2.4.9 Relative Gene Expression and Latitude

The earlier phenotypic analysis found that vernalized wild individuals dramatically reduce the number of days required to flower when compared to non-vernalized wild plants, with populations from more northerly latitude of origin showing vernalization sensitivity. Here we tested latitude against relative gene expression for the cultivars and wild samples (Figure 16).

Latitude was positively associated with relative gene expression difference in the genes; *LuFT1 (t=2.668 p=0.0257); LuFT2 (t=3.443 p=0.0076); and LuGI1.1 (t=2.320 p=0.0428*). In the cultivars, there was no correlation between any gene expression tested in this study and latitude.

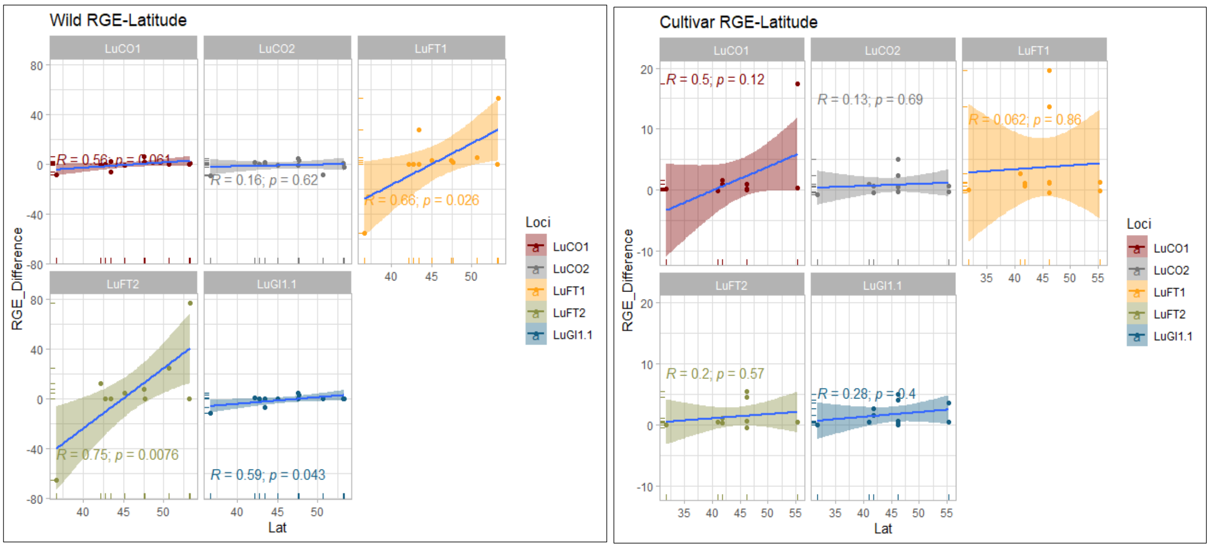


Figure 16. Scatterplots showing latitude of origin (Lat) in relation to relative gene expression difference in wild samples and cultivars separately.

## 2.4.10 Results: Relative Gene Expression and Climate

More Northerly location (higher latitude) suggests of colder, more wet, and windier climates. Wild samples were analysed for the relationship of flowering regulations genes expression differences and several climatic variables under that latitude. Climate of origin was summarised into principal components, and princliple component 1 (PC1) was then tested against relative gene expression differences as mentioned under section 2.2.1 “Samples and Experiment” of this chapter.

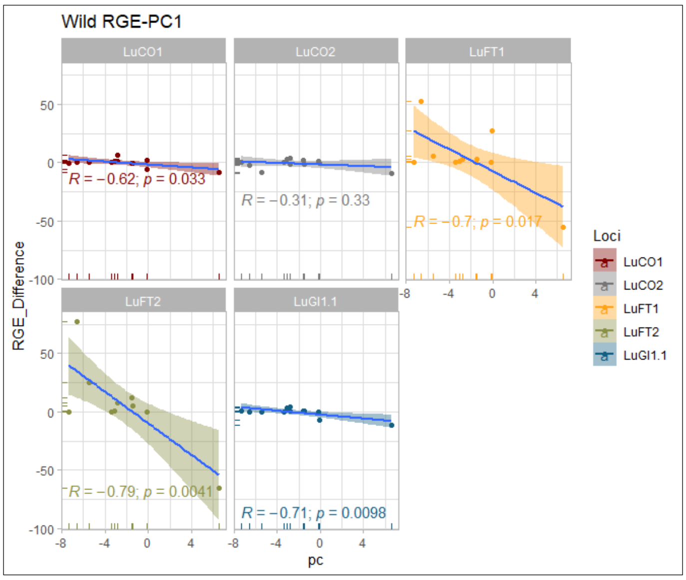


Figure 17. Scatterplots showing climate variable PC1 (pc) in relation to relative gene expression difference of wild samples.

Expression differences of *LuFT1* and *LuFT2* were significantly negatively associated with PC1. Lower PC1 values suggests colder climate (appendix 3) a summary of all the climatic variables (*LuFT1 t=-2.905, p=0.017; LuFT2 t=-3.812 p=0.004*). There was also a significant negative association between climate and *LuGI1.1* (*t=-3.180 p=0.009*). The correlation observed in the wild for *LuGI1.1* against PC1 is more significant than it is observed under latitude alone. This suggests that climatic variables represent more significant correlation to the RGE observed under *LuGI1.1* and that that locus may be more varied in the wild according to the local climate than the latitude alone.

## 2.5 DISCUSSION AND CONCLUSION

In the 2018 vernalization experiment, vernalization reduced the number of days to flowering in the wild *L. bienne*, especially in more Northern populations. This suggests that vernalization is an important mechanism in *L. bienne* and there was variation in regulation of flowering time in the wild relatives based on their latitude of origin. In addition to this, the results for *L. usitatissimum* suggests that requirements for vernalization were different between wild and cultivated *Linum*. This difference could be due to artificial selection for faster flowering of cultivars as part of domestication in *L. usitatissimum*. In the literature, genetic and association mapping have identified hundreds of genes as targets of divergence due to domestication (Smýkal P. et al., 2018). Evidence of this has been previously observed through resequencing of candidate genes related to seed sizes and weight (Guo *et al.*, 2020a). Vernalization insensitive flowering time could also be linked to productivity of *Linum* in temperate climates, and thus mechanism to vernalization is of interest in *Linum* breeding (Gutaker *et al.*, 2019).

Other studies concluded that temperature, photoperiod (day length), and light availability (amount of sun light during the day) can influence developmental rates in crop plants such as wheat and other annual crops (Craufurd and Wheeler, 2009; Cave *et al.*, 2013; McMaster *et al.*, 2008). In wild *L. bienne*, variation in environmental conditions at different latitudinal localities have selected for differences in flowering initiation due to vernalization. This variation was not observed in our cultivated *L. usitatissimum* collection, suggesting no sensitivity to vernalization. However, in the literature there are suggestion that winter type *L. usitatissimum* found in Texas, USA, are sensitive to vernalization (Darapuneni M. et al., 2014). They suggest that these winter types are different to varieties grown in more Northern areas because they are grown in Autumn due to higher spring and summer temperatures in Texas. We saw no evidence of sensitivity to vernalization in our Northern European Spring and Winter varieties of *L. usitatissimum*. We suggest that there is wider variation within *L. usitatissimum* for requirement to vernalize. Perhaps growing seasons of each variety can affect the sensitivity of *L. usitatissimum* variant to vernalization.

Another result illustrated in this study is the difference of gene expression due to vernalization conditions. After considering correction for multiple testing and variation among individuals, there was no significant difference due to vernalization in the fold change of any of the tested genes in both wild and cultivar types. Sun et al., 2019 also did not find an expression difference for *LuGI1.1*, between their cultivar lines (Sun et al. 2019). In addition to this, vernalization conditions tested in this study were not significant for *LuGI1.1* gene expression either. There is little evidence that among the genes tested in this study, cultivated and vernalized individuals showed a significant difference in gene expression due to vernalization. This suggests that fold changes in all three flowering time genes expression were not significantly affected by vernalization. However, we have not tested fold-change between populations, as we did not have enough samples over population in our samples to compare for this. To test for population variation, more individuals over each population could be tested in the future.

It is interesting to observe that in our wild *Linum*, there was a clear positive correlation between the RGE difference of both tested *LuFT* copies in relation to the number of days to flower, otherwise stated as “flowering initiation” in this thesis. These results are alike to previous photoperiodicity experiments of other cultivar types, whereby both *LuFT1* and *LuFT2* were shown to be associated with flowering initiation under photoperiodicity, but not the other flowering time genes tested (Sun et al, 2019). However, this relative gene expression difference was not observed in the cultivated *Linum* samples tested in this study. Instead, expression of *GIGANTEA* (*LuGI1.1*) was observed to be positively correlated to the number of days to flowering. There are suggestions that the flowering time gene *GIGANTEA* is involved in flowering and maturity development in heterozygous lines of the plant species *Glycine max* (Watanabe *et al.*, 2011). This becomes particularly interesting when considering the population genetics of a given study species. The illustration that the gene *GIGANTEA* is involved in flowering development of heterozygous *Glycine max* varieties, suggests that heterozygosity could become a variable to test against *GIGANTEA* expression under future studies involving cultivated *Linum*.

The difference in *LuGI* expression was not observed in the wild *Linum* as was observed in the cultivars. The differences in relative gene expression found in the current study, reveal that there is a difference in gene regulation, perhaps due to different vernalization requirement between the wild and cultivated *Linum* and their population genetic variability. It is worth considering that flowering initiation is very different in the cultivated types when looking at the effects of vernalization to the number of days to flower between the two species. This study reveals that the expression of the flowering integrator gene *LuFT* have lost the ability to respond to vernalization in developed cultivars, and instead expression of another flowering time regulator gene may have taken its role in terms of initiating flowering. A cause could be changes in genetic expression of flowering regulators throughout developments of plants. Changes in flowering time gene roles and developmental switches from vegetative to reproductive stages of plants is not new. In the model *Arabidopsis thaliana*, it has been shown that flowering time regulator (*FT*) changes expression drastically in relation to developments (Blümel *et al,* 2015). Upstream genes in the flowering time network, such as *CRY2* (Cryptochrome Circadian Regulator 2) and *LHY* (Late Elongated Hypocotyl) (Park M et al,

2016) may influence the non-significant relative expression differences of *LuFT* to flowering initiation and may have a major role to play in terms of this loss of vernalization requirement in the cultivars. Determining how vernalization affects the flowering time gene network, especially in *Linum* cultivars would be useful for agriculture to find genes which are important for specific growing conditions and exploring the potential loss of certain genetic functions due to selective breeding and human cultivation.

For most of the tested genes, vernalized, and non-vernalized expression changes were not significantly different in wild and cultivated samples. In the literature however, there are some studies in other non-model organisms, which reveal that *GIGANTEA* homologous genes were differently expressed under vernalization treatments. This was true in a study using Ryegrasses (Paina *et al.*, 2014), and white lupins (Rychel *et al.*, 2019). However, studies regard *GIGANTEA* to be a photoperiodic response gene. This comes with implication that *GIGANTEA* interacts with another photoperiodic gene, the *CONSTAN* (CO). However, some studies have illustrated that *GIGANTEA* may be independently mediating photoperiodic control of flowering (Jung *et al.*, 2007). Although, there is evidence the *GIGANTEA* may be variably expressed when vernalized in other non-model organisms, the experiment in this case was limited in the number of paired samples that experienced both vernalization and non-vernalization treatments in 2021 (9 wild and 8 cultivated). This is due to availability of DNA materials (only available for 2021 experiments) and restrictions caused by the Covid-19 pandemic. There were 62 germinating individuals in the vernalization 2021 chamber. However, only around 20 individuals germinated in the non-vernalization 2021 chamber. Some of these have not had enough time to produce leaf materials for RNA extraction. In this study, there were only 38 individuals which experienced vernalization in both 2018 and 2021 experiments for which there was both expression and phenotypic data. In the 2018 experiment (pre-Covid-19) there were 197 vernalized individuals. This gives an idea about the limitation of data available for genetic study under the 2021 experiment due to restrictions caused by the Covid-19 pandemic. However, as in Figure 4, flowering time results from the 2018 experiment revealed correlation in flowering time and latitude in wild relatives and in addition suggests the loss of vernalization in cultivars. In the future, time permitting, an experiment involving vernalization, looking at a larger set of samples, may in turn reveal a stronger difference between the gene fold-change in vernalized and non-vernalized individuals.

Another aspect that could be of interest from the genetic point of view is epigenetic regulation. Epigenetics is related to the idea that genetic regulation may be affected by different environments, which can include temperature changes. This was studied in the model *Arabidopsis,* looking at genetic regulation of flowering time (Khan, Ai and Zhang, 2014) and epigenetic responses to heat (Liu *et al.*, 2015). In the current study, certain genes in the wild samples showed a mix of increased and decreased expression across treatments. For example, the Southern samples Lla_A and Tor_4 tended to show increased expression (upregulation) when vernalized, and some Northern samples such as Tym_30 showed decreased expression (downregulation) of multiple flowering initiation genes when vernalized. Across wild sample populations, there were differences in genetic regulation based on their locality, or in this case latitude for *LuFT1*, *LuFT2*, and *LuGI1.1*. This further adds to the point that different population may regulate flowering time genes differently based on adaptation to their local environment which often affect downstream processes such as biosynthesis of plant hormones leading to developmental differences (Jaakola and Hohtola, 2010). A study of the model *Arabidopsis* comparing differences in altitude reveals that there was variation in genetic regulation of flowering time regulators, for populations under different altitude. These include a tested flowering regulator in this study, the Flowering Locus T (*FT*), with addition of vernalization genes such as VERNILIZATION INSENSITIVE 3 (*VI3*). Regulations of flowering regulator have been revealed to regulate MADS protein downstream and this differed between low and high altitudes genotypes (Suter *et al.*, 2014). Test of more *Linum* samples in consideration of altitude of different populations could help verify these results for *Linum* in future studies.

Regarding the cultivars, most of the expression differences were reduced expression following vernalization. Reduced expression of flowering time genes such as the *LuCO1* and *LuCO2* (CONSTANS 1 and 2) could be a response to the expression of downstream flowering meristem identity such as SUPPRESSOR OF OVEREXPRESSION OF CONSTANS (*SOC*) to initiate flowering following vernalization as has been found in the model *Arabidopsis* (Sasaki *et al.*, 2017)(Li *et al.*, 2015a)(Valentim *et al.*, 2015a). Therefore, further studies relating expression of homologues to *SOC* genes in *Linum* to expression changes of *CO* genes in response to vernalization could help confirm this vernalization pathway in *Linum*.

Flowering time gene expression of *LuFT1, LuFT2, and LuGI1.1* were found to be associated with environmental variables related to latitude and climate of origin of wild samples. This result suggests that in the wild, phenological adaptation to the local environments could be due to different gene expression responses of these three flowering time genes. This is supported by the suggestion that *FT* was found to be a major locus contributing to local adaptation in flowering of perennial plants (Wang J. et al., 2018). The strongest associations between genetic expression as well as phenotypic responses (flowering time) to the local environments was shown by both *LuFT* genes. This contrasts with the result for wild flowering time that was not significantly associated with *LuGI1.1* relative gene expression. However, latitude and environmental variables were associated with *LuGI1.1* relative gene expression. In addition, in the cultivar set, *LuGi1.1* was the only gene that showed an association with flowering time. It is of interest to look at genes that maybe influencing the expression of *LuGI1.1*. It may be that expression of those genes were more related to flowering time than *LuGI1.1* itself. For example, there are multiple variants of the gene *GI* within *Linum* that were not tested in this study. These genes could be tested as part of future to further explore the flowering time network in *Linum* and implications of environmental variables. In addition, other vernalization related genes have also been observed to be related to local adaptation. The *FRI* (*FRIGIDA*) gene are related to local adaptation to drought tolerance by controlling flowering time in *Arabidopsis thaliana* (Lovell J. et al., 2013; Tigano and Friesen., 2016). In future studies, *FRI* can be another gene of interest when suggesting local adaptation to vernalization responses in *Linum*. In conclusion, evidence of local adaptation can be inferred from the phenotypic variation observed in the wild *L. bienne* tested here. This is further supported by the variation in expression of *FT*, linked to correlation towards environmental variables.

One issue in this study was the decision to use of only one HKG. The use of multiple HKGs is regarded as superior to using only one HKG for quantification purposes as it gives a better picture of background expression (Remans *et al.*, 2008) (Manoli *et al.*, 2012). However, in the case of this experiment, the efficiency of *LuUBI2*, one of the tested HKG primers was very low (51%). This means that using both genes may skew the results due to the less efficient HKG so the Ct-values for *LuUBI2* were dropped from the analysis to conserve for this. In addition, the efficiency of some of the genes of interest (GOI) could be improved as part of future research. Quantification of relative gene expression using two HKGs was also explored as part of the analysis, with the less efficient *LuUBI2*. The results showed a drastic difference in relative gene expression when using two HKGs, with no significant differences between the treatments, potentially skewing relative expression. Using two-HKG indicate that the genes *LuFT1* and *LuCO2* showed the greatest difference in expression between vernalized and non-vernalized (non-vernalized) samples, but these differences were not statistically significant (appendix 8). This result suggests that using the less efficient primer may skew our results. We suggest improving the efficiency of *LuUBI2* for future research looking on the relationship of flowering time gene expressions with vernalization, using multiple HKGs. This would strengthen our results suggesting *LuFT1, LuFT2, and LuGI1.1* associations with environmental variables as found using one HKG in this study. Until such a time, our results are to be treated cautiously as a signal for the effects of environmental variables in relative gene expressions of *LuFT1, LuFT2, and LuGI1.1*.

The relationship between relative gene expressions (RGE) difference calculated with two HKGs was tested against latitude of origin of wild samples (appendix 9) and no significant relationships were found. These results contrast to the results found above, using one HKG and suggest that the low efficiency value of *LuUBI2* could have contributed to masking some significance seen with only one HKG. In the future, using two HKGs with a higher efficiency value are recommended to test the relationships between gene expression to latitude, and climatic variables. We reflected that using one HKG is not ideal as most recent relative gene expression studies recommends using multiple HKGs. However, limited results on the expression of the floral integrator *FT* using one HKG are suggestive. To robustly test the expression of *FT* in *Linum* under vernalization, observation of HKG with higher efficiencies are recommended for future research.

To further this research, other important regulatory flowering time genes in the network could be tested such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* (*SOC*) and as well as homologues of further downstream meristem identity genes found in the model Arabidopsis such as *LEAFY* (*LFY*), *APETALA1* (*AP1*), *SEPALLATA3* (*SEP3*) and *FRUITFULL* (*FUL*) (Schopf *et al.*, 1996) (Gregis *et al.*, 2009). Meristem development in *Linum* could be measured directly as a phenotypic comparison as has been done in other studies (Kayes and Clark, 1998) (Heisler and Jönsson, 2007). The complex network of flowering time genes could also imply that there is no single gene responsible for the phenotypic differences seen when both wild and cultivar type are vernalized. It may occur that several genetic regulations contribute, including the five major genes tested here. However, real-time primers that are specific to the two paralogues of each gene present in *Linum* would first need to be designed and tested, which is not a trivial task (Sun et al.). More genes which are specific to the vernalization pathway should be tested. Outside of the main floral pathway integrators, which *FT* (Flowering locus T) is a part of, this study looked only at previously designed flowering time genes which were related to photoperiodicity, such as *CO* and *GI*. It is interesting therefore, that the two *FT* loci showed significant relationships with latitude and climate of origin when observing the wild types. Vernalization-specific genes such as *FLC* (Flowering locus C) and *FRI* (FRIGIDA) have been shown to also play a role in the regulation network to suppress flowering locus T in the model *Arabidopsis* (Flowers *et al.*, 2009a). It would be of interest then to look at these vernalization pathway genes with appropriately designed primers specific for those genes in *Linum* as part of future studies.

For further studies, it would also be of interest to measure traits such as seed yield. *Gutaker et al.,* have talked about potential link of flowering time to fruiting. Expression data for genes which may affect flowering time could be linked to whether seed yield from these measured individuals are affected by the changes of expression. Furthermore, variation in seed germination due to vernalization has been illustrated in the model *Arabidopsis* (Auge *et al.*, 2017). This may potentially link to productivity in *Linum* that may vary due to the requirement of vernalization and therefore affecting flowering time.

# CHAPTER 3: THE QUANTIFICATION OF POLLEN AND POLLEN TUBES TO REVEAL POLLEN GERMINATION VIABILITY IN *LINUM* UNDER DIFFERENT TEMPERATURE TREATMENTS.

Pollen is the vehicle in which the male gamete is found. It is considered as a microgametophyte (sperm-producing gametophyte) in plants, a powder-like substance where the male gamete is produced and transferred (Johnstone, 2001). Pollen grains needs to disperse and travel to the stigma to fertilize egg cells which are found in the ovule and are non-dispersal. Pollen will often encounter many ecological and environmental pressures as they disperse (Loveless & Hamrick., 1984). They are considered as one of the major contributaries to the ability of flower to turn into seeds and produce progenies, a transformation from plant haploid gametophytes into diploid sporophytes (Pacini & Dolferus, 2016). Pollen is essential for seed production and thus agriculturally significant in crop plants such as *Linum*. As well, wild flowering plants often deploy different flowering strategies to optimise germination by pollen, and along with this, have evolved different germination mechanisms useful for artificial selections of male genotypes (Williams, 2012) (Tel Zur, et al., 2020). In this chapter we quantified pollen and pollen tube counts in *Linum* individuals treated under different temperature stress.

Pollen must fertilize the ovary through the stigma-style-ovary system to form diploid sporophytes and form seeds (Pacini & Dolferus, 2016). The receptive ability of the stigma-style-ovary system is another aspect that could be observed with regards to the success of pollen germination. In this study however, we will focus on the pollen’s ability to germinate and fertalize ovaries. Assessment of *in vivo* pollen germination success can be done by counting the number of pollen tubes in the style (Alonso et al., 2013; Williams & Reese, 2019). We decided to observe pollen initially because pollen is relatively easier to stain and study under confocal microscopy than the stigma. In a book published in 1998, Delph F, *et al,* reviewed variation in pollen growth, where there is an observation that pollen growth varies between species due to pollen/sperm competition whereby pollen traits such as, how quickly pollen reach the ovules, or how many pollen grains are transferred influence the success of germination. In addition, pollen competition in plants is described as gametophytic selection, potentially affected by environmental cues such as ambient temperature. This competition may cause pollen traits to differ in different flowering plants (Delph and Havens, 1998). In our *Linum* samples, we observed the viability of pollen and pollen tubes for the germination into seeds. We are particularly interested in the effect of temperature towards pollen viability under both wild and cultivated species.

## 3.1. Pollen Structure

In terms of their physical appearance, pollen grains varies in size across species (2-200µm) and contains parts where the generative cells are stored, often with a tube nucleus, within an intine, exine, and an aperture pore whereby the gametes can be omitted from using a structure widely known as pollen tubes (Johnstone, 2001; Pacini E, 2008). Pollen grains are microscopic. The smallest pollen size in diameter is found in *Myosotis*, with a diameter of 2.4-5µM (Sporemex, 2022), this often makes microscopy an ideal tool for observing pollen structures. The structure of a typical pollen is illustrated as below (Unacademy, 2022):

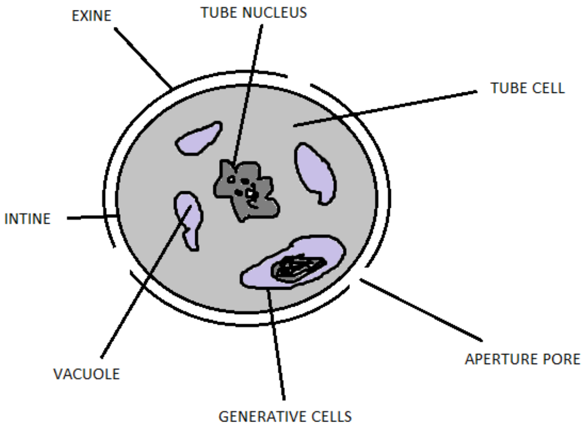


Figure 1. A diagram depicting the well-known structure of a mature pollen.

During germination, the tube nucleus part of the pollen expands into a pollen tube whereby the male gamete of the pollen can be transferred to the ovary. Pollen tube expands out of the pollen through the "aperture pore". The number of aperture pores can vary between species. It is thought that increased number of apertures in angiosperm pollen grains can offer the species a selective advantage as it increases the number of germination sites and facilitates the chances of contact between the pollen tubes and the female ovule (Furness & Rudall, 2004).

For germination, pollen often require active RNA synthesis. This was evident in the conifers, where RNA synthesis for pollen germination was studied for its dependence on transcription and translation (Breygina et al., 2021). When pollen is germinating, the tube nucleus will extend towards the aperture pore and will then form a pollen tube. These pollen tubes are organelles which can elongate and are responsible for the transfer of pollen male gametes into the stigma and down to reach the ovule (Adhikari et al., 2020). When pollen tubes reach ovules, male gametes within the tube nucleus enter the ovaries to form zygotes. The zygotes are the progenitor stem cells which can form embryos that later form into seeds (Kanday and Sundaresan, 2021). With this knowledge, it is then important to consider pollen viabilities in terms of their ability to form zygotes. This includes, amount of pollen, amount of pollen tube , and the rate which pollen tube reaches ovary. This can lead to factors influencing seed formation success (Iwazumi and Takahashi, 2012).

Pollen viability is therefore of interest related to the breeding strategies of *Linum* as both wild and cultivated types are present under different environmental conditions. In *L. usitatissimum*, pollen can be used as a tool to measure geneflow using pollen as a mediator, which may in turn can detect different strategies and local adaptations related to the rate of pollen germination under different conditions (Jhala, et al., 2010). It is therefore of interest, to look at pollen germination rate and pollen viabilities when looking at local adaptations as well as strategies of breeding systems. In the literature, study species such as cotton (*Gossypium hirsutum*) shows a high tolerance to heat. Optimum temperature for pollen growth and pollen tube length in cotton have been observed to be above 32$^{o}$C (KAKANI et al., 2005). Other observations in temperate cultivars such as apricots and sweet cherries (*Prunus spp.)* suggests that pollen germination and pollen tube length are higher in some cultivars treated with colder temperature (5°C), although varying within different lines (Pirlak., 2002). Previous observations of pollen tubes were made in another agriculturally significant plant, peach (*Prunus persica*), by Hedhly A. *et al* (Hedhly & Herrero., 2008). This study found that drastic increase in temperature had a negative effect on stigmatic reception of the pollen tubes, thus significantly slowing pollen tube growth. These results suggest that stigmatic cells' ability to sustain and adhere the pollen tube cells were reduced because of drastic temperature effect and the kinetics of pollen tube growth was also affected. This behaviour was also observed in the field (Hedhly & Herrero., 2008). A similar negative effect of temperature on pollen kinetics and stigmatic cell adhesion was also previously observed by the same group in cherries (Hedhly et al., 2004).

It has been determined that pollen germination and pollen tube growth could be observed *in vitro* and semi *in vivo* for the model *Arabidopsis thaliana* (Dickinson et al., 2018), making them versatile for observations in the laboratory or the wild. In angiosperms, the dynamics between the pollen and its pollen tube formation right into the ovaries represents some of the most environmentally sensitive part of their sexual reproduction. Temperature is considered an important variable, currently studied under various environmental research because it is a "fragile" climatic variable, and temperature represents one of the main environmental variables which could affect the health of living organisms, especially in plants (Reynolds & Casterlin, 1980) (Nievola et al., 2017) (Qaderi et al., 2019).Pollen tube growth dynamics have been highlighted to potentially be controlled by $Ca2^{+}$ signalling and sensitive to heat stress, making it vulnerable to changes to the environment (Johnson et al., 2019). It has been suggested that $Ca2^{+}$ signalling acts as a signal transduction factor under heat stress conditions in the model *Arabidopsis thaliana* (Xu et al., 2022). This may affect pollen tube dynamics in *Linum*.

This part of the chapter will test several growing temperatures to establish whether pollen germination is affected by using quantification and analysis of pollen structures. One of the interests in pollen observation is whether different temperature conditions impact the performance of pollen tubes to reach the ovule and form a gamete. To do this, pollen tube would need to be able to travel down the style into the ovule which contains the female egg cells. Whether temperature presents a significant stress for pollen tubes to be able to reach the ovule is of important agricultural interest in crops including *Linum*. A further interest is if there would be a difference in species and population level across cultivated *Linum* (*L. usitatissimum*) and its wild relatives (*L. bienne*). Another interest in terms of pollen viability is whether different climate of origin variables is associated with pollen viability. For the purposes of the study of this chapter, the vulnerability to changes in temperature are hypothesized to be of key importance for seed set in *Linum*. Pollen viabilities were previously tested in *L. usitatissimum* and were found to be little affected by heat stress. However, findings suggest the formation of bolls and seeding were negatively affected by heat stress (CROSS et al., 2003). Temperature effectson pollen have not been tested in *L. bienne*. Other cultivated species showed a negative effect on pollen viability and germination post high temperature exposure (Aloni et al., 2001).

In this study, temperature as an environment variable was evaluated against pollen viability in the wild and cultivated *Linum* samples which were also used across other studies in this thesis. Observations were made of pollen viability at various stages. These were pollen count, pollen tubes count, and whether pollen tubes were able to reach the ovaries as a sign of pollen gamete being “successfully” transferred. Different temperatures were tested *in vivo* for pollen germination strategies comprising heat and cold treatments as well as a control typical temperature treatment. The main hypothesis assessed in this chapter was that pollen viability is affected by temperature as an environmental variable in wild and cultivated *Linum*. We also expect wild and cultivated flax to show variability within their pollen viabilities under different treatments. With local adaptation in mind, we expect differences in pollen responses to cold and heat treatments between populations originating in warmer and colder climates. This may be presented as correlations between pollen viability and local climate variables. More Northern plants are expected to perform better under colder temperatures in terms of pollen viability and vice versa for Southern plants.

## 3.2 STUDY MODELS AND METHODS

Under this study, 51 individuals representing 18 wild populations, and 14 cultivar varieties were observed (appendix 10). These individuals were harvested from the vernalization experiments and so constitutes for S1 (first selfing generation). Fully grown plants were able to be used for pollen observations in their subsequent flowerings. The sample plants were tested using different controlled environments inside a Weiss Gallenkamp controlled chamber: models A3655 (Arctic) and A3658 (Tropical). Each of the chambers were connected to a main computer and were equipped with numerous fluorescent tube lighting units, which were "Philips Master TL5 HE". These light tubes are 14 Watts in power requirements for each tube and emits 4000 Kelvin light temperature in colour by specifications. The artic chambers were used to provide colder temperatures while the tropical chambers were used to provide the warmer treatments. During the growing stages, plants were grown in a pot with F2S compost (Levington Advance Seed and Modular + Sand) combined with dried rice husks (3:1 ratio) for aeration purposes. The same 750ml black square plastic pots were used for each sampled individual. In each pot, five seeds from the same maternal individual were sown. Plants were grown under a controlled glasshouse condition (16:8 daylight ratio, 25$^{o}$C daytime temperature and 13$^{O}$C night-time temperature) until first flowering was observed.

For the different temperature treatments, individuals were randomized under controlled glass-house conditions. Plants were grown until first flowering at 25°C during daytime before treatments. Plants were then treated with either a heat (+5°C) or a cold (-5°C) treatment under the corresponding tropical or arctic chambers as the first flower were observed. Plants were treated for at least 72 hours before data were collected, to allow for acclimatation. Post acclimatation, flowers were collected at an "open stage". As described in Schewe *et al* 2011*,* flowers at an "open stage" are categorised as flower at anthesis, showing an opened 5-part whorl where the stamens are found with the stigma (stage 11 in Schewe *et al*, 2011) (Schewe, et al., 2011). Aniline blue staining protocol was utilized along with a confocal microscopy. Full protocol and information of the Aniline blue are available under section "3.2.3 Pollen Observation" .

Whole flowers were harvested per individual and preserved in a 70% ethanol (70%EtOH) solution, inside separate 1.5ml Eppendorf tubes. Floweres were preserved at 4$^{o}$C. Observations found 70%EtOH solutions to be adequate for floral and pollen preservation after 24-hour periods (appendix 11). Flowers were submerged in at least 1ml of 70% EtOH before storing in a 4$^{o}$C fridge. Before microscopy, petals and anthers were removed. This resulted in the boll and the stigma exposure. Some errors in pollen measures are expected as pollens may or may not be washed out during preservation stages, but the same procedures were applied to all samples under all treatments. Counts of pollen, pollen tube and pollen reaching ovary were recorded over several weeks of flowering, using a combination of pollen staining and confocal microscopy techniques (described below). Between the different temperature treatments, samples were randomly swapped over a 72 hour rotation period with varying number of flowers at "open stage" able to be harvested per-observation.

### 3.2.1 Pollen Viability *In Vitro*

Studies have previously shown several ways to test pollen viability. These often consist of staining by a fluorescent dye and assessing the presence of pollen and pollen tubes under a confocal microscope, as performed previously in *Panicum* (Ge et al., 2011). Optimization of pollen fluorescence microscopy is also of interest in the model *Arabidopsis* (Bou Daher et al., 2008) and thus this study involving the non-model *Linum* will explore staining methods for fluorescent microscopy.

As a precursor to quantification, tests were done using *in vitro* germination under heat stress which uses different germination media to *in vitro* germinate pollen from different temperature treatments. At the aim was to find out whether *in vitro* germination using germination medium was possible in *Linum* to determine heat stress (Rodriguez-Enriquez et al., 2012). In the model *Arabidopsis thaliana*, it is suggested that *in vitro* germination is possible, with a varied germination levels at different treatments, however pollen germination and tube growth were dependent on pollen density in both liquid and solid medium (Boavida & McCormick, 2007). Recent past studies have tested the effect of germination media with extra sucrose, and boric acid for *in vitro* pollen tube germination in the non-model Chinese fir (Fragallah et al., 2019). Then, *in vitro* methods were used to germinate different pollens under different environmental conditions. We assessed different solutions in *Linum in vitro* germination. Different germination solutions were made using Sucrose as the main sugar (Table 1).

| Components | Solution 1 | Solution 2 | Solution 3 |
|---|---|---|---|
| 12% (w/v) Sucrose | 50% | 40% | 60% |
| 5mM $CaCl_2$ | 20% | 20% | 20% |
| 0.01% (w/v) Boric Acid | 15% | 20% | 10% |
| 1mM $MgSO_4$ | 15% | 20% | 10% |

Table 1. A table of the three different *in vitro* germination solution which were tested in the laboratory.

The pollens were firstly place within the different solutions for at least 3 hours before observation. To assess germination success, a control was also provided with just nuclease free water as the solution. 20% Toluidine blue was used to dye the pollens after the treatments and pollens were observed under a Leica DM500 microscope with a camera model MC190HD connected to a computer running Leica's LAS software. Results for the *in vitro* germination were not very convincing (see the results part of this chapter). This result would mean that *in vitro* germination procedures in *Linum* pollen needed progress. Therefore, it was decided that pollen would need to germinate *in vivo* on flower pistils, with measures at different pollen maturity stages to measure heat stress under different temerpature conditions.

For *in vitro* germination attempts, three solutions were prepared according to Table 1 and was adjusted to pH of between seven and eight. Three to five drops were put onto a glass microscope slide with a Pasteur pipette. Wild *Linum* flowers from control 25 C conditions that were still not opened (to avoid *in vivo* germination) were collected by cutting from the stem and all petals were carefully removed using forceps to expose the stigmas, and anthers. These parts were then dipped onto the solution on the microscope slides several times to release pollen into the solution. At least 3 Hours have passed to allow germination before the staining procedure was undertaken. For the stain, 20% Toluidine blue were prepared within a mixture of 50% nuclease free water and 50% glacial acetic acid. Two drops of the dye solution were added to the slide and the pollen samples. Observation took place using a Leica DM500 microscope fitted with MC19HD camera, connected to a computer able to run Leica's LAS software. In this study 30×10 and 50×10 magnifications were used for the pollen *in vitro* germination observations.

### 3.2.2 Pollen Treatments

For the different temperature treatments, individuals were randomised under a controlled glass-house conditions before measurements under confocal microscopy. Plants were initially grown at 25°C during daytime. Plants were then treated with either a heat (+5°C) or a cold (-5°C) treatment under the corresponding tropical or arctic chambers as the first flower were observed. Plants were treated for at least 72 hours before further progress to allow for acclimatation. Post acclimatation, flowers were collected at an "open stage.

Aniline blue staining protocol was deployed along with a confocal microscope for visual aid. Full protocol and information of the Aniline blue are available under the "3.2.3 Pollen Observation" part found below in this chapter.

As described in *Schewe et al,* flower at an "open stage" are categorised as flower at anthesis, showing an opened 5-part whorl where the stamens are found with the stigma (stage 11) (Schewe, et al., 2011). Whole flowers were harvested and preserved with 70% EtOH inside separate 1.5ml Eppendorf tubes for each line. Two solutions were evaluated for pollen preservation. One solution is Formaldehyde-acetic acid alcohol (50ml of >95% ethyl alcohol, 2.5 ml of glacial acetic acid, 5.5 ml of formaldehyde, and 42 ml of $dH_2O$) (FAA), and the other solution was 70% ethanol with $dH_2O$

(70%EtOH). Observations found both solutions to be adequate for floral and pollen preservation after 24-hour periods (appendix 12). After collection, flowers were submerged in at least 1ml of 70% EtOH inside a 1.5ml Eppendorf tube for preservation. Petals were removed as well as anthers which results in the boll and the stigma exposure. Some errors in pollen measures are expected as pollens may or may not be washed out during preservation stages, but the same treatments were applied to all samples under all treatments. Measures of pollen and pollen tube counts have been accumulated over several weeks. Between the different temperature treatments, individuals were randomly swapped, and different flowers were also measured per treatment.

### 3.2.3 Pollen and Pollen Tube Observation

For the purposes of pollen and tubes observations, a modified protocol based on the protocol by Lu Yongxian (Lu Y, 2011) were used. The fluorescent staining dye applied into stigmas to observe pollen number and observe pollen tubes was 0.5% (w/v) Aniline blue in 0.1M $KH_2PO_4$. Glycerol was recommended in some fluorescent dye protocols, but it was not used in this study as fluorescence dye were observed better in non-glycerol solutions. Optimisation of aniline blue dilution was of interest, as too dilute will limit penetration into the pollen cells and too concentrated will act as a background noise, which may interfere with the images surveyed (Herburger & Holzinger, 2016). Preliminary optimisation found that 0.5% (w/v) Aniline blue in Potassium Phosphate ($KH_2PO_4$) provided images that were useful for pollen and pollen tubes observations (Appendix 13).

Bolls and stigma attached were examined onto a 25mm x 75mm glass microscope slide with a thickness of ≤1.2mm. 2-3 drops of 0.5% aniline blue were sufficient for dying purposes. In theory aniline blue can penetrate callose better than other cells (Herburger & Holzinger, 2016), and therefore more sensitive towards pollens and pollen tubes. This made it easier to distinguish the two pollen tissues from other cell types and to quantify them. Samples were covered using 22mm x 22mm cover slip with a thickness of 0.1mm to allow for optimal viewing under a confocal microscope. Samples were stored at room temperature for >3 hours prior to viewing to allow for the dye to penetrate the callous cells. The observations were also done in a dark room with as little light as possible.

A Zeiss model 880 Confocal Microscope was used to observe fluorescence and presence of pollen and pollen tube structures with fluorescence absorbance between 400-500nm for aniline blue absorption (Yang et al., 2007) and images were further processed by Zeiss' own software (ZEN Black edition) to optimise image settings. Zeiss' 10× magnification and Zeiss' 20× magnification eye pieces all combined with a 10× eyepiece was used to observe pollen and pollen tubes. Different focal layers were examined and in cases where pollens could not be observed under one layer of focus, a 3D image with appropriate number of Z-stacks was acquired to scan through all axes.

### 3.2.4 Pollen Counts

*Linum* pollen structures were circular vessel like grains and have high callose content and therefore would be highlighted more intensively by the aniline blue dye than surrounding cells. We observed no differences in structure for the wild and cultivar species of *Linum* observed for this study. The pollen was expected to be found adhered around the stigma. Pollen tube structure were identified as thread like structures coming out of the pollen tube and penetrating the stigma, often in a straight line down to the ovary (Figure 2, next page).

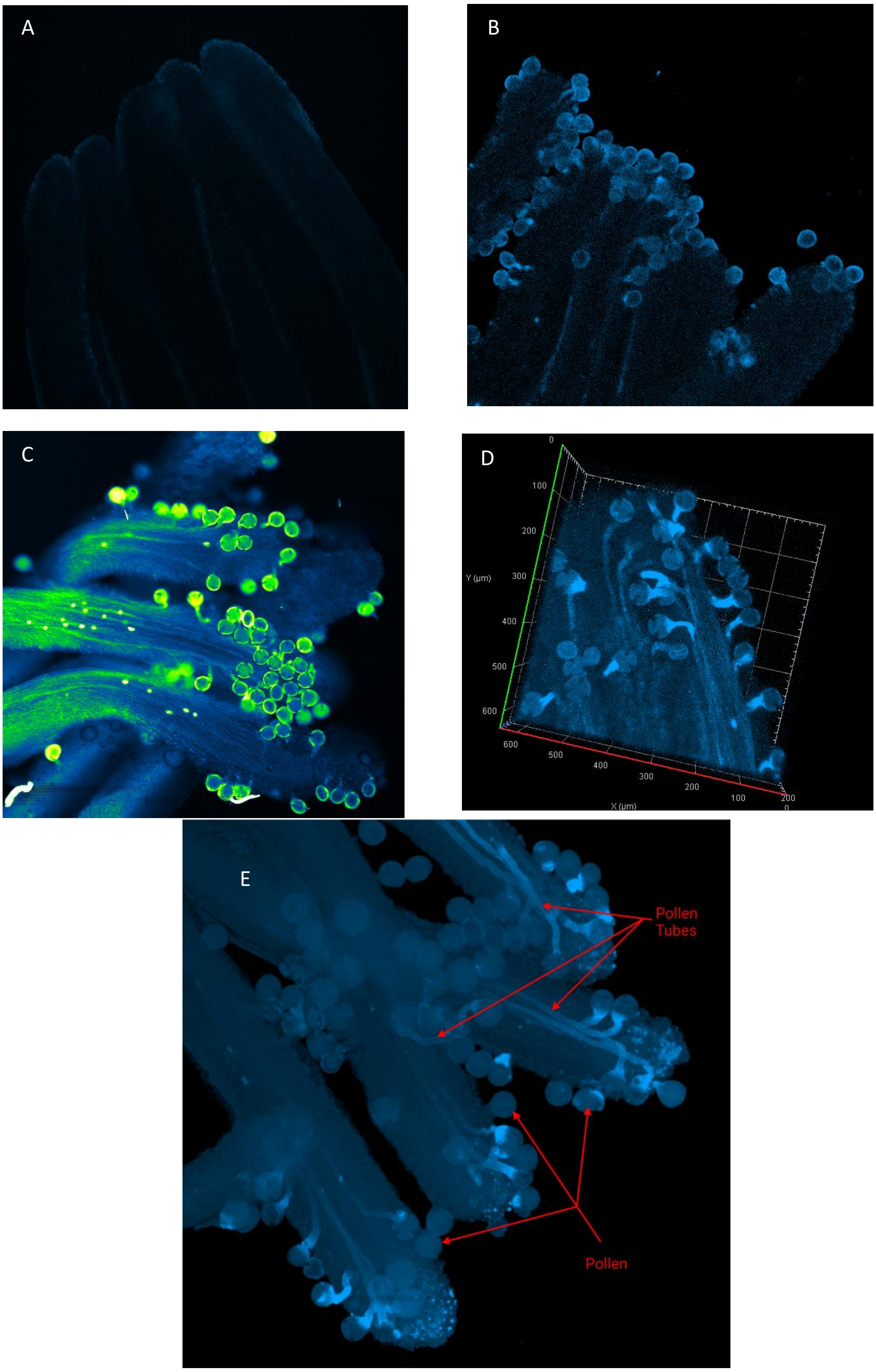


Figure 2. A. 10×20 magnified stigma, fluoresced with Aniline blue illustrating an absence of pollen, B. the presence of mature pollens on the stigmas, conveyed as blue circular shapes C. more fluorescent callose makes pollen more visible in different colour ranges. D. A 10×20 magnified 3D depiction of stigma and pollens. E. A 10×20 magnified, and labelled image of pollen structures observed.

Absence of pollen and pollen tube structures can be observed in some cases under fluorescent microscopy (Figure 2A). In other cases, prescence of pollen and pollen tubes at the stigma were also be observed (Figure 2B-E). Pollen were rounded in shape and fluoresce brighter than the stigma cells. The pollen tubes were string-like structures growing out of pollen into the stigma (Figure 2E).

Images were further processed in Zeiss' Zen Black software to apply colouration to highlight fluorescence absorption at 400-500nm wavelength and increase contrasts between the pollen and pollen tube cells and surrounding stigma/style cells. Callose in pollen and pollen tubes can be seen as more fluorescent (Figure 2C). A 3-Dimensional observation can also be made using the Z-stacks function of Zeiss' Zen Black software (Figure 2D). This feature enabled layers to be taken stack by stack under a new Z dimension, thus building a 3-dimensional image. This tool represents an aid, as pollen and pollen tubes may have not been fully observed on a 2-dimensional image. After optimisation for each slide, sample images were captured, and pollen and pollen tubes were manually counted to build a dataset.

### 3.2.5 Pollen Tube Counts

Observation was made for three different temperature treatments to observe whether pollen tubes were able to reach the ovule at the open flower stage. Aniline blue staining and confocal microscopy methods were used as before, but for this observation, we observed pollen tubes inside pistils and stigma cells. Colouration of the aniline blue were referred to "cyan" and contrast were increased with using Zeiss' Zen Black software. For stigma and styles with little pollen tube growth, wavelength was contrasted from fluorescent green to blue, highlighting dyed cells fluorescing green. The processed images could then be observed to see whether pollen tubes were able to reach down to the bottom of the style and reach the ovule. This was observed for each of the temperature treatments and recorded in binary, 0 representing pollen tube not reaching ovary and 1 representing pollen tube reaching ovary for every individual tested. The ovary as well as the pollen tube fluoresce when dyed with the Aniline blue. An observation of a pollen tube structure going into the ovary at the bottom of the pistils would suffice the question of whether the pollen tube have managed to reach the ovary or not. The resulting difference in images with pollen tubes and without pollen tubes could be observed on the Figure below:

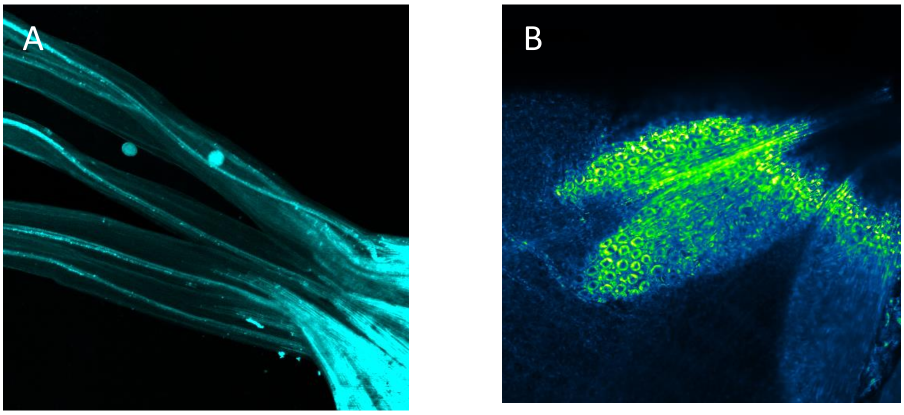


Figure 3. A. 10×20 magnification of a flower pistil with observed pollen tube growth fluorescence highlighted in cyan. B. 10×20 magnification of fluoresced ovary cells highlighted in “fluorescent green”, depicting no pollen tubes going into the ovary.

Figure 3 compares 10×20 microscopic images between individuals with pollen tubes growing through the pistils and ovaries, and individuals with no observable pollen tube in the ovaries. Pollen tubes growth were highlighted in cyan when dyed with aniline blue (Figure 3A). Long thread like structure could be seen going through the pistils into the ovary parts of the flower. Both pollen tube cells and ovary cells were highlighted and dyed in aniline blue as observed here. Absence of ovary cells can be observed just after the end of the style, which were highlighted in fluorescent green (Figure 3B). This reveals that for this individual, no pollen tube was observed going into the ovary structure and therefore pollen tube did not successfully reach the ovary. As an additional measure, this proportion of pollen forming pollen tubes was also calculated for further analysis.

## 3.3 ANALYSIS

ANOVA (Girden ER, 1992), linear modelling (Yan and Su, 2009) and tests for binomial models for this study were undertaken using the statistical programme R (R Core Team, 2022). The graphical interface used to run R was RStudio (Rstudio Team, 2020). Raw data for pollen counts, pollen tube counts as well as whether pollen tubes were able to reach the ovary for each of the treatment were firstly processed and organised in Microsoft Excel and converted to a tab-delimited CSV format for R to read. Raw data exist in both binomial and count format. Each of the linear modelling stages need to take account the data format in such cases.

### 3.3.1 Analysis: Pollen and Pollen Tube Counts

Analysis was performed to determine whether pollen and pollen tube counts for treatments differ significantly. First, the data was checked for normal distribution with Shapiro-Wilk's normality tests (SHAPIRO & WILK, 1965). This test uses a modified analysis of variance, to calculate a *p-value*. A histogram of the data counts was also observed to show whether the data is normally distributed.

General linear modelling (GLM) was utilized to test the effect of changing temperature on the number of pollen counts (McNeil *et al,* 1996). For non-fractional count data, Poisson's regression was used (Hayat & Higgins, 2014) with "family=poisson()" in the GLM command under R. A Poisson Regression model is a Generalized Linear Model (GLM) that is used to model count data (Hayat & Higgins, 2014). The output Y (count) is a value that follows the Poisson distribution. The model assumes logarithm of expected values (mean) that can be modeled into a linear form by some unknown parameters. For a post-hoc analysis comparing pairs of treatments, a Tukey's adjustment was added to the general linear model by using the R package "Multcomp" (Hothorn et al., 2008). The full command is available under appendix 14. Data were summarized using a bar chart for the pollen count against different treatments with the "barplot" command in R. For interactions between treatment and species, an R function "interaction()" can be used to interpret interactions between two categorical variables (Chambers and Hastie, 1992) to be integrated into the linear modelling.

### 3.3.2 Analysis: Proportion of Pollen Tubes

The number of pollen tubes relative to the total number of pollen counts were tested for interaction against the different treatments and between the two wild and cultivated species within the *Linum* samples. Data was modelled using general linear model with a quasibinomial function using the R package "lmerTest" (Kuznetsova et al., 2017). A 'logit' link was added using the command "(link = 'logit')" as an addition to the quasibinomial family specification which attempts to describe additional variance in the data that cannot be explained by a Binomial distribution alone. After the linear modelling, analysis of variance (ANOVA) for the different variables (Girden ER, 1992) was applied to the model in R using the command "anova(model,test=F)". Interaction between treatments and species was measured using the "interaction()" command of R. The Tukey's contrast treatments were added to the model using the R package multcomp (Hothorn et al., 2008), using the command: "model<-glht('modelprefix',mcp(interact="Tukey"))". The full model commands can be found under appendix 14.

To evaluate whether pollen tube was able to reach the ovary, a binomial "yes or no" test was utilized. To test for this, a general linear model with a binomial adjustment was run in R with the function "*family=binomial(link='logit')*" added to the general linear model command (Gelman, A. and Hill, J., 2006).

## 3.3.3 Analysis: Population of Origin Variables

To consider environmental effect, we gathered geographical data for the wild *L. bienne*. Climatic variables from the origin of the wild individuals were gathered from WorldClim database in *Chapter 2* of this thesis. Latitudinal and climatic data was only reliably available for the wild *L. bienne* samples. This is because location data for our cultivars was not precise. Therefore, for the analysis against local variables, only wild *L. bienne* individuals were considered. Climatic variables were processed into summary principle components (PCs) as was described previously in chapter 2 (see section 2.2.1 "Samples and Experiment"). Latitude and climatic variables were used to further describe temperature variables of the local areas which the wild *L. bienne* were collected from. More Northern origin plants are expected to perform better under colder temperatures in terms of pollen viability and vice versa for Southern origin plants. The loadings of the principal component analysis are as in appendix 3. The temperature loadings of climate PCs will be of particular interest in this chapter. Since the temperature loadings were positive, a higher PC1 value indicated a higher temperature. For example, the PC1 value for the Southern population '3' is *5.367* while the PC1 value for the Northern population 'Tym' is *-6.616*. The average temperature loading for population '3' collected from the WorldClim database is at *12.07$^o$C* in winter (December, January, February) while for population 'Tym' it is *5.02$^o$C*. For the summer (June, July August) the average temperature for population '3' is *22.6$^o$C* and for population 'Tym' it is *15.05$^o$C*. A linear model with a Pearson's correlation coefficient can be used to test whether latitude of locality and climatic variables affects either pollen count, pollen tube counts and or the proportion of pollen able to reach the ovary. A scatterplot of relationships was generated using the “ggscatter” command of the package ggplot2 built for R (Villanueva & Chen, 2019).

---

# 3.4 RESULTS

## 3.4.1 Results: *In vitro* pollen germination

The processed microscopic images reveal that in all cases of the *in vitro* solutions 1-3, the pollen structure burst. We found that, the *in vitro* treatments of the *Linum* pollens were not up to standard and resulted in pollen materials bursting. Comparison between pollen treated in nuclease free water and the germination solution 1 can be seen (Figure 4A-C). A 500x magnification of one of the sites of the aperture pore whereby bursting occured were also observed (Figure 4C). This bursting was observed using all germination solutions.

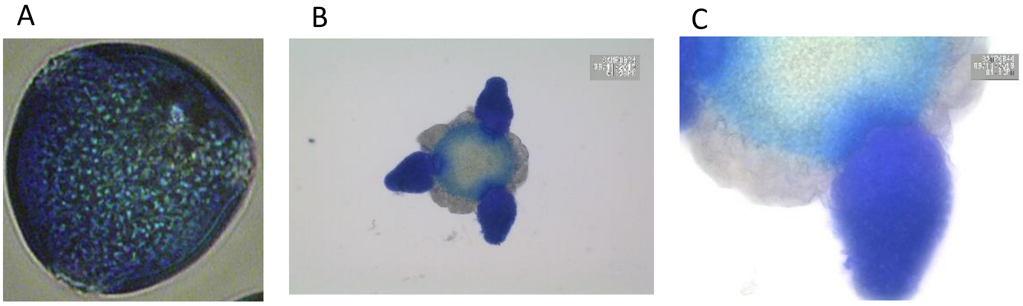


Figure 4. A. 10×50 cropped magnification of a mature pollen dyes with 20% Touludine blue, observed in nuclease free water and no added germination solutions. B. A 10×20 magnified image of a 20% Touludine blue dyed pollen cell which is undergoing bursting from all the aperture pore sites due to the germination solutions. C. 10×50 magnification of one of the aperture pore sites of a recently burst pollen.

## 3.4.2 Results: Test for Normality

As part of the analysis against raw pollen/pollen tube counts, tests were done to reveal whether the data is normally or abnormally distributed, to inform the statistical analysis which could be used against the data. For normality, a histogram of the raw data distribution should peak around the average with decreasing number of raw counts either side of the data to form what is widely known as a bell curve when a line of best fit is applied to the histogram. A histogram of the raw data was able to be illustrated with a non-normal distribution (Pollen count, % of pollen tubes) (Figure 5).

A Shapiro-Wilks test was applied to the raw data to statistically quantify normality revealed that the data is not normally distributed and therefore adjustments were made to statistical analysis tests to take this non-normality into account by using the appropriate Poisson or quasibinomial distribution family (pollen counts: *W = 0.89365, p-value = 4.099e-14* and % pollen tubes *W = 0.77881, p-value < 2.2e-16*).

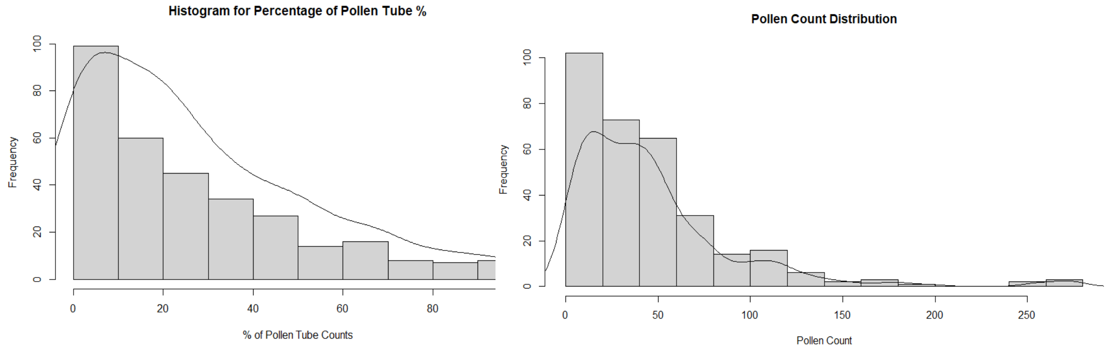


Figure 5. Histograms and density curves for the data distribution of both pollen tube % and pollen count showing a non-normal distribution.

### 3.4.3 Results: Pollen and Pollen Tubes Count

As a matter of ease, the treatments were all respectively known as 20°C (Cold), 25°C (Glasshouse), and 30°C (Heat) from here on. Summary of all the data were illustrated in a barchart (Figure 6). Both treatments (heat and cold), revealed a significant reduction on the number of pollen count (Figure 6 and Table 2). Only the heat treatments suggested significant difference in terms of tube count (Figure 6 and Table 3).

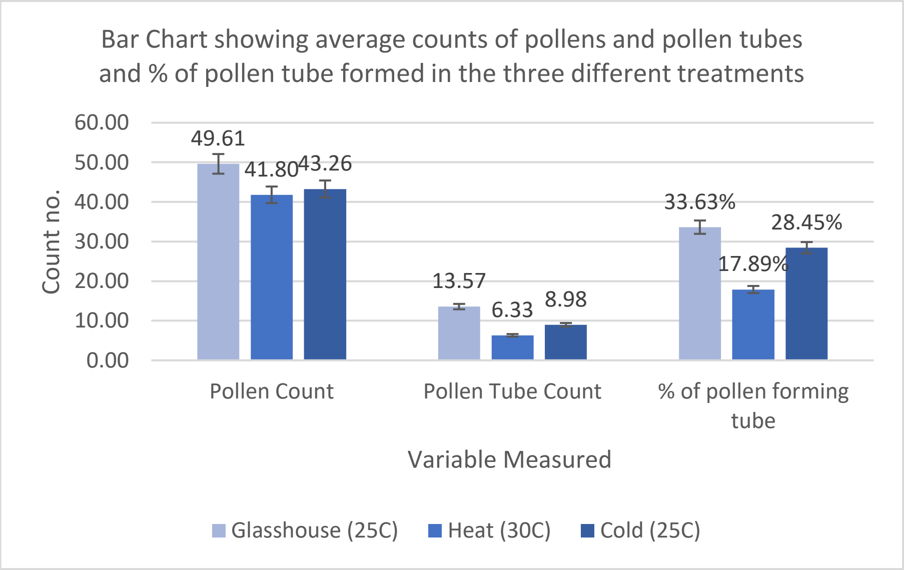


Figure 6. A bar chart of the overall average pollen data for the different heat treatments. Glasshouse treatments are highlighted in blue, heated treatment in orange and cold treatment in grey.

Pollen count of each treatment were statistically compared to check for effects on the number of pollens due to temperatures according to the hypothesis that the temperature affects the number of viable pollens within both wild and cultivated *Linum*. There was variation in the average pollen count when comparing the three treatments (Figure 6). To statistically test for this variation, we ran an ANOVA with a Tukey's post-hoc for multiple comparisons and we illustrated a model using a Poisson's regression model formula of "Pollen$Count ~ Pollen$Treatment" for the data comprising all the samples within this study. The analysies suggested that there was significant decrease between the pollen counts in the glasshouse (25°C) compared with both the cold (20°C) and heat (30°C) treatments (Table 2 below). This sugests that the treatments have a significant effect on the number of pollens within the samples tested in this study.

*Linear Hypotheses:*

| *Estimate* | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| ***25C vs 20C == 0*** | **0.104605** | **0.01972** | **5.305** | **<1e-05** | ***** |
| *30C vs 20C == 0* | 0.005315 | 0.021423 | 0.248 | 0.967 | |
| ***30C vs 25C == 0*** | **-0.09929** | **0.019749** | **-5.028** | **<1e-05** | ***** |

Table 2. A linear model with Tukey's correction for pollen count summary between the different temperature treatments. Significance (*p=<0.05*) was showing for a linear hypothesis between the 25°C – 20°C and the 30°C – 25°C treatments.

To see whether the same could be seen in the number of pollen tubes, average pollen tube count was also tested for each treatment (Figure 6). When statistically compared, the number of pollen tube on average between the treatments decreased with both cold and heat treatments. Just as were seen in the pollen count dataset. In addition, the number of pollen tubes seem to be significantly reduced (*p=<0.05*) for the heat (30°C) and cold (20°C) treatments (see Table 3).

*Linear Hypotheses:*

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| ***25C vs 20C == 0*** | **0.3805** | **0.04108** | **9.263** | **<1e-08** | ***** |
| ***30C vs 20C == 0*** | **-0.31057** | **0.05121** | **-6.065** | **<1e-08** | ***** |
| ***30C vs 25C == 0*** | **-0.69108** | **0.04588** | **-15.064** | **<1e-08** | ***** |

Table 3. A linear hypothesis with a Tukey's adjustment for the number of pollen tubes observed between the different treatments. The test shows significant differences between all the treatments (*p=<0.05*).

We observed variation in pollen tube count and pollen count when looking at different temperature treatments. However, in relation to pollen count we wanted to see if there is variation in the proportion of pollen tubes able to be formed under our temperature treatments. To observe whether the proportion of pollen tubes to pollen count is substantially affected by the temperature treatments, the average proportions of pollen tubes formed against pollen for each temperature treatments were calculated and modelled. The results suggests that against the glasshouse (25°C) treatments, the cold (20°C) treatments showed a decrease in the proportion of pollen forming pollen tubes but that the difference was not significant (Table 4). Just as the cold (20°C) treatments, the heat (30°C) treatments showed a decrease in the proportion of pollen tube forming against the pollen count (Table 4). This decrease in proportion, however, was shown to be significantly different to the glasshouse (25°C) treatments when modelled against each other (p = <0.001) (Table 4). This suggest that warmer treatments could potentially affect proportion of pollen tube able to be formed in relation to the pollen count. Below is a linear hypothesis summary (with Tukey's modification) for the proportion of pollen tube formed against the number of pollens for each treatment.

*Linear Hypotheses:*

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| *25C vs 20C == 0* | 0.2759 | 0.1184 | 2.331 | 0.0511 | |
| *30C vs 20C == 0* | -0.3159 | 0.1442 | -2.19 | 0.072 | |
| ***30C vs 25C == 0*** | **-0.5918** | **0.1298** | **-4.561** | **<0.001** | ***** |

Table 4. Linear hypothesis with a Tukey's adjustment for the proportion of pollen tube forming, showing significance between 30°C and 25°C treatments (*p=<0.05*).

To summarize the linear hypothesis significance for section 3.4.3 "Results: Pollen and Pollen Tubes Count", we illustrated the following Table:

| Treatment | Pollen Count | Pollen Tube Counts | Proportion of Pollen tubes formed |
|---|---|---|---|
| **25C vs 20C** | SIGNIFICANT | SIGNIFICANT | NOT SIGNIFICANT |
| **30C vs 20C** | NOT SIGNIFICANT | SIGNIFICANT | NOT SIGNIFICANT |
| **30C vs 25C** | SIGNIFICANT | SIGNIFICANT | SIGNIFICANT |

Table 5. Summary of linear hypothesis tests significance between different temperature treatments and counts as per section 3.4.3 "Results: Pollen and Pollen Tubes Count".

### 3.4.4 Results: Pollen Tube Reaching the Ovary

To test a model of whether any of the three treatments were significantly different in the number of pollen tubes able to reach the ovary, a binomial general linear model (GLM) was performed for “yes or no”, binomial data under the different treatments. The model was summarised using a Tukey’s adjustment for abnormally distributed data (see section “3.3.2 Analysis: Proportion of Pollen Tubes”). The results reveal that the percentage of pollen reaching the ovary is significantly different in the 30°C treatments (see Figure 7 and table 6).

The results from the general linear model suggests that between the glasshouse (25°C) and the cold (20°C) the difference in the proportion of pollen reaching the ovary was not significantly different (*p = 0.85)*. This being the case, significant difference in the number of pollen tube reaching ovaries could be observed between the heated (30°C) treatment and both the glasshouse (25°C) and cold (20°C) treatments (*p=<0.001 for both*) (see table 6).

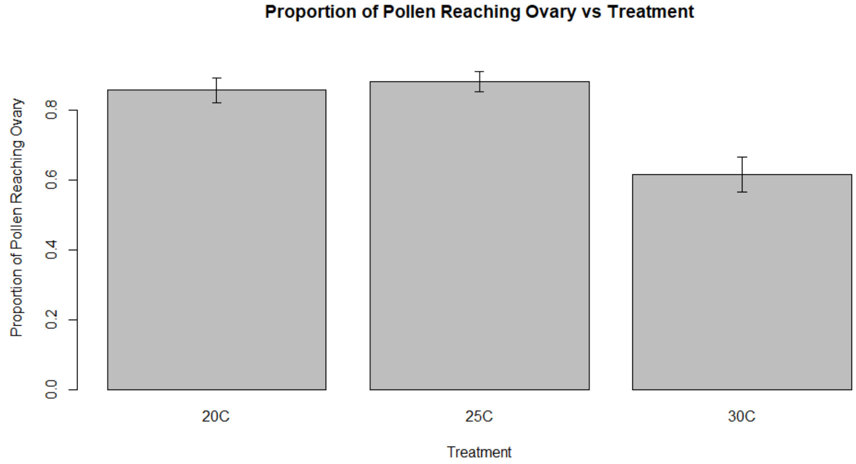


Figure 7. Bar chart showing the proportion of pollen reaching ovary according to the different temperature treatments.

*Linear Hypotheses:*

| *Treatments* | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| *25C vs 20C* | 0.2126 | 0.399 | 0.533 | 0.85457 | |
| ***30C vs 20C*** | **-1.3132** | **0.357** | **-3.678** | **0.00067** | ***** |
| ***30C vs 25C*** | **-1.5258** | **0.346** | **-4.41** | **< 1e-04** | ***** |

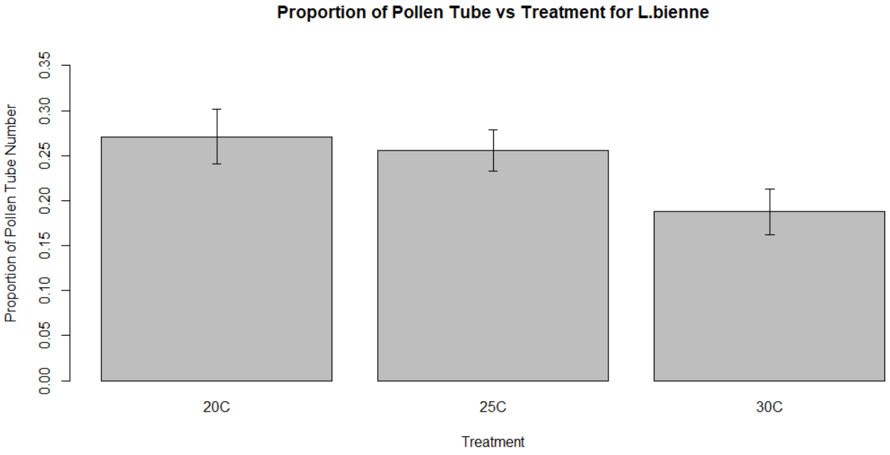
Table 6. General linear model with Tukey’s adjustment for the proportion of pollen tubes reaching ovaries. The test shows significant differences between the temperature treatments of 30°C – 20°C and 30°C – 25°C (*p=<0.05*).

### 3.4.5 Results: Pollen Tube Proportion Over Pollen Count for Different *Linum* Species

Additionally, we tested pollen tube proportion and pollen count to observe variation between species. Initial ANOVA revealed that when taking into consideration the interaction between treatment and species, the pollen tube proportion is different (*p = 0.0007* (*p = <0.05*)). The proportion of pollen tube were linearly modelled with logistic distribution against the interaction between treatment and species. Data for either wild and cultivated species were separated and analysed with a linear model hypothesis separately. Bar plots illustrates a reduction in the proportion of pollen forming pollen tubes when heat treated for both *L. bienne* and *L. usitatissimum* (Figure 8 A and 8B). The proportion of pollen tube formed was found to be highest in the glasshouse (25°C) treated for the cultivars *L. usitatissimum*. This was reduced when treated under the cold (20°C) treatment, but not as much as when heat (30°C) treated. This wasn't the case when looking at *L. bienne*. When cold (20°C) treated the proportion of pollen tube formed in *L. bienne* did not decrease.

We inferred a linear model summary when considering each species for treatments against each other. This was done with a Tukey's post-hoc correction (tables 6A and 6B ). The proportion of pollen tubes saw the largest difference when looking at the heat (30°C) treated *L. usitatissimum* against the glasshouse (25°C) treated *L. usitatissimum* (*p = <0.001*). This was followed by the glasshouse (25°C) treated *L. usitatissimum* and the cold (20°C) treated *L. usitatissimum* (*p = >0.001*). *L. bienne* showed a statistically non-significant difference between the glasshouse and the cold/heat treatments when modelled (25°C – 20°C Treatments *p=0.4556*, 25°C – 30°C Treatments *p=<0.4244*). This illustrates that the proportion of pollen able to form pollen tube is significantly affected by changes in temperature in the cultivar *L. usitatissimum* than in their wild relative.

**A – *L. bienne***



**B – *L. usitatissimum***

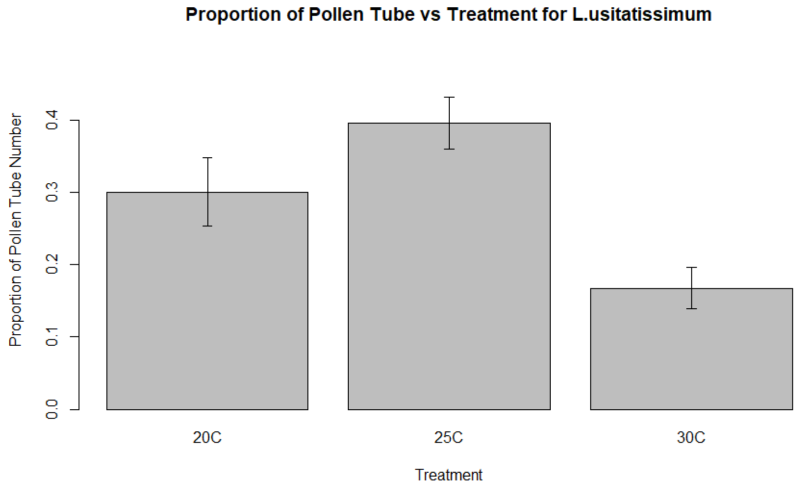


Figure 8. Bar chart to represent the proportion of pollen tube forming against temperature treatment for the wild relatives *L. bienne* (8A) and the cultivar *L. usitatissimum* (8B).

**A– *L. bienne* Summary**

*Linear Hypotheses:*

| *Treatments* | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| *25C vs 20C* | -0.158 | 0.1323 | -1.194 | 0.4556 |
| *30C vs 20C* | -0.3307 | 0.1517 | -2.179 | 0.0743 |
| *30C vs 25C* | -0.1727 | 0.1385 | -1.247 | 0.4244 |

**B – *L. usitatissimum* Summary**

*Linear Hypotheses:*

| *Treatments* | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| ***25C vs 20C*** | **0.6084** | **0.1892** | **3.217** | **0.00372** | * |
| *30C vs 20C* | -0.346 | 0.2434 | -1.422 | 0.32649 | |
| ***30C vs 25C*** | **-0.9544** | **0.2178** | **-4.382** | **< 0.001** | * |

Table 7 A and B. Summary tables of the linear model with Tukey's adjustments for the proportion of pollen tube formation between each *L bienne* (A) and *L. usitatissimum* (B). No significant differences were observed between the different treatments for *L. bienne*. There were significant differences between the 25ºC – 20ºC and the 30ºC – 25ºC treatments for the *L. usitatissimum* (*p=<0.05*).

To summarize the linear hypothesis for section 3.4.4 Results: “Pollen Tube Reaching the Ovary” and section 3.4.5 “Results: Pollen Tube Proportion Over Pollen Count for Different *Linum* Species”, a Table is illustrated in below (Table 8):

| **Treatments** | **proportion of pollen reaching ovary (*L. bienne*)** | **proportion of pollen tube forming (*L. bienne*)** | **proportion of pollen tube forming (*L. usitatissimum*)** |
|---|---|---|---|
| **25C vs 20C** | NOT SIGNIFICANT | NOT SIGNIFICANT | SIGNIFICANT |
| **30C vs 20C** | SIGNIFICANT | NOT SIGNIFICANT | NOT SIGNIFICANT |
| **30C vs 25C** | SIGNIFICANT | NOT SIGNIFICANT | SIGNIFICANT |

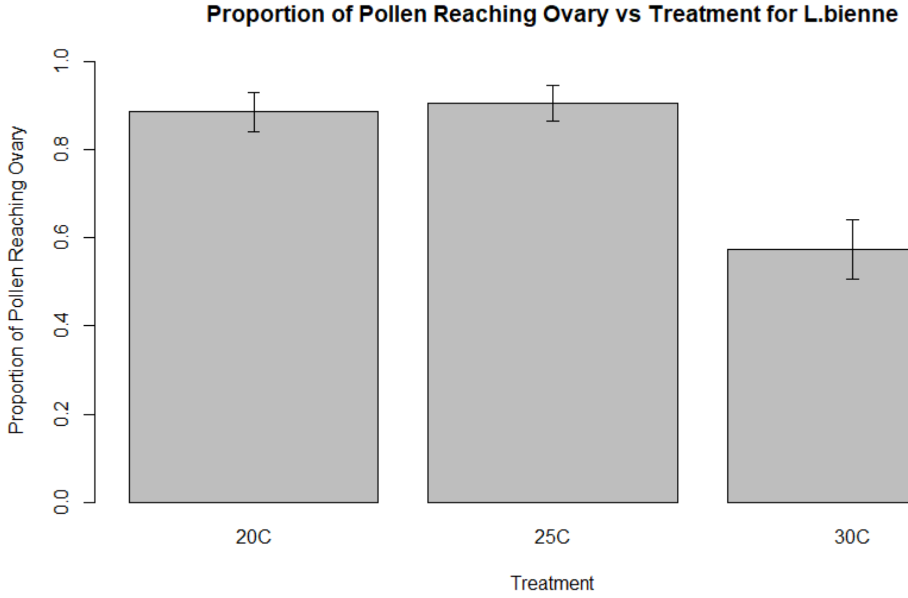
Table 8. Summary of linear hypothesis tests significance between different temperature treatments, ability of pollen reaching ovary and proportion of pollen tube forming for each species.

## 3.4.6 Results: Propotion of Pollen Tubes Reaching Ovary for Different *Linum* Species

We evaluate the ability of pollen tube reaching the ovary for the different species. Evaluation of whether pollen tube was able to reach the ovary was applied as described under section “3.3.2 Analysis: Proportion of Pollen Tubes” of this chapter. An initial Anova, revealed that the interaction between species and the treatments together were not significant (Treatment/Species $p = 0.2684$). The bar plot below (see Figures 9A) reveals a reduction in the proportion of *L. bienne* pollen able to reach the ovaries, under the heat (30°C) treatment.

Summaries of the general linear models reveals that for both wild and cultivated species, the proportion of individuals with pollen tube able to reach the ovary is significantly affected by the heat treatments (30°C) (Tables 9A (*L. bienne*) and 9B (*L. usitatissimum*)). This was significant when compared against the cold (20°C) treatment (for *L. bienne* $p=<0.00179$, for *L usitatissimum* $p=0.00195$), and the glasshouse (25°C) treatments (for both species $p=<0.001$ (<α of 0.05)). The result also shows some reduction on the proportion of individuals with pollen tube able to reach the ovary when observing the glasshouse and the cold treatments. However, this difference isn’t statistically significant ($p=>0.05$ in both cases).

**A – *L. bienne***



**B – *L. usitatissimum***

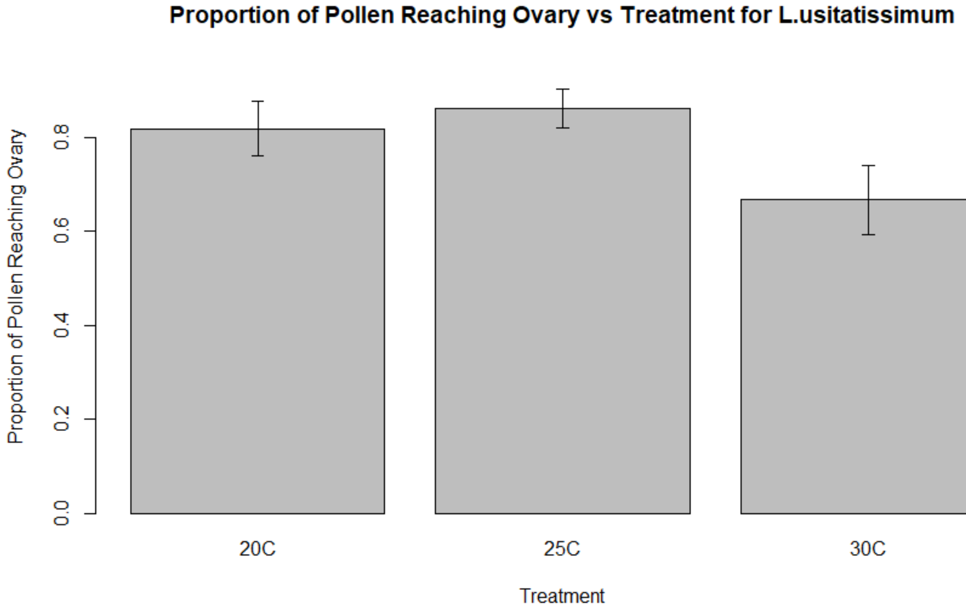


Figure 9 . Bar chart for the proportion of pollen tube reaching the ovaries between different treatment for *L. bienne* (A) and *L. usitatissimum* (B).

**A – *L. bienne* Summary**

*Linear Hypotheses:*

| *Treatments* | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| *25C vs 20C* | 0.2034 | 0.6393 | 0.318 | 0.94515 | |
| ***30C vs 20C*** | **-1.7599** | **0.5135** | **-3.427** | **0.00179** | ***** |
| ***30C vs 25C*** | **-1.9633** | **0.5445** | **-3.605** | **< 0.001** | ***** |

**B – *L. usitatissimum* Summary**

*Linear Hypotheses:*

| *Treatments* | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| *25C vs 20C* | 0.2034 | 0.6393 | 0.318 | 0.94515 | |
| ***30C vs 20C*** | **-1.7599** | **0.5135** | **-3.427** | **0.00195** | ***** |
| ***30C vs 25C*** | **-1.9633** | **0.5445** | **-3.605** | **< 0.001** | ***** |

Table 9A and 9B. Summary tables of the general linear model with Tukey's adjustment for the proportion of pollen tubes reaching the ovary between the different treatments for the different species. In both species the test shows a significantly reduced proportion of pollen reaching the ovary ($p=<0.05$).

To summarize our results for section 3.4.6 "Results: Propotion of Pollen Tubes Reaching Ovary for Different Linum Species", a Table is illustrated below (Table 10):

| **Treatments** | **proportion of pollen tube reaching the ovaries (*L. bienne*)** | **proportion of pollen tube reaching the ovaries (*L. usitatissimum*)** |
|---|---|---|
| **25C vs 20C** | NOT SIGNIFICANT | NOT SIGNIFICANT |
| **30C vs 20C** | SIGNIFICANT | SIGNIFICANT |
| **30C vs 25C** | SIGNIFICANT | SIGNIFICANT |

Table 10. Summary of linear hypothesis tests significance between different temperature treatments and proportion of pollen tube reaching the ovary for the two species.

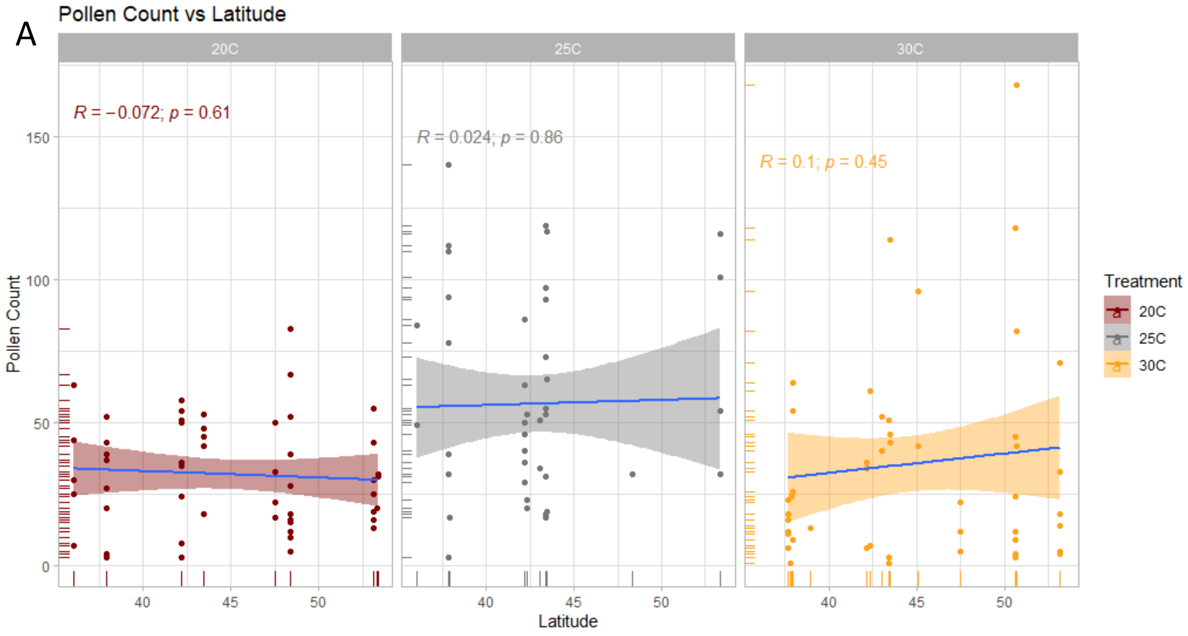
### 3.4.7 Results: Pollen and Latitude of Origin

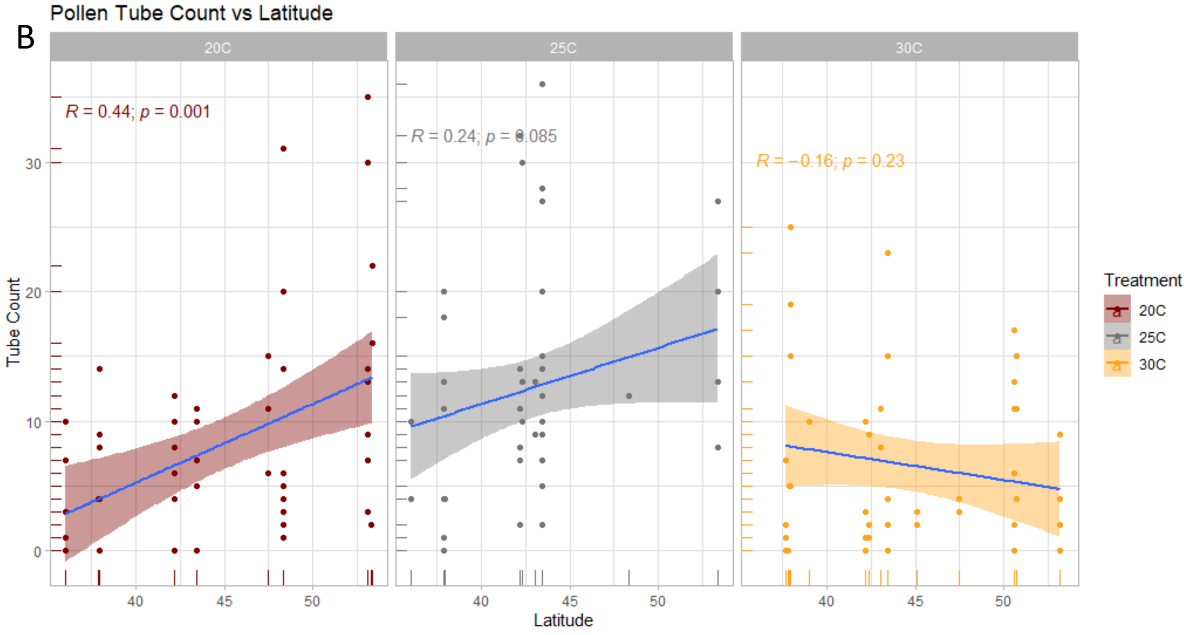
For the wild *L. bienne*, there was interest in looking at whether locality of origin of the wild plant had any roles to play on either the number of pollen count, the number of pollen tube observed and or the proportion of pollen tubes reaching the ovary. This is of interest, as this would add to the argument that local adaptation is present within our samples. Using Person's correlation and a linear modelling (represented by the blue line in Figure 10A), pollen count had no correlation with the latitude of locality of the different wild *L. bienne* samples within the study.

The number of pollen count in the glasshouse and heat-treated plants seems to increase, the more northern an individual is localized (For glasshouse (*R=0.024, p0.86*), for heat (*R=0.1 p=0.45*)), but none of the correlations were significant (*all p values = >0.05*).

Both glasshouse and cold treated individuals showed an increase in the number of observed pollen tubes, the more northern they are localized, but only the cold treatment was significantly correlated (*R=0.24, p=0.085* (glasshouse), *R=0.44 p=0.001* (*p=<0.05*)) (Figure 10B). This illustrates that, for our *L. bienne* individuals, northern individuals are more likely to form more pollen tube under cooler temperatures. When observing the heat treatment, it can be observed that correlation is opposite to that of the glasshouse and cold treatments (*R=-0.16 p=0.23*), all be it not significantly correlated. This change in correlation trend, shows some signs of the Northern individuals in this sample test, being more sensitive in their pollen tube formation to increases in temperature. This illustrates that the more Northern an individual within our *L. bienne* sample is localized, the more likely that they would favour cooler temperature for pollen tube formation.

In Figure 10C, we plotted the proportion of pollen tube observed over the pollen count (in percentage) and against latitude. The trend illustrates the positive correlation in the cold treatment, of pollen tube amount against latitude is even more significant when considering the proportion of pollen tube able to form against the pollen count per individuals (*R=0.58, p=<0.0001*). This strengthens the point that the more Northern an individual is, the more able they are to form pollen tubes under the cold treatments.


A
Pollen Count vs Latitude
20C
25C
30C
R = −0.072; p = 0.61
R = 0.024; p = 0.86
R = 0.1; p = 0.45
Pollen Count
0
50
100
150
40
45
50
Latitude
Treatment
20C
25C
30C


B
Pollen Tube Count vs Latitude
20C
25C
30C
R = 0.44; p = 0.001
R = 0.24; p = 0.085
R = −0.16; p = 0.23
Tube Count
0
10
20
30
40
45
50
Latitude
Treatment
20C
25C
30C

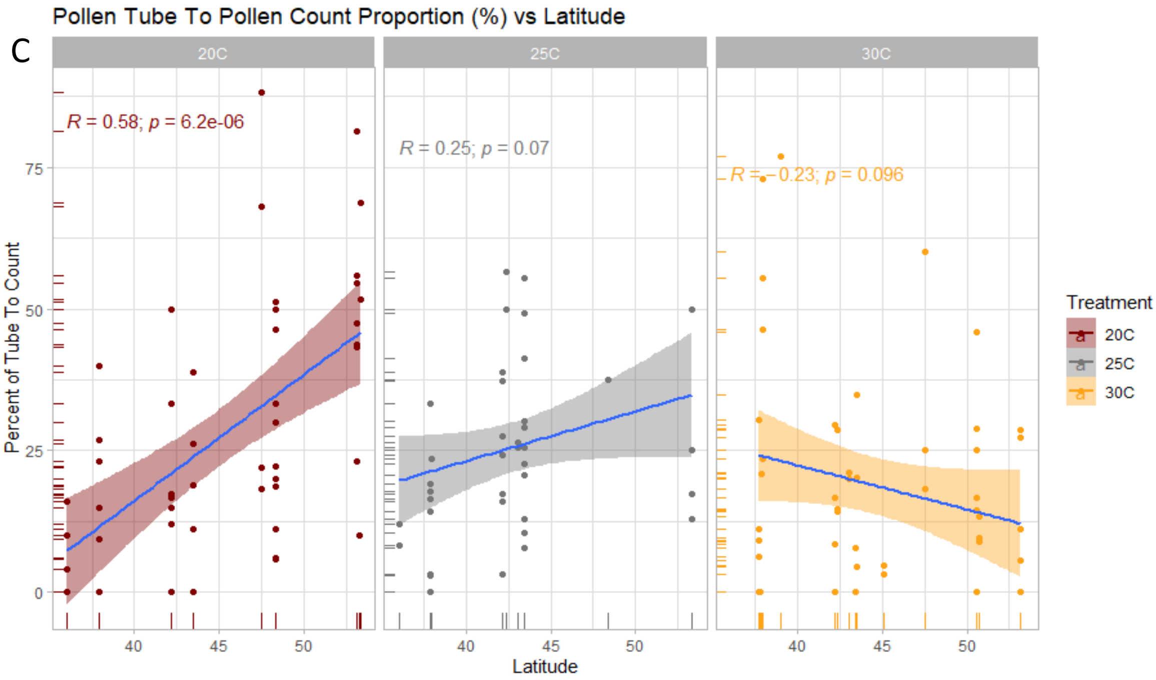


Figure 10. Pearson's correlations between pollen counts (10A) and pollen tube counts (10B) and pollen tube to pollen count proportion (10C) in relation to the latitude of localities. The only significant correlation seen was between pollen tube count and latitude of localities as seen in Figure 10B. Pollen tube counts under cold treatment (*R=0.44, p=0.001*). Fig 10C. Represents the proportion of pollen tube (%) to pollen count against latitude of individuals representing correlation under cold treatment (*R=0.58, p=0.000006*).

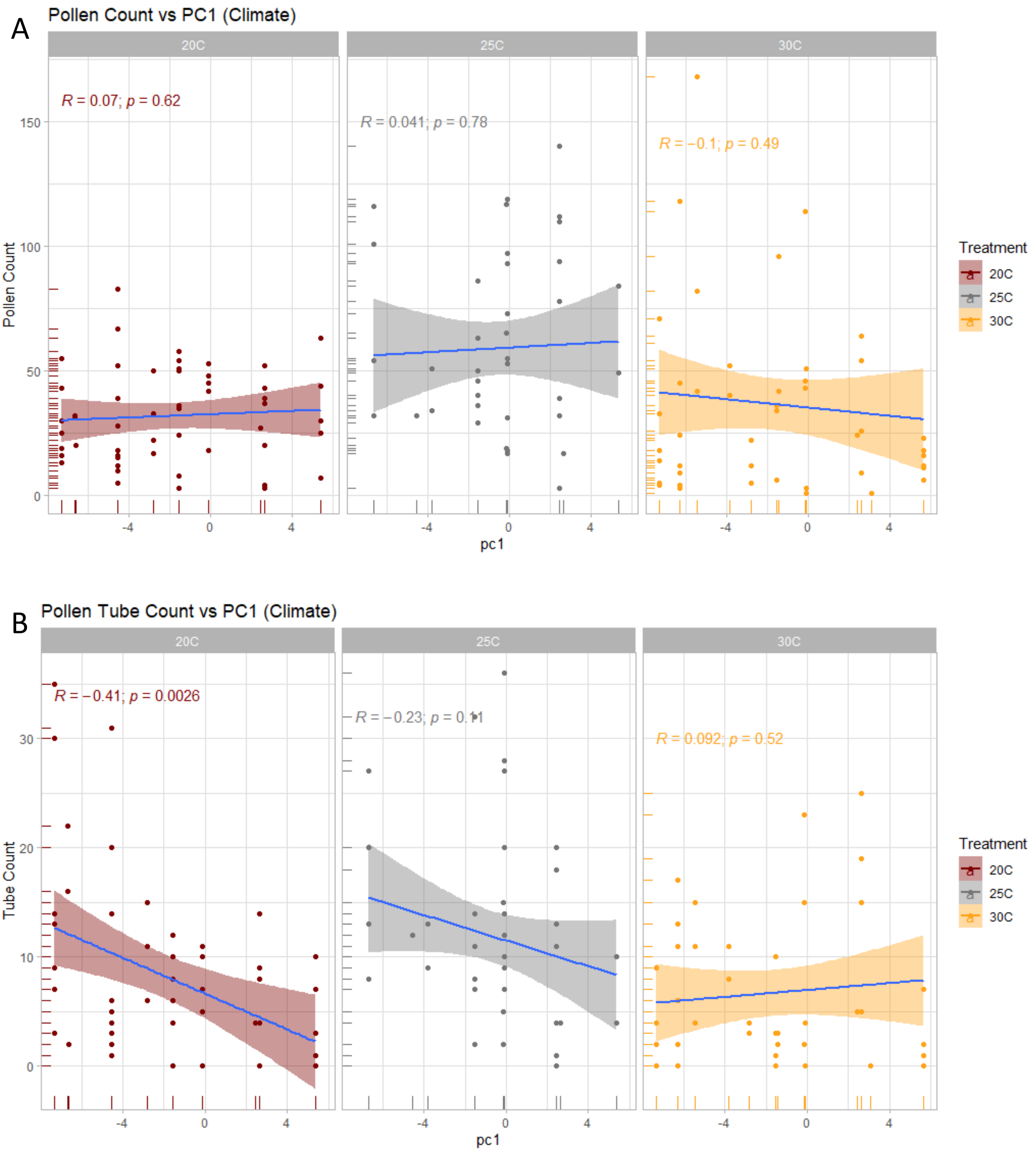
### 3.4.8 Results: Pollen and Local Climate

With the latitude of localities showing significance on the ability of pollen to form pollen tubes under the cold treatment, it is possible to link pollen viability with climatic variables, especially variables such as average temperature. We suggest that more Northern individuals are varying sinigicantly in terms of pollen tube formations to the Southern individuals. We used summary PC for climate variables from chapter 2 to compare against pollen variables and climate on our wild samples.

The number of pollen counts observed (Figure 11A) illustrates that there was no significant correlation between the pollen count and the principle component 1 (pc1) of the climatic variables (*p=>0.05* for all temperature treatments). This agrees with the findings for pollen count against latitude (see above section) whereby pollen count doesn't correlate with latitude of locality. The number of pollen tubes, reveals a difference in trend against the cold and glasshouse treatments compared with the heat treatment. With higher levels of PC1 observed, the less pollen tube is observed within the cold and the glasshouse treatment while under the heat treatment, the higher the level of PC1 for climatic variables the more pollen tube observations were made. However, just like under the findings against the latitude model, only the cold treatment showed a significant decrease in the number of pollen tubes (*R=-0.41 p=0.0026*).

More significant correlations between pollen tube proportions and climatic variables can also be summarised by PC1 (Figure 11C). The correlations reveal the proportion of pollen tube decreases more significantly with an increase in the PC1 value under the cold treatment (*R=-0.54, p=<0.0005*). However, in contrast with the pollen tube count alone; when considering pollen tube formation in relation to the pollen count, the glasshouse treatment also shows a significant decrease in proportion in the event of an increase in PC1 value, a trend which was not significant under pollen tube count only. This reveals that proportion (in percentage) of pollen tube able to form were significantly correlated with climatic variables in the form of PC1 under both cold and glasshouse treatments. There was still a positive trend when considering the heat treatment. However, this remained unsignificant (*R=0.12, p=0.42*).

A higher level of PC1 reveals a loading of warmer average temperatures for the local areas where the population were collected from in the wild (see section "3.3.3 Analysis: Population of Origin Variables"). Therefore, the significant decrease in the number of pollen tube count against a higher PC1 value reveals that at least in the cold treatment, the number of pollen tube able to be formed in our wild samples were significantly correlated with the higher temperature of their local climate. This result illustrates that under cold temperature treatments, individuals that are localised to a warmer climate are less able to form pollen tubes than that of individuals originating from colder climate.



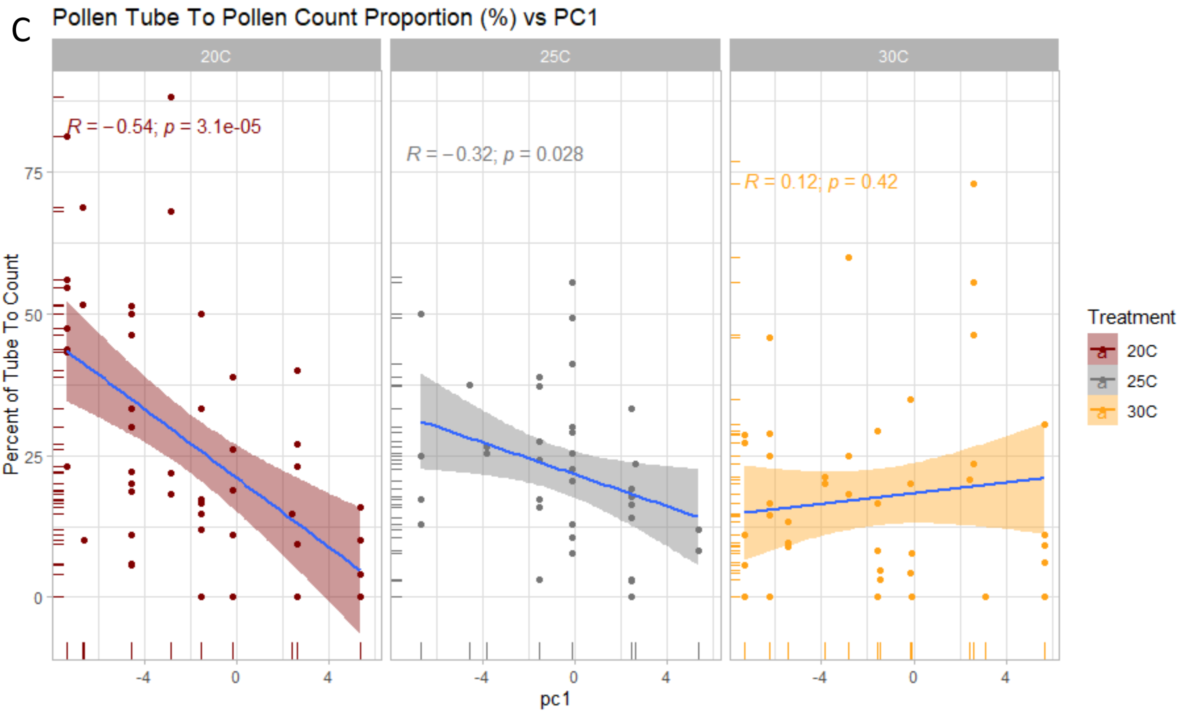


Figure 11. Pearson's correlations and linear modelling between pollen counts (A) and pollen tube counts (B) and climate PC1 values. The only significant correlation seen was between pollen tube count and PC1 values at cold temperature as seen in Fig 17 B (R=-0.41, p=0.0026, <α of 0.05). Pearson's correlations and linear modelling between pollen tube proportion (in percentage) to the count against the PC1 climatic variable were also illustrated (Figure 11C).

## 3.5 DISCUSSION AND CONCLUSSION

For pollen germination, *in vivo* germination worked better than *in vitro* germination. This is because the solutions and conditions that were tested under this study did not provide an ideal *in vitro* condition for *Linum* pollen germination. This caused the pollen's content to "burst out" of the pollen exine and thus making pollen germination impossible. This "bursting" was perhaps due to the osmotic pressure applied by the solutions. Some studies have found that "bursting" of pollen could be possible due to mechanisms of osmosensory regulations (Shachar-Hill et al., 2013). Other research in Tobacco (*Nicotiana tabacum*) pollen, revealed that pollen cell volume changed "rapidly" in response to extracellular osmotic potential (Zonia & Munnik, 2004). Another study found that swelling and bursting of pollen grains was caused by the effect of the environment in protein production and protein infolding to the otherwise stiff exines of pollens (Božič & Šiber, 2022). Another study revealed that pollen is fragile to water status and associated processes once they leave the environment of the anthers (Firon et al., 2012). Future *Linum* studies could explore how osmotic pressure of different *in vitro* germination solutions could affect the integrity of *L. bienne* pollen.

In this study, the pollen count was reduced by the different temperature treatments (cold and heated) relative to standard growing temperature of 25$^{o}$C. In the overall data, averages of measures such as the pollen count, % of pollen forming tubes and the pollen tube counts all revealed a reduction in the heat and cold treatments relative to standard glasshouse treatment (Figure 6). In tomato (*Lycopersicon esculentum*) it has been previously shown that the effect of lower temperatures had no significance reduction in the counts of pollens observed, and even in cases increased pollen tube proportion (Peet & Bartholemew., 1996). In contrast, a temperate species borage (*Borago officinalis*) showed reduction in pollen grains when treated with heat treatment (Descamps et al., 2021). This species, is heavily reliant on insect pollinators. This is not the case for *L. usitatissimum* which is known to be a better self-pollinator (Williams et al., 1990). Perhaps, differences in breeding strategies and environmental variables could lead to different sensitivity in pollen production for different species, since selfing species doesn't necessarily interact with other organisms to germinate, making it less desirable to produce as much pollen as possible for success of pollination due to reliance on environmental pollinators. Considering the number of pollen tubes, this measure was always significantly different between the different treatments (Table 3). In all cases, there was a reduction of the number of pollen tubes observed in the cold and heated treatment. This may suggest that in our samples, there was an optimal temperature in which the number of pollen tube observed was maximized. In a study involving several *Rosaceae* species, it was suggested that pollen germination was optimized differently for different species at temperatures between 15$^{o}$C to 30$^{o}$C (Beltrán et al., 2019). However, when looking at the different in average pollen tube counts between the cold and the heated treatments, this difference was also significant. This suggests perhaps pollen tube formation was better under cold stress than it was under heat stress. One example which observed a better pollen tube growth under lower temperature was one that was conducted with *Citrus* (Montalt et al., 2019).

When the number of pollen tubes was combined with the pollen counts to consider a proportion of pollen tube measure, the proportion of pollen tube decreased significantly with the heat treatment (Table 4). This illustrates that in addition to the decreasing number of pollen counts, under increased temperatures, the proportion of pollen able to form pollen tubes also decreased when temperatures were increased. For the colder temperature treatment however, there was a decrease in the proportion of pollen tube number, but this wasn’t significant. This reveals that pollen tube formation in *Linum* is more sensitive to an increase in temperature rather than a decrease in the temperature. There seems to be some agreement in other temperate species. In cultivars of cherry (*Prunus avium* L.), pollen germination reduced while pollen tube growth accelerated when there was an increase in temperature. Therefore, pollen kinetics seemed to be affected with higher temperatures (Hedhly et al., 2004). Other studies also point to changes in stigma reception ability to the pollen tubes. In a study of the cultivated peach (*Prunus persica* L.), stigmatic receptivity was affected by temperature. However, accelerated pollen tube growth was observed, which the study explained as the opposite effect of temperatures on the male and female side in peach (Vuletin Selak et al., 2013). Another study of pollen performance in olives (*Olea europaea L.*) found that temperature affected pollen tube viability and pollen tube growth was reduced under increased temperatures. The study also suggests that temperature and genotype interaction was significant for pollen performance in olives (Vuletin Selak et al., 2013). There are various suggestions when it comes to pollen tube growth and temperatures in the literature. For different species and cultivars, it maybe that pollen tube growth is affected differently under different temperature treatments. This was evident when looking at species level (wild vs cultivars) in our study. Our result suggests that in cultivated *L. usitatissimum,* the proportion of pollen tubes was significantly affected by increasing and decreasing temperature treatments relative to the glasshouse treatments. It seems that the optimal temperature for the cultivar in our samples was the glasshouse treatment at 25$^{o}$C daytime and 13$^{o}$C nighttime temperature. In the wild *L. bienne* samples however, the greatest pollen tube proportions were observed under the cold treatment Although, this was not significantly different to the other temperature treatments for *L. bienne*, it showed a trend towards a lower temperature optimum to that of their cultivar relatives.

When looking at the proportion of pollen tubes for each species, it was evident that cultivated *L. usitatissimum* was more sensitive to temperature than the wild relative *L. bienne*. This was evident because under the different temperature treatments, pollen tube proportions were significantly smaller in both cold and heat treatments with the cultivars. There was a more significant reduction in the proportion of pollen tubes than their wild ancestors. The effects of higher temperature stress negatively affected pollen tube formation more so than lower temperatures, especially in the cultivars. This effect of higher temperature stress on pollen viability is not new in plants. In chickpea (*Cicer arietinum*) for example, the pollen grains have been shown to be more sensitive to heat stress than their stigma counterparts (Devasirvatham et al., 2012). The fact that the crop wild relative *L. bienne* showed a potentially more resilient trait to rising temperature could be of interest in crop development. This importance has been highlighted in durum wheat (*Triticum*) before, which shows wild relatives yielding more grain under temperature stress when compared to the cultivar relatives (El Haddad et al., 2020). The fact that lower temperature affects pollen tube formation has been observed in other plants. Studies using pear also suggests disruption in pollen tube formation. They suggest that this is mediated by mitochondrial metabolic dysfunctions (Gao et al., 2014). The CBF/DREB1 proteins have also been identified to control expression of cold-induced genes (Thomashow, 1999). In Barley, novel alleles to frost resistance have also been identified as *FR-H2* amongst others (Sallam et al., 2021). There is potential that these cold resistance genes identified in other study models may play a role in higher tolerance to cold temperature observed in *L. bienne* pollen tube formation ability. In *Linum* this phenotypic difference in temperature stress pollen tube sensitivity may represent more diverse genetic resources for the improvement of crop heat stress resilience, as was reviewed in various wild relatives for crop improvements (Dempewolf et al., 2017).

In terms of the ability of pollen tube to reach the ovary, the heat treatments revealed a significant reduction of pollen tubes that were able to reach the ovary. This was not the case in the cold temperature treatments. This suggests that an increase in temperature had a significantly negative effect on the ability of *Linum* pollen tubes to reach the ovary (Table 5). This result illustrated that *Linum* pollen's abilities to reach ovaries are more sensitive to an increase in temperature rather than a decrease. In a study with wheat (*Triticum aestivum* L.) pollen tube growth into the ovary was reduced due to an increase in temperature. It was suggested that this was because heat causes abnormal conditions for the ovary (Saini et al., 1983). In peach (*Prunus persica* L.), stigmatic receptivity to pollen tube was negatively affected by a rise in temperature (Hedhly et al., 2005). Suggestions in the literature often reveal negative effects of temperature increase on the female parts such as stigma and ovary that results in the observed reduced ability for pollen to reach the ovary. In this study however, we did not look at stigmatic and/or ovary 98arlierpment. In line with what is seen with the proportion of pollen tube reaching ovaries in this study, it would be of further interest to see how this negative response to the increase in temperature correlates to stigmatic and/or ovary development in our *Linum* samples.

Furthermore, when analysed separately, wild *L. bienne* showed a more significant reduction in the number of pollen tube reaching the ovaries than those of their cultivar relatives. This trend was not expected since the proportion of pollen tubes to pollen count is significantly more reduced in the cultivar *L. usitatissimum* compared to the wild *L. bienne*. In the proportion of pollen tube able to reach ovary/species the wild *L. bienne* seems to have responded more strongly than their cultivar relatives. There is a suggestion here that even though wild pollen tube to pollen count proportion were higher than the cultivars, their ability to reach the pollen was more sensitive towards warmer temperature. Perhaps the pollen's female counterparts such as the stigma and the ovary were more sensitive to temperature changes in wild relatives (Fábián et al., 2019). A smiliar observation was also made in peach whereby stigmas tend to lose their capacity to support pollen tubes (Hedhly A. et al., 2005). Hedhly A. et al., 2005 also suggests that there are contrasting effect of temperature on the male and female parts of the flower. It maybe the case that, the amount of pollen tubes is affected by the ability of female counterparts, such as the stigma, to sustain pollen tubes at different temperatures. This aspect is yet to be explored in *Linum.* In this study, we showed that the heat (30°C) treatments cause a significant reduction in the number of *Linum* pollen tubes, and pollen tubes reaching the ovaries. However, this can be due to the sensitivity of female counterparts to rising temperature, as is evident in other species (Pan et al., 2018; Fábián et al., 2019). In future studies, it would be interesting to observe the effects of temperature on the growth rate of *Linum* pollen tube inside the stigma. This can be done using live images from confocal fluorescence microscopy on *in vitro* germination studies.

A major aim in this thesis was to look at local adaptations, that was done here by a comparison between pollen measures and latitude of locality and climatic variables. We were interested in whether there was any correlation between the pollen count, pollen tube count and/or the proportion of pollen tube forming against local geographic and climatic variables. None of the different treatments shows a correlation in latitudinal and climatic variables when considering pollen counts. This suggests that the number of pollen in our sample set is neither positively nor negatively correlated with local climates. Other study observing percentages of pollen across the Mediterranean region found ecological trends in not only pollen data but also other plant traits to climatic gradients across the region (Barboni et al., 2004). We expected to see a trend in *Linum* pollen count with climatic variables, however, we didn't observe a significant trend in our sample set. Although this was the case for our study, there was some evidence of a change in direction of correlation trend in the number of pollens observed and local climatic variables, when looking at the different temperature treatments. This correlation was not statistically significant but perhaps with a larger sample set and from a more varied local environment, the trend would be more significant and more obvious, and a larger sample set would help in statistical power as well. Another study suggests that ‘fresh’ pollen should be used for observation studies as pollen conservation may impact germination abilities of pollen (Beltrán et al., 2019). Although laborious, it would be of interest to use ‘fresh’ pollen instead of a preservation method, as was done in this study. The number of pollen observed under different climatic variables is important as this may influence the possibility of gene flow between different populations and species. As *Linum* is a self-pollinating species, it is perhaps plausible that environmental segregation of populations by pollen production could influence gene flow dynamics as well as pollen/pistil incompatibility as was shown post-pollination in *Polemoniaceae* (Ruane & Donohue, 2007).

When observing pollen tube count however, there was a significant association when looking at the cold treatments. This observation was true for both the latitudinal and climatic variables. Even within our limited sample sets, there was a latitudinal trend that, the more northern our sample was collected from, the more pollen tube it can produce under colder temperatures. The correlation with climatic variables also agrees with this observation. Local climate may, in the case of wild *L. bienne, influence* pollen viability, not just in terms of temperature, but also other climatic viables which may have played a role. In oak, pollen viability and sunlight availability in the local areas possibly influence pollen mediated gene flow between populations (Schueler et al., 2004).

In addition, more northern wild population showed potential local adaptation with higher proportions of pollen tube forming observed in the cold treatments (Table 6A). This suggests that wild *L. bienne* are perhaps more adapted to colder temperature. There is potential for GWAS study here for Northern wild *L. bienne* to identify novel genes for cold tolerance. This can have implication towards improvement of cultivars. There was also a difference in the trend with heat treatments. When heated, the correlation changed direction between the latitude of localities and climatic variables and pollen and pollen tube counts. This was not significant but could be tested with a larger wild individual from a wider range of local variables in the future.

In conclusion, different temperatures affected male reproductive function in terms of pollen number, pollen tube number, proportion of pollen tubes forming against pollen count, and proportion of pollen tubes reaching the ovary. In all cases there seemed to be an adverse effect of increases of temperature, be it pollen, and pollen counts and the amount of pollen tubes able to reach the ovary. Therefore, in our samples, an increase in temperature lessened pollen viabilities for *Linum*. There was also evidence that in the colder temperature treatments, pollen viability is correlated with latitude of localities and climatic variable. Local adaptation in pollen viability may have resulted in the differences seen. Additionally, we would like to observe the female counterpart of these *Linum* samples do in relation to the pollen viability trend that was seen in this study. Further study on the reception of female structures, such as the the ovary may reveal that, in line with the trend seen in pollen viabilities in *Linum*, the female counterparts also play a major part. Illustration in Sorghum have observed deterioration of ovary under heat stress (Chiluwal A. et al., 2020). There is also interest in looking at further colder treatments for the wild population to see whether there is optimal temperature for pollen viability and to generate results for more samples from more diverse latitudes. The differences seen here are evidence of local adaptation in the wild *Linum* species. More northern individuals seem to be more locally adapted to colder temperatures and could be more adversely affected by an increase in temperature in terms of their ability to form pollen tube and the ability for pollen tube to reach ovaries.

# CHAPTER 4: POPULATION STRUCTURE AND DIVERGENCE WITHIN LINUM SAMPLES ORIGINATING ACROSS WESTERN EUROPE LATITUDE RANGE.

There is ample evidence that climate change has the potential to drive distributions of population within both plants and animals, often having significant changes to ecosystems (Hampe & Petit, 2005; Mori et al., 2022). When looking at distribution of wild populations, it is with interest that latitudinal and longitudinal ranges for local populations could return insights into genetic diversity and conservation ecology. Genetic diversity is important in living organisms for adaptation to changing environments. Thus, conservation of species often depends on genetic diversity in each population (Alcala et al., 2013). In the wild, events such as genetic bottlenecks, genetic insertion/deletion, selection, and genetic duplication are widely known to shape patterns of genetic diversity in each population (Alcala et al., 2013). The central-marginal hypothesis predicts that at range margins there is decline in genetic diversity but an increase in differentiation towards speciation, due to variation which are caused by events such as genetic drift and geneflow (Langin et al., 2017).

*Linum* is a plant genus with more than 150 species (Muravenko et al., 2010). There is a wider interest establishing the extent of genetic diversity and structure within different populations in wild *Linum* populations. A wide interest in plant research is looking at genetic implication of different phenotypes within plants, especially for crop development. In crop development, genetic analysis tool such as quantitative trait locus (QTL) analysis are often used to pinpoint loci which are responsible for a desired trait within a plant type (Asíns M., 2002). Once desired traits are identified, further cultivar developments could be approached with marker assisted selection. However, in these types of studies a whole genome sequence is often required. More recently new sequencing approaches are being developed which enables genetic analysis studies to be done relatively faster and cheaper (Behjati & Tarpey., 2013). Studies involving genetic marker systems to genotype wild and cultivated types were able to identify genetic structuring to further reveal diversity and consequences of selective breeding in the cultivars without the need of a whole genome (Bacilieri et al., 2013). This can be achieved using sequence repeats such as simple sequence repeats/microsatellites (SSRs). SSRs are sequences which are ubiquitous in a specific eukaryotic genome and are amongst the most common genetic markers to be developed (Goldwin et al., 2005). SSRs were developed along with PCR procedures to amplify sequences, often looking at genetic diversity within set populations, and between plant types (Grapin et al., 2005). Studies looking at non-model species have been able to identify reduction of genetic diversity due to hybridization and genetic bottlenecks events which revealed consequences of cultivation of the species (Guan et al., 2021). This is also of interest in terms of molecular ecology. Genetic analysis has previously identified localization and invasion histories of non-model plants (Hernández et al., 2019). To look at differences across population, it is of interest to establish a structure and diversity between the local population.

In *Linum,* previous research used Single Sequence Repeats (SSR) techniques. In the research, SSRs were developed for 34 Turkish wild flax (*Linum bienne*) accessions and accessions from different cultivated flax (*Linum usitatissimum*) lines. The research found variation in polymorphisms between the accessions and clustering of 493 individuals according to their respective types, suggesting wild flax are more closely related to the deshiscent type. Along with this, the research also suggests genetic distancing among the wild types are significantly related to their geographical distances as well as their elevation (Uysal et al., 2010). The research revealed potential for marker studies using *Linum* types to understand genetic structuring, domestication, and genetic diversity of *Linum*. For this purpose, short reads within the whole sequence were used as a genetic marker for different individuals. This is substantially easier than obtaining a whole genome, as a whole genome for the wild *Linum* species is not necessary for observation of population structure based on these markers. This being the case for the wild species, a chromosomal-level whole genome is readily available for the cultivar relatives *L. usitatissimum* (Sa et al., 2021). In this chapter, we use ddRAD sequencing to investigate whether there are structuring, diversity and or divergence within our *Linum* samples.

We have a collection of wild *L. bienne* and several oil and fibre cultivar variety of *L. usitatissimum* originating from Western Europe. Within the *Linum* samples available under this project, there were a variety of individual plant populations that originated from different places across the latitude of Western Europe. We hypothesize that there is genetic variation due to localization within different environments. This could be implied by structuring and genetic diversity measures which may have taken place across time as different population form and localize to their local environment. When looking at the collection of *Linum* samples from across western Europe, genetic structure would be revealed as the genetic distance between the different *Linum* populations as well as between the wild and cultivar types. This would help in explaining their population genetics and as well as their breeding systems. There will be an interest in answering the question of “can we use genetics to interpret population structures and breeding systems in *Linum?*”.

Our *Linum* samples consisted of 121 different individuals which made up 15 wild population (*Linum bienne*) from across Western Europe and 12 Cultivars (*Linum usitatissimum*). In the 15 different wild populations, there were 109 individuals, and, in the cultivars, there is only 1 individual per cultivar line, which we have classified under one "cultivar" population for this study. The wild populations were collected respectively from the southern Spanish through to northern UK regions whilst cultivars were gained from IPK World Collection (Figure 7). This provides a wide latitudinal range for the temperate plants. When looking at wild types regarding their population structure, this was more likely affected by localization by the different population under the different latitudinal range. Groupings between the different populations are expected whereby groups can be distinguished by their local population. Population structures are inferred by determination of genetically related clustering (groups) observed without prior knowledge of the populations (Odong et al., 2011). The cultivars were expected to group together while the wild populations were expected to be grouped based on their region of locality. There is an expectation that populations within the sample sets group between northern and southern European populations. This would agree with morphological differences found under the vernalization experiment in this thesis (Part "2.4.1 Results: Vernalization and flowering time"). We also examined ancestry, to determine wild populations relationships to the set of cultivars within the experiment.

For the purposes of sequencing genetic markers, a double-digest Restriction-site Associated DNA Sequencing (ddRADSeq) protocol was optimized to enable a digestion of DNA which could be done with multiple restriction enzymes. The use of a second restriction enzyme allows for more precise and consistent cutting of the DNA as well as provide more combination of previously multiplexed samples, allowing for a more homologous indexing of many individuals (Peterson B. et al., 2012). This enabled the large-scale generation of short read sequences without the need to sequence a whole genome. DdRADSeq in turn enabled genomic studies in models without whole genome databases and non-model organisms alike (Arnold et al., 2013) (Peterson B. et al., 2012). A previous SNP comparison was used to estimate genetic diversity and population structure in rice (*Oryza sativa*) (Singh et al., 2013). SNPs are thought to have numerous advantages over SSRs. These include a more precise estimate of population diversity and the ability to consider local adaptation through identification of groups by clustering methods (Zimmerman., 2020).

Illustration of population structure, genetic diversity, and selection analyses were carried out through the data gained under the ddRADSeq protocol. Genetic information regarding the *Linum* population was gained in a relatively cost effective and time effective way. We hypothesize that genetic structuring is visible within our *Linum* samples as a signal of local adaptation. In addition we also investigates wether there are other genetic signal such as genetic diversity and heterozygousity in our samples for the implication of genetic distinction within our *Linum* populations.The expectation in the *Linum* sample, is that individuals would be grouped based on their population and location, and that grouping between the northern and southern populations would match what was previously observed in previous chapters which have examined the Linum samples morphologically. This would reveal that, structurally, northern, and southern populations in the Linum collection are genetically distant to each other. Further analysis could explore genetic traits such as heterozygosity and whether any population is genetically more diverse than other populations in the sample set. This would highlight potential breeding system differences in our Linum samples.

## 4.2 STUDY AND METHODS

### 4.2.1 Modified RAD Seq Protocol

For this chapter ddRADSeq protocol was used along with some modifications to the original protocol described by Peterson et al (2012). DNA was obtained through a modified DNA extraction protocol. The DNA extraction protocol was based on a method described by Doyle & Doyle (Doyle & Doyle 1987). After DNA extraction, a digestion procedure was carried out using two enzymes (we used MseI and PstI in this study) whereby the digestion was used to cut DNA restrictively into small sequences around 250-500bp long, further size selection (250-500bp) occurs during the Pipin stage of the protocol, after the ligation stages. DNA quality and quantities were observed through several quantification methods which includes gel electrophoresis and spectrophotometry. After satisfaction with DNA quality and quantity, libraries were built with sequences that are depicted as the Figure below.

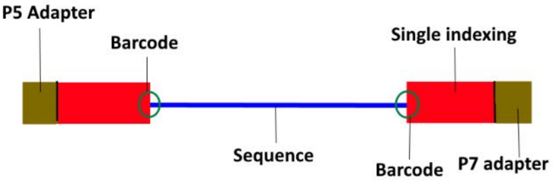


Figure 1. An illustration of the ligated DNA material per sample which are both adapted and barcoded.

The final product of the DNA sequence after the ligation and barcoding procedures can be illustrated (Figure 1). The protocol used in this experiment was based on a ddRADSeq protocol previously described by Peterson *et al* (Peterson et al., 2012). Population structure analyses were subsequently performed to group individual samples based on their genetic distance relative to each other.

### 4.2.2 DNA Extraction Procedures

For RAD-Sequencing purposes, we required young leaf materials for optimal DNA quality. Wild S0 seeds were grown in the glasshouse (16:8 day/night ratio and day/night temperatures at 25/13°C) for optimal growth condition. 121 unique individuals representing the widest possible latitudinal ranges (from 31.791° South to 53.352° North) were chosen for this purpose, covering temperate regions of western Europe. The samples have a latitudinal range which covers northern UK down to the Southern coast of Spain. These samples include wild and cultivar species with the wild species collected directly from the field. A modified CTAB DNA extraction protocol was used from methods described by Soltis lab (Doyle, 1987).

We modified the main CTAB washing buffer for DNA extractions. The presence of salt and polysaccharides could potentially interfere with downstream processes. To reduce this interference, suggestions were adapted from Clarke *et al* (Clarke J., 2009). The final modifications to the samples were made as follows:

| Reagent | Amount (For 100mL) | Final concentration |
|---|---|---|
| **10% CTAB in dH$_2$O** | 30mL | 3% |
| **5M NaCl in dH$_2$O** | 28mL | 28% |
| **0.5 M EDTA (pH 8.0)** | 4mL | 4% |
| **1 M Tris-HCl (pH 8.0)** | 10mL | 10% |
| **Polyvinylpyrrolidone (PVP) (MW 40 kDa)** | 3g | 3% |
| **β-Mercaptoethanol** | 0.2mL | 0.2% |
| **dH$_2$O** | 24.8mL | 24.8% |

Table 1. Reagents for the CTAB extraction buffer

With the CTAB buffer made, the protocol were carried out as described by Doyle (1987).

### 4.2.3 The Modified CTAB protocol

In the laboratory, dry baths were pre-heated to 60°C and RNAse were removed and ice-thawed from -21°C storage. In the glasshouse, 10-12 young, 'green' leaves were collected per individual. Samples were individually homogenised. Alternative homogenising methods were also initially used, using 3mM metal beads, a pinch of sand and a Tissue-Lyser II. However, access to Tissue-Lyser II were limited due to Covid-19 restrictions and most samples were ground using a tissue homogeniser. Samples were ground as finely as possible inside Eppendorf tubes and flash frozen in liquid Nitrogen until all samples were ready.

Per sample, DNA was cleaned as per Soltis lab protocol (Doyle, 1987). Samples were homogenised further with a vortexer until ground tissues are thoroughly mixed with the CTAB and RNAse A solution. Samples were then placed on a 60°C pre-heated dry bath for at least 60 minutes (modified to allow time for optimal DNA lysis). Samples were then thawed to chill on ice for 3 minutes before further processes.

Samples were further washed using phenol-chloroform solutions as per protocol. To wash away any potential phenol after the wash, 200µL of chloroform isoamyl were added. Samples were centrifuged at 14,000RPM for 10 minutes before carefully removing the aqueous layer into newly labelled Eppendorf tubes. Na Acetate were then added as per protocol. Salt was washed away at room temperature with isopropanol as per protocol. Centrifugation at ≥14,000 RPM for 15 minutes were applied after the addition of isopropanol. A pellet at the bottom of the tube were often observed after the last centrifugation step, although not all the time. Supernatant was subsequently removed from the tubes. Into these sample tubes, 500µL of “Ice cold” 70% ethanol were added to wash the DNA pellets. Tubes were then centrifuged at ≥14,000 RPM for 15 minutes at 4$^o$C and ethanol taken out. Samples were air dried (preferably under a laminar flow hood) for around 10 minutes, until samples were dry. Dried samples were then re-suspended in 50µL of nuclease free water as elutant. Samples were then stored overnight in a 4$^o$C fridge before measurements of both quality and quantity.

### 4.2.4 DNA Quality and Quantity

We used a NanoDrop ND1000 to quantify 1µL of samples. Sample quantity and quality were measured as DNA concentration (ng/µL) and absorbance ratios at 260/230nm and 260/280nm. A more accurate Qubit method were also explored using Invitrogen’s Qubit 2.0 fluorometer. We used Promega’s QuantiFluor ds DNA System (QuantiFluor® dsDNA System, Promega Corporation), DNA samples were able to be quantified more precisely using this protocol. In most cases, DNA were also measured using gel electrophoresis.

For further quality measures, 0.5×TBE gel electrophoresis were used. Biorad’s DNA Mini Gel tank, with a power supply unit were used to run these gels. 2µL of New England Biolab’s DNA purple loading dye and 3µL of nuclease free water were dispensed and lined up per number of loadings required. A 1Kb DNA HyperLadder (Bioline) were used to measure the size and quantity of DNA. Gels were electrophoresed and observed using a UVIDOC system.

### 4.2.5 Double-Digest RAD Sequencing

After satisfactory DNA materials were acquired (>20ng/µl), we processed the DNA through a modified ddRADSeq protocol described by Peteron et al 2012 (Peterson B. et al., 2012). Samples were digested and ligated to specific adapters and barcodes. This would be proceeded by pooling and cleaning stages of the products before further size selection and sequencing.

#### 4.2.5i RAD – Digestion:

To fragment extracted DNA materials, enzymes were used to digest said DNA materials. There are several different digestion enzymes available to use, each with their own specifications. We tested three enzymes to digest DNA materials, namely, *Mse1, Pst1, Sbf1* (manufactured by New England

Biolabs). Two different combinations were tested: *Mse1/Pst1* and *Mse1/Sbf1* with the digestion protocol using NEB's buffer 2.1.500ng genomic DNA and 5µL of buffer 2.1 was used as recommended by the supplier for optimum enzyme efficiency. We prepared a 30µL reaction. A master-mix of the buffer and enzymes were prepared for every sample. For incubation, and inactivation purposes, a Prime Techne thermal cycler was used. A 30µl digestion reaction were prepared as per manufacturer's protocol.

### 4.2.5ii RAD – Anneal adapter and Ligation

Digested DNA was subsequently ligated by using a ligase enzyme (manufactured by New England Biolabs). During ligation, digested DNA was adapted and barcoded specifically. To do this, adapter oligos were first annealed together per manufacturer's protocol.

After mixing briefly adapters were annealed using a protocol recommended by Eurofin Operon. In a thermo-cycler, adapters were heated to 95°C for 2 minutes and ramp cooled to 25°C over a period of 45 minutes. Annealed adapters were then stored in a -21°C freezer for long term and in a 4°C fridge for short term (<72 Hours). The process was repeated for every adapter available.

Samples and adapters were now ready for ligation stage. In the experiment, one forward and one reverse working stock adapters were combined with fragmented sample DNA and ligated to barcode the samples. New England Biolab's T4 ligase enzyme at 400U/µL concentration and a 10× T4 ligase buffer were used for this. The samples were loaded according to the 30µL final reaction volume from the digestion steps. The following tables describes each component for ligation. Samples were incubated at 23°C for 30 minutes and then heat inactivated at 65°C for 10 minutes. Samples were then cooled down at 2°C per 90 seconds until they reached 23°C.

### 4.2.5iii RAD – Pooling and Cleaning Ligated products

To pool and clean ligation products from primers, SeraMag magnetic cleaning beads were used. This step required all samples to be pooled into a single container. Total volume of the pooled samples was calculated as the volume of leftover ligation product after PCR (PCR product) × total amount of samples. 1.5mL Eppendorfs were used for pooling and cleaning. For each tube, the volume of pooled ligated product would be no more than 300µL, to be cleaned separately.

Before using the SeraMag beads, working solutions were prepared per manufacturer's instructions. 1mL of the working solution were then transferred into a 1.5mL Eppendorf tube and further processed as per manufacturer's instructions.

After the final magnetic stage, tube with the magnetic beads and 1mL TE were put set in a rack (non-magnetic) and PEG-8000 and Tween 20 solution was added per manufacturer's instructions. When the solution is mixed, 1mL of the previously prepared 1mL TE + SeraMag beads were pipetted into the conical flask. Conical flask was then filled up to 50mL volume with MilliQ water and mixed gently until the beads are evenly spread across the solution. The solution were then transferred into a 50mL Falcon tube, wrapped in aluminium foil, and stored in a 4°C fridge ready for use.

Before using on ligated products, the cleaning beads were tested against a 100-1Kb DNA ladder (BioLine's HyperLadder). After confirmation against DNA Ladder (different ratio of beads should cut at different fragment sizes), the prepared Sera-Mag beads were then ready to be used for the ligated products. To do this, clarification needed to be made about the ratio of beads to be used for the fragment size which are of interest. In the case of this experiment, 250-500bp regions were of interest. A 1.5× SeraMag to ligated product volume ratio were used.

After the addition of the beads working solution, solution was incubated at room temperature for 5 minutes. After this, samples were placed back into the magnetic stand and processed as per manufacturer's protocol. The produced DNA materials were then washed and dried per manufacturer's instructions. This wash step was repeated twice. When the second wash finished, beads were placed on a 37°C heat block until beads were dry. Rehydration took place with 20µL of nuclease free water as elutants. When the beads were mixed in the water, samples were placed back in the magnetic stand until all beads were pulled towards the magnet. When all beads were pulled towards the magnet, the supernatant were then extracted and transferred to a new 1.5mL Eppendorf tubes or into the next 300µL pooled tube until all were cleaned. Cleaned samples were then quantified using a NanoDrop ND1000 spectrometer.

### 4.2.5iv RAD – PCR Amplification to Generate Illumina Sequencing Libraries

High-fidelity PCRs were undergone with all the samples for ligation verification and generation of sequencing libraries for Illumina. ThermoFisher's Phusion High Fidelity DNA Polymerase were used for this protocol with P1 and P2 adapter oligos as the primers. A total volume of 20µL reactions were prepared as per manufacturer's instructions. For this protocol, 20ng of input DNA were used. The input DNA into the digestion product was 500ng. 20ng in volume of DNA were calculated as follows:

$$Input\ DNA = \frac{500\ (digestion\ DNA\ input)}{20(DNA\ required\ for\ PCR)} = 25$$

$$Final\ Volume = \frac{37\ (Final\ digestion\ volume)}{25\ (From\ input\ DNA)} = 1.5\mu\mathrm{L}$$

The input DNA from ligation were calculated as 1.5µL. The samples were then run on the 111arlier-cycler for 12 cycles of the following programme: *98°C for 2 minutes, 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds and a final 72°C for 10 minutes*. After the cycles were done samples were held at 4°C. A 1×TBE gel electrophoresis was performed to check the presence of products against a DNA ladder.

### 4.2.5v RAD – Pippin size selection Preparation

To construct a ddradseq library of 250-500bp fragments, a Pippin Prep facility was used (Sage science). A 1.5% Agarose gel cassette (with marker L) was used for Pippin preparation using manufacturer's protocol. The instrument was calibrated prior to the run as per manufacturer's protocol.

To run the Pippin size selection, the cassette was sealed with the provided seal. The automatic test was run and the current measured in each elution channel at room temperature. A size selection protocol was then manually inputted or selected from the "Protocol editor" tab with sample ID provided. Under the "Cassette" input, an appropriate cassette file was loaded and under the "Reference Lane" input, the marker well was inputted and applied to all lanes. The pippin was subsequently run for each sample well to select at 250-500bp for the purpose of this experiment. When the Pippin prep finished running, 30µL of sample was collected from the elution module. An Agilent Tapestation 2200 facility was used to identify and quantify the region of interest, with a High SensitivityD1000 ScreenTape.

### 4.2.5vi RAD – qPCR Quantification

For a final quantification, a quantitative Polymerase Chain Reaction (qPCR) procedure was explored to quantify DNA materials. In this part of the quantification a BioRad CFX96 Real-Time PCR System was used. For this purpose, 4 dilutions of library materials were explored. The dilutions were made using previous pooling from section "4.2.5iii RAD – Pooling and Cleaning Ligated products" of the RAD methods, each with the appropriate amount for the pooled DNA in 10mM Tris-HCL solution as required per dilution. All dilutions were repeated three times. No template controls (NTC) were used as a control for contamination in the qPCR runs. There were 4 standard dilutions used as a standard in the Bio-Rad system for quantification purposes for each of the dilutions. In total there was 48 wells for the samples and dilutions, 3 NTCs, and 4×3 wells for the standards. This makes up to

63 wells for the run. For the standards, we used 4 concentrations (10pMol, 1pMol, 0.1pMol, and 0.01pMol).

For qPCR purposes, a KAPA Library Quantification Kit (Roche Molecular Systems) were used. When reagents were fully made, a 96-well qPCR plate were designed. Standards and samples were loaded as per manufacturer's instructions. An appropriate cycle programme was then set up and ran.

After the protocol had finished running, a "Quantification CQ result" can be seen. CQ stands for quantification cycles, in other qPCR machines these are also specified as cycle-threshold. We used starting quantity (SQ) values to calculate the final concentration for each dilution factor. For example, for a dilution factor of 1:2000, the SQ-value were multiplied by 2000. Size corrections were able to be calculated by multiplying the average fragment length (250-500bp) by the standards (452bp). Units were able to be converted from picomolar to nanomolar by multiplying by 1000 and subsequently nanomolarity to a concentration unit (ng/µL). Using the delta-CQ values, consistencies between the triplicates were checked before calculations of concentrations.

### 4.2.5vii RAD – Sequencing

When satisfied witih quality checks, an Illumina 2500 Hi-Seq sequencer was used at Durham University's DBS Genomics facility to sequence the ddradseq library. An appropriate amount of pooled library was calculated for sequencing. All the DNA library poolings from the previous steps were combined into 1.5mL Eppendorf tube. Illumina Hi-Seq 2500 requires certain amount of genomic DNA to be read successfully. DBS Genomics at Durham University's Biosciences department recommends at least 10ng/10µl concentration of processed DNA library to process the samples through the sequencer. After quality check measures were made and 2×10µL tubes of sample were sent through the Illumina Hi-seq 2500 sequencer measuring at >10ng/µL each. An Eppendorf centrifugal vacuum concentrator were used as required to combine and concentrate pooled libraries to achieve a 10ng/µL threshold required for sequencing. After the sequencing process, two raw library files which needed further demultiplexing and mapping were outputted. The raw sequences consist of a forward and a reverse sequence for the pooled library which needed to be processed to imply information. We repeated the sequencing twice to gain a better-quality library read.

# 4.3 ANALYSIS: PROCESSING RADSEQ

## 4.3.1 Analysis: Post RAD-Sequencing

To make sure of homogeneity, we used the same pooled library from the same ddRADSeq run. The result of two sequencing processes were four raw sequences from the same pooling (two forward and two reverse sequences), for further analyses, these sequences were merged to provide one raw-forward sequence and one raw-reverse sequence. This required raw reads to be processed bioinformatically.

High Performance Computing (HPC) facilities were available for access at Durham University. The interface for the HPC facility is provided by a Linux cluster using several Intel processors. We used "Bourne-Again Shell" (Bash) to operate the programmes within the HPC facility (Ramey, 2022). The HPC facility were also remotely accessed through internet connection. Bash environments and FTP protocols were loaded using SSH clients such as PuTTY (Tatham, 2022) and MobaXTerm (Mobaxterm., 2022).

## 4.3.2 Analysis: Demultiplexing Raw Reads

Because the sequence was barcoded and adapted specifically during the ligation stages of the DDRADSeq protocol, the raw sequences were able to be demultiplexed and sorted based on their specific barcodes (Croissac et al., 2016). This enabled downstream processes to separate out the sequence based on the unique sample barcodes which were ligated during the ddRADSeq protocol. The demultiplexing protocol can be found under a pipeline prepared for *STACKS v2.61* (Catchen et al., 2013). The demultiplexing protocol were executed using the command *"process_radtags"* under *STACKS v2.61*. The two adapter sequences as found in appendix 15, were also specified in the command. The full barcode index is also provided in appendix 16. The programme was optioned to filter the data for uncalled bases, using built in default parameters. This is done by inputting *"-c"* with the *"process_radtags"* commands, it was also programmed to discard reads of low quality by inputting *"-q"*, and to rescue barcodes and RAD-Tag cutsites by inputting *"-r"* in the command line. The barcode option of the programme was specified to read barcodes which are in line with the sequence and occurs in paired end reads. The number of allowed mismatches when rescuing barcodes were set to two. The full command is listed under (appendix 17). After demultiplexing, STACKS were able to output 580,520,687 reads of which 22,639,406 were reads containing adapters and 814,583 were reads considered as "low quality". After de-multiplexing with default parameter filters, 115/121 individuals were kept, 6 individuals were filtered out due to either excessive number of uncalled bases or low read qualities.

### 4.3.3 Analysis: Mapping to a Reference Genome

Post-demultiplexing, samples were able to be mapped either without a reference genome (*de_novo*) or with a reference genome. We were interested in GWAS analysis, whereby a reference genome was essential. The sample set in this case was inclusive of cultivars (*L. usitatissimum*), and the wild pale flax (*L. bienne*) individuals which are widely regarded this wild species whereby the cultivars were domesticated from (Allaby et al., 2005). These are amongst the justification for opting to map to a whole genome which was already available for the cultivars of flax but not for the wild type.

The Genome for the cultivar is publicly available and are assembled down to chromosomal levels, making it suitable to be mapped to (Sa et al., 2021). Mapping to a reference genome were undergone using the ref_map.pl programme of the STACKS v2.61 software (Catchen et al., 2013). The programme calls on each of the STACKS components noting “gstacks” as the component which align the positions for the RADSeq reads and calls SNPs in each of the sample based on the alignments. The programme expects the data to have been previously aligned with the genome using a separate aligner. For this purpose, BWA_MEM algorithm of BWA v0.7 was used to align the RADSeq reads to the cultivar genome (Li & Durbin., 2009). The outputs were pipelined to Samtools v1.15 to be converted to .BAM files which are the input file type required for ref_map.pl to work. BWA and Samtools were piped under one command. The command can be found under (appendix 18). After obtaining .BAM files for each of the samples, ref_map.pl can be executed using the “ref_map.pl” command. The optional command to execute the “populations” programme of STACKS were also piped into the ref_map.pl command. This is a filter to accept a minimum percentage of individuals in a population to process a locus. This was set to 80% by using the command “-r 0.8”. The populations command also allows further file output options such as the PLINK and VCF formats to be processed further downstream with a population genetic specific software. A population map was also specified to map which samples belongs to which population as well as to map a geographical region specific to that population. This is useful for further analysis downstream. A population map file for the *Linum* samples is listed as in appendix 19. The full command for the ref_map.pl pipeline can be found under (appendix 20). Post ref_map.pl run, several outputs will be of interest. The first if the VCF output which can be used for analysis quality of the reads such as the read depths, allele frequency and PGD-Spider (Lischer and Excoffier, 2012), PLINK (Purcell et al., 2007) or vcftools (Danecek et al., 2011) was used to convert the SNP data into PED and MAP.

### 4.3.4 Analysis: Processing Mapped Reads

To process the mapped reads an initial step was to convert .PED and .MAP output files into .RAW and additional .BED files for read inputs into further downstream processes. To do this, PLINK v1.90 was able to be used (Purcell et al., 2007). PLINK was able to specify .PED and .MAP files in a directory given the same name. The additional command "--recodeA" is a data management options whereby both .PED and .MAP files were able to be processed into a single .RAW data file which includes formatting that are useful for further population genetics analysis. For further analysis a .BED file format were also required. This is obtained through the "--make-bed" addition to the PLINK command. There were numerous (>1) chromosomes in the .MAP files. This means that PLINK needs to acknowledge and allow for this in the output .RAW file. To do this the "--allow-extra-chr" option for PLINK were inputted into the command line for both make-bed and .RAW file options. The full command for PLINK conversions can be found under appendix 21. After both .RAW and .BED files were constructed, further population analysis could take place. In this case downstream analysis would be carried out using a programme written in "R". R was run in Durham University's HPC server for the purposes of this analysis.

### 4.3.5 Analysis: Processing Data in SambaR

To further the analysis of the mapped reads, output .RAW and .BED files were further processed through using a programme constructed under "R". The programme is called "SambaR" (Snp 115arl Management and Basic Analyses in R) and is available for open use with no requirement for a licence. SambaR functions to integrate numerous R packages such as "Adegenet", "poppr", "FactoMineR", et cetra. SambaR functions as "collections of functions which increase the power of existing R tools for population-genetic analyses" (De Jong et al., 2021). This convenience makes SambaR an ideal tool to save time with running population genetics analysis, which can include population structure, diversity, and selection analysis. SambaR were run in the "R" environment; therefore, the statistical tool R is essential for further downstream processing. For the analysis of the data within this thesis, R (version 4.1.3) were loaded from the HPC server into the environment. SambaR could be manually downloaded using from the official Github page (https://github.com/mennodejong1986/SambaR) and loaded into R manually. Alternatively, it could be loaded into R using the command "source("https://raw.githubusercontent.com/mennodejong1986/SambaR/master/SAMBAR_v1.07.txt")". This command executes a source code from the Github online server where the source code for SambaR can be found. Further downstream SambaR commands worked after the source codes were executed. R packages were able to be automatically checked and added through the "getpackages()" command. This command outputted a .txt file whereby each packages required could be checked as been successfully loaded or not. In the case of this analysis, all recommended packages were able to be loaded in addition to the essential packages to run the SambaR process. Data can be manipulated easily in SambaR using several pre-loaded R packages.

SambaR also accepts a geographical input file. This is useful for implications of the geographical range within our samples. From collection, the samples within the pooled library should have a latitudinal range across Western Europe. This provides latitudinal cline for structure analysis. SambaR uses the function 'getMap' of the R package rworldmap-1.3.6 (South, 2011). The addition of pie charts was included for mapping in SambaR, using the function 'add.pie' of the R package mapplots-1.5.1 (Gerritsen, 2018). An input .txt file, consisting for 3 tab-separated columns were created using the geographical information of the different population observed within the sample. The information for this file is available in appendix 22. As an addition to SambaR's "importdata()" command, the option "geofile=fileprefix" were included.

For different analysis, data were processed differently. For genetic structure analysis it is better to have as many individuals and populations as possible to infer better structuring between the populations in the collections. However, for population diversity and differentiation analysis, we determined that it was better to include as many SNPs as possible. Therefore, filters that enables to retain as many SNPs as possible were desirable during population diversity and population differentiation analysis. This may in turn reduce the number of individuals in the samples, thus, number of populations may be reduced under the filters for population diversity and population differentiation. For recommendations, SambaR recommends not using snpmiss=>0.05 for population structure analysis while for diversity and selection analysis, SambaR recommends higher snpmiss parameters.

### 4.3.6 Analysis: Data Management in SambaR

SNP data management and analyses were performed in R-4.1.2 (R Core Team, 2022) using wrapper functions of the R package SambaR (Github page: https://github.com/mennodejong1986/SambaR ). The data was then imported into R and stored in a Genlight object using the function 'read.PLINK' of the R package adegenet-2.1.5 (Jombart, 2008; Jombart and Ahmed, 2011). For the purposes of filtering, SambaR filter data per population and with much less individual/population, the data would be subject to stricter filtering procedure. The dataset for this study only consists of 115 individuals with 26 populations including eleven cultivars. Populations with only 1 individual will later be excluded from the population analysis. In the cultivar varieties, only one individual/population occurred and for the purposes of this analysis, they were put under one “cultivar (CUL)” population to see if any structural implication can be made against the wild *L. bienne*. In addition, having one individual per population is not recommended for population genetic analyses. Not only would it make the filtering/population impossible but also, for practical reasons, SambaR doesn’t allow populations which are only represented by one individual. This is because the mean difference between sequences (Nucleotide diversity) are estimated by mean sequence differences between individuals (Innan et al., 1999). For estimation of nucleotide diversity, population genetic analysis programmes often estimate by averaging the estimated number of nucleotide changes over all the samples (Innan et al., 1999). If a population only happen to have one individual, the mean difference is essentially only comparable to itself. For example, in heterozygous analyses, the nucleotide diversity would essentially just be the heterozygosity of the one individual as opposed to diversity amongst a population. As a result of this, cultivars would have to be merged and compared against the wild population. We included the cultivars to observe their genetic distance to the wild populations across Europe as a species. For initial data filtering option, all the populations were merged into one population using the “mergepop” command of SambaR. This was done after the data were imported. As a matter of downstream analysis, the output filtered data can be extracted into PLINK type format so that we can manually input the population prefix back in the dataset for further analyses.

After merging, the data was filtered using the function ‘filterdata’ of the R package SambaR. For population structure analysis purposes, the data were filtered with the following parameters: *indmiss=0.7, snpmiss=0.05, min_mac=2, dohefilter=TRUE, snpdepthfilter=TRUE, and min_spacing=500*. After the filtering options, 97 out of 115 individuals (97-97 per population) were retained. For genetic diversity and differentiation analysis, the data was filtered with *indmiss=0.7, snpmiss=0.2, min_mac=2, dohefilter=TRUE and min_spacing=500*. After filtering 94 out of 115 individuals (94-94 per population) were retained. After filtering 56328 out of 178847 SNPs were retained. Thinning (removal of missing data) reduced the dataset further to 2100 SNPs. The proportion of missing data per population can be summarised using Figure 2 below. Note that populations with proportion of missing data >0.07 were filtered out for genetic structure and diversity/differentiation analysis.

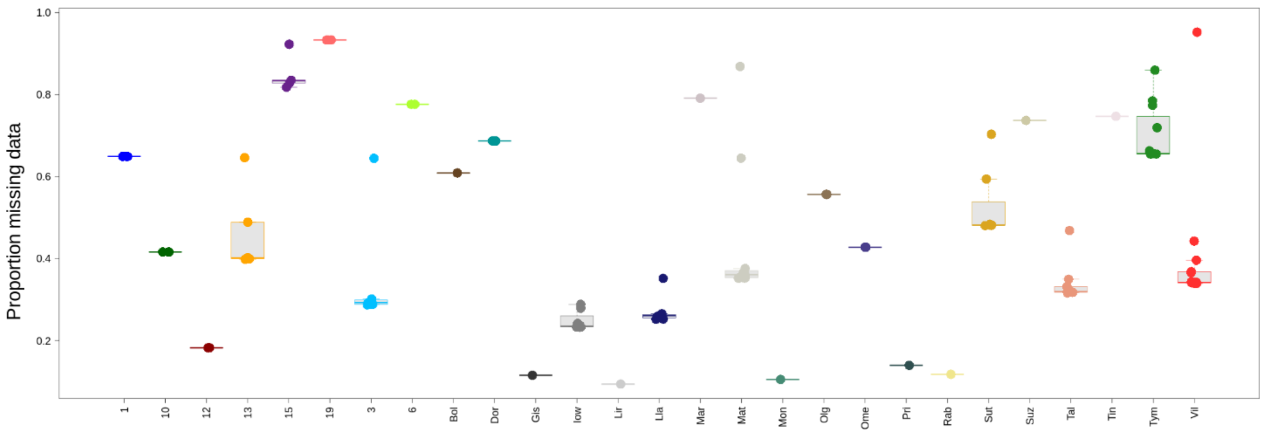


Figure 2. Boxplot to show the proportion of missing data after the filtering procedure over populations.

The "*dohefilter*" option in the "*filterdata()*" command refers to filtering of function which remove SNPs with heterozygosity levels which are potentially indicative of paralogs. Paralogs are genes that are present in a particular organism which are related to each other through gene duplication events (Koonin., 2005). Paralogous genes are often exacerbated in plant genetics because the events of gene duplications such as polyploidy is more prevalent in plants (Mastretta-Yanes et al., 2014). Paralogous genes have also been found to bias population genetic estimates, which will affect downstream population genetic type of analyses (Verdu et al., 2016). Since *Linum* was found to have undergone polyploidy, the identification of paralogs due to heterozygosity may be an issue for further population genetic type analysis (Sveinsson et al., 2014). The identification of paralogs was able to be inferred from locus specific heterozygosity against the locus specific minor allele frequency. The plot below (Figure 3) indicates these locus specific frequencies and reveals the amount of data which SambaR were able to filter out, due to the identification of paralogs: (The SNP dataset in red were subsequently removed after the "dohefilter" option was set to "TRUE").

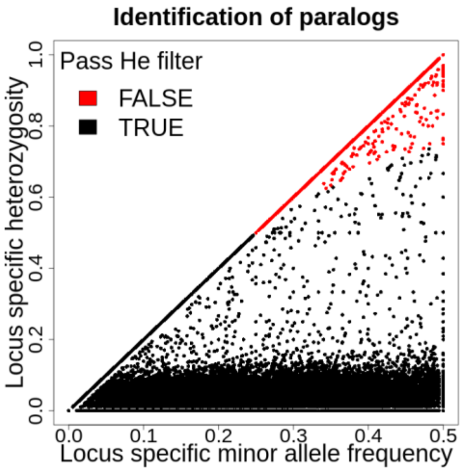


Figure 3. A scatterplot to reveal Sambar's identification and filtering of paralogs.

In this case, SambaR's default parameters for removing the paralogs were only removing a small proportion of heterozygosity (highlighted in red in Figure 3). This reveals that not all paralogs were potentially removed with SambaR's "*dohefilter*" option. This was perhaps due to SambaR's conservative use of Hardy Weinberg's Equilibrium (HWE). HWE makes assumptions that in the populations, there are: no mutation, random mating, no gene flow, infinite population size, and no selection. With the knowledge that *Linum* are highly selfing (see previous chapters), the "*dohefilter*" option of SambaR may not be optimised for removing paralogs in *Linum* sequences (Jahnke & Etterson., 2019). Heterozygousity needs to be manually addressed from the input files before processing in SambaR. SambaR's "*filterdata*" function filters individuals based on their proportion of missing datapoints considering all SNPs. As an addition to heterozygosity filter, SambaR were also instructed to filter SNPs based on SNPs depth ("*snpdepthfilter=TRUE"* command), This command filters SNP which has a high read depth. Subsequently, the function filters SNPs based on their proportion of missing datapoints considering retained individuals only. The filtering of individuals could return different estimates of missing data and, as a result, different numbers of retained SNPs and individuals occurs. In population structure filtering (i.e. more individuals and less SNPs), the GC-content of the retained dataset equalled to 0.53 and the 'transversion vs transition'-ratio equalled 0.59. For genetic diversity and selection analysis (i.e. less individuals and more SNPs), the GC-content of the retained dataset equalled 0.53 and the 'transversion vs transition'-ratio equalled 0.57. In the filtering for diversity and selection, linkage disequilibrium (LD) estimates were able to be calculated using *PLINK (-genome —r2 —ld-window-kb 1000000 —ld-window -r2 0)*. LD can infer trait-associated region(s) of a genome which may be of interest to further studies. However, for short-read sequences LD estimates may be redundant. This is because most of the reads do not cover all sites and consequently restricts LD to loci which are potentially very close (Maruki & Lynch., 2014) (Bilton et al., 2018).

## 4.3.7 Analysis: Genetic Analysis in SambaR

For the purposes of Structure analysis, several analyses took place within SambaR. The analysis is often based on *Nei's genetic distance* between populations (Nei, 1972). This includes Correspondence analyses (CA), Principal coordinate analyses (PcoA), Principal component analyses (PCA), DAPC analyses using principles such as Landscape and Ecological Association (LEA) tests (Frichot & Francois., 2015). Other genetic distance related analyses were also illustrated in SambaR. For Correspondence Analyses (CA), analysis was performed using the function "*dudi.coa*" of the R package ade4-1.7.19 (Dray and Dufour, 2007; Bougeard and Dray, 2018). Data was imputed per SNP/individual by calculating genotype probabilities from population specific minor allele frequencies. Principal coordinate analyses (PcoA) were performed using the function "*pcoa*" of the R package ape-5.6.2 (Paradis and Schliep, 2018). This occurred on distance matrices containing 3 different measures of genetic distance, with Nei's genetic distance, calculated with the function "*stamppNeisD*" of the R package StAMPP-1.6.3 (Pembleton et al., 2013), Hamming's genetic distance, calculated with the function "*bitwise.dist*" of the R package poppr-2.9.3 (Kamvar et al., 2014), and pi (pairwise sequence dissimilarity), calculated with the function "*calcpi*" of the R package SambaR. The principal component analyses (PCA) were performed using the function "*snpgdsPCA*" of the R package SNPRelate-1.28.0 (Zheng et al., 2012). DAPC analyses were performed using the function "*dapc*" of the R package adegenet-2.1.5 (Jombart, 2008; Jombart and Ahmed, 2011), both with and without prior population assignment. Multi-dimensional scaling (MDS) was performed using the function "*cmdscale*" (metric MDS) of the R package stats-4.1.3 (R Core Team, 2022) and the function "*isoMDS*" (non-metric MDS) of the R package MASS-7.3.57 (Venables and Ripley, 2002), on a Euclidean distance matrix generated with the function "*dist*" of the R package stats-4.1.3 (R Core Team, 2022). Neighbourhood joining (NJ) clustering was performed using the function "*NJ*" of the R package phangorn-2.8.1 (Schliep, 2011), using as input a Hamming's genetic distance matrix between individuals, calculated with the function "*bitwise.dist*" of the R package poppr-2.9.3 (Kamvar et al. 2014). Bayesian population assignment (BPA) probabilities were calculated and plotted using the functions 'assign2pop' and 'plotassign2pop' of the R package SambaR. The optimal number of clusters (K) was determined using the elbow method on cross-entropy scores generated by the 'snmf' function, with the assumption that the startpoint of a plateau represents the optimal K.

Divergence measures were measured using locus specific Fst estimates (according to Wright (1943), Nei (1977), and Cockerham and Weir (1987) (for all pairwise population comparisons) were subsequently calculated with the functions ‘runWrightFst’, ‘locusNeiFst’, and ‘locusWCFst’ of the R package SambaR. HWE, (2D) folded site frequency spectra (SFS), Tajima’s D and genome wide heterozygosity analyses were able to be executed using the function ‘calcdiversity’ of the R package SambaR.

Genome wide heterozygosity (genomeHe) was calculated in SambaR for each sample using the formula: *genome He = (He_seg * N_seg)/N_total*, in which: N_seg = the quantity of sites segregating within the population to which the tested individual belonged. He_seg = the proportion of Heterozygous sites within the investigated individual for those segregating sites. N_total = the total length of sequenced sites (polymorphic as well as monomorphic) which were able to pass the filter settings.

Geographical maps were generated with the function ‘getMap’ of the R package rworldmap-1.3.6 (South, 2011). Piecharts were added using the function ‘add.pie’ of the R package mapplots-1.5.1 (Gerritsen, 2018). Admixture coefficients were implied with the functions ‘snmf’ and ‘Q’ of the R package LEA-3.8.0 (Frichot and Francois, 2014). Alpha was set to 10, tolerance to 0.00001, and number of iterations to 200. Ancestry coefficients were calculated with the software Admixture-1.3 (Alexander et al., 2009) and illustrated using the ‘plotstructure’-function of SambaR.

### 4.3.8 Analysis: Ancestry Coefficients

As part of our population structure analysis, individual ancestry coefficients were able to be inferred using Landscape and Ecological Association (LEA) test. This was done using the package LEA under R (Frichot & Francois., 2015). These tests analysed population structure based on selection on a whole genome level. LEA applies landscape genomic data and identification of allele frequencies that illustrates genetic association with ecological associations. The LEA package derives adaptive alleles from large data sets, often referring to previous ecological association which is implied as ancestry coefficients (Frichot & Francois., 2015).

### 4.3.9 Analysis: Using different Genome reference

There are multiple whole genomes now publicly available for *L. usitatissimum* and none for *L. bienne*. For this study we mapped to the *L. usitatissimum* whole genomes. They utilized different sequencing tools and have different read depths. For the purposes of this analysis, these differences between the reference genome could translate into differences in *snpmiss* and *indmiss* parameters in the SambaR input. These differences translate into the number of SNPs and individuals able to be analysed. For the purposes of comparison, we used another *L. usitatissimum* (Atlant variant) genome sequenced by Nanopore and Illumina sequencing and contain another cultivar type when compared to the CDC Bethune genome used for previous analyses in this thesis. The cultivar type used for this whole-genome sequencing was found to have low variability of morphological and anatomical characteristics under stress conditions, suggesting a variant that is adept under stress conditions. After sequencing they found 8.4 Gb of sequence data with N50 of 12kbp and a read coverage of 23×, and Illumina read coverage of 30× (Dmitriev et al., 2021). The CDC Bethune (v2) contained a summary of 21.80 Gb HiFi reads generated with N50 of ≥12kbp (Sa et al., 2021). The difference in initial data quality can be summarised in the below data quality plot when compared against our short reads:

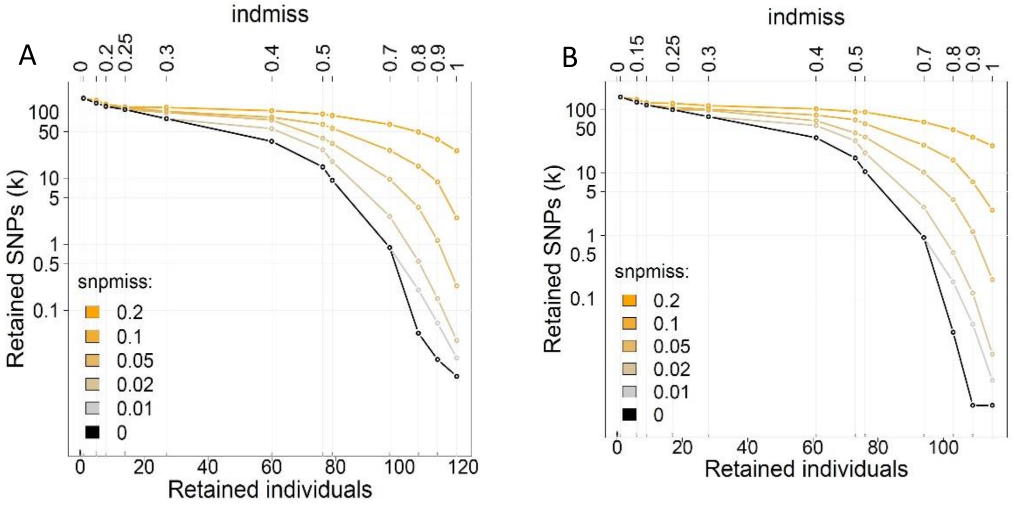


Figure 4. Sambar's data quality plots for the CDC Bethune genome (4A) and the Atlant genome (4B).

The plot above illustrates the data quality when short reads are aligned to either the CDC Bethune genome (Figure 4A) or the Atlant genome (Figure 4B). This quality plot revealed that retained individuals for CDC Bethune genome was around the same as the Atlant whole genome (120 individuals). A *snpmiss* parameter of 0.05 and *indmiss* parameter of 0.7 for example, will return ~100 individuals with ~10k retained SNPs (before thinning). When aligning to the Atlant genome, this number was reduced by 10. This reveals that at least for these two different genomes (both different in terms of sequencing methods and cultivar type), that alignment of the short read sequences to different reference genome have some effects in terms of the quality of retained SNPs and retained individuals after filtering in SambaR. The CDC Bethune genome reveals a higher preference based on the number of individual samples kept.

One of the most important qualities when looking at the quality after filtering, is the amount of heterozygosity in the data after filtering. For this, two plots can be shown to illustrate linear relationship between the proportion of Heterozygosity in all sites and segregating sites. Using the same filtering parameters, two plots were illustrated with to compare between the different reference the short reads were aligned to.

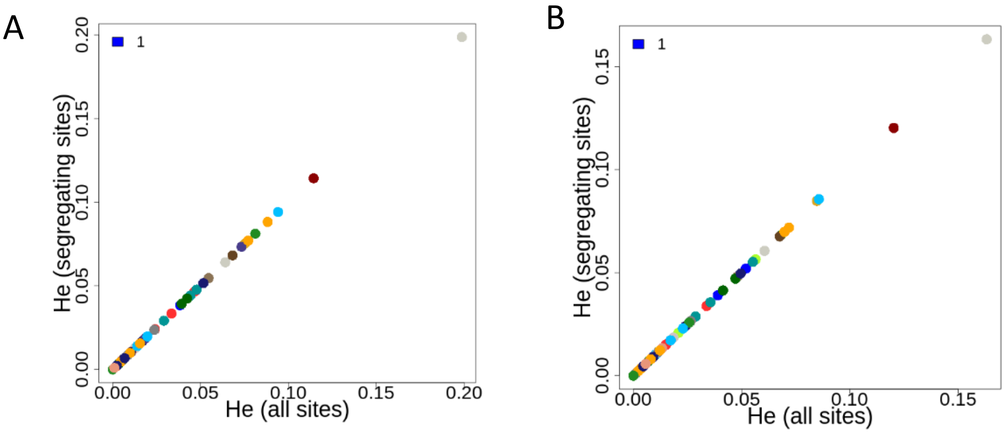


Figure 5. Plots for heterozygosity after filtering short reads with the CDC Bethune genome (5A) and the Atlant genome (5B). Each dots illustrates different individuals differentiated by population (colours).

Figure 5A is the linear plot for Heterozygosity (He) proportion on all sites and segregating sites for the data aligned to the CDC Bethune reference genome. The Figure 5B is the same plot for the data aligned to the Atlant reference genome. This revealed that aligning the short reads to either genome will result in significant amounts of Heterozygosity. There were differences in the rankings of the individuals within the population, noted by the differences in colour labelling. This reveals that proportion of heterozygosity remains mostly the same even using different reference genome. This potentially convey those heterozygous reads in our short reads are real heterozygous reads or perhaps reads in the genome due to the polyploidic nature of *Linum*. However, it is interesting how the rankings of these heterozygosity have changed within populations in our short reads. For the purposes of structure, diversity and differentiation analysis in the populations sampled here, we aligned to the more commonly used CDC Bethune genome.

## 4.3.10 Analysis: Removing Heterozygosity

As was seen in Figure 5A, all sites heterozygosity ranges from 0.000 to 0.20 within the samples and under different reference genomes. This was potentially an issue as the presence of heterozygosity might present an issue with paralogs which is especially of high importance to organisms which have previously underwent whole-genome duplication events such as polyploidy (McKinney et al., 2016). This is also revealed in the “locus specific heterozygosity” in the plot used for identification of paralogs in SambaR’s “dohefilter” function. There were potential paralogs due to these heterozygousities. SambaR’s filtering option may not remove all paralogs (Figure 6). In addition, we expected *Linum* to be highly selfing and polyploidy. As a result of this, explored options to filter for heterozygosity.

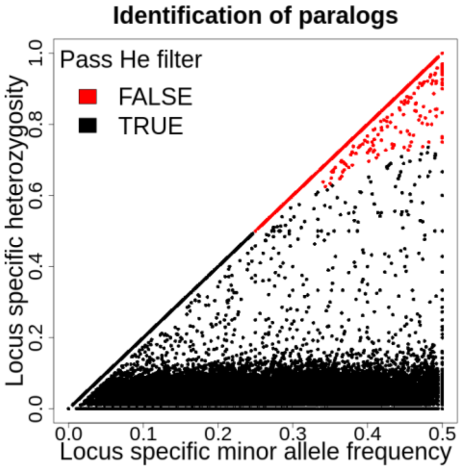


Figure 6. Scatterplots to reveal Sambar’s identification and filtering of paralogs for the default “dohefilter” option, revealing potential paralogs not filtered by SambaR (in black).

An option to avoid paralogs is to remove heterozygous alleles altogether, however the consequence of removing 100% heterozygosity is that further analysis such as the population diversity analysis will be made redundant due to the lack of heterozygous alleles (Chapman et al., 2009). In SambaR this will return an error whereby SambaR will specify that all samples are homozygous and therefore a population diversity measure is made redundant. Measures of heterozygosity, such as multi-locus heterozygosity (MLH) is useful for predicting whether populations are inbreeding or outcrossing in terms of their breeding strategies (Jensen et al., 2007). It is then preferable if the heterozygosity in the data were not entirely removed. To do this there is a requirement to specifically filter out a certain number of heterozygous reads. This can be done by removing SNPs and/or loci locations which have a higher than the threshold amount of heterozygosity, manually.

### 4.3.11 Analysis: Filtering for Heterozygous Alleles

To avoid the redundancy of a population diversity analysis, a specific genotype filter measure to exclude locations with exceptionally higher heterozygosity were applied to the original population output of STACKS. To do this, the original VCF output file were first converted from muti-allelic to biallelic genotype/allele reads. This was done using the function “*bcftools norm -m*” under bcftools V1.10. Formatting of alternative alleles were done using the command “*bcftools view -e “FORMAT/AD[:1]<2 && INFO/AD[1]<5”*” and extraction of genotype per sample is done by using the command “*bcftools query -f ‘%CHROM %POS[\t%GT]\n’*”. Formatting of alleles and extraction of genotypes were done in bcftools V1.10 (Li., 2011). Once genotyped and formatted to biallelic reads, the file were processed in R using genotyping to mark heterozygous allele as “1” and homozygous allele as “0”. Once heterozygous allele is marked as “1”, a filter was applied whereby loci location with more than extreme heterozygosity can be identified and outputted. We decided to filter only for the top 1% heterozygous alleles to avoid missing important information from these potentially true heterozygous alleles. Filtering for heterozygousity can be done using the addition of “which(meanps[,q]>quantile(meanps,0.99))” to the initial allele output in R. This will remove top 1% of loci with excessive “1s” or heterozygous alleles. R was then asked to output a *“.txt*” file whereby it listed the loci position which can then be filtered out in vcftools. The full command lines for this process can be found under appendix 23.

We used *vcftools’ “--exclude-positions”* command in Vcftools for filtering against the output heterozygousity file (Danecek et al., 2011). First, the “.txt” file was made sure to be tab-delimited and was converted into a “Unix .txt” format using the Linux command “*awk ‘{ sub(“\r$”, “”); print }’ winfileinput.txt > unixfileoutput.txt*”. These “*.txt*” formatting options were essential so that vcftools could read the loci positions without errors under *Linux* commands. In the same command, the file can be recoded using the *“—recode*” command of vcftools and can be pipelined to output a .VCF file output whereby exceptionally heterozygous loci location was filtered out. The full command for this process can be found under appendix 24. This would keep out loci with exceptionally high heterozygosity which may represent paralogous alleles while preserving an amount of heterozygosity which potentially represent outcrossing populations within our sample sets.

### 4.3.12 De_novo mapping

Mapping to L. usitatissimum was preferred in this analysis, to conserve the number of individuals and populations able to be analysed. However, using L. usitatissimum whole genome as a reference for population samples of L. bienne may not be optimal for SNP calling during further bioinformatic analysis, this may have implication on retained SNPs. We investigated how mapping “de novo” will impact the results gathered from mapping to the L. usitatissimum whole genome. After filtering for missing SNPs and individuals in SambaR, we found only 68 individuals were viable for further analysis when mapped de novo. After filtering and thinning, we found a total of 3029 kept SNPs. The number of individuals kept for de novo mapping were significantly less than that from the option of using L. usitatissimum whole genome as a reference, where we kept 100 individuals. We also lost potentially significant results because population 12 (Mediterranean population) were not kept from de novo mapping. Despite the loss of individuals using de novo mapping, we kept a significantly higher number of SNPs than mapping to the L. usitatissimum whole genome (kept only 800 SNPs with L. usitatissimum reference). This suggests that potential SNPs may have been lost due to mapping to L. usitatissimum whole genome. However, in the interest of keeping more individuals for population structure inference, we mapped to the L. usitatissimum reference. For future repeats of this population analysis, we suggest the use of L. bienne whole genome when they become publically available and/or including many more individual/population samples in the ddRADSeq protocol before mapping de novo. This may result in increases in the number of individuals retained after de novo mapping. We also suggest using more population from the Mediterranean region of Europe to observe for more variation from this region.

## 4.4 RESULTS: IMPLICATIONS OF POPULATION ANALYSIS

The final heterozygous-filtered data contained 866 SNPs (after filtering and thinning). After the filtering procedure, SambaR retained 100 individuals that makes up to 15 populations with retained grouping as Northern (> 45° North) and Southern (≤45° South) (Table 2), as described in Landoni et al (2022). We expected to see structuring between Northern and Southern populations, supporting those reciprocal responses to traits seen in Landoni et al (2022).

| Initial Population Name | Latitude | Group | Final Population Prefix |
|---|---|---|---|
| **Cul** | Not Applicable | Cultivars | CUL |
| **3** | 36.03633 | Southern | A-3 |
| **1** | 36.80044 | Southern | B-1 |
| **10** | 37.88211 | Southern | C-10 |
| **6** | 37.93551 | Southern | D-6 |
| **12** | 42.31008 | Southern | E-12 |
| **13** | 43.02902 | Southern | F-13 |
| **Lla** | 43.40738 | Southern | G-Lla |
| **Vil** | 45.09393 | Northern | H-Vil |
| **Tal** | 47.6997 | Northern | I-Tal |
| **Mat** | 48.35697 | Northern | J-Mat |
| **Dor** | 50.6 | Northern | K-Dor |
| **Iow2** | 50.68183 | Northern | L-Iow2 |
| **Tym** | 53.30307 | Northern | M-Tym |
| **Sut** | 53.35291 | Northern | N-Sut |

Table 2. List of output populations and their details of locality after the final filtering and thinning options of SambaR. Populations which are found ≤45° South were grouped as a more Southern population than those >45° in latitude. Apart from the cultivars (CUL), the populations were alphabetically ordered by latitudes with the most Southern population first. Populations were specified and grouped as per previous reciprocal results under Landoni et al (Landoni et al., 2022).

In the following results the prefix of the population would be in the format of “Latitude (in ascending order)-Population”. For example, for individuals from the most Southern population 3 the label would be “A-3” and the most Northern population Sut the labels would be “N-Sut”. For Figures depicting individual comparison in the sample (such as genetic distance trees), the data will be formatted by “population_individual”, for example individual 10 belonging to the population 12 will be formatted as “12_10”. The full individual and population details is available in appendix 25.

Our samples represent a wide range of latitude within Western Europe, representing Northern and Southern groups as was observed in Figure 7. In this results section we analysed the *Linum* samples as stated above in terms of their genetic diversity, structure, and divergence.

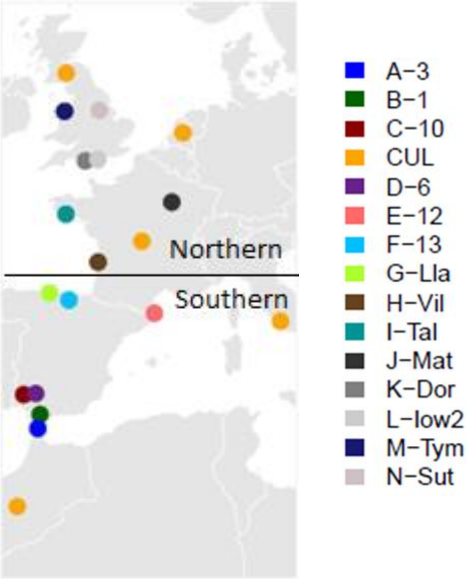


Figure 7. A colour map revealing the geography of the samples after passing filters and thinning with reference to their colours, population, and regions on the labels. For our analyses, Southern population are the wild Spanish population and Northern population are wild population found in France and the UK. This grouping aligns with previous reciprocal study groupings (Landoni et al., 2022)

## 4.4.1 Results: Genetic Diversity

### 4.4.1 i Genome-wide Diversity

We aimed to see population differentiation as well as to observe potential differences in breeding strategies. We looked at genome wide diversity to observe these. We expected to see a difference in diversity between *Northern* (> 45° North) and *Southern* (≤45° South) wild populations in our samples.

The bar plot reveals a lower proportion of segregating sites in most of the Northern populations (Figure 8A). French populations such as “I-Tal” and “H-Vil” seems to reveal the lowest proportion of segregating sites (0.0462 and 0.1039) along with the Northern UK populations of “L-Iow2” and “M-Tym” (0.0727 and 0.1016). In exception of this is the French population of “J-Mat”, which seems to have a high proportion of segregating sites. In contrast to this low segregating site proportion in most of the Northern populations, the more Southern populations have a higher proportion of segregating sites (>0.2). The highest proportion of segregating sites is observed in the

Mediterranean population “E-12” (0.3915). These observations may suggest that the Southern populations are more likely to contain genes which are not conserved and are therefore potentially more outcrossing in terms of their breeding strategy when compared to their Northern relatives.

In addition to this, the proportion of heterozygous sites can also suggest potential diversity within populations. When looking at heterozygosity within segregating sites only (Figure 8B), there wasn’t much of a pattern to be observed. The proportion of heterozygote alleles for all sites, however, suggests that most of the latter, Northern population are less heterozygous. This suggests that perhaps the Southern populations are employing a strategically different breeding system. We expected the level of heterozygousity to not differentiate as much between the Southern and Northern population should they all be “selfing” populations. It may be the case that the more Southern populations are more outcrossing in breeding strategy than those of the more Northern populations in our sample lists.

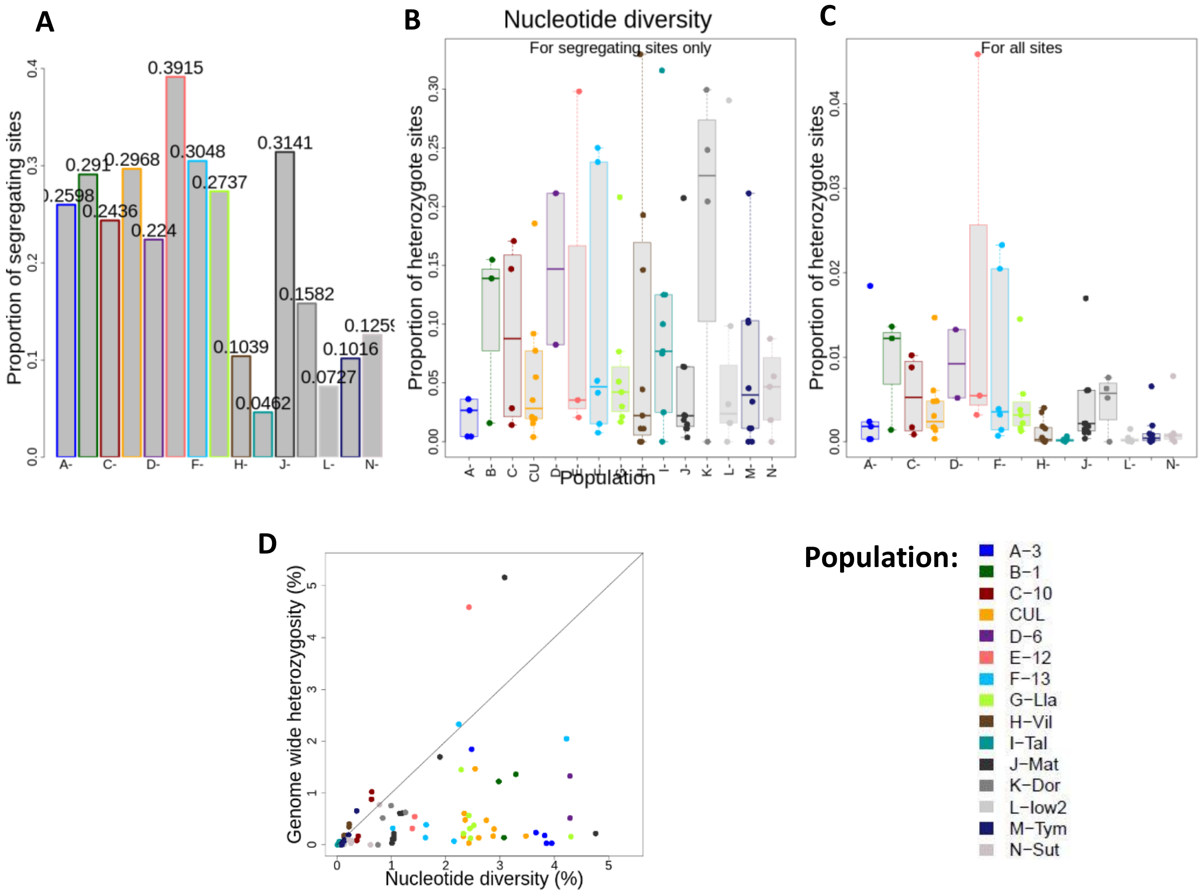


Figure 8. Summaries of nucleotide diversity measures for our *Linum* samples. 8A. Bar chart to show the proportion of segregating sites within each population levels in our samples based on regions. 8B. Box plots to show the proportion of heterozygous alleles within the segregating sites, within each population in our

samples based on regions. 8C. Box plots to show the proportion of heterozygous alleles for all sites within the genome. 8D. A scatterplot to show the nucleotide diversity or pi (%) against the genome wide heterozygosity (%) for all individuals colour coded by population. The colours on every part of this Figure corresponds to population list as inferred in the Figure.

### 4.4.1ii Multi-Locus Heterozygosity

A measure of multi-locus heterozygosity can also be observed with a boxplot (Figure 9). In the box plot, an observation of multi-locus heterozygosity was illustrated. In the x-axis, the populations were lined alphabetically according to their latitude, with the most Southern populations to the left. Southern populations such as population B-1, C-10, E-12, F-13, and D-6 were seen to have a more diverse multi-locus heterozygosity, agreeing with the sequence heterozygosity as observed in previous section. When looking at the Northern population, the box plots are forming closer to 0. This infers less multi-locus heterozygosity in the more Northern individuals. Interestingly, population 3, which is a Southern population region-wise, is trending more with the Northern population and the Northern population “K-Dor” seems to have a more diverse multi-locus heterozygosity. Less multi-locus heterozygosity in a population is thought to illustrate more inbreeding within that population and less outcrossing. It could be that the breeding strategy of population 3 is genetically influenced by the more Northern population and are therefore more inbred than the rest of the Southern population and the opposite is true for some Northern populations such as “K-Dor”.

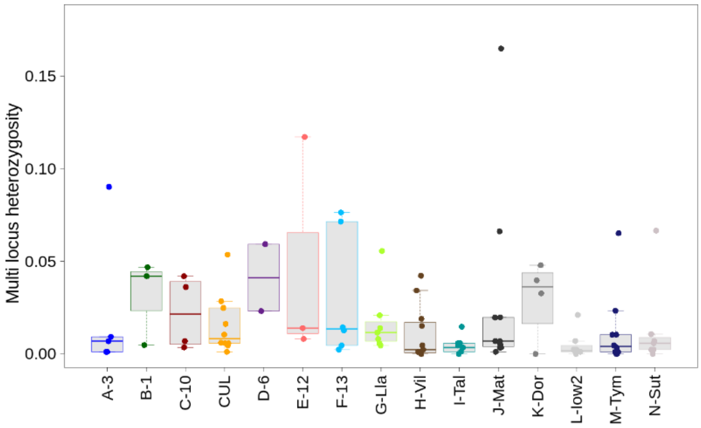


Figure 9. A boxplot to reveal multi-locus heterozygosity at a population level.

### 4.4.1iii Numberof private alleles

A measure of the number of private alleles can suggest population divergence as higher mutation rate may be implied (Szpiech & Rosenberg, 2011). We reveal a higher number of private alleles for our cultivars than most of the wild populations observed within this study except population E-12 (Figure 10).

Our analysis suggests that most of our wild population have around 228 – 246 out of 866 SNPs which are private alleles. Our cultivars have more, at 269 SNPs showing as private alleles. The Mediterannean population E-12 suggests an even higher number of SNPs showing private alleles at 374 SNPs. This suggests more discinction in this population, which suggests population divergence for the Mediterranean population.

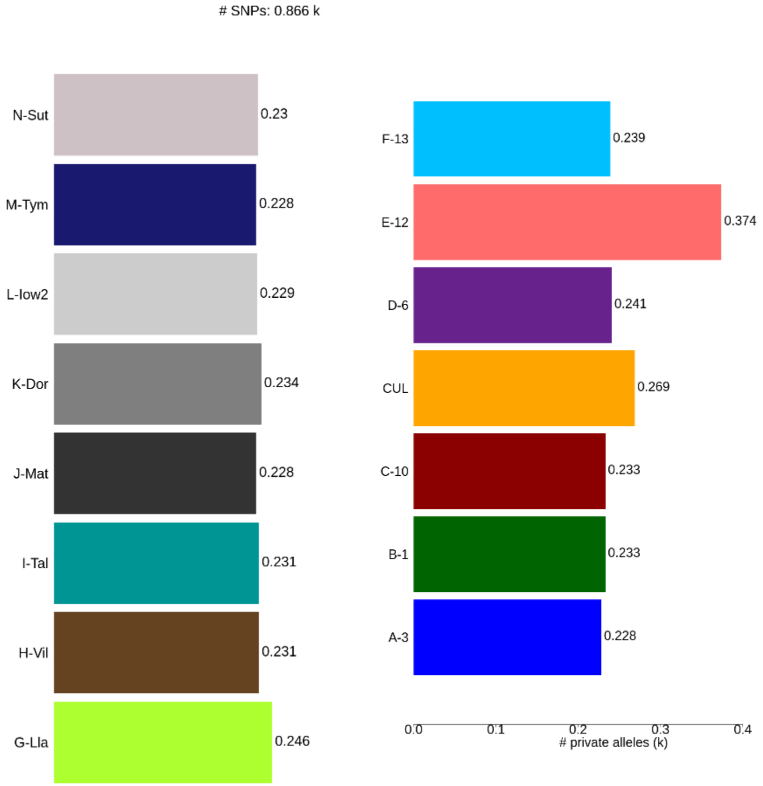


Figure 10. Bar chart to illustrate the number of private alleles over population. # SNPs = 0.866k, number of private allels are in thousands (k).

### 4.4.1iv Variation in Segregating Sites (Tajima’s D)

One way to look at population divergence is to examine variation within segregating sites for the population. This will infer whether selection have occurred which have removed variation or there is potential selection in which variation is maintained within population. To look at this, we can look at the populations and how many sites are variable and how identical individuals are within the population. Tajima's D values can be interpreted for our populations, whereby a measure of pi and pi related to the number of variable sites relative to the number of sequences can be normalized as Tajima's D (Korneliussen T. et al., 2013). Negative Tajima's D will suggest that there is selection removing variation within population and population is recently expanded. Positive Tajima's D will suggest that there is selection maintaining variation and populations are not expanding.

Tajima's D can be illustrated through a barchart for every population within our samples (Figure 11). It is with confidence that none of the population observed within our samples results in positive Tajima's D. This reveals that there are selection removing variation within all *Linum* samples in this study. This was however more observed in two populations (populations 12 and 6). There is a suggestion here that based on the more negative Tajima's D, Southern populations are more recently expanded than the Northern populations. In addition to this, summary of Tajima's D estimation was able to be obtained from SambaR (see appendix 28) and a subsequent t-test were able to be implied between the North and South wild populations. The t-test revealed that Tajima's D estimation for Northern and Southern populations were significantly different *(P(T<=t) two-tail = 0.023 (<0.05))*.

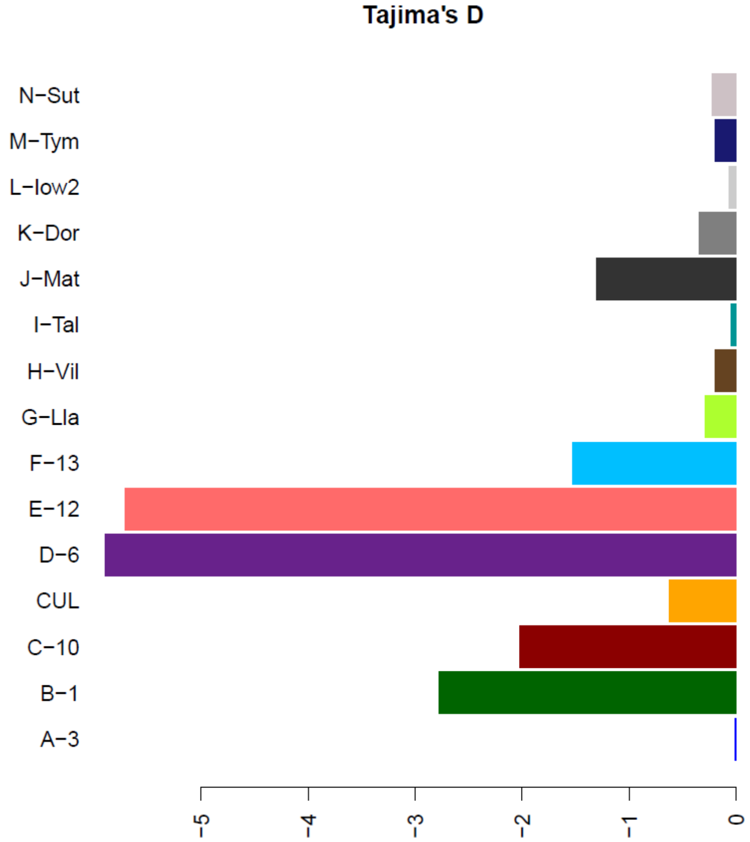


Figure 11. A bar chart to show Tajima's D statistics for the different populations.

## 4.4.2 Results: Genetic Structure

When the data was fully prepared with potential paralogs removed (top 1% most heterozygous loci removed), it is with further confidence that we can carry out structure analysis without paralogous alleles. To reveal population structure, we constructed a Ward D neighbour joining tree. When genotyped heterozygous filters were applied, the tree suggests genetic structuring between the different populations (Figure 12).

The Ward.D neighbour joining tree reveals that there are three main population clusters. The tree reveals that most of the cultivars are grouped together (Figure 12). The Southern population were grouped further away from the Northern and Cultivar individuals in our sample set. In terms of the wild, there is an observation of two groups whereby Southern population were grouping as genetically closer to each other and the Northern population grouping on together. This reveals that, genetically, in terms of their population the two Southern and Northern populations are potentially genetically distanced to each other. The Southern group are also more genetically distanced to the cultivars and may have a higher allelic richness. The cultivar “Suz” grouped with the Northern cluster. In this cluster, there also occurred some Southern individuals such as the Southern population “3”. This could be due to unexpected history of migration in population A-3. We suspect that this is due to human errors such as seed labelling and labelling in the laboratory that may have led to these individuals not grouping to their clusters as expected. Future studies could examine structuring between more individuals from these population that may infer migration or hybridisation in population A-3.

In other observations, the population highlighted in red circle are the only “Mediterranean” population observed within our samples (population E-12). On the heterozygous-filtered, unrooted tree, this group seems to be structurally different to the rest of the samples. They do not group with either Northern or Southern wild populations. This Mediterranean population seems to have a unique genetic structuring to the rest of the wild groups as well as the cultivars observed in this study. This reveals that in the samples contained within our RADSeq analysis, population E-12 is potentially unique to the rest of the wild populations found in the West, which are closely grouped to either the North or South cline of the population.

In summary, the genetic network reveals that within the Western Europe wild *Linum* poulations, there were observations of genetic structuring, generally between populations originating in the North and South of Western Europe. There are some individuals that are out of place from their expected cluster, which could indicate potential gene flow or human errors within the data. The discovery of population E-12 being structurally unique to either North or South populations is an indication that the more Eastern, Mediterranean populations are structurally different to the rest of the wild populations in the West.

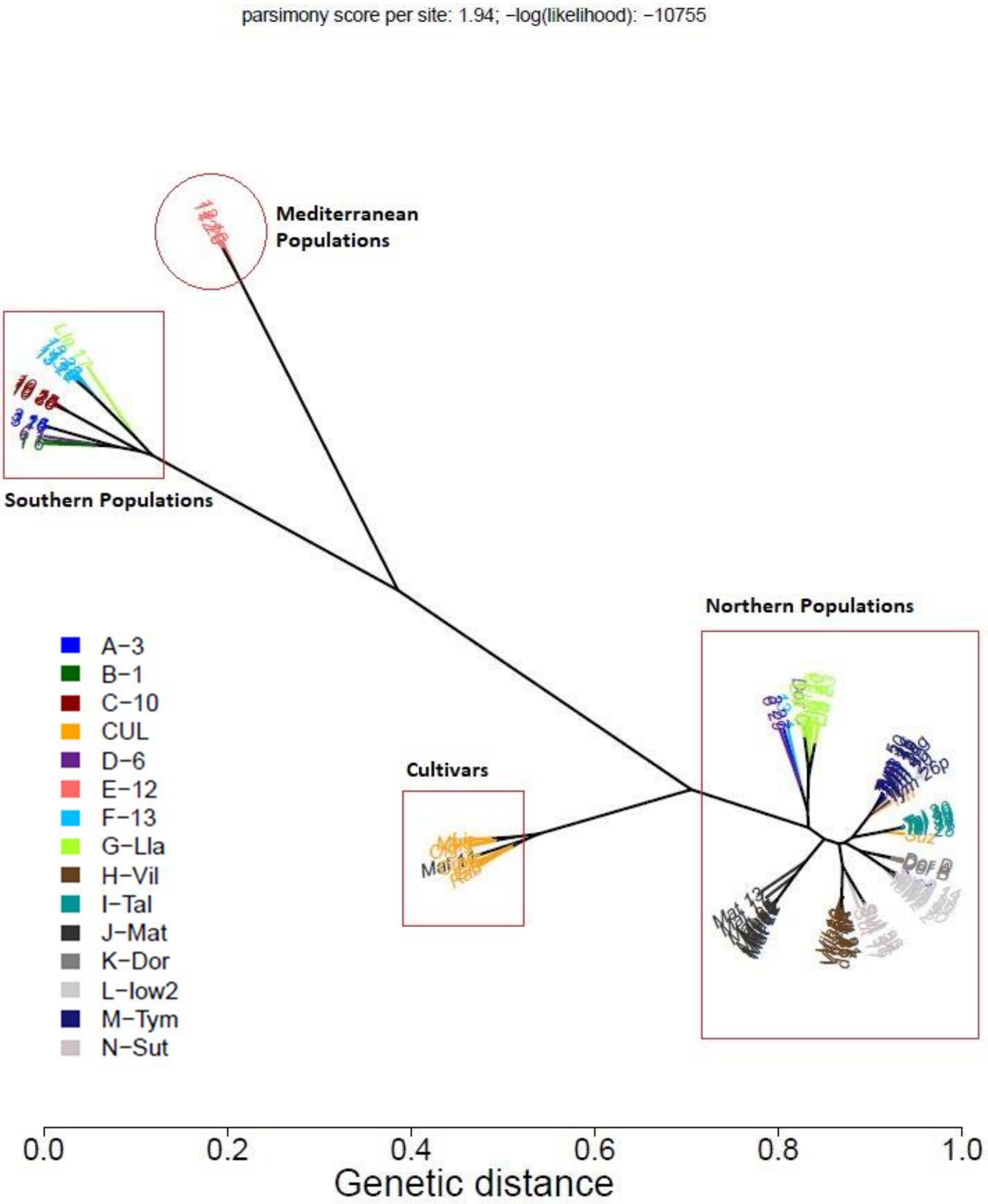


Figure 12. A Ward.D (Ward's minimum variance method) Neighbour-joining tree for the illustration of genetic distance on an individual level. The genetic tree reveals structuring in our samples with Northern and Southern main clines for the wild. Cultivars are revealed to be more closely related to the Northern cline.

### 4.4.2i Principle Coordinate Analysis (PcoA)

When looking at genetic structure, dissimilarity between the individuals based on Nei's genetic distance can also be observed. A principal coordinate analysis (PcoA) was conducted, with PC1 (representing 46.7% variance) and PC2 (representing 29% variance) plotted to examine genetic similarity (Figure 12). Another multivariate statistical technique that will go hand to hand with the PcoA analysis is the more descriptive Correspondence analysis (CA), whereby another plot can be illustrated in appendix 26. There are observations here of clusters for both PcoA and CA analysis, highlighted in red boxes. Southern and Northern populations are mostly clustering with what seem like a cline between the two clusters highlighted (Figure 13 and Appendix 26). In addition, population E-12 seem to be clustering separately in what we have suggested as "Mediterranean population".

Some Southern individuals from populations such as F-13 and A-3 can be observed clustering closer to the Northern cluster, suggesting potential gene flow between Northern and Southern populations. Some Northern populations such as the French population "Mat", can also be observed to have an individual outside of the cluster, closer to the Southern populations. The Spanish population of "Lla" are clustering with their relatives in the North. This suggests that this population may have implication of gene flow through past genetic events.

Another point of interest is the separation of the population E-12 which we suggest as being the "Mediterranean population". In this analysis, this was the only "Mediterranean population" able to be analysed due to DNA material availability. They seem to be structuring differently to both Northern and Southern groups of the Western European *L. bienne* samples in this study (Figure 13 and Appendix 26).

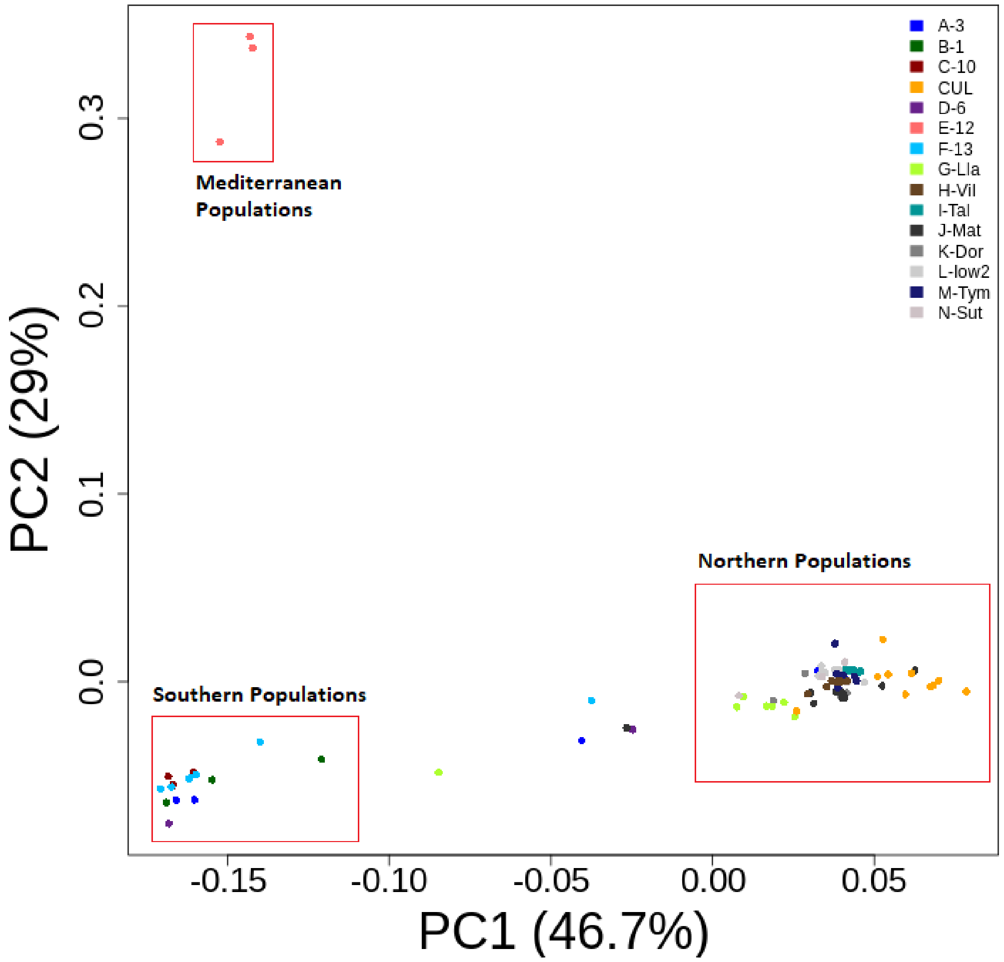


Figure 13. A PcoA scatter plot with PC1=47.3% of total variance and PC2=34.9% of total variance. The red boxes highlight potential clustering of populations. The plot agrees with the structuring suggested by the genetic distance tree, in addition, a few intermediate individuals between the Northern and Southern populations were also suggested, forming a potential cline between the Southern and Northern population.

### 4.4.2ii Sequence Dissimilarity

One way to illustrate genetic distance is to look at the proportion of dissimilarities between individual sequences or otherwise known as pi. This is  the average pairwise difference between individuals. A sequence dissimilarity measure (pi) was illustrated as a matrix to illustrate dissimilarities between individuals and within their sequences, with Southern populations showing more dissimilarities (Figure 14).

In terms of genetic structuring, there is an observation that most of the Southern populations are showing a degree of sequence dissimilarities when compared to the Northern populations (Figure 14). The sequence dissimilarity measures agree with the finding observed with the genetic tree, that the Southern and Northern populations are genetically distanced. Most of the Northern individuals are showing little signs of dissimilarity between themselves. For example, sequence dissimilarity between individuals in the Northern population “N-Sut” and “I-Tal” is close to 0. Dissimilarities are observed to occur on the more Southern population with our *Linum* sample set (Figure 14).

The sequence dissimilarity (pi) matrix (Figure 14) also reveals that population E-12 (pink strip on the matrix) has the highest level of sequence dissimilarity (pi) comparatively between all the other individuals analysed in this study. This is highlighted by the more intense shading of yellow when comparing individuals belonging to population E-12 to the rest of the individuals in the sample set (revealing pi values between 0.385 to 0.44). What’s more interesting is that this dissimilarity is even more than the observed difference for cultivated and wild individuals between in our samples, when genotyping for highly heterozygous genes were implied to the data.

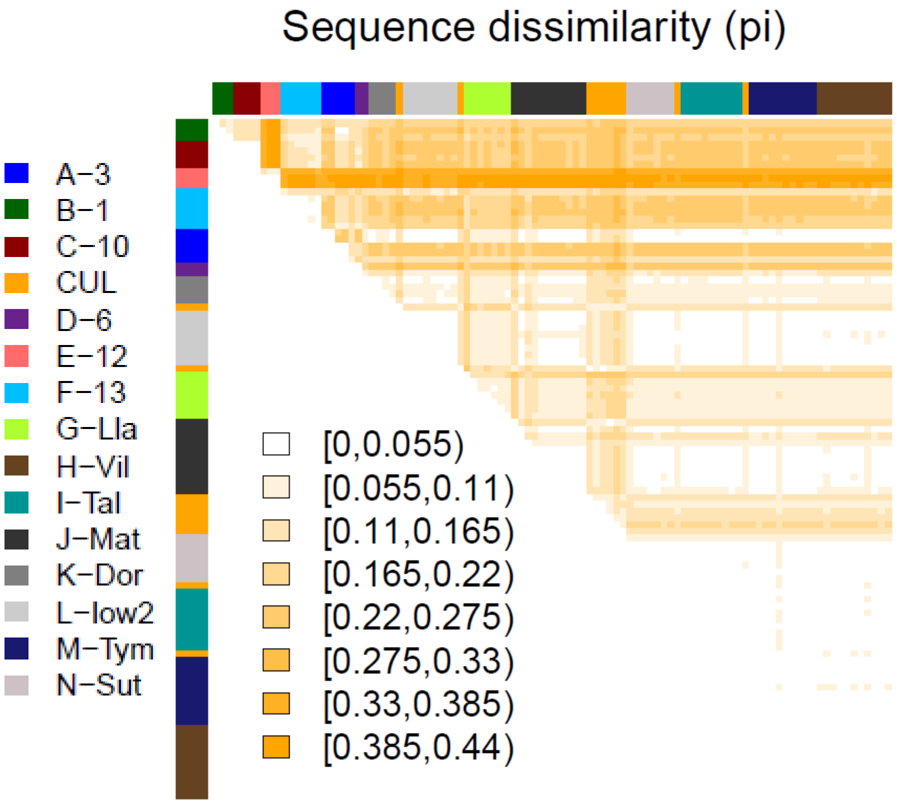
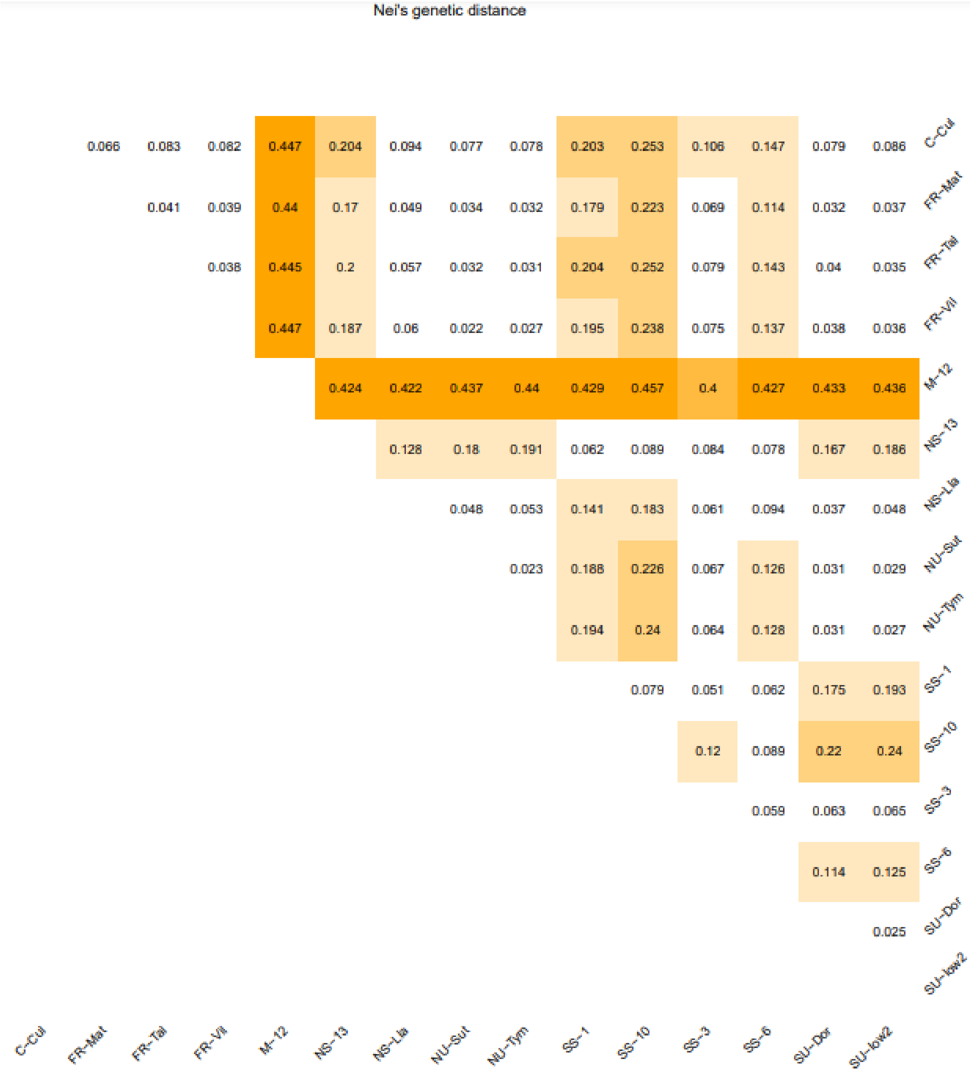


Figure 14. A sequence dissimilarity (pi) matrix for individuals summarised into populations. The colour chart on the left indicates the population each colour strip is representing in the matrix.

Similar observations were also seen in Nei's genetic distance. The higher the Nei's genetic distance is the more distinct the sequence is when compared to the rest of the samples. A matrix can be illustrated to observe Nei's genetic distance (Figure 15). One major observation in this matrix is the higher number of Nei's genetic distance highlighted in shades of orange for the Southern population (B-1, C-10, E-12, F-13, A-3 and D-6) when compared to the cultivars. This number was observed to be lower for the Northern individuals. This futher supports the findings that cultivars are more genetically related with our Northern wild population as seen in our previous analyses. Population A-3 is an exception from this trend seen with the Northern and Southern population. This is perhaps a signal of genetic flow with the Southern population. Population A-3 is perhaps more genetically derived from the more Northern population observed in our sample set, making them more of a "Northern" population in terms of their genetic make-ups. A significantly higher Nei's genetic distance was also observed with population "E-12" when compared with the rest of the samples (>0.4), which can be inferred as Nei's genetic distance revealing more sequence distinction on population 12 when compared with other populations. To put to perspective, the next highest

Nei's genetic distance would be at 0.201 between the cultivated population and population 13. In the literature, a Nei's genetic distance of more than 0.250 is considered as significant genetic distance between populations observed (Nei M., 1972; Wright S, 1978). This finding suggests that the Mediterranean population E-12 is considered as genetically more distanced to the rest of the samples observed within this study.


Nei's genetic distance

| | C-Cul | FR-Mat | FR-Tal | FR-Vil | M-12 | NS-13 | NS-Lla | NU-Sut | NU-Tym | SS-1 | SS-10 | SS-3 | SS-6 | SU-Dor | SU-low2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C-Cul | | 0.066 | 0.083 | 0.082 | 0.447 | 0.204 | 0.094 | 0.077 | 0.078 | 0.203 | 0.253 | 0.106 | 0.147 | 0.079 | 0.086 |
| FR-Mat | | | 0.041 | 0.039 | 0.44 | 0.17 | 0.049 | 0.034 | 0.032 | 0.179 | 0.223 | 0.069 | 0.114 | 0.032 | 0.037 |
| FR-Tal | | | | 0.038 | 0.445 | 0.2 | 0.057 | 0.032 | 0.031 | 0.204 | 0.252 | 0.079 | 0.143 | 0.04 | 0.035 |
| FR-Vil | | | | | 0.447 | 0.187 | 0.06 | 0.022 | 0.027 | 0.195 | 0.238 | 0.075 | 0.137 | 0.038 | 0.036 |
| M-12 | | | | | | 0.424 | 0.422 | 0.437 | 0.44 | 0.429 | 0.457 | 0.4 | 0.427 | 0.433 | 0.436 |
| NS-13 | | | | | | | 0.128 | 0.18 | 0.191 | 0.062 | 0.089 | 0.084 | 0.078 | 0.167 | 0.186 |
| NS-Lla | | | | | | | | 0.048 | 0.053 | 0.141 | 0.183 | 0.061 | 0.094 | 0.037 | 0.048 |
| NU-Sut | | | | | | | | | 0.023 | 0.188 | 0.226 | 0.067 | 0.126 | 0.031 | 0.029 |
| NU-Tym | | | | | | | | | | 0.194 | 0.24 | 0.064 | 0.128 | 0.031 | 0.027 |
| SS-1 | | | | | | | | | | | 0.079 | 0.051 | 0.062 | 0.175 | 0.193 |
| SS-10 | | | | | | | | | | | | 0.12 | 0.089 | 0.22 | 0.24 |
| SS-3 | | | | | | | | | | | | | 0.059 | 0.063 | 0.065 |
| SS-6 | | | | | | | | | | | | | | 0.114 | 0.125 |
| SU-Dor | | | | | | | | | | | | | | | 0.025 |
| SU-low2 | | | | | | | | | | | | | | | |

Figure 15. A Nei's genetic matrix between the populations observed in this sample. The matrix reveals a higher genetic distance between the Northern and Southern populations. Darker orange shades reveals a more significant genetic differences.

## 4.4.2iii DAPC Analysis

As an addition, we used a multivariate method to illustrate population structure. One method is the Discriminant Analysis of Principal Component (DAPC) (Miller et al., 2020). The DAPC is calculated in SambaR by scoring the number of retained Principle Components (PCs), interpolation of this scoring and scoring cumulative variance explained by the Principle Component Analysis (PCA)

(De jong M. et al., 2021). The summary DAPC results for our genotype filtered data can be found under Appendix 27. A geographical map can be inferred using DAPC whereby population clustering can be illustrated under a population based on their latitude and longitude of origin (Figure 16).

SambaR outputs pie charts on a map based on our population's locality to illustrate DAPC scoring (Figure 16). The Figure is separated by the number of clustering per population or otherwise known as 'K'. The cultivar individual "Mar" was excluded from this analysis as this was an exclusive Canadian population in the sample and we are interested only on Western European Latitudes for this study. From the Figure, we can observe groupings based on a DAPC analysis. When *K=2* there was already some grouping of the Southern population. However, the cultivars are not distinctive yet. When *K=3*, the population 'E-12', highlighted in red became distinct. When *K=4* populations showed further grouping whereby the four groups (Northern, Southern, Cultivars, Mediterranean) were observed as can be implied by the genetic distance tree. Here, the cultivars are seen as grouping in yellow, Northern populations in 'colourless', Southern population in brown and the population 'E-12' in red. Higher K's can also be inferred whereby some of the individuals from Southern populations are grouping closer to the Northern populations as inferred in K=5. When K=6 the Southern populations are further split, separating away from each other. When K=4, it can be inferred that some cultivars populations such as those in France and Netherlands are grouping with the Northern populations. The Netherlands population is the cultivar population "Suz", which if we refer to the genetic distance tree, is also grouping with the Northern clines, suggesting genetic relationship with Northern population and perhaps cultivation from these populations.

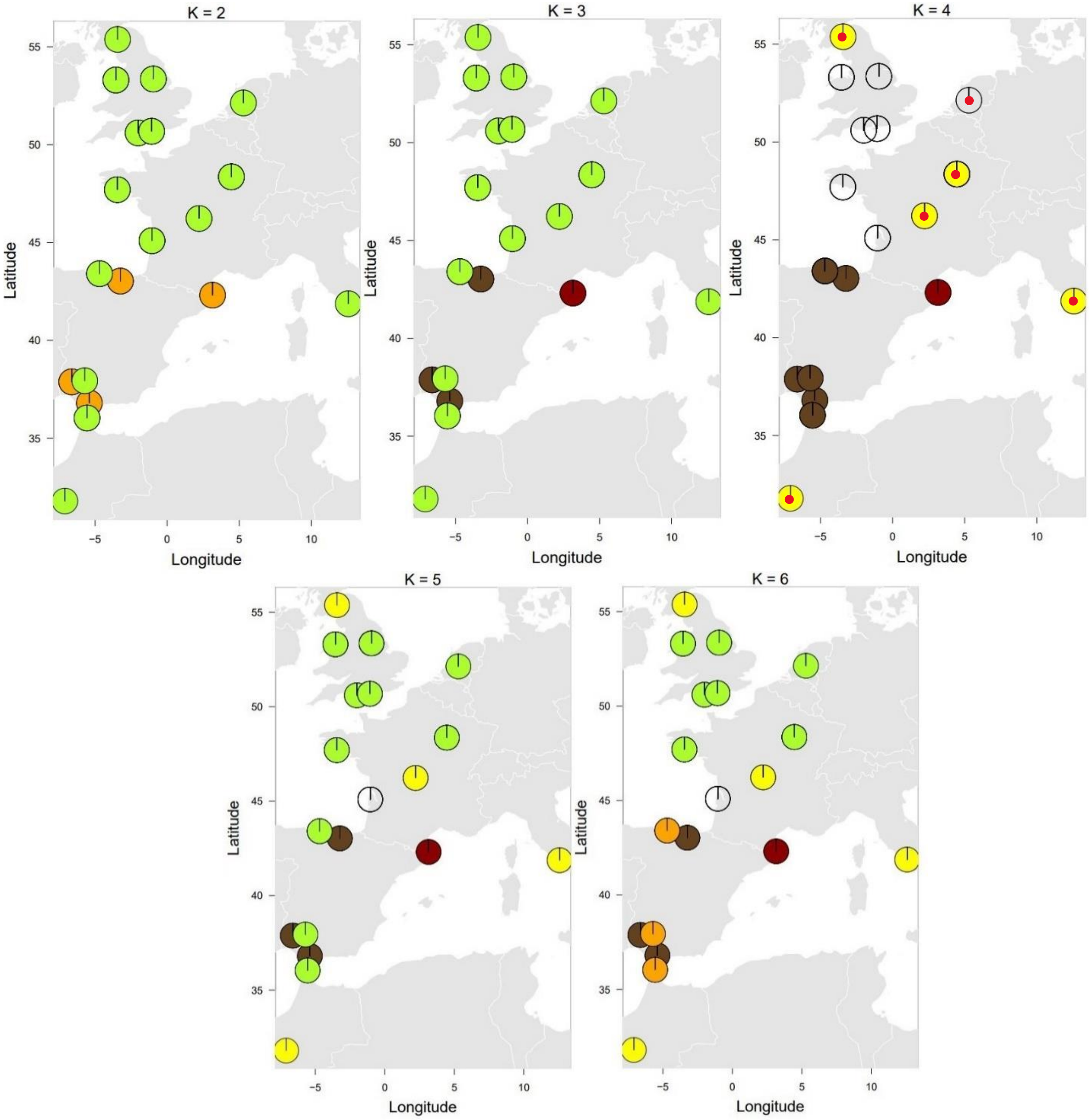


Figure 16. A geographical map of all the populations conveying genetic differences based on DAPC analysis and separated by K=2-6. Groupings was observed better when K=4. When K=4 Southern wild individuals were grouping in brown, Northern wild individuals colourless and the Mediterranean population in red. The cultivars were observed to be more spread out and are marked with a red dot in the middle of the respective pie charts under K=4.

### 4.4.2iv Landscape and Ecological Association

As part of the population structure analysis, a Landscape and Ecological Association analysis (LEA) were utilised whereby ancestry coefficients can be implied between populations. Number of sub-populations (K) was described at 1-6. A Cross-entropy criterion can be used to determine the best run for a fixed value of K. The plot below illustrates the cross-entropy criterion for our dataset. The lower the cross-entropy value the better prediction capability a K value has.

The Cross-entropy criterion plot reveals that for our dataset, at higher K-values the minimal cross-entropy was the lowest (Figure 17). This starts to level off when K=4. This levelling off minimal cross-entropy suggests that the optimal number of populations of K would be equal to 4 for further LEA analysis. An LEA bar plot with ancestry coefficients can be constructed for K=1-6 to imply any admixture within our dataset.

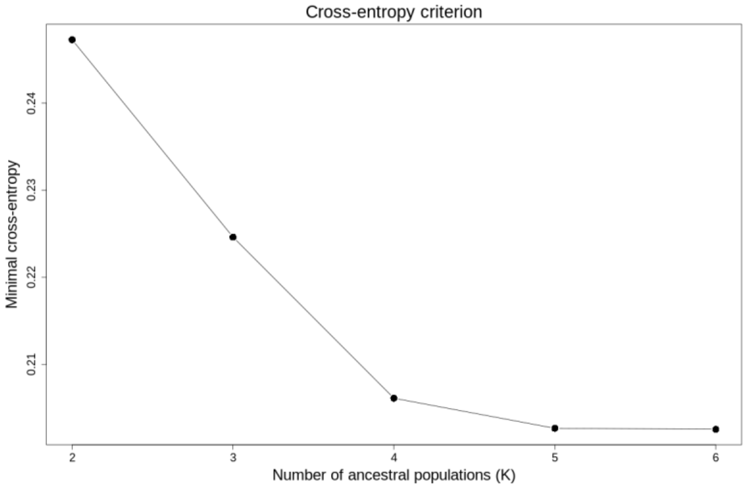


Figure 17. Cross-entropy criterion graph illustrating the best number of population (K) is K=4. When the trend levels off.

We further illustrated genetic relationships between our sample using an LEA barplot revealing some admixture within the Southern and Northern populations when *K≥4* (Figure 14). Admixture is suggested by the presence of sequences/SNPs from multiple genetic clusters for an individual. The Northern population were less admixed. The cultivar populations were more similar to the Northern accessions (Figure 18) when *K≤3*. They begin to differentiate when *K=≤4*. Admixture was also observed with the Northern accessions showing most of the coefficient changes when *K=≤4*.

In the LEA plot there was also observation of divergence in population E-12. This is observed by the diverging bars on population E-12, observed when *K=≥3* (Figure 18). The LEA agrees with the distinction of population 12 from the other groups observed in the other genetic structure analysies above.

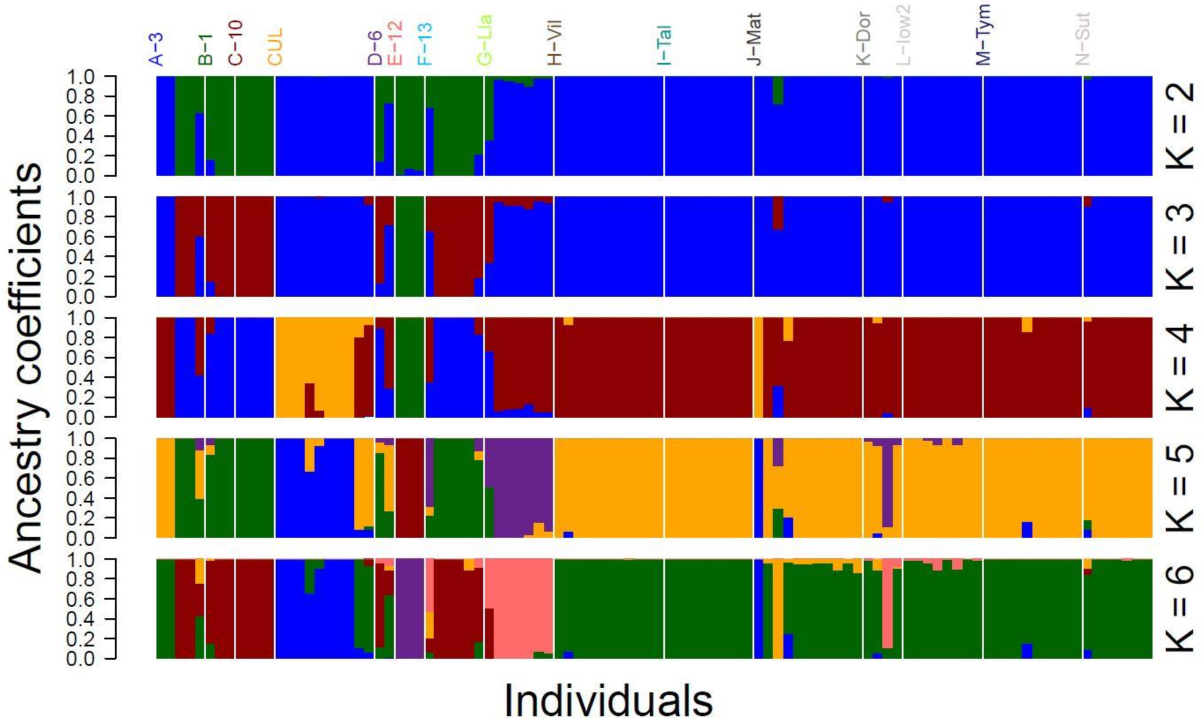


Figure 18. A LEA bar chart for every individual, separated into populations for K=2-6. The population names are in alphabetic order of the most South latitude to the most North.

To observe this admixture even further, a map can be drawn with pie charts that reveals the ancestry coefficients of different population based on their location (Figure 19). The pie charts on the geographical map represents the genetic relatedness of each population to another based on an LEA analysis. Just as the bar charts on Figure 14, it is revealed that admixture occurs within some of the populations observed within this analysis. When K=3, admixture within the Southern wild population can already be seen, with some of the more Southern population revealing LEA association with the more Northern population highlighted in green for K=3. When K=4, the cultivars was differentiated further. This is marked by the colour yellow on the pie chart. There seems to be admixture in the supposedly cultivated population originating from France and in the Netherlands. This was shown in K=4 whereby these populations are observed to have contained admixture from every population but the population E-12 (Figure 19). Admixturing occurred in the cultivar populations “Suz” and “Tin”, suggesting genetic relatedness of these cultivars it’s wild ancestors (Figure 19).

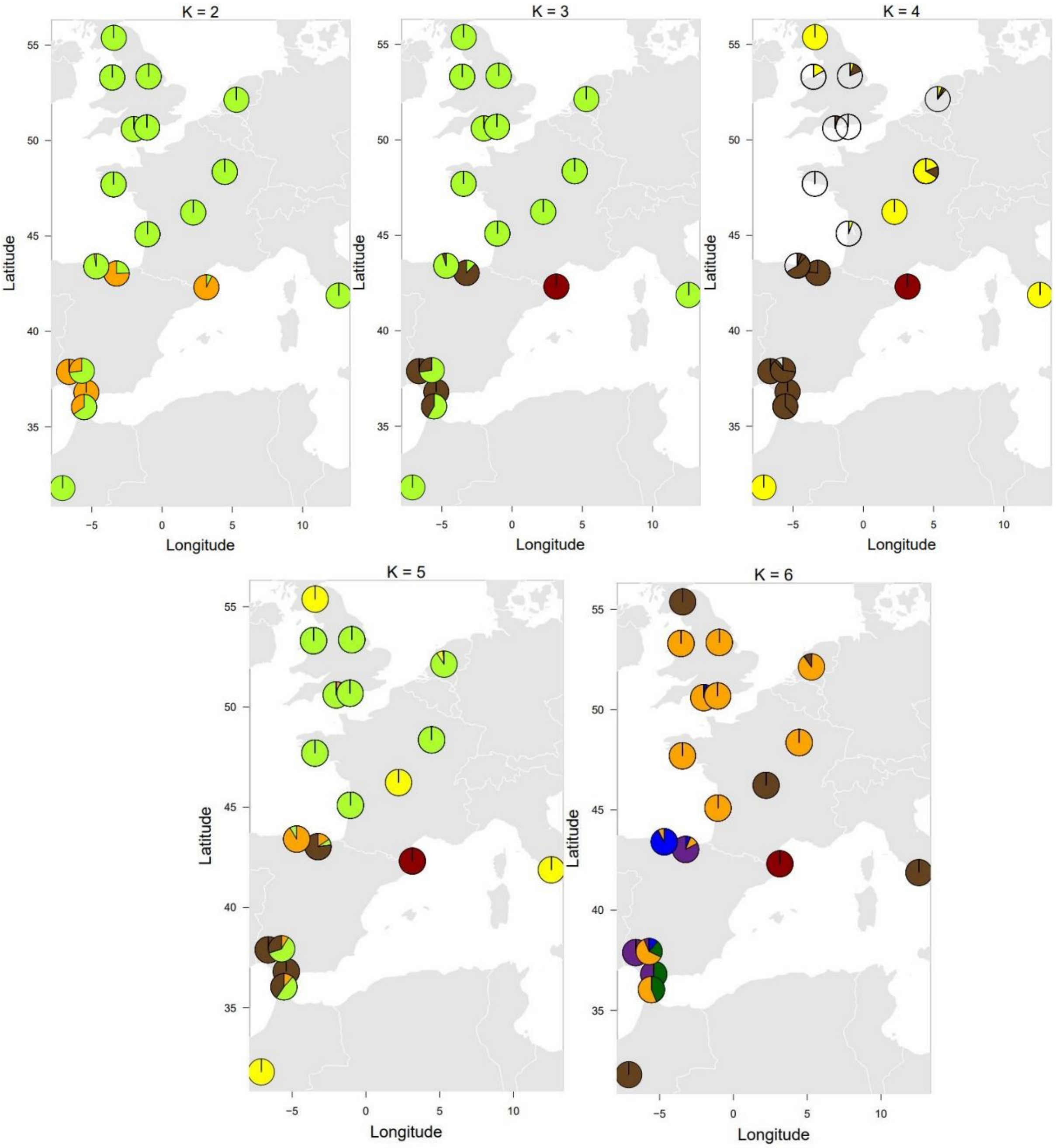
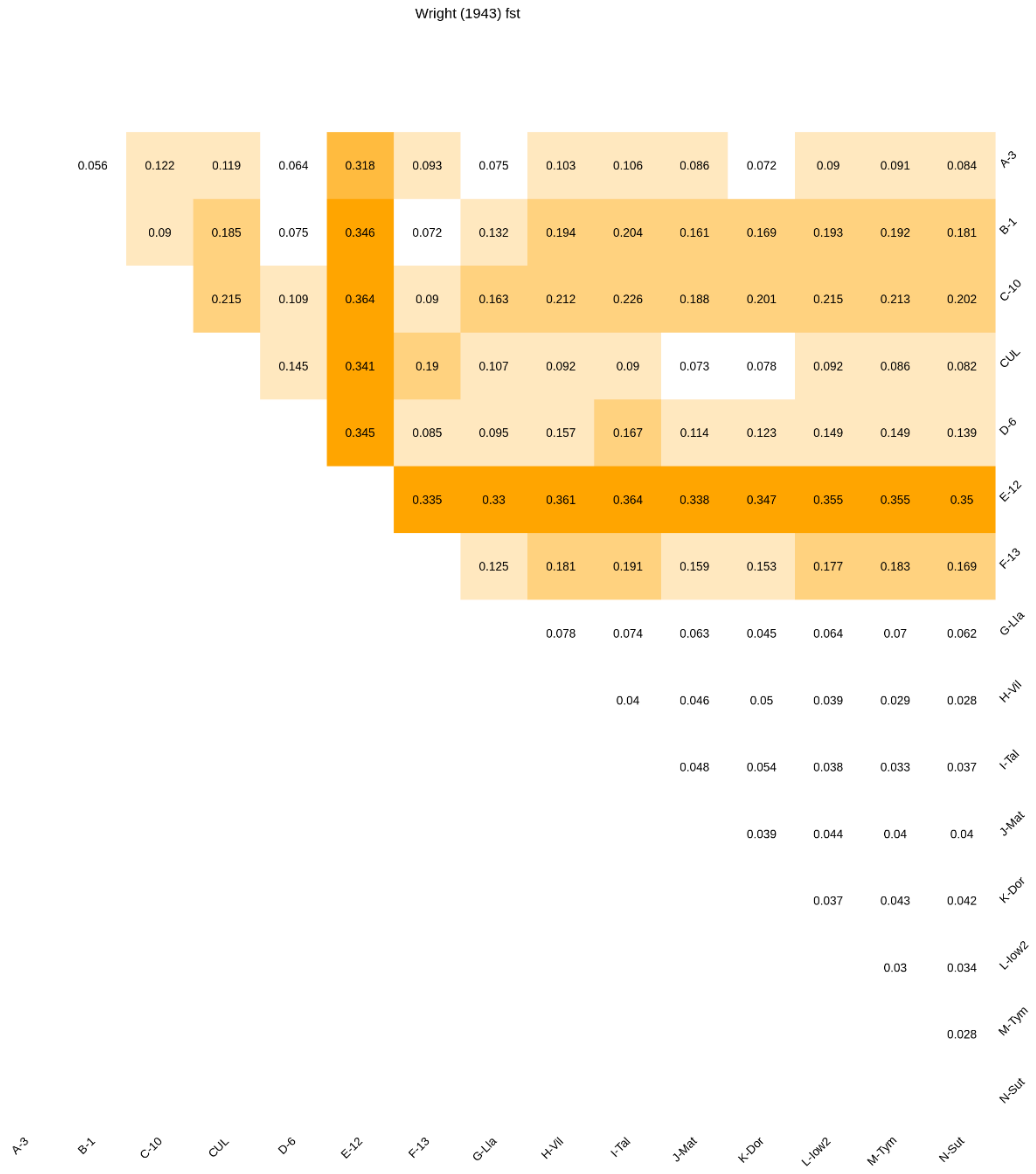


Figure 19. A geographical map of all the populations conveying ancestry differences based on LEA for K=2-6.

### 4.4.3 Results: Population Divergence

As observed in the LEA analysis, there was some potential divergence occurring within one or more population of our samples. To infer this further, we can look at measures such as population dissimilarity based on their SNPs. Another way of looking at diverging population is to observe their various allele frequencies. Wright's Fst values suggest differentiation between populations. Fst value of "0" infers no variance between the population whilst Fst value of "1" suggests complete variance between the population compared, illustrating potential differentiation and divergence (Wright., 1965; Weir., 2012; Bird K. et al., 2017).

A matrix can be drawn from Wright's Fst values (Figure 20 on the next page). The matrix reveals a summary of Wright's Fst values against each of the population summarised within this study. As with the Nei's genetic distance matrix, the darker shade of orange reveals a higher Wright's Fst value. The matrix suggests higher Wright's Fst values between Southern and Northern Population in contrast with Northern to Northern population. This supports further the structuring observed within the genetic distance tree, the DAPC analysis and the LEA analysis. Variation in their Wright's Fst values suggest that population is more diverse and different than initially thought. In this analysis, it is further suggested that the Mediterranean population E-12 have a higher total genetic variance at a population level against other wild population. A high Wright's Fst value suggest a considerable degree of differentiation and divergence in population E-12 when compared to the rest of the population in this study.



Wright (1943) fst

| A-3 | B-1 | C-10 | CUL | D-6 | E-12 | F-13 | G-Lla | H-Vil | I-Tal | J-Mat | K-Dor | L-low2 | M-Tym | N-Sut | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.056 | 0.122 | 0.119 | 0.064 | 0.318 | 0.093 | 0.075 | 0.103 | 0.106 | 0.086 | 0.072 | 0.09 | 0.091 | 0.084 | A-3 |
| | | 0.09 | 0.185 | 0.075 | 0.346 | 0.072 | 0.132 | 0.194 | 0.204 | 0.161 | 0.169 | 0.193 | 0.192 | 0.181 | B-1 |
| | | | 0.215 | 0.109 | 0.364 | 0.09 | 0.163 | 0.212 | 0.226 | 0.188 | 0.201 | 0.215 | 0.213 | 0.202 | C-10 |
| | | | | 0.145 | 0.341 | 0.19 | 0.107 | 0.092 | 0.09 | 0.073 | 0.078 | 0.092 | 0.086 | 0.082 | CUL |
| | | | | | 0.345 | 0.085 | 0.095 | 0.157 | 0.167 | 0.114 | 0.123 | 0.149 | 0.149 | 0.139 | D-6 |
| | | | | | | 0.335 | 0.33 | 0.361 | 0.364 | 0.338 | 0.347 | 0.355 | 0.355 | 0.35 | E-12 |
| | | | | | | | 0.125 | 0.181 | 0.191 | 0.159 | 0.153 | 0.177 | 0.183 | 0.169 | F-13 |
| | | | | | | | | 0.078 | 0.074 | 0.063 | 0.045 | 0.064 | 0.07 | 0.062 | G-Lla |
| | | | | | | | | | 0.04 | 0.046 | 0.05 | 0.039 | 0.029 | 0.028 | H-Vil |
| | | | | | | | | | | 0.048 | 0.054 | 0.038 | 0.033 | 0.037 | I-Tal |
| | | | | | | | | | | | 0.039 | 0.044 | 0.04 | 0.04 | J-Mat |
| | | | | | | | | | | | | 0.037 | 0.043 | 0.042 | K-Dor |
| | | | | | | | | | | | | | 0.03 | 0.034 | L-low2 |
| | | | | | | | | | | | | | | 0.028 | M-Tym |
| | | | | | | | | | | | | | | | N-Sut |

Figure 20. A matrix depicting Wright's Fst value. Higher value is highlighted in darker orange shades

## 4.4.3i Population Dissimilarity

We observed dissimilarity of Sequence between individual population. It is with interest for this study to observe the most divergence populations by observing the population which has the most dissimilar SNPs with the other populations. As a starting point, variation and divergence can be looked at for the individuals within the same population.

The boxplot (Figure 21) reveals the overall sequence (Figure 21A) and SNP only ( Figure 21B) dissimilarity of individuals within a population. The dissimilarity is low overall for all populations (>5% for overall sequence dissimilarity and >0.2 proportions for SNPs only). Although this was the case, there was some differences which can be observed. In the earlier more Southern populations, the SNP dissimilarity within populations were more diverse than that of most of the Northern populations, except for the French population "J-Mat". Population 3 is showing the most dissimilarity within the Southern population. There is potential here that within population 3, either they are more diverging than other populations, or there was contamination within the sampling processes whereby other population may have been unintentionally selected as population 3.

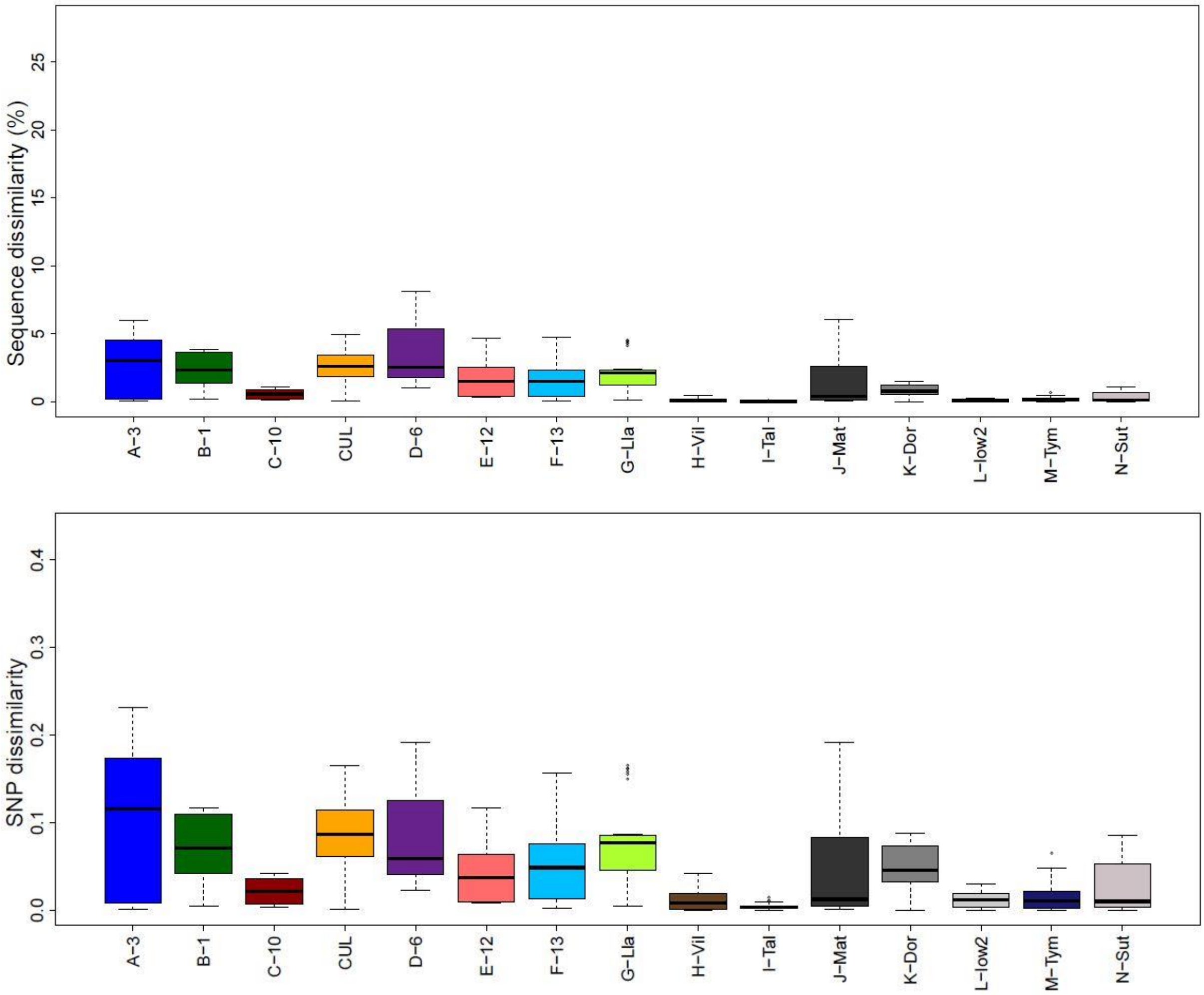


Figure 21. A boxplot to show Sequence (A) and SNP (B) dissimilarity between the populations in this study.

For completion, a measure of nucleotide diversity (pi), as was mentioned under section "*4.4.2 i Genome-wide Diversity*" of this chapter can be compared and mean pi for each population can be extracted from SambaR (Figure 22). The mean pi was tested using a two sample t-test. The t-test revealed that there was a significant mean pi difference between the wild population groups

(Northern and Southern populations) *(P(T<=t) two-tail = 0.019 (<0.05))*. This illustrates divergence in the genetic diversity of Southern and Northern populations.

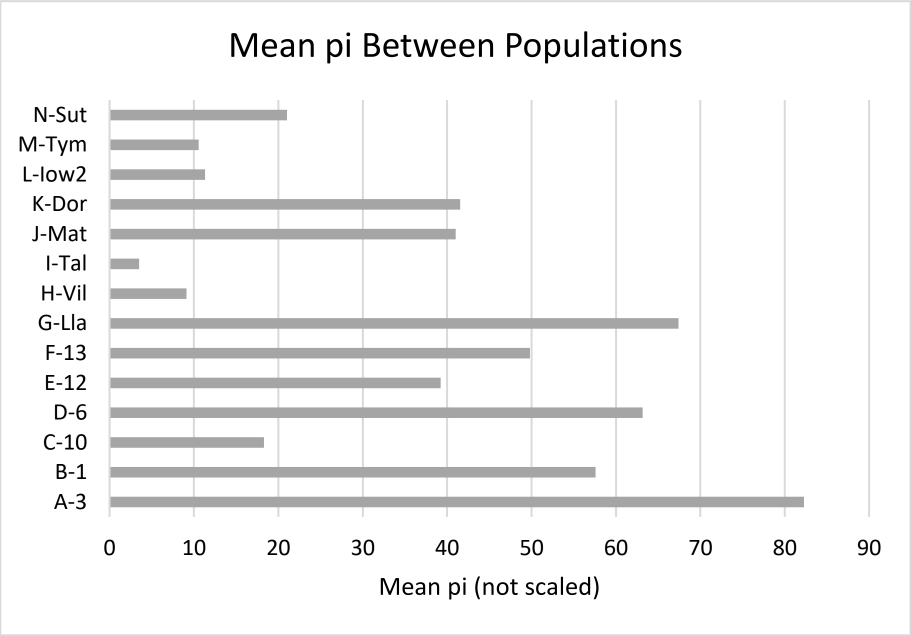


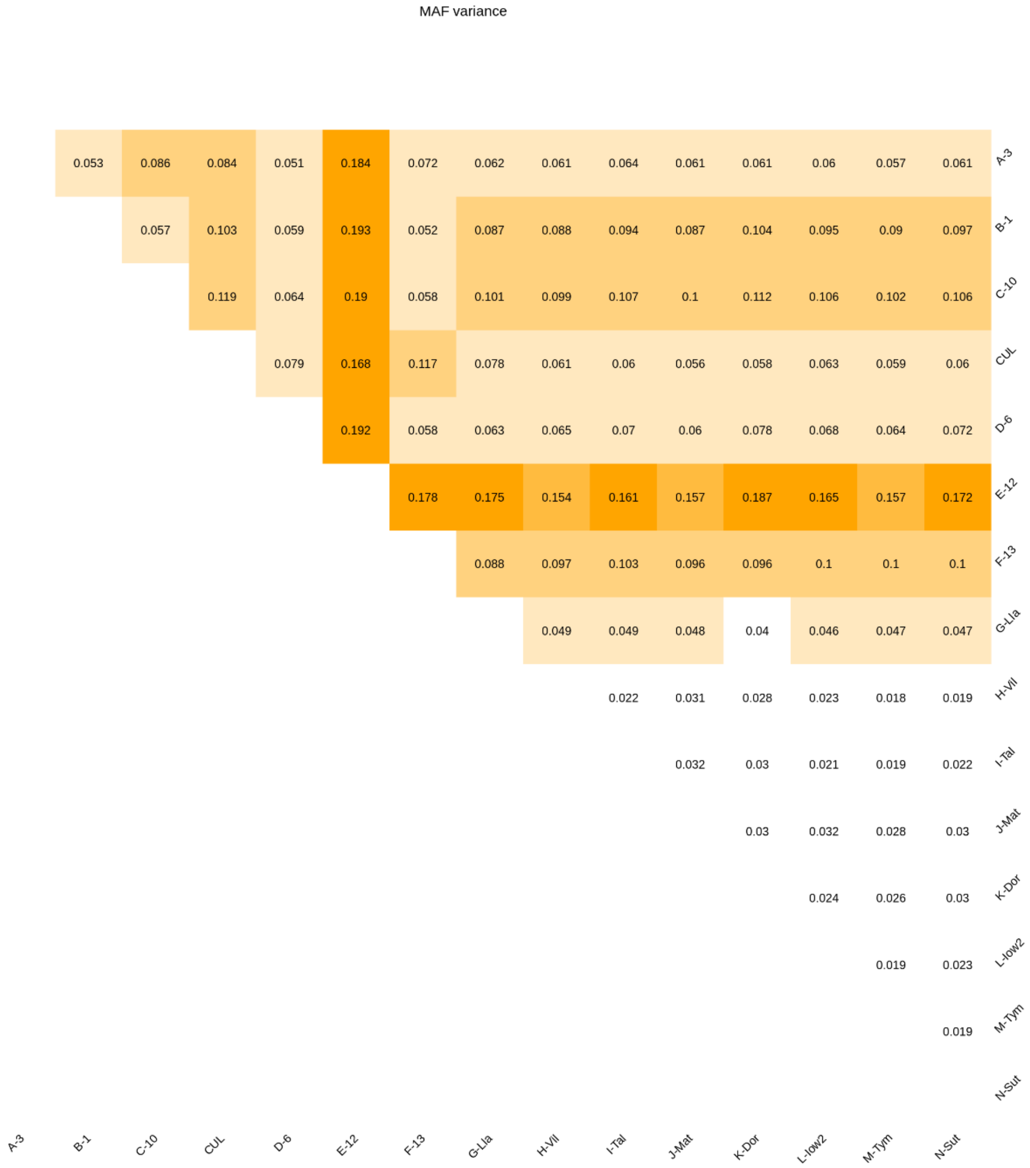
Figure 22. A bar chart illustrating mean pi for each population in our sample as outputted by SambaR.

### 4.4.3ii Minor Allele Frequency

Another measure of population divergence can be done using the minor allele frequency (MAF). Minor Allele Frequency (MAF) is widely known as the frequency on which the second most common allele occurs in a sequence of a given population. They have been shown to play a role in population selection and divergence because MAF variants occurs once, and they drive a significant amount of selection (Hernandez et al., 2019). This would give an idea of how varied a genotype is for a given SNP. This can be used to differentiate between common and rare variants in the population (Linck & Battey., 2019). If the MAF is low, it may imply that the major allele for the SNP is conserved. A high presence of common alleles may also reflect signs of genetic bottlenecks (Marth et al., 2004). Alternatively, a high presence of rare alleles may suggest that a population is expanding (Marth et al., 2004).

A matrix was drawn to illustrate MAF variance between the population samples within this study (Figure 23 on the next page). Higher variance in MAF variance is revealed in the more Southern population highlighted in shades of orange when MAF variance = >0.04. This reveals that major alleles maybe conserved more between the Northern population. This difference may suggest that certain genes are conserved in more Northern individuals. MAF variance are also highest in the Mediterranean population “E-12”.

For further comparisons, a mean MAF of each population (see appendix 29) were able to be extracted from SambaR and a two-sample t-test were applied to the mean MAF dataset (Figure 24). The t-test revealed that the difference in mean MAF values between Southern (A-G) and Northern (H-N) populations were significant *(P(T<=t) two-tail = 0.009 (<0.05))*.



| | A-3 | B-1 | C-10 | CUL | D-6 | E-12 | F-13 | G-Lla | H-Vil | I-Tal | J-Mat | K-Dor | L-low2 | M-Tym | N-Sut |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A-3 | | 0.053 | 0.086 | 0.084 | 0.051 | 0.184 | 0.072 | 0.062 | 0.061 | 0.064 | 0.061 | 0.061 | 0.06 | 0.057 | 0.061 |
| B-1 | | | 0.057 | 0.103 | 0.059 | 0.193 | 0.052 | 0.087 | 0.088 | 0.094 | 0.087 | 0.104 | 0.095 | 0.09 | 0.097 |
| C-10 | | | | 0.119 | 0.064 | 0.19 | 0.058 | 0.101 | 0.099 | 0.107 | 0.1 | 0.112 | 0.106 | 0.102 | 0.106 |
| CUL | | | | | 0.079 | 0.168 | 0.117 | 0.078 | 0.061 | 0.06 | 0.056 | 0.058 | 0.063 | 0.059 | 0.06 |
| D-6 | | | | | | 0.192 | 0.058 | 0.063 | 0.065 | 0.07 | 0.06 | 0.078 | 0.068 | 0.064 | 0.072 |
| E-12 | | | | | | | 0.178 | 0.175 | 0.154 | 0.161 | 0.157 | 0.187 | 0.165 | 0.157 | 0.172 |
| F-13 | | | | | | | | 0.088 | 0.097 | 0.103 | 0.096 | 0.096 | 0.1 | 0.1 | 0.1 |
| G-Lla | | | | | | | | | 0.049 | 0.049 | 0.048 | 0.04 | 0.046 | 0.047 | 0.047 |
| H-Vil | | | | | | | | | | 0.022 | 0.031 | 0.028 | 0.023 | 0.018 | 0.019 |
| I-Tal | | | | | | | | | | | 0.032 | 0.03 | 0.021 | 0.019 | 0.022 |
| J-Mat | | | | | | | | | | | | 0.03 | 0.032 | 0.028 | 0.03 |
| K-Dor | | | | | | | | | | | | | 0.024 | 0.026 | 0.03 |
| L-low2 | | | | | | | | | | | | | | 0.019 | 0.023 |
| M-Tym | | | | | | | | | | | | | | | 0.019 |
| N-Sut | | | | | | | | | | | | | | | |

Figure 23. A matrix showing Minor Allele Frequency (MAF) between the different populations tested in this thesis.

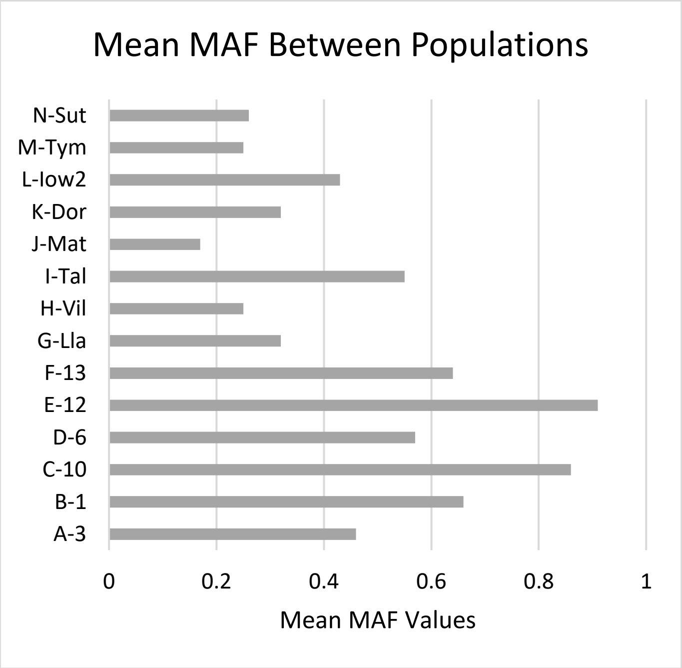


Figure 24. A Bar chart to represent the mean MAF values over each population.

---

# 4.5 DISCUSSION AND CONCLUSSION

## 4.5.1 Genetic structure

We hypothesized that within our population samples, there are population groupings based on their geographic origin. The Ward.D neighbour-joining tree revealed this, showing genetic structuring between northern and southern populations, and between wild and cultivars within our sample set. The PCoA based on Nei's genetic distance also supports this. This genetic structuring suggests possible genetic variation within Western European *Linum bienne* populations. The PcoA and CA analysis revealed grouping based on northern and southern regions, with some intermediary individuals in between the clusters. In addition, the pie-charts based on the DAPC agreed that for up to four clusters (K=4), there was structuring related to the geography of the *Linum* populations of this study. The sequence dissimiliarity matrix revealed that the northern populations were differentiated from those of the southern populations. These results confirm structuring according to the geography of the wild *Linum*. Although this is the case, filtering for heterozygosity may impact measures such as the dissimilarity matrix and genetic distancing that maybe of true nature as opposed to paralogs. In this thesis, we were conservative towards avoidance of paralogs, based on the knowledge that our study models are highly-selfing and prone to gene duplication events such as polyploidy, which can complicate assembly of our sequences (Mastretta-Yanes et al., 2014). With a 1% heterozygous filter we managed to remove the most heterozygous individuals whilst retaining heterozygousities that may suggest dissimilarities important for population diversity and divergence analysies. We also filtered at 5% heterozygousity where we begin to see populations dropping off for further analyses. Future studies may look at a range of heterozygous filter for comparison to this study.

The geographic distribution of a plant population can lead to variation within species, so it is not surprising to find this in our wild *Linum* samples. In the model *Arabidopsis thaliana*, it has been illustrated that polymorphisms revealed differences within the population genetics of northern and southern populations (Fodorenko et al., 2001). In addition, studies of the model *Arabidopsis thaliana* have revealed patterns of genetic structuring that signals evolutionary processes such as migrations (Shirsekar et al., 2021). Additionally, Mediterranean populations of *Arabidopsis thaliana* revealed relationship that are closer to their relatives further South in Morocco and North Africa (Brennan et al., 2014). In that study, there was genetic structuring between northern and southern accessions. The more southern Mediterranean population were revealed to be related to populations further south of the range of the samples studied. There could be similar case made here when looking at the northern and southern clustering of wild *Linum* populations within our samples, with the Southern population suggesting genetic distinction from the Northern clusters. When looking at genetic structuring, these could be influenced by interaction of ecological and genetic processes such as local adaptation to seed dispersal. Ecological barriers to seed dispersal, due to the biogeography of Northern and Southern Europe may also play a part in the structuring seen in these Western European wild *Linum* populations. Seed dispersal maybe limited in Southern and Norhern regions of Europe thus limiting gene flow between the two wild groups, causing this genetic structuring. This barrier to disperal was observed in the Southeast Asian mangroves (Wee et al., 2020) and in peatmosses (*Sphagnum*) (Kyrkjeeide et al., 2016). Studies suggests that climate is the dominant determinant of plant range in Europe but in addition, species dispersal plays an important role in the genetic flow of a population (Normand et al., 2011).

Another case made in the arctic-alpine plant species *Lloydia serotina*, was that reproductive biology can result in population structuring (Jones and Giddeon., 1999). Genetic differentiation that suggests variation in breeding system can also be observed in studies using three orchid species (Sun and Wong., 2001). In this study we observed signals of geneflow between Northern and Southern populations. We also saw that Southern population maybe more outcrossing in terms of breeding strategy. It maybe plausible that there is variation in breeding strategies between Northern and Southen clines of wild *L. bienne* observed in this study, this may be resulting from a seed dispersal barrier (ecological and geographical) ultimately resulting in genetic structuring. Studies on breeding strategies of wild *L. bienne* is scarce. We think that Southern population are more outcrossing based on higher heterozygousity. However, more samples needed to be studied in the future before this can be confirmed.

Genetic structure analysis suggested that some individuals did not cluster with either northern or southern groups. This is potentially a signal for geneflow. If this signal is gene flow between Southern and Northern population, this would impede  local adaptations within the Northern wild *Linum* populations. In *Arabidopsis*, it has been suggested that different populations under different environments showed this genetic structure (Hämälä et al., 2019). Another possibility is that historically, Northern populations diverged from these populations with some individuals not clustering under PCoA. However, the number of these individuals not clustering is limited to a few individuals. We saw no individuals forming a cline with the Mediterranean population E-12. This suggest that this population is divergent, and perhaps more possibility of local adaptation occured in the Mediterranean populations.

We saw in our neighbour joining tree, PcoA and DAPC analysis that the Northern and Southern wild population are more divergent from each other than the cultivars are to either population groups. The cultivars are more closely related to the Northern wild individuals. If these cultivars are genetically closer to the Northern groups, we expect them to have some phenotypic similarities observed in the other chapters such as their pollen viabilities and the relative gene expression to flowering initiation. We didn’t find sensitivity to vernalization requirement in our cultivars as we did in the Northern wild populations. There wasn’t any phenotypic case where it may suggest that the cultivar population are closely related to the wild Northern populations. However, it may be that the cultivars are genetically more related to the wild Northern populations than they are to the Southern populations due to cultivation from the Northern populations. Artificial selection in the cultivars may have implications on the variation on phenotypes between the two species.

The cultivars “Suz” and “Tin” were found to be grouping with Northern wild populations as opposed to the cultivar group (Figure 12). We think that there’s potential mislabelling, either in previous seed collections or in the laboratory, with regards to the cultivars “Suz” and “Tin” as seen in our Ward.D tree. These cultivars were grouped with the Northern population, however, are very unlikely to be doing so since there was no phenotypic case suggesting this. We take caution with any results regarding these cultivars.

### 4.5.2 Population Diversity and Divergence

Genome-wide diversity measures looking at proportion of segregating sites suggests a lower level in most Northern populations and a higher population in the Southern population. A higher amount of segregating site can suggest a higher mutation rate for the Northern populations and expansion further North (Fu Y., 1995). Segregating sites can suggest evidence of positive selection in a population (Gilad Y. et al., 2002; Przeworski M., 2002; Booker T. et al., 2017). Suggestion of more occurrence in positive selection could be interpreted by a higher number of segregating sites within our Northern groups, potentially driven by the local adaptation to the environment (Figure 11). Positive selection is often associated with selective sweep, suggesting less genetic diversity in our Northern populations (Booker T. et al., 2017). Less genetic diversity could imply more inbreeding in these Northern population.

Furthermore, studies using *Camellia sinensis* (Tea) and their wild relatives suggests genetic divergence in the wild relatives by looking at proportion of segregating sites and in addition, the proportion of heterozygous sites (Yang H. et al., 2016). We looked at the proportion of heterozygous sites for all sites in our study model. Most of these sites were filtered due to filtering of paralogs. However, for all sites, surviving heterozygous alles were higher in the Southern populations, suggesting allelic richness (Figure 8C). This suggest that the Southern populations are genetically more diverse than the Northern population. Northern population may be more inclined to be conservative towards local adaptation to colder seasons. This was perhaps less desirable in the Southern populations due to warmer climates.

The Mediterranean population E-12 was the most distinct among our samples. Based on Nei's genetic distance these Mediterranean individuals are clustering on their own as opposed to either Northern or Southern clines. It is thought that this Mediterranean population is genetically distinct from the rest of the Western European wild population. It is possible that the more genetically distinct population E-12 was closer to wild relatives in the eastern Europe region, but this was not tested in this study. We did not have materials for the wild *Linum* population originating in eastern European region such as Turkey and Italy, in which both wild and cultivated flax have been observed. There is even a possibility that population E-12 is more related to population as far as the eastern Mediterannean and the Middle East regions, because the cultivated flax (*L. usitatissimum*) is thought to be native to these regions before they were introduced to the more Northern climates (Sen, T. and Reddy, H.J., 2011). It would be of interest for future research to include more Mediterranean and Middle Eastern populations, to determine their genetic relationships with our samples. The distinction of population E-12 was also supported by the Tajima's D statistics, whereby population E-12 had a more negative Tajima's D than the rest of the populations sampled in this study. This may suggest that this population were diverged in the past, and this was a result of population expansion in wild *Linum*. In past studies, low Tajima's D in the fir species *Cunninghamia konishii* have been inferred as a post-glacial expansion with their populations (Hwang et al., 2003). A study in *L. flavum* suggested a post-glacial migration history in *Linum,* which may result in a population expansion during the post-glacial period (Plenk et al., 2017).

### 4.5.3 Breeding Strategies in Western European *L. bienne*

Population diversity measures such as the measures of heterozygosity along with the multi locus heterozygosity also differed between the southern and northern populations. When filtering for excessive heterozygous alleles, remaining heterozygousity were observed mostly in Southern populations, with most of the Northern population infering no heterozygousity (Figure 9). In addition to the structuring observed in the genetic distance tree and the PcoA, wild populations might also be differentiating in the way they breed. We observed a higher heterozygousity in Southern plants in all sites (Figure 8c and Figure 9). The more heterozygous populations suggest less inbreeding and more outcrossing than the less heterozygous northern populations. Sharry and Lord (1996) also explored heterozygosity, concluding that less heterozygous populations may have been more inbred than those of the more heterozygous populations (Sherry & Lord., 1996). We think that our Southern *L.*

*bienne* populations are more outcrossing and that our Northern population are inbreeding, perhaps to conserve local adaptations or perhaps due to climatic barriers for seed dispersals. This differentiation would also be supported by the difference seen in Wright's Fst matrix. It is seen that Fst values between Northern and Southern populations are higher as highlighted in shades of orange. In the literature it is accepted that an Fst value greater than 0.15 between populations can be considered as significant in differentiating populations (Luo et al., 2019). This adds to the genetic structuring revealed by the genetic distance, PcoA and CA analysis that northern and southern populations are genetically different. In addition, there is potential that Southern population are more outcrossing in terms of their breeding strategy. This agrees with the findings on genetic structuring observed in section 4.41 "Results: Genetic Structure" and can so therefore observation in genetic diversity supports the finding that there is potential breeding strategy variation in Northern and Southern clines of *L. bienne*, causing genetic structuring between the two Western European wild *Linum* populations.

### 4.5.4 Genetic variation in Western European *L. bienne*

Molecular diversity and association analyses in the past have revealed the potential of genetic variation in wild flax for the development of cultivated flax (Soto-Cerda et al., 2014). It is interesting that the cultivar populations within our sample sets seemed to be more genetically closely related to the Northern populations than to the Southern populations. The genetic distance tree and the ancestry coefficients on the LEA bar plot agreed that the cultivars in our samples were more closely related to the more northern wild accessions than so the southern accessions. Variation in genetic makeup within the *L. bienne* gene pool represents the potential of genetic variation for the improvement of cultivated flax in Northern Europe. Traits that are conserved in Northern populations in response to colder climate can be useful for flax cultivation in different growing seasons.

To add to the potential genepool resources, we can infer some ecological interpretation for our wild populations. This includes selection and bottlenecks events that may have arisen in the past. Tajima's D suggest that there is less variation compared to what the population can sustain. In population genetics studies, negative Tajima's D have previously suggested population expansions and differentiation after a bottleneck (De Jong et al., 2011; Gunther et al., 2016). This negative Tajima's D was strongest in the more southern populations, particularly in population E-12. This may suggest that some southern populations, such as population E-12 have diverged through past expansions and potentially still expanding after a bottleneck. High and low divergent populations revealed by the Fst values could suggest population availability for breeding programmes (Baiakhmetov et al., 2021). The higher Fst values among the more southern populations and lower among the northern population suggested that there were potential differences in gene flow between populations. The minor allele frequency (MAF) measures between the northern and southern populations also revealed a significant divergence between the two groups. The lower observed MAF within the Northern populations might suggest that these populations are more conserved in terms of their genetic strategies, therefore may have resulted in the smaller sequence and nucleotide diversity whilst Southern population may have a larger diversity to provide a wider genepool for cultivar breeding programmes.

In conclusion, the SNPs data revealed genetic structuring among the Western European wild relatives of flax as implied by the genetic distance tree. This was further supported by the PcoA, DAPC, and LEA analyses. There was also potential that this genetic differentiation could mean further differences in terms of local adaptations. Differences observed in heterozygosity and Fst values revealed that there maybe a difference in breeding strategies of these wild *Linum* populations. Population divergence measures also revealed that these northern and southern populations were divergent. It would be with further interest to observe SNPs diversity from a wider region such as Eastern Europe and the Eastern region of the Mediterranean. The only Mediterranean population observed in this case was population 12 and it was interesting to observe the difference in genetic structuring and genetic divergence here. However, we make these suggestions with caution as the number of individuals that were analysed was relatively small, especially in the southern region. There may also be potential human errors as was observed in some odd individuals, not grouping where they should be expected within their local range. It would be with greater interest to run this type of analysis with a wider sample range, more genotype data, and a more individuals in our sampled populations to strengthen these initial findings within our *Linum* populations.

# CHAPTER 5: VARIATION IN TRAITS BETWEEN VERNALIZED AND NON-VERNALIZED *LINUM*.

Environmental factors can affect more than one plant trait (Campetella et al., 2020). Plant traits can correlate differently between populations and geographic location even down to individual levels in the wild (Jiang et al., 2021). *Linum* is an oilseed and natural fiber crop, with a significant agricultural value. Oilseed and Fibre crops have different target traits in the interest of plant breeders. For oilseed types, oil content and number of the capsules and seed sizes are the most important traits for breeding (Çopur et al., 2006), whereas in fibre types, plant height and number of stems are the essential traits for fibre flax breeding (Xie D. et al., 2018). We already established that wild *L. bienne* could provide a wider genepool for breeding purposes. In this chapter we explored traits such as plant height, stem number and bud numbers due to vernalization as an environmental variable in wild our *L. bienne* samples. This can provide insights to how vernalization may affect traits in wild *Linum*.

In chapter two, we saw how vernalization affected the number of days to flowering in both *Linum* species, suggesting variation in number of days to flowering between wild individuals from across Western Europe. In this chapter we explored how other traits could be affected by vernalization. In *Linum*, Genome-Wide Association (GWAS) have identified candidate genes fo traits such as plant height, number of branches and seed weight, revealing a biological basis for improvement amongst these traits for agricultural interests (Xie D. et al., 2018). In this chapter, phenotypic measures, such as plant height, plant stem numbers, and plant bud numbers will be quantified, analysed, visualized, and discussed further as an additional finding to support implication of wild *L. bienne* as a wider genepool. These traits are of importance, especially in terms of agriculture because the candidate genes for them have been identified which reveals potential for improvement in *L. usitatissimum*. We will test whether other traits were significantly affected due to vernalization treatments and investigate their relation to environmental variables such as latitude and local climate variables. We hypothesize that traits are significantly affected by vernalization treatments.

A plant stem is one of the main structures of a vascular plant, the other being roots. They often support other major important floral organs such as the leaves, flowers, and fruits of the plants. They are also required for support, nutrient storage, and productions of new living tissues (Raven, 1982). Studies in the pepper species *Capsicum annuum*, and maize identified changes in stem morphology due to varying light wavelength and soil temperature (Schuerger A. et al., 1997;

Walker J., 1969). These early findings suggest that stem morphology is affected by light and temperatures as environmental variables. Due to wide latitudinal range in our *L. bienne* samples across Western Europe, there are possibility that stem formation can vary under different environment perhaps because of local adaptation. We hypothesized that the number of stems are varies under different vernalization treatments, and we further suggest that there is variation to correlation to the environment and to the number of days to flower when comparing vernalized and non-vernalized conditions. Reduction in stem numbers for vernalized individuals will suggest that vernalization acts as an abiotic stress and that control for stem numbers could act as a local adaptation for requirement to vernalize.

The buds of a plant are widely known as the undeveloped embryonic shoot which is often found on the axil of a plant leaf (Walters, Keil and Walters, 1996) A *Linum* bud is a small lateral or terminal protuberance on the stem of a vascular plant that may develop into a flower, and subsequently form seeds (Trelease, 1931). This is therefore an important factor to the production of seeds and of interest to oilseed breeding. In *Arabis alpina*, it is suggested that exposure to colder environments initiates formation of flowering buds (Lazaro et al., 2018). This suggests potential adaptations and differentiation within this species due to vernalization mechanisms (Toräng et al., 2015). In this chapter we observed whether development of buds is affected by vernalization for our *Linum* samples. We hypothesized that the number of buds able to form varied due to vernalization treatments. We further suggest that there is variation on bud formation under local climatic variables.

In addition, we will observe seed areas of wild *L. bienne* seed samples from our vernalization experiment as well as seeds from the wild. Under abiotic stress, it is widely understood that seed sizes have important consequences for germination in plants. In *Anthoxanthum odoratum* it is observed that population with larger seeds had a higher probability of germinating (Roach D., 1987). In the winter annual plant *Dithyrea californica*, amount of precipitation (climatic variable) had a beneficial effect on plant fecundity and influenced seed-size survival selection (Larios E. et al., 2014). This suggests seed-size natural selection due to environmental factors in wild populations. Environmental factors during seed development have also been illustrated to influence seed germination in *Lotus tenuis* (Clua and Gimenez., 2003). On a species level, initial results using mountain alpines indicates elevation-dependence seed production is specific to each species (Olejniczak P. et al., 2018). This reflects different resource allocation strategies for different species due to environmental variables introduced under different altitudes. More recently, research in naturally occurring sand rice (*Agriophyllum squarrosum*) suggests variation in seed sizes due to local environments. They suggest that large-seeded individuals were more competitive in semi-arid regions (Zhao P. et al., 2022). With all these in mind, it is evident that in the literature, variation in seed sizes occurs in many plant study species with local environmental variables, potentially playing a role in natural selection of seed sizes in plant populations.

Seed size is a trait of interest in *Linum*. Variation in seed size due to vernalization treatment in the wild may indicate local adaptation to the environment in terms of seed sizes. In addition to this, seed sizes of *Linum* wild relatives can be useful in their cultivar relatives. This is because linseed is one of the biproduct of *Linum*. We will investigate if our wild *L. bienne* vary in seed sizes when under different vernalization treatments. Germination of plants are suggested to be related to their seed sizes, so seed size can also count for population fitness (Keddy & Constable., 1986). A study on onions (*Allium cepa l.*) also suggests that vernalization temperatures, duration and bulb size significantly influenced seed yield of all cultivars (Muthamia., 1994). This suggests that in many plant species requirements of vernalization will have a significant effect on seed health which suggests seed sizes and seed yield which is of interest in agriculture. We hypothesize that seed sizes for different *Linum* populations will have been affected by vernalization and this will have represented a correlation between seed size and environmental variables such as latitude of localities and climatic variables.

With the above in mind, it is of both ecological and agricultural interest to investigate the consequences of vernalization in *Linum* traits. Findings here can be linked to local adaptation in the wild *L. bienne* and implications of potential loss of vernalization in cultivated *L. usitatissimum*. In this chapter we investigate whether vernalization treatments affects the fitness of wild *L. bienne*. For traits, we expect to find variation between vernalized and non-vernalized individuals. We expect seed sizes to also vary between Northern and Southern individuals, suggesting local adaptation in seed sizes. Implication of requirement of vernalization can then suggest that any adverse traits observed within the more Northern individuals due to vernalization stress are avoided by the potential loss of vernalization as was observed under chapter 2 of this thesis.

## 5.2 METHODS

Regarding experimental setup, vernalization methods are the same as found in chapter 2 of this thesis (see section 2.2.1 “Samples and Experiment” for vernalization setup). For measures regarding plant height, plant stem number, and plant bud numbers, observations were made according to when an individual is observed as flowering for the first time. This is when the individual has had their first flower, which is fully opened, with petals revealing the flower’s sexual organs. This will coincide with the measure of the number of days a plant takes to flower and will reduce biases based on plant age. Measurement for flowering initation and plant trait measures such as plant heights can be directly compared against each other. A measure of plant height is done using a 2-sided measuring tape, using Centimetres (Cm) as a unit of measurement. The height measurement was taken from where the plant has emerged from the soil onto the tallest upper stem of the plant, measuring the total observable height of plant material for the individual. Stems are stretched until a straight measure can be observed using the measuring tape. The measure of plant stems and buds were based on observations of the number of either plant stems or plant buds per individual at first flowering. Conditions for the vernalization and non-vernalization treatments can be found under section 2.2.1 “Samples and Experiment” of chapter 2 in this thesis.

For measures regarding plant heights, stem numbers and bud numbers, data was only observed for the 2018 vernalization experiment. There was phenotypic measurement for both 2018 and 2021 vernalization experiment underwent in chapter two of this thesis. However, the 2021 vernalization experiment data for these were limited because of Covid-19 restrictions in place at the department of Biosciences during the time of the experiment. This caused phenotypic measures to be lost in most individuals during the 2021 vernalization experiment. There will also be a bias on earlier flowering plants, as plants with later flowering will have not had a chance to flower before the Bioscience departmental lockdown in March 2020 (see section 2.4 “Results: flowering time between experiments from different years” of this thesis). For this reason, the phenotypic measures for the 2020-21 vernalization experiment were discarded for future analysis. Environmental data such as the latitude of origin and climatic variables were used as was suggested in chapter 2 of this thesis.

In addition, we observed seed sizes. For this observation we only used the wild *L. bienne* individuals as we were mainly interested in relationship between seed sizes and environmental variables. We observed seed sizes of vernalized and non-vernalized wild individuals along with wild S0 individuals. We did this observation using non-fluoresence microscopy. A 10×10 objective was used to observe *Linum* seeds using a colour camera mounted on a Leica DMI-3000 Microscope. The Microscope was hardwired to a Windows computer running Leica's LAS X software, where measurements of seed length and width can take place. Measurement of the seed area is done by using ImageJ to measure pixels and calculate area. To do this, images of 5 seeds per-individual were captured using a microscope with a 10×10 objective. We used Leica's LAS X software to capture and save images before processing in ImageJ. In ImageJ, the images were individually measured. We calibrated ImageJ to each image by drawing a 1mm line and setting a scale to that line as 1mm. The image is then converted into 8-bit format and made binary (Process → Binary → Make Binary). Although ImageJ usually outputs binary images without holes, there are occasions where holes in the seed can be observed. Potential holes were able to manually be filled or automatically via the "fill holes" function (Process → Binary → Fill Holes). Images were then adjusted by threshold (Image → Adjust → Threshold) and analyzed by using the "Analyze particles" function of ImageJ. This function can analyze particles (pixels) and calculate area of a of each binary colour (black and white). The seeds were measured as the area in black. Values were inputted into Microsoft office and formatted for further analysis in R.

### 5.2.1 Data Analysis

All phenotypic measures were statistically tested during this chapter and all additional data on this chapter were analysed using R v 4.1.2 (R Core Team, 2022) and Rstudio (Rstudio Team, 2020) as a graphical interface. For modelling and correlation purposes the R package "ggpubr" (Kassambara, 2022), "ggplot2" (Wickham and Sievert, n.d.), and "car" (Fox and Weisberg, 2011) were used to execute general-linear modelling (GLM). There are several assumptions made for this GLM; the data is normally distributed, and that values of trait measurements, and other variables are independent of each other. We also assumed that the variance in data residual is the same. For summary of statistics, model can be summarised in terms of adjusted $r^2$, F-statistics, and p-values. The full R commands for the analysis during this chapter is available in appendix 30. Barcharts can be illustrated using the "barplot" function of R. A two-tailed t-test were then used to compare data under vernalized and non-vernalized treatments.

For correlation analysis to environmental variables, data for latitude were collected for each population sampled from the wild at the place of collection by their respective collectors, as per chapter two of this thesis. Principle component values variables were able to be gathered for different climatic variables as was described under chapter 2 of this thesis. Scatterplot for correlation between traits and environments were illustrated using "ggscatter" function of R with a

GLM modelling with a Pearson’s correlation method. The same is applied for analyses of seed sizes against environmental variables.

# 5.3 RESULTS

## 5.3.1 Plant Traits vs. Days to Flowering

In chapter two of this thesis, we looked at the relationship between the number of days to flowering and latitude for our 2018 vernalization experiments. The results demonstrates a positive relationship between the number of days a plant takes to flower and the latitude of locality to each plant. During this experiment, we also examined other plant traits such as plant height, and the number of stems. We ask whether the number of days to flowering will have other effects in other traits. This is interesting as in chapter one we discussed that different plant traits in *Linum* may correlate with each other. We measured three different traits in the 2018 vernalization experiments (plant height, stem numbers, bud numbers). These measures were plotted against the number of days the plant takes to flower and faceted based on treatments. The measurements were made in relation to the first flowering. As individual plants were first flowering, the plant height, number of stems and number of buds were also observed and recorded. The raw count data for this can be found under appendix 31.

Correlation reveals the relationship between the flowering time initiation (number of days to flower (labelled on y-axis as “Days_to_fl”) and overall plant height (labelled on the x-axis as “Height_(Cm)”) for no-vernalization and vernalization treatments (Figure 1). A Pearson’s correlation reveals that the number of days to flowering is negatively correlated with the plant height at first flowering, under the no-vernalization treatments (*R=-0.28, p=<0.001*). This contrasted with what is observed in the vernalization treatment, whereby, there were no trends observed (*R=-0.022, p=0.79*) (Figure 1). This illustrates that, with no-vernalization the number of days to flowering were negatively affecting plant height (i.e., the longer a plant takes to flower the smaller (in height) they tend to be at first flowering). This effect was reduced when the plants were vernalized.

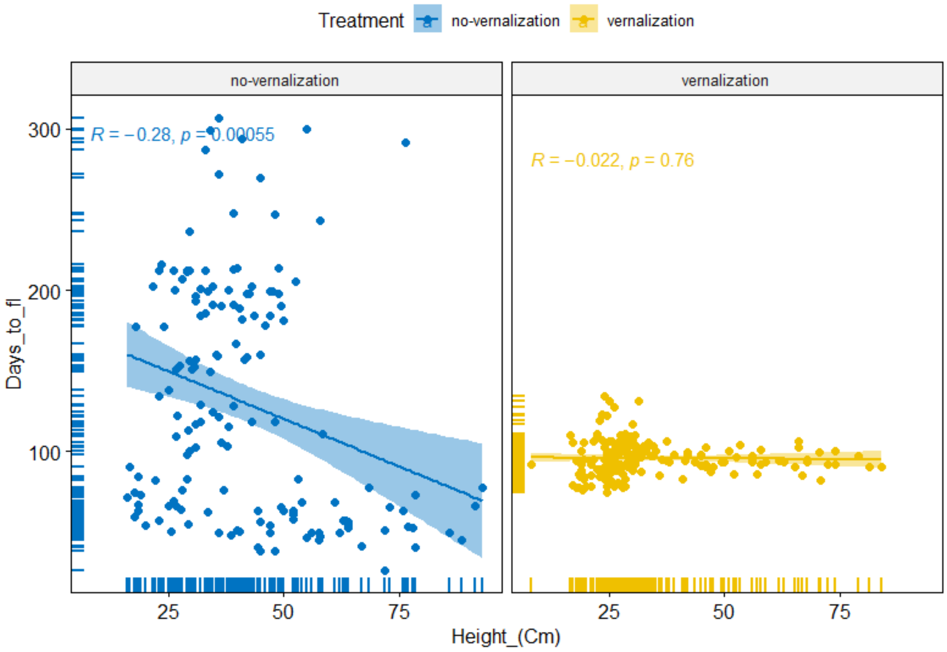


Figure 1. Two scatterplots to illustrate the relationship between the number of days a plant takes to flower (Days_to_fl) and he plant height (Height_(Cm)). The plot revealed a contrast between no-vernalization and vernalization treatments, with a significant negative correlation under the no-vernalization treatment (*R=-0.28, p=<0.001*).

Furthermore, a comparison between the plant height and the different treatments was illustrated using a two-sample t-test on Figure 2 (assuming unequal variance). The t-test suggests that there was a significant difference on the plant height between the two treatments (*P(T<=t) two-tail=<0.001*).

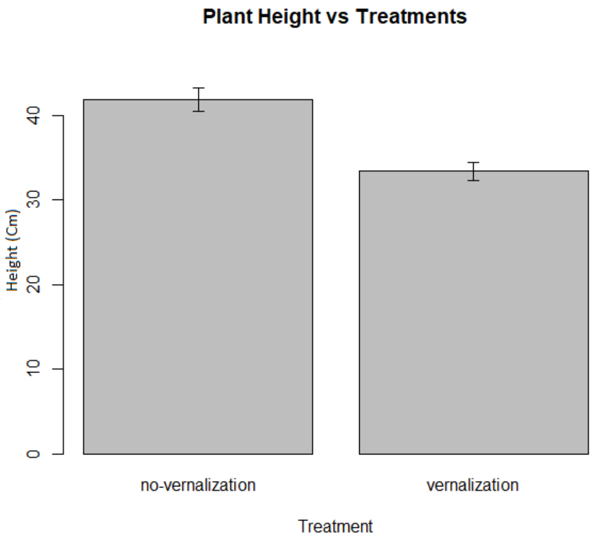


Figure 2. A bar chart to illustrate the significant difference between no-vernalization and vernalization treatments on mean plant height (Height (Cm) (*P(T<=t) two-tail=<0.001*).

Another trait we measured during the study is the number of stems at first flowering. The result suggests that when no-vernalization and vernalization treatments were applied to the *Linum* samples, a positive correlation can be observed between the number of days a plant takes to flower and the number of observed stems on first flowering (Figure 3). However, the strength of the relationship seems to have been reduced when plants were vernalized. R values for the vernalized treatments were reduced from *R=0.69* to *R=0.18.*

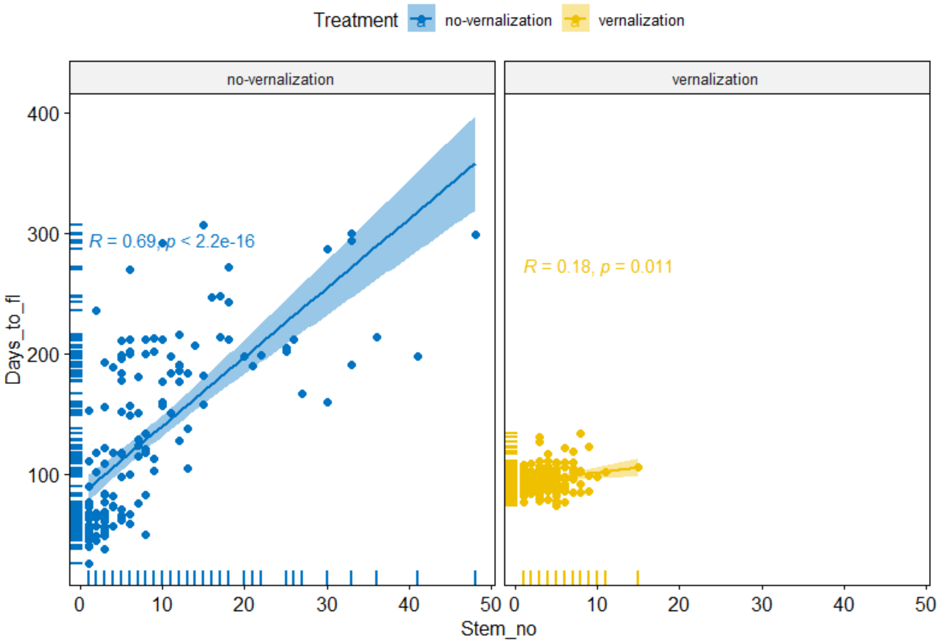


Figure 3. Scatterplots to illustrate the relationship between the number of days a plant takes to flower (Days_to_fl) and the number of observed stems at first flowering (Stem_no) for no-vernalization and vernalization treatments. For both treatments, the correlation was significantly positive (no-vernalization *R=0.69, p=<0.001;* vernalization *R=0.18, p=0.011*).

To observe whether this change of stem number was significant, a two-sample t-test (assuming unequal variance) can be implied for the two sets of data (Figure 4). The t-test revealed that the number of stems observed after first flowering was significantly lower when individuals were vernalized (*P(T<=t) two-tail=<0.001*). These findings suggest that the number of stems observed when vernalized is significantly reduced, and the positive correlation between the number of days to flowering and the number of stems observed was also reduced. However, the correlation was still significant between the two traits, even when vernalized (*p=<0.05* for both treatments).

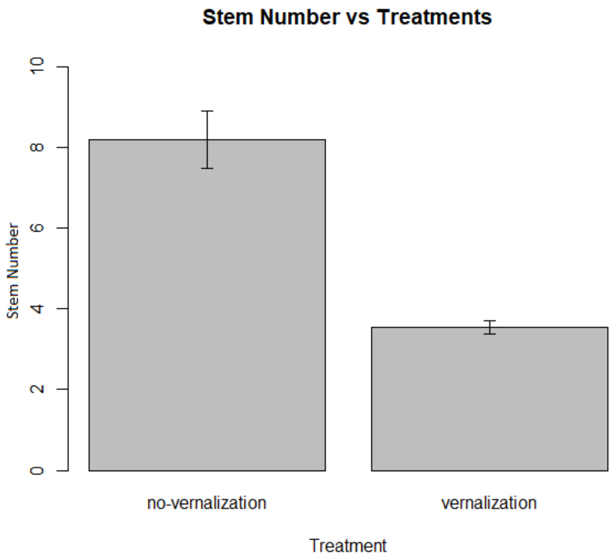


Figure 4. A bar chart to illustrate the mean stem number differences between the two treatments, showing a significant reduction in stem number when vernalization occurred (*P(T<=t) two-tail=<0.001*).

Another plant traits which were observed during the vernalization study is the number of buds. The relationship between the number of days to flowering and the number of observed buds on the first flowering suggests that there was a relationship when plants were introduced to vernalization treatments (Figure 5). Interestingly, this reduction corresponds with the of number of days to flower due to vernalization where it was also reduced. Perhaps vernalization is also affecting the ability of buds to be formed. The correlation between the number of days a plant takes to flower and the observed number of buds on first flowering suggests that when vernalized they're negatively correlating with each other (*R=-0.25, p=<0.001*). The result in Figure 5 demonstrates that under vernalization, plants that flowers longer tend to have less buds. When non-vernalized this trend was not observed.

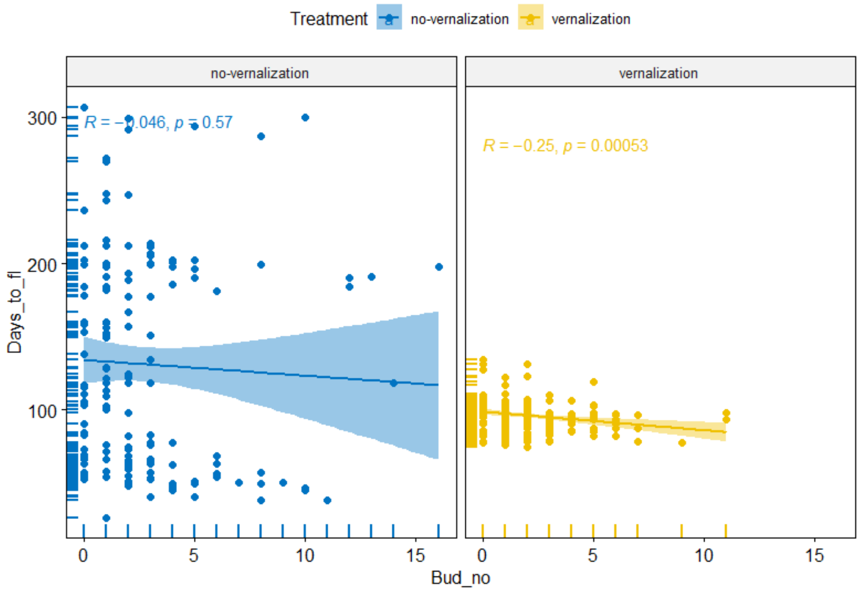


Figure 5. Scatterplots with a linear model to illustrate the relationship between the number of days a plant takes to flower (Days_to_fl) to the number of observed buds on first flowering (Bud_no). The correlation suggests that there was no significant relationship under the no-vernalization treatments (*R= 0.046, p=0.57*). When vernalized, there was a negative correlation observed (*R=-0.25, p=<0.001*).

When samples were introduced to vernalization, it seems that the number of days a plant takes to flower is less sporadic. However, when looking at the difference between the two treatments and the number of buds observed on first flowering, it is suggested that they are significantly different under a two-sample t-test (*P(T<=t) two-tail=0.021*) (Figure 6).

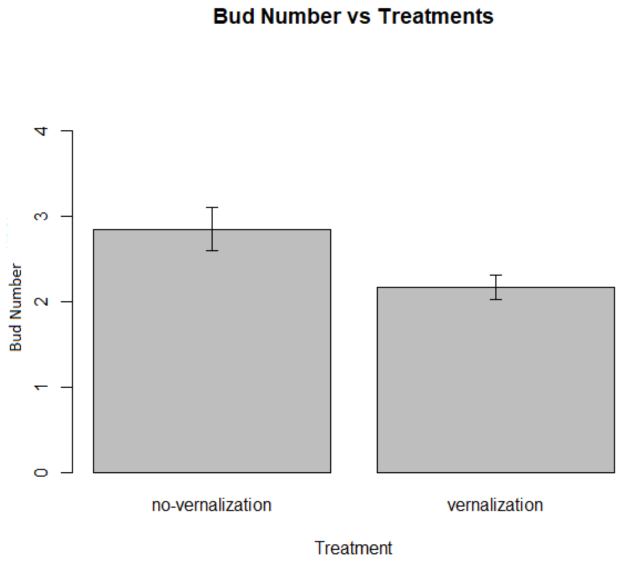


Figure 6. A barchart to illustrate the difference in mean bud number under the two vernalization treatments, revealing a significant difference between the two treatments (*P(T<=t) two-tail=0.021*).

### 5.3.2 Plant traits vs. Latitude

To look at whether any of the plant traits under the vernalization treatments were correlated with the local environmental variables such as latitude, we modelled the relationship between the three traits (plant height, number of stems, and the number of buds) against latitude and climatic variables. For this purpose, only the wild individuals from either the no-vernalization or vernalization treatments were examined. This is because environmental data for the cultivars were not reliable for our dataset, as was in chapter 2 of this thesis.

A linear model fitted on a scatterplot suggests there was correlation between the height of the plants (Cm) and latitude (°N) (Figure 7). The scatterplot was split by either non-vernalization or vernalization treatments. They revealed that with both treatments, there were no significant correlation between plant height and the latitude of locality of each of the sampled plants (for no-vernalization *R=0.1, p=0.31,* for vernalization *R=0.13, p=0.11*). This result suggests that in terms of plant height, neither no-vernalization treatments and/or vernalized treatments is correlated with the latitude of localities of each respected sample.

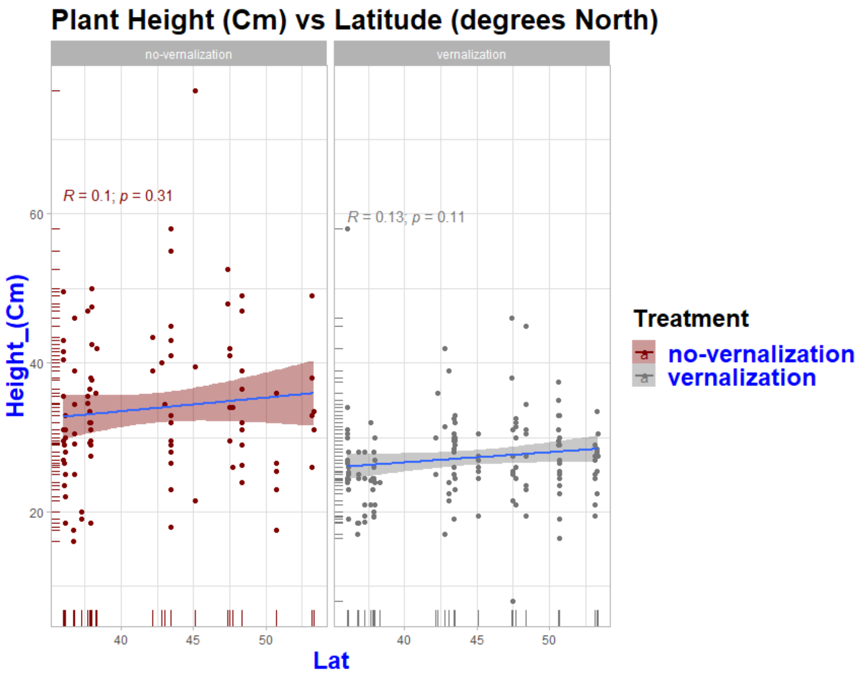


Figure 7. Scatterplots with a linear model to illustrate the relationship between the height of individual plants and their latitude of locality, showing no significant correlation between plant height and latitude (for no-vernalization R=0.1, p=0.31, for vernalization R=0.13, p=0.11).

The number of stems were also observed in relation with the latitude of the individuals in the two separate experiments. The stem numbers were observed to be less strongly correlated with the latitude of individuals when *Linum* samples were vernalised.

The result suggests relationship between the number of stems observed and the latitude locality for individuals under both no-vernalization and vernalization treatments (Figure 8). Under both treatments the relationship between stem number and the latitude is positive, with significance under no-vernalization treatment (*R=0.4, p=<0.001*). However, the significance of this relationship was not seen under the vernalization experiment (*R=0.15, p=0.06*). This suggests that when *Linum* samples are treated with vernalization, the stem number observed in some individuals is reduced to the point that significant correlation with larger stem number observed in more Northern individuals were marginally observable under vernalization.

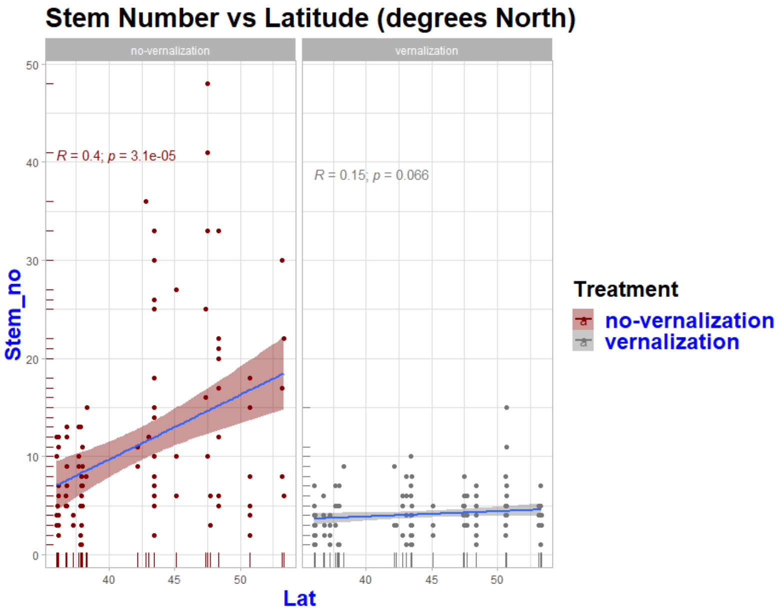


Figure 8. Scatterplots with linear model to illustrate the number of stems observed over the two treatments. The model reveals that when under no-vernalization treatments, the number of stems observed, is positively correlated with latitude (*R=0.4, p=<0.001*). This significance was not seen in the vernalization treatments (*R=0.15, p=0.06*).

Another trait which was observed during this study was the number of buds for every individual. This was also observed based on the treatments of the individuals, to see whether there was a trend between the number of buds observed and the latitude of different individuals.

Linear modelling on a scatterplot illustrates that the number of buds observed was significantly correlated with the latitude of locality under the no-vernalization treatments (*R=0.23, p=0.017*) (Figure 9). However, this relationship was contradicted when *Linum* samples were vernalized. There was a negative correlation between the number of observed buds and the latitude of individuals. Although this was the case, the negative correlation observed was not significant (*R=-0.086, p=0.29*). This result suggests that correlation between the bud number and latitude differs in the two treatments. There is a positive trend, that the more northern an individual is localized, the more buds were observed when treated with no vernalization. However, when vernalized, this trend was not observed,

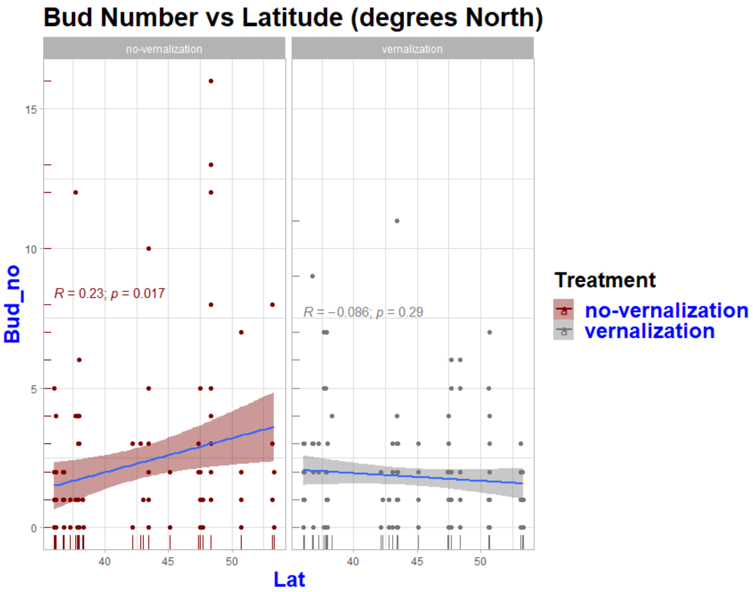


Figure 9. Scatterplots with a linear model to reveal the relationship between the number of buds observed and latitude between the individuals under no-vernalization and vernalization treatments. The Figure revealed that there is a positive correlation between the number of bud and latitude on the no-vernalization treatments (*R=0.23, p=0.017*). The trend became negative when vernalized, although not significantly correlated (*R=-0.086, p=0.29*).

### 5.3.3 Plant traits vs Climate Variables (pc1)

To relate the latitude to environmental variables, it is viable to relate it to climatic variables of the local climates to latitude. This is to see whether local climatic variable may affect plant traits which were measured during this study. This can then be related to whether any vernalization treatments may have affected traits that are measured in this study.

We saw that plant height is significantly correlated to climate variables, only in the no-vernalization treatments (*R=-0.17, p=0.006*) (Figure 10). This correlation was reduced when plants were vernalized, and the correlation became statistically insignificant (*R=-0.15, p=0.062*). This reveals that vernalization treatments have effects on the height of the overall plants in that it is reducing correlation with local climate variables seen under no-vernalization treatment. This is suggesting that plant height is affected by vernalization.

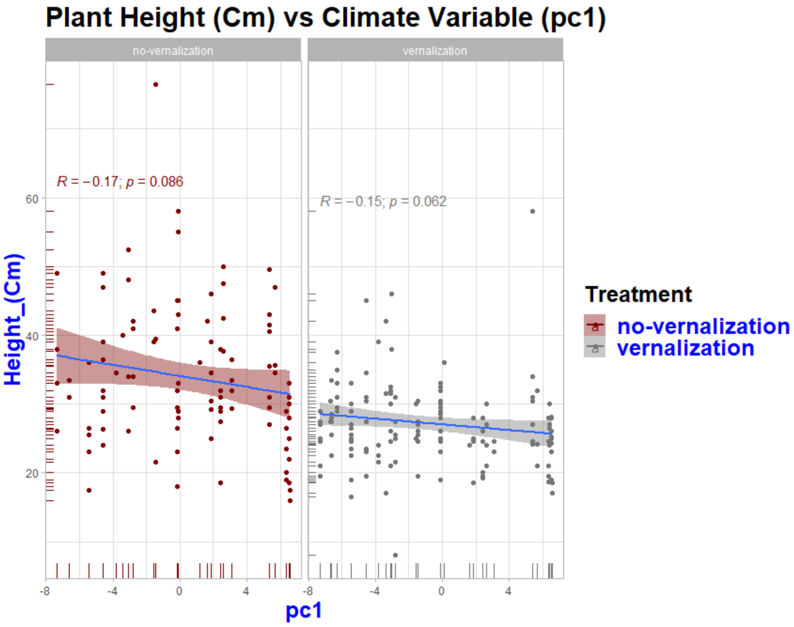


Figure 10. Scatterplots with a linear model fitted to illustrate relationships between plant height (Cm) against climate variable (pc1) for different vernalization treatments. There were negative correlations between the no-vernalization and vernalization treatments with the correlation on the no-vernalization treatments showing a significant correlation (*R=-0.17, p=0.006*).

Another measured trait was the stem number of individuals. This was also measured and modelled against climate variables (pc1). The model suggests that there was a negative correlation between stem numbers and climatic variables under the no-vernalization treatments. This correlation was statistically significant (*R=-0.4, p=<0.001*). When compared against the vernalization treatments, the significance seen under the no-vernalization treatments was observed less, although still significant. Under vernalization, it seems that stem number of individuals were drastically reduced for individuals experiencing more negative pc1 climatic variable values. The correlation between the stem number and the climatic variable are still found to be statistically significant under α=0.05 (*R=-0.18, p=0.028*).

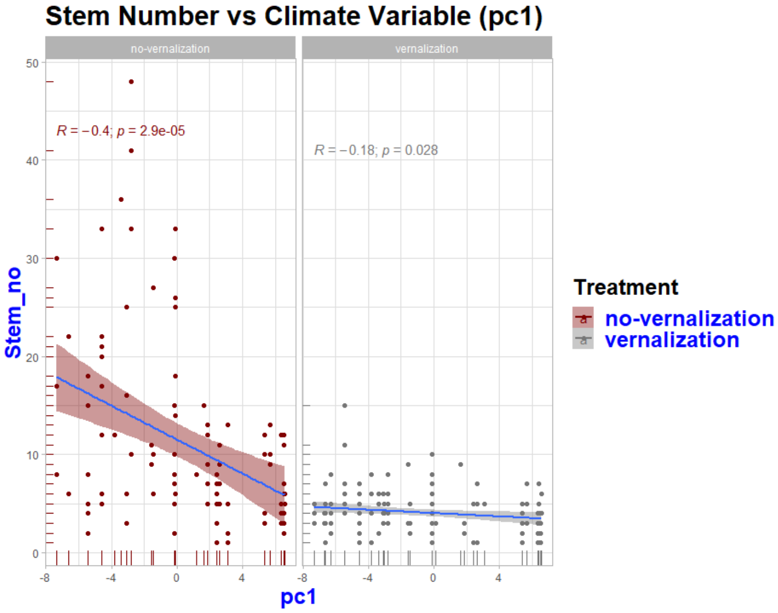


Figure 11. Scatterplot with a fitted linear model to suggest the correlation between stem number and the climatic variables of their local environments (in terms of pc1). The correlation reveals negative correlation between the no-vernalization and vernalization treatments, with a bigger significance under the no-vernalization treatments (*R=-0.4, p=<0.001*).

The number of buds observed was also seen to be negatively correlated with climatic variables (pc1) under the no-vernalization treatment (Figure 12). This corelation was found to be statistically significant when modelled (*R=-0.22, p=0.028*). In contrast with what was observed with the stem number though, the correlation between bud numbers and pc1 seems to have changed directions under the vernalization treatments. Albeit, not statistically significant (*R=0.12, p=0.15*). This difference suggests that vernalization affects the number of buds observed during this study to a point that it no longer correlates with local environments as observed under the no-vernalization treatments. This reveals when all plants were vernalized, the individuals with the most numbers of buds observed under the no-vernalization treatments, reduced their number of buds drastically through the observed effects of vernalization.

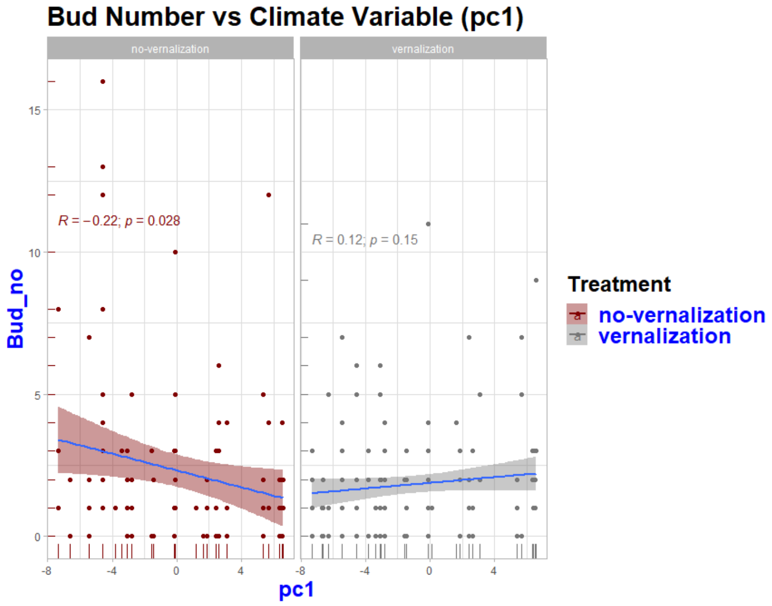


Figure 12. Scatterplots with fitted linear models which suggests negative correlation between observed bud numbers and climatic variables (pc1) under the no-vernalization treatments (*R=-0.22, p=0.028*). The correlation changed directions when vernalization was introduced. However, this correlation was not statistically significant (*R=0.12, p=0.15*).

## 5.3.4 Seed Area vs Latitude

As a part of this chapter, we examined seed size variation in wild *Linum* based on different latitudes. It is of interest to see whether seed sizes of the different wild *Linum* populations suggests significant difference to each other. We investigated whether seed size (seed area) forms a correlation to environmental variables. We investigate the question of; "to what extent can localization to a local environment in wild *Linum* be of an affect to their seed size?"

To explore this, a preliminary seed size measurements were able to be carried out, using data from non-vernalization, vernalization and as an addition, wild individuals. Microscopic images of 5 seeds were taken. The images were taken with the same light levels, making sure all 5 seeds were visible on the image. After the image were captured, it was saved into a drive and were further processed using ImageJ (Schneider C. et al., 2012). Seed areas were able to be highlighted and calculated in ImageJ. A macro was written to semi-automate the process. The macro can be found under appendix 32. In this case, vernalized individuals were labelled as "Vern18", and non-vernalized individuals were labelled as "nonvern18". Wild S0 individuals were also able to be counted observed, labelled as "Wild".

When looking at the correlation between seed area and latitude, it is found that in all treatments, seed areas are significantly correlated for all treatments (nonvern18 *R=0.51, p=<0.001*; vern18 *R=0.44, p=<0.001*; Wild *R=0.61, p=<0.001*) (Figure 13). The wild seeds were also correlating stronger than the vernalized seeds. This suggests in all cases seed area is positively correlated to latitude and that initial vernalization treatments will result in no significant change to this correlation.

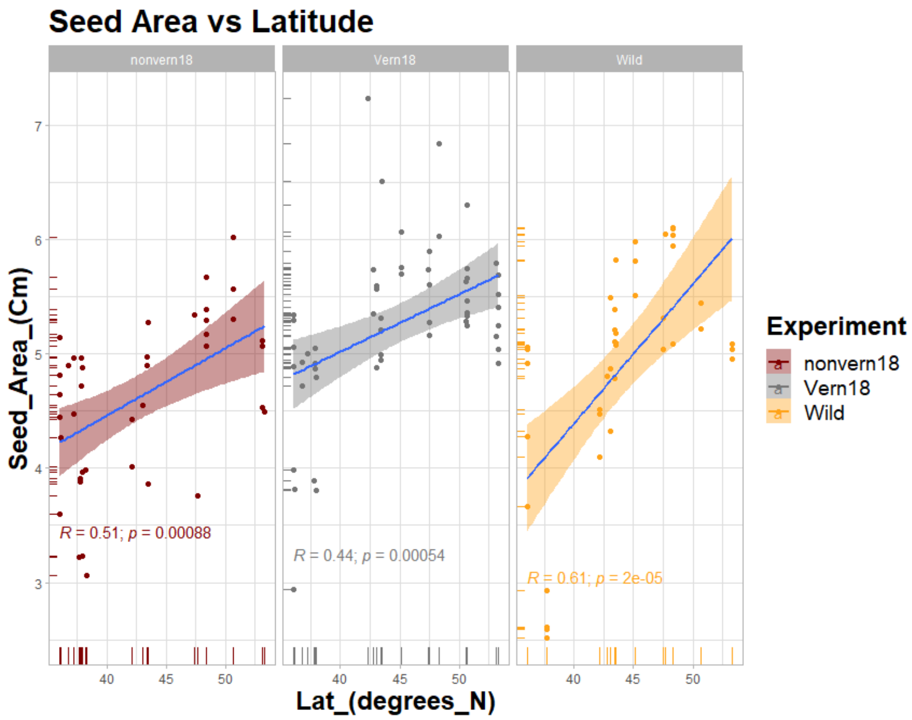


Figure 13. Scatterplots with a linear model to show the seed area (Cm) in relation to latitude for our individual *Linum* samples. The correlation illustrates that seed area is significantly correlated with latitude of locality for all our treatments, the correlation was considerable higher in the wild (nonvern18 *R=0.51, p=<0.001*; vern18 *R=0.44, p=<0.001*; Wild *R=0.61, p=<0.001*).

## 5.3.5 Seed Area vs Climate Variable (pc1)

For completion, it is with interest to further see whether seed size in terms of total seed area observed for different individual sampled during this study. The scatterplots with linear modelling suggest negative correlations between seed area and climatic variables (Figure 14). Negative pc1 values represent loading values for colder climates (see chapter 2, under section 2.2.1 Samples and Experiment”). This data suggests that in all treatments, the seed sizes seem to be larger in area in individuals found in a colder climate and that vernalization have little effects to change this significance. In the wild individuals, the correlation seems to be stronger.

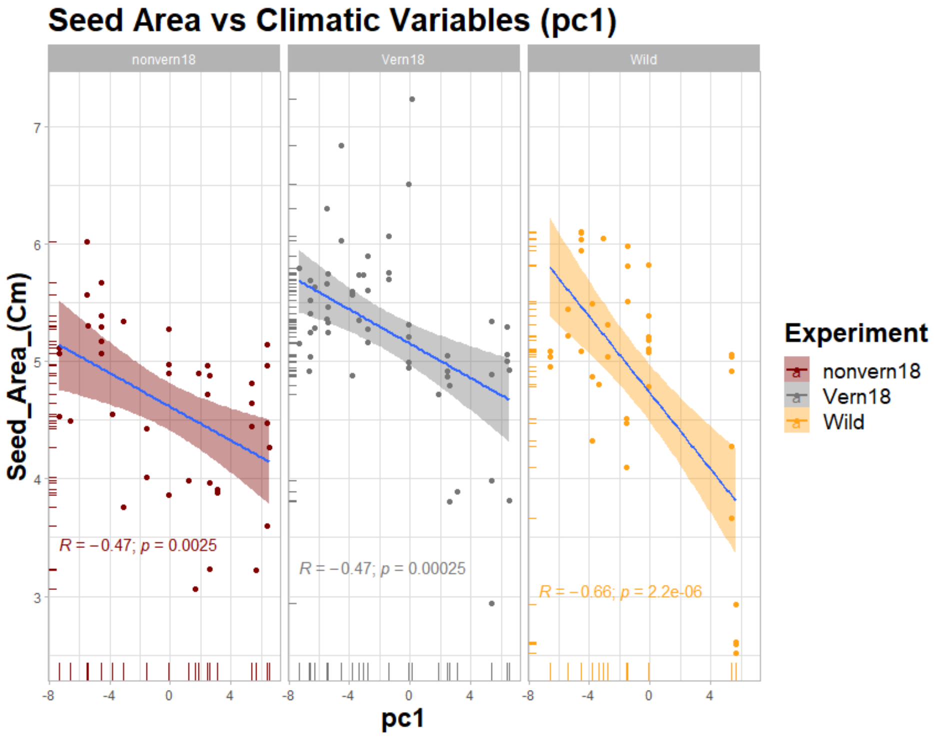


Figure 14. Scatterplot with a linear model to illustrate the correlation between the seed area (Cm) and the climatic variables (pc1). The correlation suggests that there is a negative correlation between seed area and climatic variables in all treatments (nonvern18 *R=-0.47, p=0.0025*; vern18).

---

## 5.4 DISCUSSIONS AND CONCLUSIONS

Additional results to investigate whether traits maybe associated with the number of days a plant may take to flower suggests that there is a relationship between the number of days individual plants takes to flower and plant height and stem numbers. It seems that, under no-vernalization, when individuals flowered earlier, they tend to be taller at first flowering. In the literature, there are many cases where number of days a plant takes to flower is positively correlated with plant height (Singh N. et al., 1995; Gupta S. et al., 2021). In *L. bienne* we expected that this would be the case when they are under no vernalization stress. We hypothesized that under vernalization this correlation between number of days to flowering and plant height would be reduced. We suggest that vernalization acts as an abiotic stress in wild *L. bienne*.

### 5.4.1 Number of Days to Flower and Plant Heights

When vernalized we observed that, correlation between number of days to flower and plant height was not significant, as the number of days for some individuals to flower were reduced greatly. This illustrates that under overall vernalization treatments, correlation between the number of days to flower and plant height at first flowering were not observed. When comparing average plant heights, there is a significant difference between vernalized and non-vernalized individuals. This suggest that under vernalization, both number of days (as found in chapter two) and plant heights were significantly reduced. This illustrates vernalization as an abiotic stress in *L. bienne* samples found across Western Europe. In literature, there are many examples where vernalization reduces plant heights (Dole & Wilkins, 1994; Clough et al., 2001; Wu et al., 2016). In other suggestions, some of the shorter plants during first flowering was able to sense vernalization and flower much quicker when compared to under no-vernalization treatments. A study using a variety of grass species found that flowering phenology is associated with the overall plant height under different environments (Soto-Cerda et al., 2014). This suggests that plants which can flower quicker may have different overall height to plants that are slower to flower. In our *Linum bienne* samples, it is illustrated that reduction of the number of days to flower is also reflected by the reduction in plant height observed, which was significantly different between vernalized and non-vernalized individuals. We accepted that under vernalization, both number of days to flowering and plant height are significantly reduced in *L. bienne* found across Western Europe.

### 5.4.2 Number of Days to Flower and Stem Numbers

There was also a positive correlation between the number of days individuals takes to flower and stem number under the no-vernalization treatments. It seems that the longer an individual takes to flower the more stem it can develop before first flowering. This correlation was reduced when vernalization was implied to the wild individuals, with individuals under vernalization treatments significantly forming a smaller number of stems. However, the correlation between the number of days to flower and the number of stems was still significant. This shows that stem number is highly correlated to the number of days individuals can flower initially even when wild individuals were exposed to vernalization. When testing against different treatments, individuals exposed to the vernalization experiments have significantly a smaller number of stems. This suggests an effect of vernalization in the number of stems wild *Linum* individuals can form. There is suggestion in the literature that different plant populations may have different number of stems according to their local environment (Sultan, 2000; Szakiel et al., 2010; ATES., 2011). Vernalization suggests stark differences in stem number when compared to non-vernalization treatments, which suggests that in *L. bienne*, this is also true. In other cases, vernalization affects the morphology of stem formation in *Phleum pratense* . (Seppänen et al., 2010) This illustrates the potential of vernalization as an abiotic stress to affect developments of stems in wild *L. bienne* found across Western Europe.

### 5.4.3 Number of Days to Flower and Bud Numbers

In terms of the number of buds forming, our results suggest that under the no-vernalization treatment this measure was not significantly correlated to the number of days an individual is able to form their first flower. When exposed to vernalization however, the number of buds observed seems to negatively correlate with the number of first flowering as they were significantly reduced due to vernalization as an abiotic stress. This suggests that under vernalization requirements, there may be groupings of plants which can form a higher number of buds with a shortened number of days to first flowering. In all three cases of the trait measures, including number of buds, variation in correlation between the number of days to flower and traits measured in this chapter occurred . In all cases though, vernalization reduced the number of observed bud numbers significantly. This suggests that vernalization is an abiotic stress that wild population are potentially locally adapting into.

### 5.4.4 Effects of Vernalization in Plant Traits

With the above additional results for the plant traits (plant height, stem number, bud number) considered, it is suggested that the effect of vernalization leads to significantly lower values in all three trait measures. This suggests that under vernalization, wild *Linum* samples in this study experienced significant reductions in desired plants traits such as height and plant stem number. In the model *Arabidopsis thaliana*, it has been observed that traits such as leaf sizes were reduced under vernalization (Hopkins., 2008). The study also found a latitudinal cline to this response in vernalization. This suggests that morphology of plant traits may be affected by vernalization and that environmental variables may also play a part. In the study for this chapter, it is found in wild *Linum bienne* samples across Western Europe that plant height, stem number, and bud number is negative affected by vernalization. Findings in these traits under vernalization, adds to the suggestion that vernalization is an abiotic stress and that variation in sensitivity observed can be seen as potential local adaptation mechanisms between wild *L. bienne* populations found across Western Europe. To thrive in a colder niche there could be a case that some wild *L. bienne* populations are more locally adapted to colder climates through requirement of vernalization. In addition, manipulation to flowering under different seasons/environments can also be studied further as suggested under other studies observing vernalization in *Calandrinia* (Cave & Johnston, 2010). Reduction in desirable traits also suggests that the loss of vernalization under *L. usitatissimum*, observed in chapter two of this thesis could be a benefit in terms of agriculture. Loss of vernalization could mean that negative effects on traits such as plant height, stem numbers and bud numbers observed under vernalization in the wild *L. bienne* is selectively avoided in Northern European Winter and Spring *L. usitatissimum* types. Although this was the case, we assumed that the data is normally distributed, and that values of trait measurements, and other variables are independent of each other. Traits observed may be dependent of each other and that the same genes may influence multiple traits observed in this chapter. These analyses do not consider these possibilities.

The response to plant height in both vernalization treatments, suggests there was no correlation with the latitude of locality for these plants. However, when looking at local climatic variables, there was negative correlation between the local climate variable against plant height when no-vernalization occurred, with individuals from warmer climate (bigger pc1 values) being taller. This correlation became less significant under vernalization treatment. This suggests that perhaps climatic variables are having stronger effects to plant height, in our case, individuals under no-vernalization seems to be taller in colder climate (smaller pc1 values). Vernalization also plays a part in this, when plants were exposed to vernalization, it seems that this correlation became less significant. It would be viable to suggest that the process of vernalization induces stress to wild *L.*

*bienne* individuals. To undergo vernalization and flower at the correct time, individual requiring vernalization may have less favourable traits. Loss of vernalization requirement in the cultivated *L. usitatissimum* in Northern Europe (as discussed in chapter two), could mean that this process is bypassed in some cultivar types and stress induced by cold for plants to require vernalization is not a factor. This may form as one of the factors enabling the cultivars to have form better in traits of agricultural interests for their biproducts. We didn't compare traits between cultivars and wild species when vernalized as cultivated samples were not included in this chapter for analysis. This will be of interest for future study.

### 5.4.5 Effects of Climate in Plant traits for wild *L. bienne*

When looking at other traits against latitude, such as the stem number, there was also suggestions that the number of stems observed is correlated with the latitude of locality for the individuals for the no-vernalization treatment. In this case the correlation seems to be positive, with more Northern plants revealing a greater number of stems. This was also supported when looking at the climatic variable. A negative correlation was observed which suggests that individual from warmer climates (higher pc1 values) having a smaller number of stems. When looking at individuals experiencing vernalization, the stem number able to be formed was not significantly correlated to either latitude of locality or local climatic variables. This suggests varied morphology in response to vernalization for our *Linum* collection. The bud number observations suggest there was a positive correlation between the observed bud number and latitude of localities under no-vernalization treatments, this suggests more Northern individuals were able to form more buds under no-vernalization. When vernalized however, the correlation became negative, with more Northern individuals suggesting for be forming less buds. However, this change of direction in correlation, was not observed to be significant. When looking at bud number and pc1, it is supporting the traits seen with latitude, with the two treatments showing contrasting correlation. In this case negative correlation between bud number and climatic variable (pc1) suggests that individuals found in warmer climate seems to form less buds. The correlation changed direction under vernalization treatments, however, as seen with the latitudinal correlation, this change in correlation direction was not statistically significant.

In all three cases of the measured traits, there is suggestion of differentiation in strategy for different individuals and when no-vernalization and vernalization treatments were implied. This suggests there may be genetic interest for looking at these differences in morphology, as was found under chapter two of this thesis when looking at relative gene expression of flowering time genes and the number of days an individual takes to flower. In a study with the majorly cultivated durum wheat, a study has found different phenology and trait strategies in different landraces. This was also linked to differences in genetic variation responses for the vernalization gene (*VRN1*) and photoperiod sensitivity gene (*PPD1*) (Royo et al., 2020). There is suggestion in the literatures which highlight different morphology and phenotype of species of plants due to vernalization and suggests that this is also linked to genetic variation. It is likely that differences seen in plant height, stem number, and bud number in wild *Linum* is a result of genetic variation due to vernalization responses. It will be with great interest in the future to look at specific genes which may play a part in plant growth, specifically for stems and buds' formation to see whether variation between individuals still correlate and to see whether variation seen between no-vernalization and vernalization treatments can be linked genetic factors.

### 5.4.6 Effects of Climates in Seed Areas for wild *L. bienne*

We also observed seed sizes in terms of seed areas, against latitude of locality for every individual as well as local climatic variable for them. It is suggested that there was significant variation between seed area, and these were all correlated with latitude and local climatic variables (pc1). The correlation was still retained under no-vernalization and vernalization treatments. Positive correlation with latitude suggests that in all treatment cases, the more Northern individual's seeds were bigger than those of their Southern counterparts. This also translates to the correlation seen in climatic variables, with negative correlations, suggesting that individuals from colder regions having the bigger seeds. Bigger seeds may imply better fitness in plant reproduction, smaller seeds are however, easier to produce, suggesting this trade-off between seed size and the number of seeds a plant can make for a given amount of energy (MOLES et al., 2004). In a study with wild barley (*Hordeum vulgare ssp.*), there is a suggestion that variation in seed sizes contributes to plant growth and reproduction, with larger seeds often found to have better fitness (Giles., 1990). A study on different *Glycine* species found that seed mass tends to be larger at Northern latitudes. They also found that the major contributing climatic variable to this are temperature and daylight availability (Murray et al., 2004). These agree with our finding that *Linum* seeds also found to be larger in Northern populations. When looking at the wild seeds in our observations, the correlation between seed size and latitude and climate variables were found to be stronger. In sand rice (*Agriophyllum squarrosum*), there was also geographic variation of seed size due in wild populations. In *Linum*, the seed sizes are strongly correlated to local variation in climatic variables. It would be of interest to genetically evaluate this to see whether the strong correlation found in this thesis had genetic implication for genetic basis which may reveal evidence of genetic variance. In the past, studies looking at genetic variance for variation in seed sizes in different plant species have found no correlation between genetic variation and variation in seed sizes, suggesting that seed size could not have evolved due to natural selection (Wolfe., 1995) (Schwaegerle & Levin., 1990). It would be of good interest to genetically evaluate the seed size variation under different environmental variables as illustrated in this thesis.

# GENERAL DISCUSSION

Experiments in this thesis were set out to investigate adaptive strategies in *Linum*, using two species (*L. usitatissimum* and *L. bienne*). We were interested in variation in traits and genetic materials of both wild and cultivated species to observe signals of local adaptation. We were particularly interested in flowering mechanisms. We found that as a temperate plant, requirement to vernalize varies in wild *L. bienne* found across Western Europe. However, we saw no significant changes to the flowering initiation when we vernalized the cultivars *L. usitatissimum*. Examples in the literature suggests that there is variation in vernalization sensitivity between different *L. usitatissimum* varieties grown under different seasons (Darapuneni M. et al., 2014) (see section 2.5 of chapter two). We suggest that there is no sensitivity to vernalization in winter and spring types *L. usitatissimum* found in the Northern part of Western Europe. Comparison against lines from Southeastern European regions such as Turkey, where *L. usitatissimum* is cultivated in different seasons, will be of interest for observation of vernalization sensitivity variation in *L. usitatissimum* types found across two different growing climates across Europe.

In *L. bienne*, we observed a positive trend on the number of days individuals took to flower and their latitude of origin. The finding suggests that wild individuals further North are more sensitive towards the requirement to vernalize, with substantial reduction in the number of days to flower when vernalized. This was further reinforced by genetic expression studies of 3 flowering time genes (*FT, CO,* and *GI*). We found no significance between the expression response of each of the tested genes when looking at vernalized and non-vernalized treatments. However, we observed the relative expression of *FT* to be positively correlated with latitude of origin in wild *L. bienne* individuals. With this observation, we accepted the hypothesis that variation in flowering initiation occurs in wild *Linum* across Western Europe, and this is reflected in the relative gene expression of the floral integrator *FT* against latitude and climate of origin of samples. This trend was not observed in any of the other flowering time genes tested.

We also acknowledge that the sample size for the relative gene expression study is relatively small. When we removed the Southern-most individual, we no longer observed a positive correlation between *FT* expression and latitude. Interestingly, flowering initation were still positively correlated with latitude, suggesting that perhaps expression of other floral integrators and/or repressors could be affecting the flowering initiation. Furthermore, we also saw a relationship between the number of days individuals took to flower and the local climate. We saw that expression of *FT* were correlated with climatic variables. This introduces the possibility that, wild population are locally adapting to the environmental variables such as precipitation, solar radiation, and temperatures. Wang J. et al., 2018 illustrates that for perennial plants, the locus *FT* forms part of the genomic region responsible for local adaptation (Wang J. et al., 2018). We make a case that evidence of local adaptation can be observed in wild *L. bienne* across Western Europe, based on the relationship seen between *FT* expression and latitude and climatic variables. Although this is the case, there is a possibility that variation in requirement to vernalize can arise from mechanisms such as genetic drift. We didn't have precise location data for our cultivars. It would be of further interest to compare the relationship of *FT* expression with latitude/climatic variables for the different *Linum* species. We expect that, due to domestication and the loss of requirement to vernalize observed in this study, our *L. usitatissimum* variants would not form a relationship with latitude/climatic variables. This would suggest a contrast between the wild *L. bienne* and the cultivated *L. usitatissimum*. Whilst the requirement to vernalize may be artificially selected in the cultivars, the wild suggests that flowering initation and requirement to vernalize may be driven by local adaptation under various climate. Although this was our finding, we suggest taking the results with precautions. Along with a relatively small sample size, relative gene expressions were comparative only against one HKG. We propose a future study with bigger sample sizes from a wide range of latitude, specifically more Southern populations and their floral gene relative expression compared against multiple HKG's. This is to robustly reflect the correlation with *FT* expression and latitude. In addition to this, it would be ideal to observe expression of genes directly in the vernalization pathway such as *FRI*, *VRN1* and *FLC*. We expect that the relative expression of these genes is related to the latitude and local climate in which the wild population are found since they affect *FT* expression.

We saw in chapter two that the number of days wild *L. bienne* takes to flower is related to the latitude. This implies that temperature, as a variable, may control for the number of days to flower. To investigate whether temperature had further effects on flowering mechanisms, we continued to observe responses in floral organs. We explored pollen viability in response to reduction (-5$^{o}$C from initial temperature) and addition (+5$^{o}$C from initial temperature). We see a response in reduction of pollen count for both reduction and addition of temperature. When looking at pollen tube count, this varied within all our temperature treated samples. However, when scaling this with the pollen (proportion of pollen tube forming to pollen count) only the addition in temperature shows significant reduction in the proportion of pollen tube forming. This suggests that *Linum* pollen tubes are potentially sensitive to increases in temperature. To reinforce this, we found the proportion of pollen tube able to reach the ovary to be significantly reduced under the +5$^{o}$C temperature treatment. This was observed in both species. This may suggest that pollen ability to germinate is limited to temperature or the female floral organs are not able to sustain pollen tubes under heat stress. This suggests the interest in looking at ability of female counterparts to sustain pollen tube growth under different temperature for future studies in this area.

We concluded this chapter by observing whether latitude/climate variables are correlated with the amount of pollen and pollen tubes observed in *L. bienne* when treated to the different temperature treatments. We found no correlation between the number of pollen and latitude/climate. However, the number of pollen tube and its proportion to the number of pollen observed a correlation with latitude and climate in the 20°C (-5°C) treatment. Furthermore, correlation with latitude is found to be positive, illustrating that more Northern individual can form more pollen tubes. This may suggest potential adaptation to the colder temperature for the more Northern individual in relation to its ability to form pollen tubes and thus increasing its chances of germinating. We suggest that this also formed because of local adaptation found in the Northern population of Western European *L. bienne*. From this chapter we reinforced the suggestion that local adaptation between the North and South wild *L. bienne* in Western Europe occurs, as was illustrated in chapter two. For future research, it would be of interest to look at this from the female counterpart's perspective. We suspect that ovary would be sensitive to heat-stress and stigma will have varying ability to sustain pollen tube growth. In addition to this, it would also be beneficial to conduct a GWAS study to identify the genes which may play a role in cold resistance in the wild which maybe linked to the higher amount of pollen tube able to be observed under wild individuals.

In chapter four we analysed 100 Western European *Linum* individuals and their population genetics. One of the main results from this chapter was the observation of Northern and Southern groups in the wild *L. bienne*. There seems to be structuring between the Western European wild *L. bienne* population as was inferred in section 4.4.1 "Results: Genetic Structure". We observed initial signal of potential variation in breeding strategies of *L. bienne* found across Western Europe in our population diversity analysis. There is potential that the Southern group of wild *L. bienne* is more outcrossing in terms of their breeding strategy and thus causing this genetic structuring we see when illustrating genetic distance tree. In the literatures, it is found that various mode of seed dispersal often affects a plant's breeding strategy. They found that animal-dispersed plant species exhibited higher levels of genetic diversity and lack of inbreeding (Nazareno et al., 2021). This further suggests that there is potentially more variation within Southern *L. bienne* inidividuals in Western Europe. All measures of population diversity suggests that heterozygousity is higher in the Southern cline, with relatively little amount of heterozygousity observed in the Northern group. Measures of population divergence such as Tajima's D also suggest that Southern population is more recently expanded and are expanding more than the Northern group. These findings suggests that there is potential variation in breeding strategies and dispersal of North and South groups of *L. bienne* found in Western Europe. The implication of variation in breeding strategies could result in the variation we saw in chapter two whereby Northern individuals tends to be more sensitive to vernalization as was expressed by their variation in relative ge expression under different latitudes. Furthermore, we saw that ability of pollen tube forming and germination processes is correlated with individual's latitude when treated to "cold" treatments (see section 3.4.8 "Results: Pollen and Latitude of Origin"). Genetic structuring and diversity analysis in chapter four supports the findings that Northern and Southern clines of *L. bienne* found in Western Europe is potentially locally adapted to their environment.

In addition to the observation between Northern and Southern clusters, we observed a Mediterranean population which is significantly diverged from other populations. This could be a result of increased rate of outcrossing in nature we saw in the Southern clines when looking at population diversity measures. It would be of interest to include population from the Eastern European region and see if this cline between North and South population could also be observed between Western and Eastern European wild *L. bienne*, with Eastern population potentially more outcrossing in breeding strategy. In future studies investigating Western and Eastern *L. bienne* population, wild *L. bienne* seeds needs to be collected from Eastern European regions such as Turkey and the Eastern Mediterranean. There is no availability of S0 seeds from these wild population to date.

We also acknowledge that we may have lost some valuable data by mapping to the *L. usitatissimum* whole genome. This was inferred from the higher number of SNPs gained from *de novo* mapping (see section 4.5.4 "*De_novo* mapping"). In addition, it is worth exploring other NGS techniques to imply further structuring based on a  larger coverage of the genome. This is because of disadvantages to using ddRADSeq to sequence our short reads. Our ddRADSeq only studies 250-500bp of sequences per individual cut by the digestion process of ddRADSeq and require high quality DNA which may impact the number of individuals able to be analysed and outputted from the ddRADSeq procedures. We also looked at two species with potential divergence in our wild populations. Other recent short-reads NGS techniques that can be looked at for observation of mulptiple species is Sequence capture. Sequence capture holds more promise for obtaining data sets that are comparable across species and for calibrating parameter estimates for demographic or phylogenetic studies (Harvey et al., 2016). Sequence capture also doesn't require a high quality of DNA as opposed to RAD Sequencing processes. There is potential that individuals with a relatively lower quality of DNA could be analysed further with Sequence capture (Harvey et al., 2016). In addition, studies using Sequence capture can capture larger sizes of sequences (300-1200bp) which covers 972 genes (Sanderson et al., 2020). Future studies can look at the potential of using Sequence capture for population genetic analysis in *L. bienne* across Western Europe.

When looking at traits in chapter five, we found evidence that plant height, stem number, and bud number is significantly reduced by the introduction of vernalization. This suggest that vernalization is an abiotic stress variable in wild *L. bienne* found across Western Europe. When under no-vernalization, we saw positive correlation between stem and buds formed at first flowering and latitude. This illustrates that under no vernalization stress, more Northern individuals can form more stems and buds. However, under vernalization there was no observed correlation between traits and latitude/climate. This suggest that, in the wild, vernalization acts as abiotic stress and as was found in chapter two, vernalization responses varies across the wild *L. bienne*. In terms of local adaptation in wild *L. bienne*, the ability Northern individuals to form more stems and buds, under no-vernalization can suggest that seasonal queues, especially regarding vernalization, may occur in the Northern cline. This ability to form more stems and buds under no-vernalization was seen less in the Southern cline, suggesting the potential for local adaptation regarding vernalization requirement in the Northern cline, which are expected to be more sensitive to vernalization due to colder climates. This was also supported by higher expression of the floral integrator *FT* as was suggested in chapter two. In future studies, it is of interest to look at expression of genes directly affecting number of stems and buds to reinforce the variation seen within this thesis with genetic expression studies.

We further support the potential local adaptation between Northern and Southern clines of *L. bienne* population using seed sizes. Under three growth conditions, seeds areas are positively correlated with latitude and negatively correlated with climatic variables, mostly attributed to colder temperatures. This suggest that Northern individuals seems to have a bigger seed area to the Southern individuals, which can infer that trade-off between seed size and potential to produce more seeds between the two clines. In the wild S0 seeds, this correlation between seed area and latitude is more strongly observed, which may suggest that seed area is a variable for measuring seed size variation in wild *L. bienne* for future studies regarding local adaptation and breeding strategies of wild *L. bienne*.

With the above five chapters concluded, we illustrated potential signal of local adaptation in wild *L. bienne* to Northern climates, requiring vernalization. We observed no sensitivity of *L. usitatissimum* in our sample set to vernalization. We expect there are variation between various *L. usitatissimum* variety regarding vernalization sensitivity, especially for non-winter types. As was seen in *L. bienne*, this vernalization requirement can suggest reduction in traits that are beneficial to argriculture, and so loss of vernalization in winter types maybe beneficial for these traits of interest. We imply that the loss of vernalization is agriculturally beneficial, however, this may limit growth time and cycle of plants under different seasons, thus resulting in seasonal types of *L. usitatissimum* varieties. Wild Northern population could be studied to understand mechanisms of vernalization in *Linum*, this may add to breeding purposes, providing *Linum* cultivars that are perhaps more resistant to colder climates an can be grown in different seasons.

In terms of wild *L. bienne* population found across Western Europe, population structure suggests that gene flow between the Northern and Southern group in our samples are limited due to the grouping suggested by the Ward's D tree based on Nei's genetic. Genetically, we saw little evidence of geneflow between the wild population which is a signal of locally adapting populations (Boshier et al., 2015). Phenotypically, we saw variation for flowering traits for Nortern and Southern group of population within our sample set as a signal of this local adaptation.

Although this was the case, in our PCoA we saw a few outliers to these groupings suggesting a cline between the Southern and Northern populations. This suggests a limited geneflow between the two groups, albeit we think that this geneflow is limited to one or two individuals in our sample set. We suggest future studies to look at more wild population originating in Northern Spain, as this is where we saw potential outliers which may suggest gene flow between the Southern and Northern groups of wild *Linum* observed in this study. Further divergence can be seen in our Mediterranean population. Population divergence analysies suggests that the population E-12 in our sample set is distinct to both Northern and Southern group with a higher number of private alleles and heterozygousities. This suggest that perhaps this population is part of a diverging group of wild *L. bienne*. Higher allelic richness can also suggest a different approach to breeding strategies than seen in the rest of the Western European wild population which may allude to differentiation in mechanisms such as seed dispersals.

Less heterozygous alleles in the Northern group of population also infer less allelic richness and potentially more inbreeding, suggesting that these Northern group have a more conserved genetic makeup, potentially counting for local adaptation to a colder climate as we initially observed under flowering mechanisms such as vernalization. However, it is interesting to find that there are intermediate individuals forming a cluster between the Southern and Northen population in our PCoA analyses. This suggests that perhaps several populations based in Northern Spain acts as a geneflow meditator, connecting the two distinct Western European wild flax population. This could suggest that, although the Northern group of wild *L. bienne* population are more conserved and locally adapted to Northern climates, they have diverged from a Southern population. This suggests implications for *Linum* to diverge into colder climates and are locally adapting to these climates. This implies the importance of colder climates for natural selection of wild plants and their diversification. This could impact availability of a wider genepool for study on mechanisms to counter cold climates in plants, vernalization to name one. These wider genepool are useful for future breeding purposes agriculturally. In a warming climate, these wider genepool may be threatened as they are found to be more conservative and inbred than their Southern relatives.

Other studies suggests that the Mediterranean is a “hotspot” for recent plant diversification (Buira et al., 2020). We think that wild *L. bienne* population originating further in the Mediterranean as opposed to Western Europe may be more diverse and distinct than previously thought for *L. bienne* across Western European region. Diverging Mediterranean populations may also adapt at a different rate to their environment and are more outcrossing (as was observed in population E-12). We think that Mediterranean wild *L. bienne* population may provide a wider genepool for genetic study of *Linum* and have the potential to provide diversification to breeding strategies in *Linum*. To develop this further, we suggest collection of further wild *L. bienne* from across the Mediterranean region to confirm this genetic diversity from the Western European relatives, seen in this study.

We take these implications within this thesis with precautions as the number of sample sets are relatively small and minimum representative for population genetics purposes was only seen to three individuals. For future relative gene expression studies, a confirmation using multiple HKG’s needs to be conducted as future research towards the results found within this thesis. However, signals of local adaptations can still be observed even with these limitations in the wild *L. bienne* found across Western Europe. This would imply potential traits of interest in relation to local adaptation which maybe playing a role towards resistance to abiotic factors such as vernalization. There are also implications for further studies observing cold-resistance genes and pollen viability in wild *L. bienne* to observe mechanisms of cold resistance in future *Linum* GWAS studies that maybe beneficial for implication to *Linum* agriculture.

# APPENDICES

## Appendix 1: A list of *L. bienne* and *L. usitatissimum* samples collected for this thesis

### *L. bienne*

| Pop | Lat | Long | collected | collector | details |
|---|---|---|---|---|---|
| 1 | 36.80044 | -5.39258 | 04/06/2016 | RPB | Llanos del Rabel trail |
| 2 | 36.08092 | -5.62553 | 05/06/2016 | RPB | Virgen de la Luz Santuary Facinas |
| 3 | 36.03633 | -5.55589 | 05/06/2016 | RPB | Guadalmesi |
| 4 | 36.15083 | -5.70494 | 05/06/2016 | RPB | El Nene, facinas |
| 5 | 37.25853 | -6.09722 | 07/06/2016 | RPB | Puebla del Río-Aznalcazar, Sevilla |
| 8 | 37.70194 | -5.83206 | 08/06/2016 | RPB | Road 432, Km 12 Road to El Pedroso |
| 6 | 37.93551 | -5.71117 | 08/06/2016 | RPB | Constantina-Cazalla de la Sierra, Sevilla (trail) |
| 7 | 38.25336 | -4.31739 | 08/06/2016 | RPB | Cardeña-Villa del Río, Cortijo Tejoneras, Córdoba |
| 9 | 37.8085 | -6.42881 | 09/06/2016 | RPB | N-433 before exit to Zufre-La Granada de Riotinto, Huelva |
| 10 | 37.88211 | -6.61781 | 09/06/2016 | RPB | Linares de la Sierra, Huelva |
| 11 | 38.33105 | -3.58086 | 14/06/2016 | RPB | La Aliseda, Finca La Inmmediata (Km 3), Jaen |
| 12 | 42.31008 | 3.151889 | 17/06/2016 | RPB | Palau-Savereda |
| 13 | 43.02902 | -3.24035 | 19/06/2016 | RPB | Quincoces de Yuso-Relloso, Burgos |
| 14 | 42.79404 | -3.42474 | 20/06/2016 | RPB | Tartales de Cilla |
| 15 | 43.46278 | -3.65331 | 20/06/2016 | RPB | Cantabria-Carriazo.Galizano |
| 16 | 42.17089 | -8.68382 | 21/06/2016 | RPB | Universidad de Vigo |
| 17 | 50.69147 | -1.0954 | 24/09/2016 | RPB | Bembridge, 1st stop, Isle of Wight |
| 18 | 50.68183 | -1.07492 | 24/09/2016 | RPB | Bembridge, 2nd stop, Isle of Wight |
| 19 | 42.17089 | -8.68382 |  | RPB | Universidad de Vigo |
| Tor | 36.73956 | -3.92635 | 11/07/2013 | ACB | Torrox Costa, Malaga |
| Lla | 43.40738 | -4.68753 | 21/07/2014 | ACB | Llanes, Asturias |
| Mat | 48.35697 | 4.458631 | 29/06/2016 | ACB | Mathaux, Aube |
| Vil | 45.09393 | -1.05034 | 02/07/2016 | ACB | Villeneuve, Charente Maritime |
| Bro | 47.16704 | -0.20705 | 04/07/2016 | ACB | Brossay, Maine et Loire |
| Roc | 47.38702 | -0.52574 | 04/07/2016 | ACB | Domaine de Rochambeau, Maine et Loire |
| Saf | 47.4961 | -1.59254 | 05/07/2016 | ACB | Saffre, Loire Atlantique |
| Tal | 47.6997 | -3.45465 | 06/07/2016 | ACB | Pointe du Talude, Morbihan |
| Fro | 53.30782 | -2.71877 | 02/09/2016 | ACB | Frodsham, Cheshire |
| Tym | 53.30307 | -3.55328 | 02/09/2016 | ACB | Tyr_Mawr_Holiday_Park, Denbighshire |
| Sut | 53.35291 | -0.95927 | 09/09/2016 | ACB | Sutton_Cum_Lound, Nottinghamshire |
| Man | 53.13731 | -1.14367 | 10/09/2016 | ACB | Mansfield, Nottinghamshire |
| Dor | 50.6 | -2.01 |  | Emorsgate Seeds | Dorset |

### *L. usitatissimum*

| Shortname | Name | Type | Subsp | Country | Source | Code | centrelat | centrelong |
|---|---|---|---|---|---|---|---|---|
| Pri | Primus | oilseed | mediterraneum | Italy | IPK | 247707 | 41.87194 | 12.56738 |
| Rab | Raba0189 | oilseed | mediterraneum | Morocco | IPK | 247713 | 31.7917 | -7.09262 |
| Gis | Gisa | oilseed | caesium | Italy | IPK | 260080 | 41.87194 | 12.56738 |
| Tin | TineTammesLila | oilseed | caesium | Netherlands | IPK | 236553 | 52.13263 | 5.291266 |
| Mon | Monarch | fibre | elongatum | UK | IPK | 255846 | 55.37805 | -3.43597 |
| Ome | Omegalin | oilseed spring | | France | TDL | | 46.22764 | 2.213749 |
| Lir | LiralCrown | oilseed | caesium | UK | IPK | 231715 | 55.37805 | -3.43597 |
| Ble | Blenda04C | fibre | elongatum | Netherlands | IPK | 225632 | 52.13263 | 5.291266 |
| Ara | Aramis | fibre spring | | France | TDL | | 46.22764 | 2.213749 |
| Ari | Ariane | fibre | elongatum | France | IPK | 254727 | 46.22764 | 2.213749 |
| Bol | Bolchoi | fibre spring | | France | TDL | | 46.22764 | 2.213749 |
| Boo | Boothby Grafoe | oilseed | | UK | own | | 55.37805 | -3.43597 |
| Ede | Eden | fibre spring | | France | TDL | | 46.22764 | 2.213749 |
| Mar | Marmalade | oilseed | | Canada | Flaxland | | 56.13037 | -106.347 |
| Olg | Olga | fibre winter | | France | TDL | | 46.22764 | 2.213749 |
| Suz | Suzanne | fibre | | Netherlands | Flaxland | | 52.13263 | 5.291266 |
| Vol | Volga | oilseed winter | | France | TDL | | 46.22764 | 2.213749 |

## Appendix 2: Sample lists for vernalization study

| Individual | Population | Species |
|---|---|---|
| 2_29 | 2 | Bienne |
| 2_3 | 2 | Bienne |
| 3_13 | 3 | Bienne |
| 3_15 | 3 | Bienne |
| 4_23 | 4 | Bienne |
| 4_27 | 4 | Bienne |
| 5_4 | 5 | Bienne |
| 6_1 | 6 | Bienne |
| 6_26 | 6 | Bienne |
| 6_29 | 6 | Bienne |
| 7_17 | 7 | Bienne |
| 9_27 | 9 | Bienne |
| 9_23 | 9 | Bienne |
| 9_24 | 9 | Bienne |

| | | |
|---|---|---|
| 10_26 | 10 | Bienne |
| 10_30 | 10 | Bienne |
| 11_23 | 11 | Bienne |
| 13_12 | 13 | Bienne |
| 14_21 | 14 | Bienne |
| 14_6 | 14 | Bienne |
| 15_27 | 15 | Bienne |
| 15_28 | 15 | Bienne |
| 15_29 | 15 | Bienne |
| 15_32 | 15 | Bienne |
| 19_26 | 19 | Bienne |
| 19_30 | 19 | Bienne |
| Dor_B | Dor | Bienne |
| Iow1_10 | Iow1 | Bienne |
| IOW1_11 | Iow1 | Bienne |
| Iow1_17 | Iow1 | Bienne |
| Iow2_2 | Iow2 | Bienne |
| Iow2_25 | Iow2 | Bienne |
| Iow2_26 | Iow2 | Bienne |
| Iow2_30 | Iow2 | Bienne |
| Lla_17 | Lla | Bienne |
| Lla_22 | Lla | Bienne |
| Lla_25 | Lla | Bienne |
| Lla_20 | Lla | Bienne |
| Lla_43_A | Lla | Bienne |
| Lla_B | Lla | Bienne |
| Lla_33 | Lla | Bienne |
| Man_5 | Man | Bienne |
| Man_6 | Man | Bienne |
| Man_4 | Man | Bienne |
| Man_8 | Man | Bienne |
| Mat_17 | Mat | Bienne |
| Mat_2 | Mat | Bienne |
| Mat_23 | Mat | Bienne |
| Mat_24 | Mat | Bienne |
| Mat_14 | Mat | Bienne |
| Roc_12 | Roc | Bienne |
| Saf_10 | Saf | Bienne |
| Saf_19 | Saf | Bienne |
| Saf_9 | Saf | Bienne |
| Saf_16 | Saf | Bienne |
| Tal_10 | Tal | Bienne |
| Tal_25 | Tal | Bienne |
| Tal_28 | Tal | Bienne |
| Tal_4 | Tal | Bienne |

| | | |
|---|---|---|
| Tor_4 | Tor | Bienne |
| Tym_26 | Tym | Bienne |
| Tym_5 | Tym | Bienne |
| Vil_21 | Vil | Bienne |
| Vil_25 | Vil | Bienne |
| Vil_36 | Vil | Bienne |
| Vil_27 | Vil | Bienne |
| Ara | Ara | Usitatissimum |
| Bey | Bey | Usitatissimum |
| Ble | Ble | Usitatissimum |
| Ede | Ede | Usitatissimum |
| Gis | Gis | Usitatissimum |
| Lir | Lir | Usitatissimum |
| Mon | Mon | Usitatissimum |
| Olg | Olg | Usitatissimum |
| Ome | Ome | Usitatissimum |
| Rab | Rab | Usitatissimum |
| Tin | Tin | Usitatissimum |
| Suz | Suz | Usitatissimum |

## Appendix 3: Principal component for each population and its climatic variable loading values

Principal Components

| Population | PC1 | PC2 | PC3 |
|---|---|---|---|
| 3 | 5.367 | 3.009 | 0.541 |
| 2 | 6.383 | 3.188 | -0.130 |
| 4 | 6.533 | 2.432 | -0.611 |
| 1 | 1.846 | -1.332 | -0.190 |
| 5 | 6.354 | -1.009 | 0.370 |
| 8 | 5.655 | -1.622 | 0.537 |
| 9 | 3.087 | -2.111 | 0.259 |
| 10 | 2.429 | -2.026 | 0.143 |
| 6 | 2.619 | -2.285 | 0.324 |
| 7 | 1.175 | -3.868 | 0.203 |
| 11 | 1.613 | -4.139 | 0.844 |
| 19 | -1.551 | 2.828 | -5.173 |
| 12 | 0.120 | 1.792 | 3.345 |
| 14 | -3.398 | -3.411 | -1.836 |
| 13 | -3.830 | -2.266 | -2.315 |
| 15 | -0.135 | 1.706 | -2.983 |
| IOW2 | -5.447 | 2.161 | 1.706 |
| IOW1 | -5.463 | 2.189 | 1.743 |
| Tor | 6.561 | 1.674 | 2.282 |

|  |  |  |  |  |
|---|---|---|---|---|
|  | CGa1 | 5.845 | 2.724 | 2.514 |
|  | Lla | -0.093 | 0.232 | -1.601 |
|  | Vil | -1.449 | 0.359 | -0.876 |
|  | Bro | -3.291 | -0.895 | 0.205 |
|  | Roc | -3.057 | -0.802 | 0.147 |
|  | Saf | -2.820 | -0.663 | -0.468 |
|  | Tal | -3.097 | 3.614 | 1.631 |
|  | Mat | -4.574 | -2.090 | -0.497 |
|  | BH | -5.397 | 3.754 | 2.368 |
|  | Lil | -5.769 | 1.930 | 0.624 |
|  | CR | -5.348 | 0.097 | -0.455 |
|  | Tym | -6.617 | 1.478 | 1.481 |
|  | Sut | -6.676 | -0.878 | 1.388 |
|  | Man | -7.323 | -0.454 | 0.932 |
|  | LJLb1 | 0.054 | -4.368 | 1.209 |
|  | Dor | -6.279 | 2.961 | 2.316 |
| **Loadings** | prec_DJF | 0.174 | 0.529 | -0.681 |
|  | prec_JJA | -0.869 | 0.209 | -0.273 |
|  | prec_MAM | -0.094 | 0.445 | -0.849 |
|  | prec_SON | -0.233 | 0.615 | -0.694 |
|  | srad_DJF | 0.930 | -0.227 | -0.008 |
|  | srad_JJA | 0.894 | -0.269 | 0.051 |
|  | srad_MAM | 0.930 | -0.224 | 0.020 |
|  | srad_SON | 0.940 | -0.200 | -0.032 |
|  | tavg_DJF | 0.922 | 0.352 | -0.029 |
|  | tavg_JJA | 0.923 | -0.298 | 0.158 |
|  | tavg_MAM | 0.991 | 0.035 | 0.024 |
|  | tavg_SON | 0.992 | 0.088 | 0.046 |
|  | tmax_DJF | 0.964 | 0.125 | -0.042 |
|  | tmax_JJA | 0.807 | -0.523 | 0.128 |
|  | tmax_MAM | 0.947 | -0.253 | 0.003 |
|  | tmax_SON | 0.974 | -0.167 | 0.025 |
|  | tmin_DJF | 0.783 | 0.591 | -0.012 |
|  | tmin_JJA | 0.959 | 0.111 | 0.183 |
|  | tmin_MAM | 0.913 | 0.387 | 0.042 |
|  | tmin_SON | 0.911 | 0.384 | 0.068 |
|  | vapr_DJF | 0.827 | 0.501 | -0.030 |
|  | vapr_JJA | 0.638 | 0.589 | 0.049 |
|  | vapr_MAM | 0.810 | 0.523 | -0.018 |
|  | vapr_SON | 0.805 | 0.532 | -0.002 |
|  | wind_DJF | -0.545 | 0.684 | 0.434 |
|  | wind_JJA | -0.377 | 0.625 | 0.544 |
|  | wind_MAM | -0.453 | 0.690 | 0.489 |
|  | wind_SON | -0.517 | 0.653 | 0.504 |

## Appendix 4: List of Primers and its sequences for flowering time experiments in Linum

**Appendix 5: Primer efficiency comparisons: (We used #3)**

| Primers | Sequence | Conc. |
|---|---|---|
| LuGAPDH_for | AGGTTCTTCCCGCTCTCAAT | 25nm |
| LuGAPDH_rev | CCTCCTTGATAGCAGCCTTG | 25nm |
| LuUBI2_for | CCAAGATCCAGGACAAGGAA | 25nm |
| LuUBI2_rev | GAACCAGGTGGAGAGTCGAT | 25nm |
| LuCO1-pr | AGGCTCCGGTCATGATGAATGACCACTG | 25nm |
| LuCO2-pr | TGCTCCCGTCATGAATATGAATGACCAC | 25nm |
| LuCO-rev | AGATACGCTGTGGCTCAAG | 25nm |
| LuGI1.1-pr | CTCTACTCTTCCGCATCCTGTCA | 25nm |
| LuGI1.2-pr | TGATGGAGTTGAAGTACAGCATGAACC | 25nm |
| LuGI2-pr | CACTACGCCAAGTTGATTGCATCG | 25nm |
| LuGI-rev | GTATGTACAAGTTCCATGACA | 25nm |
| LuFT1-pr | AACTCTACAACTTAGGTTCCCCCGTTG | 25nm |
| LuFT2-pr | AACTCTACAACTTAGGTCCGCCTGTTG | 25nm |
| LuFT-rev | GTCTCTCGTTGGCAGTTAAA | 25nm |

#1

**Primer Efficiency**

| *Sample* | *Primer* | *R^2* | *Slope* | *Efficiency (%)* | *Converted* | *Value* |
|---|---|---|---|---|---|---|
| | *LuGAPDH* | *0.97378* | *-3.86418* | *81.4625* | *1.81463* | 1.814625 |
| | *LuUBI2* | *0.93874* | *-2.83438* | *125.326* | *2.25326* | 2.253256 |
| | *LuGI1.1* | *0.99999* | *-0.60742* | *4329.12* | *44.2912* | 44.29123 |
| | *LuCO1* | *0.97491* | *-3.12638* | *108.862* | *2.08862* | 2.088616 |
| | *LuCO2* | *0.9966* | *-2.68462* | *135.772* | *2.35772* | 2.357722 |
| | *LuFT1* | *0.93924* | *-2.75478* | *130.678* | *2.30678* | 2.306777 |
| | *LuFT2* | *0.9502* | *-2.94433* | *118.593* | *2.18593* | 2.185926 |

#2

| **Primer Efficiency (factor 10)** | | | | | | |
|---|---|---|---|---|---|---|
| *Sample* | *Primer* | *R^2* | *Slope* | *Efficiency (%)* | *Conversion* | *Value* |
| | *LuGAPDH* | *0.97378* | *-3.86418* | *88.6802* | *1.8868* | 1.886802 |
| | *LuUBI2* | *0.93874* | *-2.83438* | *160.188* | *2.60188* | 2.601882 |
| | *LuGI1.1* | *0.99999* | *-0.60742* | *4834.91* | *49.3491* | 49.34912 |
| | *LuCO1* | *0.97491* | *-3.12638* | *-99.8688* | *0.00131* | 0.001312 |
| | *LuCO2* | *0.9966* | *-2.68462* | *97.5082* | *1.97508* | 1.975082 |
| | *LuFT1* | *0.93924* | *-2.75478* | *207.128* | *3.07128* | 3.071284 |
| | *LuFT2* | *0.9502* | *-2.94433* | *214.621* | *3.14621* | 3.146207 |

#3 –

| **Primer Efficiency** | | | | | | |
|---|---|---|---|---|---|---|
| *Sample* | *Primer* | *R^2* | *Slope* | *Efficiency (%)* | *Converted* | *Value* |
| | *LuGAPDH* | *0.99505* | *-3.67007* | *87.2728* | *1.87273* | 1.872728 |
| | *LuUBI2* | *0.90322* | *-5.54602* | *51.464* | *1.51464* | 1.51464 |
| | *LuGI1.1* | *0.33336* | *-0.55935* | *6034.64* | *61.3464* | 61.34644 |
| | *LuCO1* | *0.55246* | *-3.56356* | *90.8175* | *1.90818* | 1.908175 |
| | *LuCO2* | *0.96165* | *-3.61371* | *89.1143* | *1.89114* | 1.891143 |
| | *LuFT1* | *0.79856* | *-3.08473* | *110.949* | *2.10949* | 2.109493 |
| | *LuFT2* | *0.98005* | *-2.38037* | *163.09* | *2.6309* | 2.630896 |

## Appendix 6: Relative Gene Expression R commands:

```
library(ggpubr)
library("gridExtra")
library("ggplot2")
library("cowplot")
library("dplyr")
```

```
library("ggpubr")
library("viridis")
library(readr)
combined_vern_same_ind_clean1 <- read_csv("combined_vern_same_ind_clean1.csv",
                                          col_types = cols(Days_to_flower = col_number(),
                                                           Experiment = col_character()))
View(combined_vern_same_ind_clean1)

ggplot(combined_vern_same_ind_clean1, aes(x=Experiment, y=Days_to_flower, fill=Experiment)) +
geom_boxplot() + labs(title = "Difference in Days to Flower Between Vernalization Years", x = "Vernalization
Year", y = "Number of Days to Flower")

#multi-gene:
library(ggplot2)
ggplot(rge_input_luco1, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene
Expression - LuCO1", x = "Sample", y = "Relative Gene Expression")+  stat_compare_means(method = "anova")
ggplot(rge_input_luco2, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene
Expression - LuCO2", x = "Sample", y = "Relative Gene Expression")
ggplot(rge_input_luft1, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene
Expression - LuFT1", x = "Sample", y = "Relative Gene Expression")
ggplot(rge_input_luft2, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene
Expression - LuFT2", x = "Sample", y = "Relative Gene Expression")
ggplot(rge_input_lugi11, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene
Expression - LuGI1.1", x = "Sample", y = "Relative Gene Expression")


#delta-delta:
setwd("G:/Linum Project/rdir")
#luco1:
library(readr)
```

```
rge_input_luco1_deltadelta <- read_csv("rge_input_luco1_deltadelta_withoutoutlyer.csv",
                                       col_types = cols(RGE = col_number()))
View(rge_input_luco1_deltadelta)
library(ggplot2)
ggplot(rge_input_luco1_deltadelta, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative
Gene Expression - LuCO1", x = "Sample", y = "Relative Gene Expression")+ stat_compare_means(label =
"p.signif", method = "t.test",ref.group = ".all.")
#luco2:
library(readr)
rge_input_luco2_deltadelta <- read_csv("rge_input_luco2_deltadelta_withoutoutlyer.csv",
                                    col_types = cols(RGE = col_number()))
View(rge_input_luco2_deltadelta)
library(ggplot2)
ggplot(rge_input_luco2_deltadelta, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative
Gene Expression - LuCO2", x = "Sample", y = "Relative Gene Expression")+stat_compare_means(label =
"p.signif", method = "t.test",ref.group = ".all.")
#lugi1.1:
library(readr)
rge_input_lugi1_1_deltadelta <- read_csv("rge_input_lugi1_1_deltadelta_withoutoutlyer.csv",
                                      col_types = cols(RGE = col_number()))
View(rge_input_lugi1_1_deltadelta)
library(ggplot2)
ggplot(rge_input_lugi1_1_deltadelta, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative
Gene Expression - LuGI1.1", x = "Sample", y = "Relative Gene Expression")+stat_compare_means(label =
"p.signif", method = "t.test",ref.group = ".all.")
#Luft1:
#without Saf_10:
library(readr)
rge_input_luft1_deltadelta_withoutoutlyer <- read_csv("rge_input_luft1_deltadelta_withoutoutlyer.csv",
                                                  col_types = cols(RGE = col_number()))
```

```
View(rge_input_luft1_deltadelta_withoutoutlyer)

library(ggplot2)
ggplot(rge_input_luft1_deltadelta_withoutoutlyer, aes(x=group, y=RGE, fill=group)) + geom_boxplot() +
labs(title = "Relative Gene Expression - Luft1", x = "Sample", y = "Relative Gene
Expression")+stat_compare_means(label = "p.signif", method = "t.test",ref.group = ".all.")

#luft2:
library(readr)
rge_input_luft2_deltadelta <- read_csv("rge_input_luft2_deltadelta_withoutoutlier.csv",
                                       col_types = cols(RGE = col_number()))
View(rge_input_luft2_deltadelta)
#without saf_10:
library(readr)
rge_input_luft2_deltadelta_withoutoutlier <- read_csv("rge_input_luft2_deltadelta_withoutoutlier.csv",
                                                      col_types = cols(RGE = col_number()))
View(rge_input_luft2_deltadelta_withoutoutlier)
library(ggplot2)
ggplot(rge_input_luft2_deltadelta_withoutoutlier, aes(x=group, y=RGE, fill=group)) + geom_boxplot() +
labs(title = "Relative Gene Expression - Luft2", x = "Sample", y = "Relative Gene
Expression")+stat_compare_means(label = "p.signif", method = "t.test",ref.group = ".all.")


#luco1:
rge_input_luco1$group <- factor(rge_input_luco1$group, levels = c("control_cul", "vern_cul", "control_wil",
"vern_wil"))
meanRGEluco1 <- tapply(rge_input_luco1$RGE, rge_input_luco1$group, mean)
sdevRGEluco1 <- tapply(rge_input_luco1$RGE, rge_input_luco1$group, sd)
numberRGEluco1 <- tapply(rge_input_luco1$RGE, rge_input_luco1$group, length)
data.frame(mean=meanRGEluco1, std.dev=sdevRGEluco1, n=numberRGEluco1)
```

```

#anova:
rge_luco1_anova <- lm(RGE ~ group, data = rge_input_luco1)
anova(rge_luco1_anova)

#luco1 cultivar comparison:
library(readr)
rge_input_luco1_cul <- read_csv("rge_input_luco1_cul.csv",
                    +    col_types = cols(RGE = col_number()))

#t-test:
t.test(RGE~group, data = rge_input_luco1_cul)

#luco1 wild comparison:
library(readr)
rge_input_luco1_wil <- read_csv("rge_input_luco1_wil.csv",
                   col_types = cols(RGE = col_number()))
View(rge_input_luco1_wil)

#t-test:
t.test(RGE~group, data = rge_input_luco1_wil)


#luco2:
rge_input_luco2$group <- factor(rge_input_luco2$group, levels = c("control_cul", "vern_cul", "control_wil",
"vern_wil"))
meanRGEluco2 <- tapply(rge_input_luco2$RGE, rge_input_luco2$group, mean)
sdevRGEluco2 <- tapply(rge_input_luco2$RGE, rge_input_luco2$group, sd)
numberRGEluco2 <- tapply(rge_input_luco2$RGE, rge_input_luco2$group, length)
```

```
data.frame(mean=meanRGEluco2, std.dev=sdevRGEluco2, n=numberRGEluco2)

#anova:
rge_luco2_anova <- lm(RGE ~ group, data = rge_input_luco2)
anova(rge_luco2_anova)

#luco2 cultivar comparison:
library(readr)
rge_input_luco2_cul <- read_csv("rge_input_luco2_cul.csv",
                              col_types = cols(RGE = col_number()))
View(rge_input_luco2_cul)

#t-test:
rge_input_luco2_cul$group <- factor(rge_input_luco2_cul$group, levels = c("control_cul", "vern_cul"))
rge_input_luco2_cul_ttest <- lm(RGE ~ group, data = rge_input_luco2_cul)
t.test(RGE~group, data = rge_input_luco2_cul)

#luco2 wild comparison:
library(readr)
rge_input_luco2_wil <- read_csv("rge_input_luco2_wil.csv",
                              col_types = cols(`1.053672824` = col_number()))
View(rge_input_luco2_wil)

t.test(RGE~group, data = rge_input_luco2_wil)

#luft2:
rge_input_luft2$group <- factor(rge_input_luft2$group, levels = c("control_cul", "vern_cul", "control_wil",
"vern_wil"))
meanRGEluft2 <- tapply(rge_input_luft2$RGE, rge_input_luft2$group, mean)
```

```
sdevRGEluft2 <- tapply(rge_input_luft2$RGE, rge_input_luft2$group, sd)
numberRGEluft2 <- tapply(rge_input_luft2$RGE, rge_input_luft2$group, length)
data.frame(mean=meanRGEluft2, std.dev=sdevRGEluft2, n=numberRGEluft2)

#anova
rge_luft2_anova <- lm(RGE ~ group, data = rge_input_luft2)
anova(rge_luft2_anova)

#luft2 cultivar comparison:
library(readr)
rge_input_luft2_cul <- read_csv("rge_input_luft2_cul.csv",
                          col_types = cols(RGE = col_number()))
View(rge_input_luft2_cul)

#t-test:
rge_input_luft2_cul$group <- factor(rge_input_luft2_cul$group, levels = c("control_cul", "vern_cul"))
rge_input_luft2_cul_ttest <- lm(RGE ~ group, data = rge_input_luft2_cul)
t.test(RGE~group, data = rge_input_luft2_cul)

#luft2 wild comparison:
library(readr)
rge_input_luft2_wil <- read_csv("rge_input_luft2_wil.csv",
                           +    col_types = cols(Ind = col_character(),
                                            +        RGE = col_number(), group = col_character()))
View(rge_input_luft2_wil)


#luft1:
```

```
rge_input_luft1$group <- factor(rge_input_luft1$group, levels = c("control_cul", "vern_cul", "control_wil", "vern_wil"))
meanRGEluft1 <- tapply(rge_input_luft1$RGE, rge_input_luft1$group, mean)
sdevRGEluft1 <- tapply(rge_input_luft1$RGE, rge_input_luft1$group, sd)
numberRGEluft1 <- tapply(rge_input_luft1$RGE, rge_input_luft1$group, length)
data.frame(mean=meanRGEluft1, std.dev=sdevRGEluft1, n=numberRGEluft1)

#anova
rge_luft1_anova <- lm(RGE ~ group, data = rge_input_luft1)
anova(rge_luft1_anova)


#t-test:
rge_input_luft2_wil$group <- factor(rge_input_luft2_wil$group, levels = c("control_cul", "vern_cul"))
rge_input_luft2_wil_ttest <- lm(RGE ~ group, data = rge_input_luft2_wil)
t.test(RGE~group, data = rge_input_luft2_wil)

#luft1 cultivar comparison:
library(readr)
rge_input_luft1_cul <- read_csv("rge_input_luft1_cul.csv",
                              col_types = cols(...1 = col_character(),
                                               RGE = col_number(), group = col_character()))
View(rge_input_luft1_cul)

#t-test:
t.test(RGE~group, data = rge_input_luft1_cul)

#luft1 wild comparison:
library(readr)
```

```
rge_input_luft1_wil <- read_csv("rge_input_luft1_wil.csv",
                              col_types = cols(RGE = col_number()))
View(rge_input_luft1_wil)

#t-test:
t.test(RGE~group, data = rge_input_luft1_wil)

#lugi11:
rge_input_lugi11$group <- factor(rge_input_lugi11$group, levels = c("control_cul", "vern_cul", "control_wil",
"vern_wil"))
meanRGElugi11 <- tapply(rge_input_lugi11$RGE, rge_input_lugi11$group, mean)
sdevRGElugi11 <- tapply(rge_input_lugi11$RGE, rge_input_lugi11$group, sd)
numberRGElugi11 <- tapply(rge_input_lugi11$RGE, rge_input_lugi11$group, length)
data.frame(mean=meanRGElugi11, std.dev=sdevRGElugi11, n=numberRGElugi11)

#anova
rge_lugi11_anova <- lm(RGE ~ group, data = rge_input_lugi11)
anova(rge_lugi11_anova)

#lugi11 cultivar comparison:
library(readr)
rge_input_lugi11_cul <- read_csv("rge_input_lugi11_cul.csv",
                              col_types = cols(RGE = col_number()))
View(rge_input_lugi11_cul)

#t-test:
t.test(RGE~group, data = rge_input_lugi11_cul)

#lugi11 wild comparison:
```

```
library(readr)
rge_input_lugi11_wil <- read_csv("rge_input_lugi11_wil.csv",
                 col_types = cols(RGE = col_number()))
View(rge_input_lugi11_wil)

#t-test:
t.test(RGE~group, data = rge_input_lugi11_wil)

```

## **Appendix 7: Relative Gene Expression GLM R Commands:**

```
library(readr)
combined_vern_same_ind <- read_csv("combined_vern_same_ind.csv",
                  col_types = cols(Lat = col_number(),
                           Alt = col_number(), Height = col_number(),
                           Stem_no = col_number(), Bud_no = col_number(),
                           Days_to_fl = col_number(), pc_1 = col_number(),
                           pc_2 = col_number(), pc_3 = col_number()))
View(combined_vern_same_ind)

#plot:
x <- combined_vern_same_ind$Height
y <- combined_vern_same_ind$Days_to_fl
plot(x, y, main = "Height against Days to Flowering",
     xlab = "Plant Height (Cm)", ylab = "Days to Flowering",
     pch = 19, frame = FALSE)

#scatterplot
install.packages("car")
library("car")
#height:
```

```
View(combined_vern_same_ind)
scatterplot(Days_to_fl ~ Height | Experiment, data = combined_vern_same_ind,
            smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Plant Height (Cm)", ylab = "Days to Flowering")
abline(lm(Height ~ Days_to_fl data = combined_vern_same_ind),col="red")

ggscatter(combined_vern_same_ind, x = "Height", y = "Days_to_fl", size = 2.0,
          rug = TRUE,                                # Add marginal rug
          color = "Experiment", palette = "jco", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Experiment), method = "pearson")

#stem no:
scatterplot(Days_to_fl ~ Stem_no | Experiment, data = combined_vern_same_ind,
            smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Stem number", ylab = "Days to Flowering")
abline(lm(Stem_no ~ Days_to_fl data = combined_vern_same_ind),col="red")


ggscatter(combined_vern_same_ind, x = "Stem_no", y = "Days_to_fl", size = 2.0,
          rug = TRUE,                                # Add marginal rug
          color = "Experiment", palette = "jco", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Experiment), method = "pearson")

#bud no:
scatterplot(Days_to_fl ~ Bud_no | Experiment, data = combined_vern_same_ind,
            smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Bud number", ylab = "Days to Flowering")
abline(lm(Stem_no ~ Days_to_fl data = combined_vern_same_ind),col="red")
```

```
ggscatter(combined_vern_same_ind, x = "Bud_no", y = "Days_to_fl", size = 2.0,
          rug = TRUE,                          # Add marginal rug
          color = "Experiment", palette = "jco", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Experiment), method = "pearson")

#pc1:
scatterplot(pc_1 ~ Days_to_fl | Experiment, data = combined_vern_same_ind,
            smooth = FALSE, grid = FALSE, frame = FALSE, xlab = "Days to First Flower", ylab = "Climate (PC1)")
abline(lm(pc_1 ~ Days_to_fl data = combined_vern_same_ind),col="red")

plot(pc_1 ~ Days_to_fl, data = combined_vern_same_ind)
abline(lm(pc_1 ~ Days_to_fl, data = combined_vern_same_ind),col="red")
summary(abline)

#Alt:
scatterplot(Days_to_fl ~ Alt | Experiment, data = combined_vern_same_ind,
            smooth = FALSE, grid = FALSE, frame = FALSE, xlab = "Days to First Flower", ylab = "Altitude (M)")
abline(lm(Alt ~ Days_to_fl data = combined_vern_same_ind),col="red")

plot(Alt ~ Days_to_fl, data = combined_vern_same_ind)
abline(lm(Alt ~ Days_to_fl, data = combined_vern_same_ind),col="red")
summary(abline)


#Model summary
summary1 <- summary(glm(Stem_no ~ Days_to_fl, data = combined_vern_same_ind))
adjRsq <- summary1$adj.r.squared
fStat <- summary1$statistic
```

```
pValue <- pf(fStat[summary1])
summary(summary1)
#PC1 - Days to fl LM:
pc1daysmod<-cbind(combined_vern_same_ind$Days_to_fl, combined_vern_same_ind$pc_1)
pc1daysmod2<-pc1daysmodlm<-lm(pc1daysmod~Experiment,data=combined_vern_same_ind)
summary(pc1daysmodlm)


#Alt - Days to fl GLM:
altdaysmod<-cbind(combined_vern_same_ind$Days_to_fl, combined_vern_same_ind$Alt)
altdaysmod2<-altdaysmodlm<-lm(altdaysmod~Experiment,data=combined_vern_same_ind)
summary(altdaysmodlm)

#Lat - Days to fl GLM:
latdaysmod<-cbind(combined_vern_same_ind$Days_to_fl, combined_vern_same_ind$Lat)
latdaysmod2<-latdaysmodlm<-lm(altdaysmod~Experiment,data=combined_vern_same_ind)
summary(latdaysmodlm)

#RGE - Days to fl GLM:
library(readr)
combined_loci_wild_deltadelta_ft2clear <- read_csv("combined_loci_wild_deltadelta_ft2clear.csv",
                                      col_types = cols(`RGE_c` = col_number(),
                                                   `RGE_v` = col_number(), `RGE_Difference` = col_number(),
                                                   Days_to_fl = col_number(), Lat = col_number(),
                                                   Alt = col_number(), pc = col_number(),
                                                   pc2 = col_number(), pc3 = col_number()))
View(combined_loci_wild_deltadelta_ft2clear)

#Cultivars:
```

```
library(readr)
combined_loci_cult_deltadelta_fl <- read_csv("combined_loci_cult_deltadelta_fl.csv",
                                  col_types = cols(`RGE_c` = col_number(),
                                             `RGE_v` = col_number(), `RGE_Difference` = col_number(),
                                             Lat = col_number(), Days_to_fl = col_number(),
                                             `Height_(Cm)` = col_number()))
View(combined_loci_cult_deltadelta_fl)


#Gene expression influence in flowering time:
#Wild
#subset data to loci first:
Luco1_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuCO1")
Luco2_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuCO2")
Luft1_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuFT1")
Luft2_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuFT2")
Lugi11_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuGI1.1")

#GLM:
modluco1wild<-glm(Luco1_data$Days_to_fl ~ Luco1_data$"RGE_Difference")
modluco2wild<-glm(Luco2_data$Days_to_fl ~ Luco2_data$"RGE_Difference")
modluft1wild<-glm(Luft1_data$Days_to_fl ~ Luft1_data$"RGE_Difference")
modluft2wild<-glm(Luft2_data$Days_to_fl ~ Luft2_data$"RGE_Difference")
modlugi11wild<-glm(Lugi11_data$Days_to_fl ~ Lugi11_data$"RGE_Difference")

summary(modluco1wild)
summary(modluco2wild)
summary(modluft1wild)
summary(modluft2wild)
```

```
summary(modlugi11wild)

#Cultivars
Luco1_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuCO1")
Luco2_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuCO2")
Luft1_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuFT1")
Luft2_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuFT2")
Lugi11_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuGI1.1")

#GLM:
modluco1cul<-glm(Luco1_cult_data$Days_to_fl ~ Luco1_cult_data$"RGE_Difference")
modluco2cul<-glm(Luco2_cult_data$Days_to_fl ~ Luco2_cult_data$"RGE_Difference")
modluft1cul<-glm(Luft1_cult_data$Days_to_fl ~ Luft1_cult_data$"RGE_Difference")
modluft2cul<-glm(Luft2_cult_data$Days_to_fl ~ Luft2_cult_data$"RGE_Difference")
modlugi11cul<-glm(Lugi11_cult_data$Days_to_fl ~ Lugi11_cult_data$"RGE_Difference")

summary(modluco1cul)
summary(modluco2cul)
summary(modluft1cul)
summary(modluft2cul)
summary(modlugi11cul)

#charts:
View(Luco1_data)

write.csv(Luco1_data,"G:/Linum Project/rdir/Luco1_data_only.csv", row.names = FALSE)

#
#Delta-delta data:
```

```
#Load packages
library(ggplot2)
library(ggpubr)
library(viridis)
library(readr)
#Cultivars
setwd("G:/Linum Project/rdir")


ggscatter(combined_loci_cult_deltadelta_fl, x = "Days_to_fl", y = "RGE_Difference", size = 1.0,
          rug = TRUE,                                # Add marginal rug
          color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.95, method =
"pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Cultivar RGE-Flowering")

#Lat:
ggscatter(combined_loci_cult_deltadelta_fl, x = "Lat", y = "RGE_Difference", size = 1.0,
          rug = TRUE,                                # Add marginal rug
          color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.95, method =
"pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Cultivar RGE-Latitude")

#GLM:
modluco1cullat<-glm(Luco1_cult_data$Lat ~ Luco1_cult_data$"RGE_Difference")
modluco2cullat<-glm(Luco2_cult_data$Lat ~ Luco2_cult_data$"RGE_Difference")
modluft1cullat<-glm(Luft1_cult_data$Lat ~ Luft1_cult_data$"RGE_Difference")
modluft2cullat<-glm(Luft2_cult_data$Lat ~ Luft2_cult_data$"RGE_Difference")
modlugi11cullat<-glm(Lugi11_cult_data$Lat ~ Lugi11_cult_data$"RGE_Difference")
```

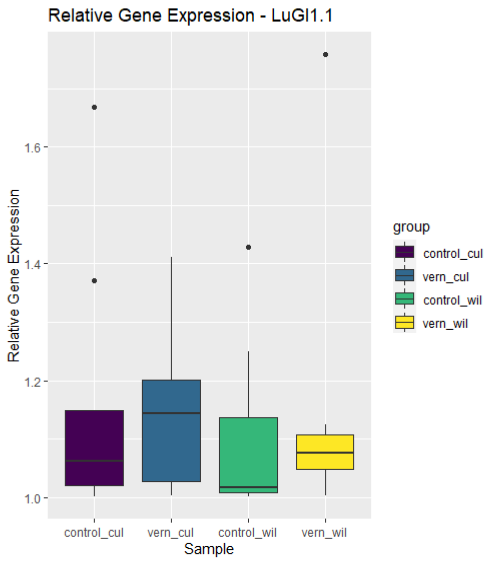
```
summary(modluco1cullat)
summary(modluco2cullat)
summary(modluft1cullat)
summary(modluft2cullat)
summary(modlugi11cullat)

#Wild:
ggscatter(combined_loci_wild_deltadelta_ft2clear, x = "Days_to_fl", y = "RGE_Difference", size = 1.0,
          rug = TRUE,                          # Add marginal rug
          color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.32, method =
"pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Wild RGE-Flowering")

#Lat
ggscatter(combined_loci_wild_deltadelta_ft2clear, x = "Lat", y = "RGE_Difference", size = 1.0,
          rug = TRUE,                          # Add marginal rug
          color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc="centre", method =
"pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Wild RGE-Latitude")

#GLM:
modluco1willat<-glm(Luco1_data$Lat ~ Luco1_data$"RGE_Difference")
modluco2willat<-glm(Luco2_data$Lat ~ Luco2_data$"RGE_Difference")
modluft1willat<-glm(Luft1_data$Lat ~ Luft1_data$"RGE_Difference")
modluft2willat<-glm(Luft2_data$Lat ~ Luft2_data$"RGE_Difference")
modlugi11willat<-glm(Lugi11_data$Lat ~ Lugi11_data$"RGE_Difference")

```

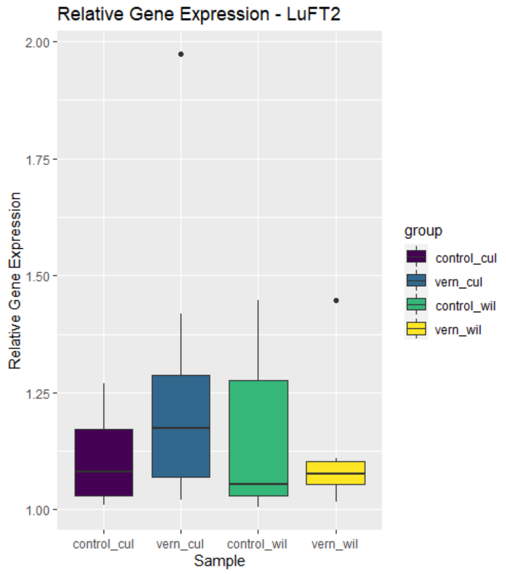
```
summary(modluco1willat)
summary(modluco2willat)
summary(modluft1willat)
summary(modluft2willat)
summary(modlugi11willat)


#PC1:
ggscatter(combined_loci_wild_deltadelta_ft2clear, x = "pc", y = "RGE_Difference", size = 1.0,
          rug = TRUE,                          # Add marginal rug
          color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.38, method =
"pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Wild RGE-PC1")

#GLM:
modluco1wilpc<-glm(Luco1_data$pc ~ Luco1_data$"RGE_Difference")
modluco2wilpc<-glm(Luco2_data$pc ~ Luco2_data$"RGE_Difference")
modluft1wilpc<-glm(Luft1_data$pc ~ Luft1_data$"RGE_Difference")
modluft2wilpc<-glm(Luft2_data$pc ~ Luft2_data$"RGE_Difference")
modlugi11wilpc<-glm(Lugi11_data$pc ~ Lugi11_data$"RGE_Difference")

summary(modluco1wilpc)
summary(modluco2wilpc)
summary(modluft1wilpc)
summary(modluft2wilpc)
summary(modlugi11wilpc)

#scatterplot
```

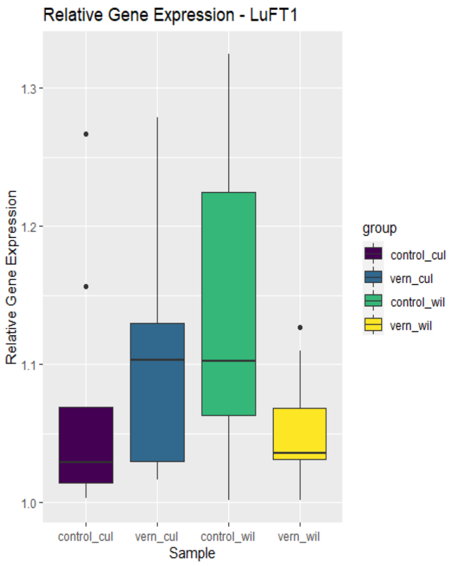
```
install.packages("tidyverse")
library(tidyverse)
library(ggplot2)

ggplot(combined_loci_cult_deltadelta_fl, aes(x=RGE_Difference, y=Days_to_fl, shape=Loci, color=Loci)) +
  geom_point(size=2.5, shape=18) + ggtitle("Cultivar Flowering initiation")+labs(y= "Days to Flowering", x = "Relative Gene Expression Difference")

#wild
ggplot(combined_loci_wild_deltadelta_ft2clear, aes(x=RGE_Difference, y=Days_to_fl, shape=Loci, color=Loci)) +
  geom_point(size=2.5, shape=18) + ggtitle("Wild Flowering initiation")+labs(y= "Days to Flowering", x = "Relative Gene Expression Difference")

#
s3d <- scatterplot3d(combined_vern_same_ind, pch = 16, color=colors)
legend(s3d$xyz.convert(7.5, 3, 4.5), legend = levels(combined_vern_same_ind$Experiment),
      col =  c("#999999", "#E69F00", "#56B4E9"), pch = 16)

```

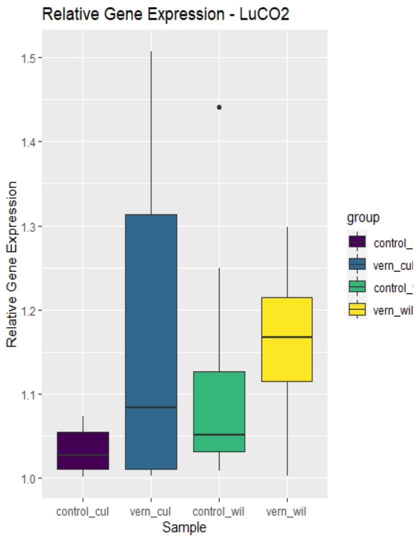
## Appendix 8: Relative gene expression difference based on using two HKGs

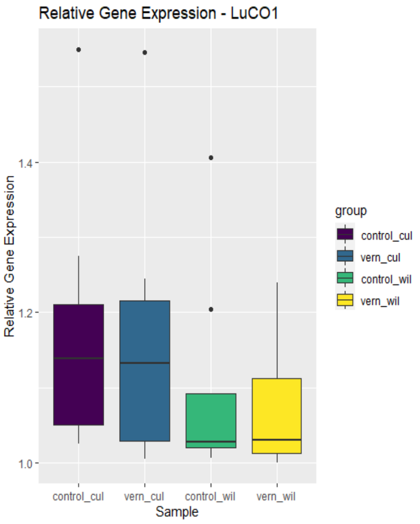
Relative gene expression comparison for the different genes tested in relation to each species and treatment:







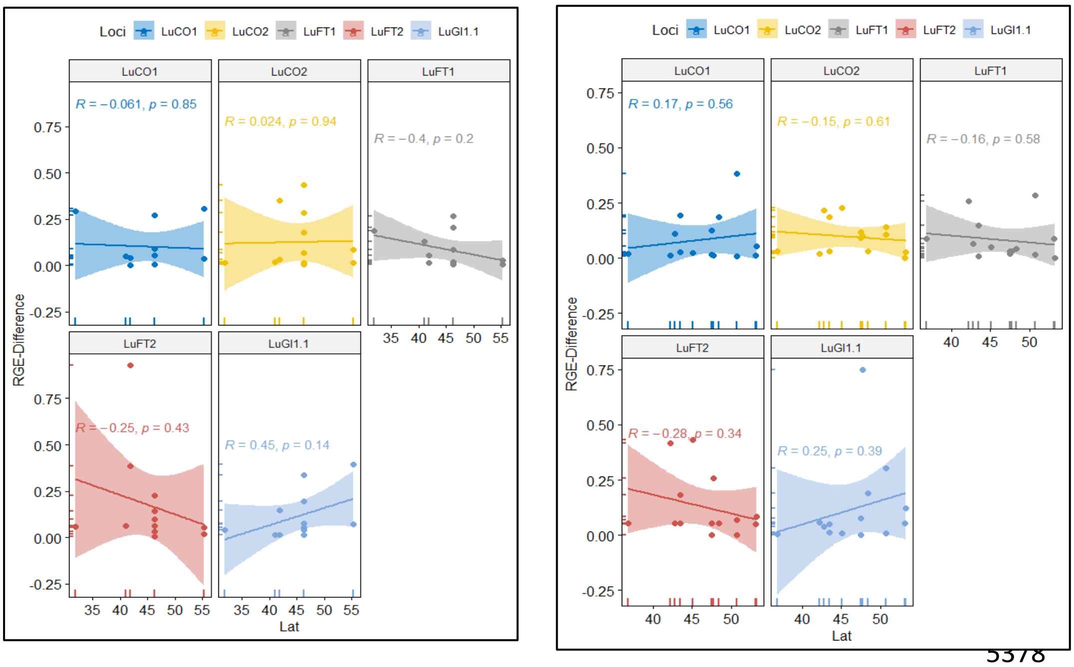




## Appendix 9: Scatterplot for gene expression and latitude correlation under two Housekeeping genes

Scatterplot showing relationships between latitude and relative gene expression (RGE) differences for five flowering time genes.  A = wild types, B = Cultivars. RGE differences were calculated using 2 HKGs:

A B



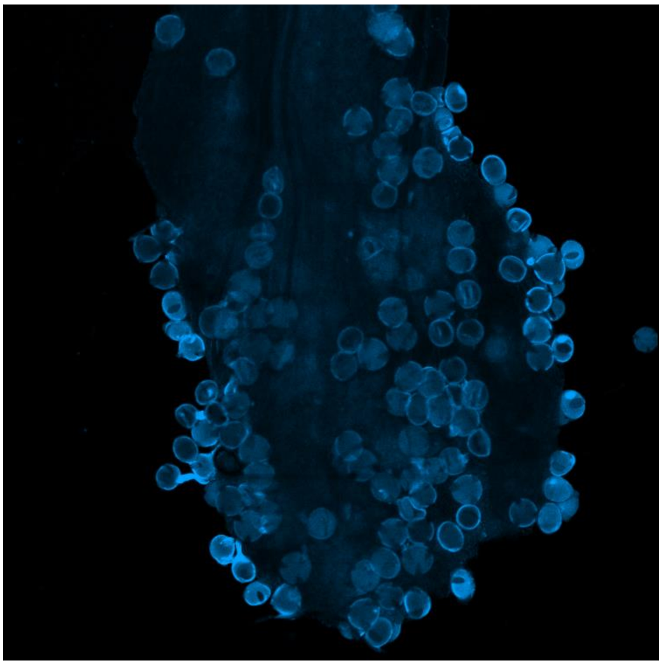
| Gene | Group test | Cultivar t-test | Wild t-test |
|---|---|---|---|
| **LuCO1** | 0.4672 | 0.9027 | 0.7722 |
| **LuCO2** | 0.194 | 0.09186 | 0.5341 |
| **LuGI1.1** | 0.9434 | 0.9163 | 0.6533 |
| **LuFT1** | 0.2373 | 0.08303 | 0.4711 |
| **LuFT2** | 0.354 | 0.2029 | 0.6945 |

**Appendix 10: List of *Linum* individuals for pollen viability germination experiments in chapter 3 of the thesis with number of individuals in each treatments for each individuals.**

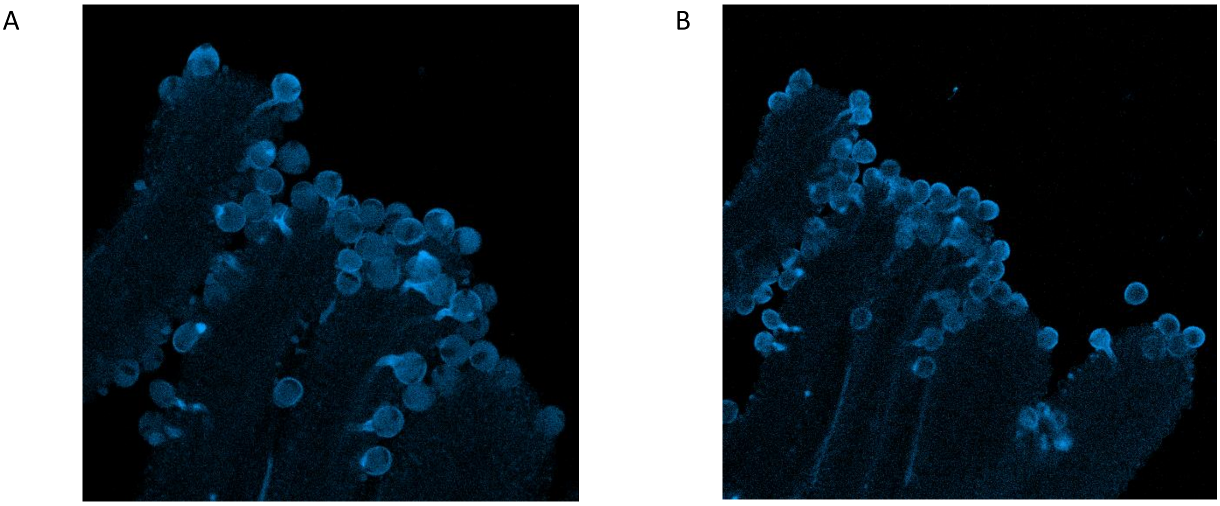
| Individual | Population | Species | Numbers in 20C | Numbers in 25C | Numbers in 30C |
|---|---|---|---|---|---|
| Ara | Ara | Usitatissimum | 0 | 0 | 13 |
| Bey | Bey | Usitatissimum | 0 | 3 | 0 |
| Ble | Ble | Usitatissimum | 4 | 0 | 0 |
| Bol | Bol | Usitatissimum | 0 | 0 | 0 |
| Boo | Boo | Usitatissimum | 0 | 22 | 0 |
| Ede | Ede | Usitatissimum | 0 | 4 | 0 |
| Kar | Kar | Usitatissimum | 3 | 0 | 10 |
| Olg | Olg | Usitatissimum | 0 | 16 | 3 |
| Ome | Ome | Usitatissimum | 30 | 11 | 0 |
| Rab | Rab | Usitatissimum | 0 | 7 | 0 |
| Sar | Sar | Usitatissimum | 2 | 6 | 0 |
| Suz | Suz | Usitatissimum | 0 | 3 | 0 |
| Tin | Tin | Usitatissimum | 6 | 0 | 7 |
| Vol | Vol | Usitatissimum | 0 | 0 | 8 |
| 3_12 | 3 | bienne | 5 | 2 | 0 |
| 6_17 | 6 | bienne | 7 | 0 | 6 |
| 10_24 | 10 | bienne | 1 | 12 | 0 |
| 15_31 | 15 | bienne | 4 | 0 | 0 |
| 15_32 | 15 | bienne | 1 | 0 | 5 |
| 19_22 | 19 | bienne | 3 | 0 | 2 |
| 19_30 | 19 | bienne | 1 | 0 | 0 |
| 19_28 | 19 | bienne | 3 | 0 | 2 |
| Man_6 | Man | bienne | 7 | 0 | 0 |
| Mat_2 | Mat | bienne | 12 | 2 | 0 |
| Saf_19 | Saf | bienne | 4 | 0 | 0 |
| Sut_41 | Sut | bienne | 2 | 6 | 0 |
| Tym_21 | Tym | bienne | 1 | 0 | 0 |
| 6_30 | 6 | bienne | 0 | 1 | 0 |
| 13_6 | 13 | bienne | 0 | 3 | 0 |
| 15_17 | 15 | bienne | 0 | 3 | 0 |
| 19_23 | 19 | bienne | 0 | 7 | 0 |
| 19_32 | 19 | bienne | 0 | 1 | 0 |
| Ezc_7 | Ezc | bienne | 0 | 5 | 0 |
| Lla_33 | Lla | bienne | 0 | 1 | 0 |
| Lla_13 | Lla | bienne | 0 | 1 | 0 |
| Lla_31 | Lla | bienne | 0 | 5 | 0 |
| Lla_18 | Lla | bienne | 0 | 4 | 0 |
| Lla_22 | Lla | bienne | 0 | 1 | 0 |
| 8_16 | 8 | bienne | 0 | 0 | 6 |
| 9_27 | 9 | bienne | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 10_30 | 10 | bienne | 0 | 0 | 1 |
| 13_12 | 13 | bienne | 0 | 0 | 2 |
| Dor_B | Dor | bienne | 0 | 0 | 8 |
| Ezc_9 | Ezc | bienne | 0 | 0 | 3 |
| low2_26 | low2 | bienne | 0 | 0 | 2 |
| low2_25 | low2 | bienne | 0 | 0 | 1 |
| Lla_20 | Lla | bienne | 0 | 0 | 1 |
| Lla_29 | Lla | bienne | 0 | 0 | 2 |
| Man_8 | Man | bienne | 0 | 0 | 6 |
| Saf_9 | Saf | bienne | 0 | 0 | 3 |
| Vil_36 | Vil | bienne | 0 | 0 | 2 |

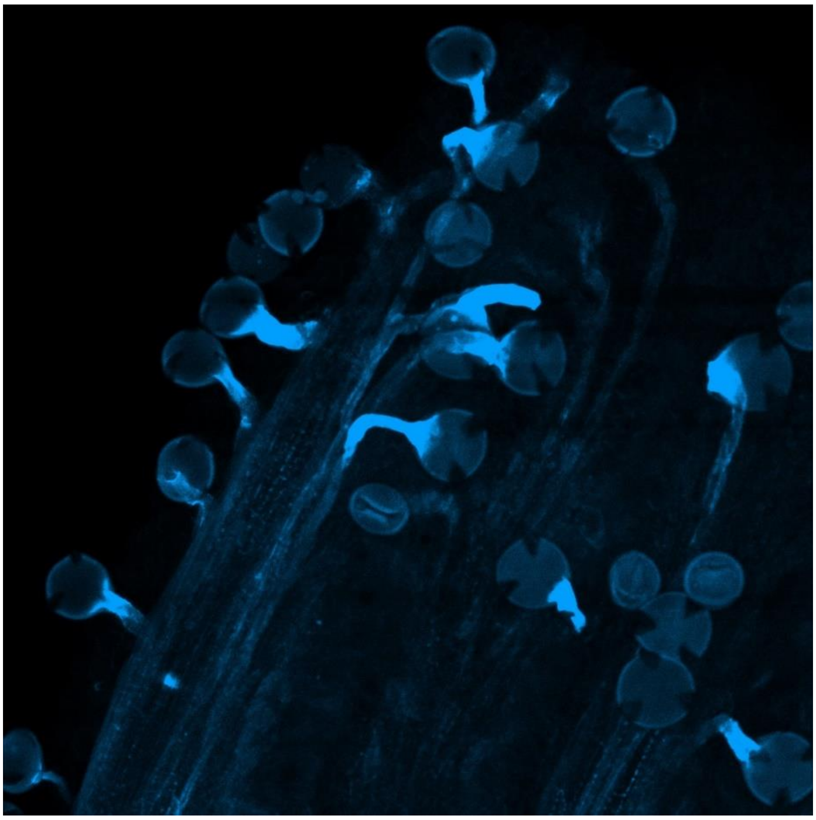
**Appendix 11: Fluorescent microscopy observation of a *Linum* flower conserved at 4°C in 70%EtOH solution, showing pollen and tube observations after 72 hours preservation.**



**Appendix 12: 10×20 magnification of fluorescence observed after flower preservation using both FAA (A) and 70% EtOH (B) solutions. The flowers were from the same individual observed in the glasshouse conditions.**



**Appendix 13: Pollen Tube Observation Using 0.5% (w/v) Aniline blue in Potassium Phosphate (KH2PO4)**



## Appendix 14: Pollen Modelling Commands

```
#test for normality (pollen tube)
plot(density(Pollen$Percent))
hist(Pollen$Percent, main="Histogram for Percentage of Pollen Tube %", xlab="% of Pollen Tube Counts")
densityPollen<-density(Pollen$Percent)
lines(densityPollen$x,densityPollen$y*5000)
shapiro.test(Pollen$Percent)
#data is not normally distributed
#adjust for non-normally distributed data (tukey's)

#Data for 20C - Cold (pollen tube)
setwd("G:/Linum Project/Pollen germination")
library(readxl)
Pollencold <- as.data.frame(read_excel("Pollen_and_tube_count2.xlsx", sheet = "20c", na = "NA"))
str(Pollencold)
View(Pollencold)
Pollencold$Treatment<-as.factor(Pollencold$Treatment)

#test for normality - Cold
plot(density(Pollencold$Percent))
hist(Pollencold$Percent, main="Histogram for Percentage of Pollen Tube %", xlab="% of Pollen Tube Counts")
densityPollencold<-density(Pollencold$Percent)
lines(densityPollencold$x,densityPollencold$y*5000)
shapiro.test(Pollencold$Percent)
#data is not normally distributed
#adjust for non-normally distributed data (tukey's)

#Pollen number:
plot(density(Pollen$Count))
```

```
hist(Pollen$Count, main = "Pollen Count Distribution", xlab = "Pollen Count")
densityCount<-density(Pollen$Count)
lines(densityCount$x,densityCount$y*5000)
shapiro.test(Pollen$Count)

#test pollen count versus treatment, quasibinomial model - TOTAL
count<-Pollen$Count
modcount<-lm(Pollen$Count~Treatment, data=Pollen)
summary(modcount)
tukcount<-glht(modcount,mcp(Treatment="Tukey"))
summary(tukcount)
Props<-Pollen$Tube/Pollen$Count
Pollen<-data.frame(Pollen,count)
str(Pollen)
barcentres<-barplot(tapply(Pollen$Count,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Count,Pollen$Treatment,mean)+5)), ylab = "Pollen Count", xlab = "Treatment", main = "Number of Pollen vs Treatment")
means<-tapply(Pollen$Count,Pollen$Treatment,mean)
ses<-tapply(Pollen$Count,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Count,Pollen$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#test pollen count versus treatment, Poisson correction - TOTAL
count<-Pollen$Count
modcountpoisson<-glm(Pollen$Count~Treatment, family=poisson(), data=Pollen,)
summary(modcountpoisson)
tukcount<-glht(modcountpoisson,mcp(Treatment="Tukey"))
summary(tukcount)
Props<-Pollen$Tube/Pollen$Count
Pollen<-data.frame(Pollen,count)
```

```
str(Pollen)
barcentres<-barplot(tapply(Pollen$Count,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Count,Pollen$Treatment,mean)+5)), ylab = "Pollen Count", xlab = "Treatment", main = "Number of Pollen vs Treatment")
means<-tapply(Pollen$Count,Pollen$Treatment,mean)
ses<-tapply(Pollen$Count,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Count,Pollen$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#test pollen tube count versus treatment, Poisson correction - TOTAL
tubecount<-Pollen$Tube
modtubepoisson<-glm(Pollen$Tube~Treatment, family=poisson(), data=Pollen,)
summary(modtubepoisson)
tuktube<-glht(modtubepoisson,mcp(Treatment="Tukey"))
summary(tuktube)
Props<-Pollen$Tube/Pollen$Count
Pollentube<-data.frame(Pollen,tubecount)
str(Pollentube)
barcentres<-barplot(tapply(Pollen$Tube,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Tube,Pollen$Treatment,mean)+2)), ylab = "Pollen Tube Count", xlab = "Treatment", main = "Number of Pollen Tubes vs Treatment")
means<-tapply(Pollen$Tube,Pollen$Treatment,mean)
ses<-tapply(Pollen$Tube,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Tube,Pollen$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#test reached versus treatment, binomial model
mod1<-glm(Reached~Treatment,family=binomial(link='logit'),data=Pollen)
summary(mod1)
library(multcomp)
tuk1<-glht(mod1,mcp(Treatment="Tukey"))
summary(tuk1)
```

```
barplot(tapply(Pollen$Reached,Pollen$Treatment,mean))
barcentres<-barplot(tapply(Pollen$Reached,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Reached,Pollen$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Reaching Ovary", xlab = "Treatment", main = "Proportion of Pollen Reaching Ovary vs Treatment")
means<-tapply(Pollen$Reached,Pollen$Treatment,mean)
ses<-tapply(Pollen$Reached,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Reached,Pollen$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#test prop tube versus treatment, quasibinomial model
Prop<-cbind(Pollen$Tube, Pollen$Count)
mod2<-glm(Prop~Treatment, family=quasibinomial(link = 'logit'), data=Pollen)
summary(mod2)
tuk2<-glht(mod2,mcp(Treatment="Tukey"))
summary(tuk2)
Props<-Pollen$Tube/Pollen$Count
Pollen<-data.frame(Pollen,Props)
str(Pollen)
barcentres<-barplot(tapply(Pollen$Props,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Tube Number", xlab = "Treatment", main = "Proportion of Pollen Tube vs Treatment")
means<-tapply(Pollen$Props,Pollen$Treatment,mean)
ses<-tapply(Pollen$Props,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Props,Pollen$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#test proptubes versus treatment x latitude
modlat<-glm(Prop~Treatment*Lat, family=quasibinomial(link = 'logit'), data=Pollen)
summary(modlat)
anova(modlat,test = "F")
```

```
interactlat<-interaction(Pollen$Treatment,Pollen$Lat)
Pollen<-data.frame(Pollen,interactlat)
View(Pollen)
modlatx<-glm(Prop~-1+interactlat, family=quasibinomial(link = 'logit'), data=Pollen)
summary(modlatx)
tuklatx<-glht(modlatx,mcp(interactlat="Tukey"))
summary(tuklatx)
Props<-Pollen$Tube/Pollen$Count
Pollen<-data.frame(Pollen,Props)
barcentres<-
barplot(tapply(Pollen$Props,Pollen$interactlat,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$interact,mean
)+1.0)), ylab = "Proportion of Pollen Tube Number/Pollen Count", xlab = "Treatment and Latitude", main =
"Proportion of Pollen Tube per Latitude and Treatment")

#test prop tubes/count versus treatment x species
mod3<-glm(Props~Treatment*Species, family=quasibinomial(link = 'logit'), data=Pollen)
summary(mod3)
anova(mod3,test="F")
interact<-interaction(Pollen$Treatment,Pollen$Species)
Pollen<-data.frame(Pollen,interact)
mod3x<-glm(Props~-1+interact, family=quasibinomial(link = 'logit'), data=Pollen)
tuk3x<-glht(mod3x,mcp(interact="Tukey"))
summary(tuk3x)
Props<-Pollen$Tube/Pollen$Count
Pollen<-data.frame(Pollen,Props)
str(Pollen)
barcentres<-
barplot(tapply(Pollen$Props,Pollen$interact,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$interact,mean)+
0.1)), ylab = "Proportion of Pollen Tube Number/Pollen Count", xlab = "Treatment and Species", main =
"Proportion of Pollen Tube per species and treatment")
```

```
means<-tapply(Pollen$Props,Pollen$interact,mean)
ses<-tapply(Pollen$Props,Pollen$interact,sd)/sqrt(tapply(Pollen$Props,Pollen$interact,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)



#Analyse species separately for comparison:
#bienne:
library(readr)
bienne_all <- read_csv("bienne_all.csv",
                    col_types = cols(Count = col_number(),
                                Tube = col_number(), Percent = col_number(),
                                Reached = col_number(), Lat = col_number(),
                                pc1 = col_number()))
View(bienne_all)
bienne_all$Treatment<-as.factor(bienne_all$Treatment)

Propsbienne<-bienne_all$Tube/bienne_all$Count
Propbienne<-cbind(bienne_all$Tube, bienne_all$Count)
modpropbienne<-glm(Propbienne~Treatment, family=quasibinomial(link = 'logit'), data=bienne_all)
summary(modpropbienne)
tukpropbienne<-glht(modpropbienne,mcp(Treatment="Tukey"))
summary(tukpropbienne)
Propsb<-bienne_all$Tube/bienne_all$Count
bienne_all<-data.frame(bienne_all,Propsb)
str(bienne_all)
barcentres<-
barplot(tapply(bienne_all$Propsb,bienne_all$Treatment,mean),ylim=c(0,max(tapply(bienne_all$Propsb,bienn
```

```
e_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Tube Number", xlab = "Treatment", main =
"Proportion of Pollen Tube vs Treatment for L.bienne")
means<-tapply(bienne_all$Propsb,bienne_all$Treatment,mean)
ses<-
tapply(bienne_all$Propsb,bienne_all$Treatment,sd)/sqrt(tapply(bienne_all$Propsb,bienne_all$Treatment,len
gth))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#usitatissimum:
library(readr)
usitatissimum_all <- read_csv("usitatissimum_all.csv",
                          col_types = cols(Count = col_number(),
                                         Tube = col_number(), Percent = col_number(),
                                         Reached = col_number(), Lat = col_number(),
                                         pc1 = col_number()))
View(usitatissimum_all)
usitatissimum_all$Treatment<-as.factor(usitatissimum_all$Treatment)

Propscul<-usitatissimum_all$Tube/usitatissimum_all$Count
Propcul<-cbind(usitatissimum_all$Tube, usitatissimum_all$Count)
modpropcul<-glm(Propcul~Treatment, family=quasibinomial(link = 'logit'), data=usitatissimum_all)
summary(modpropcul)
tukpropcul<-glht(modpropcul,mcp(Treatment="Tukey"))
summary(tukpropcul)
Propsc<-usitatissimum_all$Tube/usitatissimum_all$Count
usitatissimum_all<-data.frame(usitatissimum_all,Propsc)
str(usitatissimum_all)
barcentres<-
barplot(tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,mean),ylim=c(0,max(tapply(usitatissim
```

```
um_all$Propsc,usitatissimum_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Tube Number", xlab = "Treatment", main = "Proportion of Pollen Tube vs Treatment for L.usitatissimum")
means<-tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,mean)
ses<-tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,sd)/sqrt(tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)


###test reached versus treatment x species with mixed model, ind as random effect
library(lmerTest)
library("MuMIn")
mod4<-glmer(Reached~Treatment*Species+(1|Ind), family=binomial(link='logit'), data=Pollen)
summary(mod4)
anova(mod4,test="F")
r.squaredGLMM(mod4)
mod4x<-glm(Reached~Treatment*Species, family=binomial(link='logit'), data=Pollen)
anova(mod4, mod4x, test="LRT") #likelihood ratio test
library(multcomp)
tuk4<-glht(mod4,mcp(Treatment="Tukey"))
summary(tuk4)
interact<-interaction(Pollen$Treatment,Pollen$Species)
Pollen<-data.frame(Pollen,interact)
mod4y<-glmer(Reached~-1+interact+(1|Ind), family=binomial(link='logit'), data=Pollen)
tuk4y<-glht(mod4y,mcp(interact="Tukey"))
summary(tuk4y)
barcentres<-barplot(tapply(Pollen$Reached,Pollen$interact,mean),ylim=c(0,max(tapply(Pollen$Reached,Pollen$interact,mean)+0.1)),ylab = "Proportion of Pollen Tube Reached", xlab = "Treatment and Species", main = "Proportion of Pollen Tube Reaching Ovaries For Each Treatment and Species")
```

```
means<-tapply(Pollen$Reached,Pollen$interact,mean)
ses<-tapply(Pollen$Reached,Pollen$interact,sd)/sqrt(tapply(Pollen$Reached,Pollen$interact,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)
write.table(tuk4y)

#Pollen Reached for bienne:
library(readr)
bienne_all <- read_csv("bienne_all.csv",
                    col_types = cols(Count = col_number(),
                                 Tube = col_number(), Percent = col_number(),
                                 Reached = col_number(), Lat = col_number(),
                                 pc1 = col_number()))
View(bienne_all)
bienne_all$Treatment<-as.factor(bienne_all$Treatment)

modbreach<-glm(Reached~Treatment,family=binomial(link='logit'),data=bienne_all)
summary(modbreach)
library(multcomp)
tukbreach<-glht(modbreach,mcp(Treatment="Tukey"))
summary(tukbreach)
barplot(tapply(bienne_all$Reached,bienne_all$Treatment,mean))
barcentres<-barplot(tapply(bienne_all$Reached,bienne_all$Treatment,mean),ylim=c(0,max(tapply(bienne_all$Reached,bienne_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Reaching Ovary", xlab = "Treatment", main = "Proportion of Pollen Reaching Ovary vs Treatment for L.bienne")
means<-tapply(bienne_all$Reached,bienne_all$Treatment,mean)
ses<-tapply(bienne_all$Reached,bienne_all$Treatment,sd)/sqrt(tapply(bienne_all$Reached,bienne_all$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)
```

```
#Pollen Reached for usitatissimum:
library(readr)
usitatissimum_all <- read_csv("usitatissimum_all.csv",
                          col_types = cols(Count = col_number(),
                                    Tube = col_number(), Percent = col_number(),
                                    Reached = col_number(), Lat = col_number(),
                                    pc1 = col_number()))
View(usitatissimum_all)
usitatissimum_all$Treatment<-as.factor(usitatissimum_all$Treatment)

modcreach<-glm(Reached~Treatment,family=binomial(link='logit'),data=bienne_all)
summary(modcreach)
library(multcomp)
tukcreach<-glht(modcreach,mcp(Treatment="Tukey"))
summary(tukcreach)
barplot(tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean))
barcentres<-barplot(tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean),ylim=c(0,max(tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Reaching Ovary", xlab = "Treatment", main = "Proportion of Pollen Reaching Ovary vs Treatment for L.usitatissimum")
means<-tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean)
ses<-tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,sd)/sqrt(tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)


#test proportion tubes versus treatment x species with mixed model, ind as random effect
Prop<-cbind(Pollen$Tube, Pollen$Count)
```

```
mod5<-glmer(Prop~Treatment*Species+(1|Ind), family=quasibinomial(link='logit'), data=Pollen)
summary(mod5)
tuk5<-glht(mod5,mcp(Treatment="Tukey"))
summary(tuk5)
interact<-interaction(Pollen$Treatment,Pollen$Species)
mod5y<-glmer(Prop~-1+interact+(1|Ind), family=binomial(link='logit'), data=Pollen)
summary(mod5y)
tuk5y<-glht(mod5y,mcp(interact="Tukey"))
summary(tuk5y)
Props<-Pollen$Tube/Pollen$Count
Pollen<-data.frame(Pollen,Props)
str(Pollen)
barcentres<-
barplot(tapply(Pollen$Props,Pollen$interact,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$interact,mean)+
0.1)))
means<-tapply(Pollen$Props,Pollen$interact,mean)
ses<-tapply(Pollen$Props,Pollen$interact,sd)/sqrt(tapply(Pollen$Props,Pollen$interact,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3
```

## Appendix 15 : Adapter sequences used for the ddRADSeq process

Adapter_1 = AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT --

Adapter_2 = CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

## Appendix 16 : P1 and P2 Barcode sequences for each samples

| P1 | P2 | Sample |
|---|---|---|
| gcatga | tcatgc | Sample_18 |
| gcatga | cggatc | Sample_19 |
| gcatga | taagac | Sample_20 |
| gcatga | actccg | Sample_49 |
| gcatga | aacgtg | Sample_50 |
| gcatga | ctgtat | Sample_69 |
| gcatga | gtaaca | Sample_70 |
| gcatga | gagcgt | Sample_71 |
| gcatga | tgccaa | Sample_106 |
| gcatga | gttact | Sample_83 |
| gcatga | cgcttg | Sample_21 |
| gcatga | tagcat | Sample_22 |
| gatccg | tcatgc | Sample_23 |
| gatccg | cggatc | Sample_74 |
| gatccg | taagac | Sample_84 |
| gatccg | actccg | Sample_85 |
| gatccg | aacgtg | Sample_92 |
| gatccg | ctgtat | Sample_93 |
| gatccg | gtaaca | Sample_107 |
| gatccg | gagcgt | Sample_110 |
| gatccg | tgccaa | Sample_111 |
| gatccg | tagcat | Sample_87 |
| gatccg | cgcttg | Sample_86 |
| gtctta | tcatgc | Sample_88 |
| gtctta | cggatc | Sample_89 |

|  |  |  |  |
|---|---|---|---|
| 5767 | gtctta | taagac | Sample_90 |
| 5768 | gtctta | actccg | Sample_91 |
| 5769 | gtctta | aacgtg | Sample_104 |
| 5770 | gtctta | ctgtat | Sample_105 |
| 5771 | gtctta | gagcgt | Sample_108 |
| 5772 | gtctta | tgccaa | Sample_109 |
| 5773 | gtctta | cgcttg | Sample_54 |
| 5774 | gtctta | tagcat | Sample_55 |
| 5775 | cggagt | tcatgc | Sample_56 |
| 5776 | cggagt | cggatc | Sample_57 |
| 5777 | cggagt | taagac | Sample_40 |
| 5778 | cggagt | actccg | Sample_45 |
| 5779 | cggagt | aacgtg | Sample_46 |
| 5780 | cggagt | ctgtat | Sample_47 |
| 5781 | cggagt | gtaaca | Sample_48 |
| 5782 | cggagt | gagcgt | Sample_52 |
| 5783 | cggagt | tgccaa | Sample_61 |
| 5784 | cggagt | gttact | Sample_77 |
| 5785 | cggagt | cgcttg | Sample_81 |
| 5786 | cggagt | tagcat | Sample_62 |
| 5787 | cacgtt | tcatgc | Sample_63 |
| 5788 | cacgtt | cggatc | Sample_64 |
| 5789 | cacgtt | taagac | Sample_65 |
| 5790 | cacgtt | aacgtg | Sample_67 |
| 5791 | cacgtt | ctgtat | Sample_51 |
| 5792 | cacgtt | gtaaca | Sample_80 |
| 5793 | cacgtt | gagcgt | Sample_94 |

|  |  |  |  |
|---|---|---|---|
| 5794 | cacgtt | tgccaa | Sample_95 |
| 5795 | cacgtt | gttact | Sample_96 |
| 5796 | cacgtt | cgcttg | Sample_103 |
| 5797 | cacgtt | tagcat | Sample_112 |
| 5798 | atacag | tcatgc | Sample_113 |
| 5799 | atacag | cggatc | Sample_42 |
| 5800 | atacag | taagac | Sample_25 |
| 5801 | atacag | actccg | Sample_26 |
| 5802 | atacag | aacgtg | Sample_27 |
| 5803 | atacag | ctgtat | Sample_28 |
| 5804 | atacag | gtaaca | Sample_29 |
| 5805 | atacag | gagcgt | Sample_30 |
| 5806 | atacag | tgccaa | Sample_31 |
| 5807 | atacag | gttact | Sample_75 |
| 5808 | atacag | cgcttg | Sample_100 |
| 5809 | atacag | tagcat | Sample_101 |
| 5810 | tgttac | cggatc | Sample_114 |
| 5811 | tgttac | taagac | Sample_115 |
| 5812 | tgttac | actccg | Sample_24 |
| 5813 | tgttac | aacgtg | Sample_43 |
| 5814 | tgttac | ctgtat | Sample_82 |
| 5815 | tgttac | gtaaca | Sample_58 |
| 5816 | tgttac | gagcgt | Sample_60 |
| 5817 | tgttac | tgccaa | Sample_44 |
| 5818 | tgttac | gttact | Sample_12 |
| 5819 | tgttac | cgcttg | Sample_13 |
| 5820 | tgttac | tagcat | Sample_14 |

|  |  |  |  |
|---|---|---|---|
| 5821 | acgctc | tcatgc | Sample_15 |
| 5822 | acgctc | cggatc | Sample_16 |
| 5823 | acgctc | taagac | Sample_17 |
| 5824 | acgctc | actccg | Sample_73 |
| 5825 | acgctc | aacgtg | Sample_41 |
| 5826 | acgctc | ctgtat | Sample_33 |
| 5827 | acgctc | gtaaca | Sample_34 |
| 5828 | acgctc | gagcgt | Sample_35 |
| 5829 | acgctc | tgccaa | Sample_36 |
| 5830 | acgctc | gttact | Sample_37 |
| 5831 | acgctc | cgcttg | Sample_38 |
| 5832 | acgctc | tagcat | Sample_39 |
| 5833 | ttggca | cggatc | Sample_79 |
| 5834 | ttggca | aacgtg | Sample_99 |
| 5835 | ttggca | ctgtat | Sample_59 |
| 5836 | ttggca | gtaaca | Sample_6 |
| 5837 | ttggca | gagcgt | Sample_7 |
| 5838 | ttggca | tgccaa | Sample_8 |
| 5839 | ttggca | gttact | Sample_9 |
| 5840 | ttggca | cgcttg | Sample_10 |
| 5841 | ttggca | tagcat | Sample_11 |
| 5842 | agtaac | tcatgc | Sample_116 |
| 5843 | agtaac | cggatc | Sample_117 |
| 5844 | agtaac | taagac | Sample_118 |
| 5845 | agtaac | actccg | Sample_119 |
| 5846 | agtaac | aacgtg | Sample_120 |
| 5847 | agtaac | ctgtat | Sample_1 |

```
agtaac	gtaaca	Sample_2
agtaac	gagcgt	Sample_3
agtaac	tgccaa	Sample_4
agtaac	gttact	Sample_5
agtaac	cgcttg	Sample_32
agtaac	tagcat	Sample_76
caagcg	tcatgc	Sample_121
caagcg	cggatc	Sample_122
caagcg	taagac	Sample_123
caagcg	actccg	Sample_124
caagcg	aacgtg	Sample_125
caagcg	ctgtat	Sample_126
caagcg	gtaaca	Sample_150
caagcg	gagcgt	Sample_151
caagcg	tgccaa	Sample_152
```

## Appendix 17: Process_radtags command

```
process_radtags -P -1 /nobackup/mnkz72/raw/merged_rad_raw_R1.fastq.gz -2
/nobackup/mnkz72/raw/merged_rad_raw_R2.fastq.gz -b /nobackup/mnkz72/barcodes/index_p1p2new.txt -o
/nobackup/mnkz72/raw/demultiplexed -r -c -q --barcode_dist_2 2 --disable_rad_check --inline_inline --renz_1
pstI --renz_2 mseI --adapter_1 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT -
-adapter_2 CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
```

## Appendix 18: BWA-MEM Command

```
bwa mem -t16 /nobackup/mnkz72/1st_genome_asm/cultivar/GCA_000224295.2_ASM22429v2_genomic.fna
/nobackup/mnkz72/raw/demultiplexed_run2/clean/Sample_1.1.fq.gz
/nobackup/mnkz72/raw/demultiplexed_run2/clean/Sample_1.1.fq.gz | samtools view -b -h | samtools sort -
@12 -o /nobackup/mnkz72/raw/aligned_run2/Sample_1.bam
```

## Appendix 19: Population list for popmap file

| | | |
|---|---|---|
| Sample_1 | Vil | FRA |
| Sample_10 | Tym | UKN |
| Sample_100 | Mat | FRA |
| Sample_101 | Mat | FRA |
| Sample_103 | Lla | SNO |
| Sample_104 | 15 | SNO |
| Sample_105 | 15 | SNO |
| Sample_106 | 15 | SNO |
| Sample_107 | 12 | MED |
| Sample_108 | 15 | SNO |
| Sample_109 | 15 | SNO |
| Sample_11 | Tym | UKN |
| Sample_110 | 12 | MED |
| Sample_111 | 12 | MED |
| Sample_112 | Lla | SNO |
| Sample_113 | Lla | SNO |
| Sample_114 | Mat | FRA |
| Sample_115 | Mat | FRA |
| Sample_116 | Tym | UKN |
| Sample_117 | Tym | UKN |
| Sample_118 | Tym | UKN |
| Sample_119 | Tym | UKN |
| Sample_12 | Sut | UKN |
| Sample_120 | Tym | UKN |
| Sample_121 | Vil | FRA |
| Sample_122 | Vil | FRA |

|  |  |  |  |
|---|---|---|---|
| 5904 | Sample_123 | Vil | FRA |
| 5905 | Sample_124 | Vil | FRA |
| 5906 | Sample_125 | Vil | FRA |
| 5907 | Sample_126 | Vil | FRA |
| 5908 | Sample_13 | Sut | UKN |
| 5909 | Sample_14 | Sut | UKN |
| 5910 | Sample_15 | Sut | UKN |
| 5911 | Sample_16 | Sut | UKN |
| 5912 | Sample_17 | Sut | UKN |
| 5913 | Sample_18 | 1 | SSA |
| 5914 | Sample_19 | 1 | SSA |
| 5915 | Sample_2 | Vil | FRA |
| 5916 | Sample_20 | 1 | SSA |
| 5917 | Sample_21 | 6 | SSB |
| 5918 | Sample_22 | Ome | CUL |
| 5919 | Sample_23 | 6 | SSB |
| 5920 | Sample_24 | Mon | CUL |
| 5921 | Sample_25 | Mat | FRA |
| 5922 | Sample_26 | Mat | FRA |
| 5923 | Sample_27 | Mat | FRA |
| 5924 | Sample_28 | Mat | FRA |
| 5925 | Sample_29 | Mat | FRA |
| 5926 | Sample_3 | Vil | FRA |
| 5927 | Sample_30 | Mat | FRA |
| 5928 | Sample_31 | Mat | FRA |
| 5929 | Sample_32 | Vil | FRA |
| 5930 | Sample_33 | Tal | FRA |

|  |  |  |  |
|---|---|---|---|
| 5931 | Sample_34 | Tal | FRA |
| 5932 | Sample_35 | Tal | FRA |
| 5933 | Sample_36 | Tal | FRA |
| 5934 | Sample_37 | Tal | FRA |
| 5935 | Sample_38 | Tal | FRA |
| 5936 | Sample_39 | Tal | FRA |
| 5937 | Sample_40 | Bol | CUL |
| 5938 | Sample_41 | Suz | CUL |
| 5939 | Sample_42 | Mar | CUL |
| 5940 | Sample_43 | Olg | CUL |
| 5941 | Sample_44 | Rab | CUL |
| 5942 | Sample_45 | Dor | UKS |
| 5943 | Sample_46 | Dor | UKS |
| 5944 | Sample_47 | Dor | UKS |
| 5945 | Sample_48 | Dor | UKS |
| 5946 | Sample_49 | 3 | SSA |
| 5947 | Sample_5 | Vil | FRA |
| 5948 | Sample_50 | 3 | SSA |
| 5949 | Sample_51 | Lir | CUL |
| 5950 | Sample_52 | Gis | CUL |
| 5951 | Sample_54 | 19 | SNO |
| 5952 | Sample_55 | 19 | SNO |
| 5953 | Sample_56 | 19 | SNO |
| 5954 | Sample_57 | 19 | SNO |
| 5955 | Sample_59 | Tin | CUL |
| 5956 | Sample_6 | Tym | UKN |
| 5957 | Sample_60 | Pri | CUL |

|  |  |  |  |
|---|---|---|---|
| 5958 | Sample_61 | Iow2 | UKS |
| 5959 | Sample_62 | Iow2 | UKS |
| 5960 | Sample_63 | Iow2 | UKS |
| 5961 | Sample_64 | Iow2 | UKS |
| 5962 | Sample_65 | Iow2 | UKS |
| 5963 | Sample_67 | Iow2 | UKS |
| 5964 | Sample_7 | Tym | UKN |
| 5965 | Sample_70 | 3 | SSA |
| 5966 | Sample_71 | 3 | SSA |
| 5967 | Sample_73 | Sut | UKN |
| 5968 | Sample_74 | 10 | SSB |
| 5969 | Sample_75 | Mat | FRA |
| 5970 | Sample_76 | Vil | FRA |
| 5971 | Sample_77 | Iow2 | UKS |
| 5972 | Sample_79 | Tal | FRA |
| 5973 | Sample_8 | Tym | UKN |
| 5974 | Sample_80 | Lla | SNO |
| 5975 | Sample_81 | Iow2 | UKS |
| 5976 | Sample_82 | Olg | CUL |
| 5977 | Sample_83 | 3 | SSA |
| 5978 | Sample_84 | 10 | SSB |
| 5979 | Sample_85 | 10 | SSB |
| 5980 | Sample_86 | 13 | SNO |
| 5981 | Sample_87 | 13 | SNO |
| 5982 | Sample_88 | 13 | SNO |
| 5983 | Sample_89 | 13 | SNO |
| 5984 | Sample_9 | Tym | UKN |

| | | |
|---|---|---|
| Sample_90 | 13 | SNO |
| Sample_91 | 13 | SNO |
| Sample_92 | 10 | SSB |
| Sample_93 | 3 | SSB |
| Sample_94 | Lla | SNO |
| Sample_95 | Lla | SNO |
| Sample_96 | Lla | SNO |
| Sample_99 | Tal | FRA |
| Sample_97 | Tal | FRA |
| Sample_98 | Tal | FRA |

## Appendix 20: ref_map.pl pipeline command

```
ref_map.pl -T 16 -r 0.8 -X populations:--fstats --structure --plink --vcf  --samples
/nobackup/mnkz72/raw/aligned_samples --popmap
/nobackup/mnkz72/popmap/popmap/popmap_new_0322.txt -o /nobackup/mnkz72/new_refmap/run_2
```

## Appendix 21: PLINK command

```
module load bioinformatics
module load plink
#Change path to workind directory:
cd /nobackup/mnkz72/new_refmap/rdir/run3
#Allow for extra choromosomes
plink --file populations.plink --out run3_pops --recodeA --allow-extra-chr --noweb
#Make BED:
plink --file populations.plink --out run3_pops --make-bed --allow-extra-chr --noweb
```

## Appendix 22: Geographical information of the different population

| sample | longitude | latitude |
|---|---|---|
| 1_1 | -5.39258 | 36.80044 |
| 1_26 | -5.39258 | 36.80044 |
| 1_6 | -5.39258 | 36.80044 |
| 10_15 | -6.61781 | 37.88211 |
| 10_26 | -6.61781 | 37.88211 |
| 10_27 | -6.61781 | 37.88211 |
| 10_30 | -6.61781 | 37.88211 |
| 12_10 | 3.151889 | 42.31008 |
| 12_20 | 3.151889 | 42.31008 |
| 12_8 | 3.151889 | 42.31008 |
| 13_1 | -3.24035 | 43.02902 |
| 13_12 | -3.24035 | 43.02902 |
| 13_19 | -3.24035 | 43.02902 |
| 13_22 | -3.24035 | 43.02902 |
| 13_23 | -3.24035 | 43.02902 |
| 3_11 | -5.55589 | 36.03633 |
| 3_13 | -5.55589 | 36.03633 |
| 3_14 | -5.55589 | 36.03633 |
| 3_15 | -5.55589 | 36.03633 |
| Gis | 12.56738 | 41.87194 |
| Iow2_10 | -1.07492 | 50.68183 |
| Iow2_13 | -1.07492 | 50.68183 |
| Iow2_14 | -1.07492 | 50.68183 |
| Iow2_2 | -1.07492 | 50.68183 |
| Iow2_25 | -1.07492 | 50.68183 |
| Iow2_26 | -1.07492 | 50.68183 |
| Iow2_7 | -1.07492 | 50.68183 |
| Iow2_9 | -1.07492 | 50.68183 |
| Lir | -3.43597 | 55.37805 |
| Lla_17 | -4.68753 | 43.40738 |
| Lla_20 | -4.68753 | 43.40738 |
| Lla_23 | -4.68753 | 43.40738 |
| Lla_33 | -4.68753 | 43.40738 |
| Lla_A | -4.68753 | 43.40738 |
| Lla_B | -4.68753 | 43.40738 |
| Lla_C | -4.68753 | 43.40738 |
| Suz | 5.291266 | 52.13263 |
| Mat_12 | 4.458631 | 48.35697 |
| Mat_13 | 4.458631 | 48.35697 |
| Mat_2 | 4.458631 | 48.35697 |
| Mat_20 | 4.458631 | 48.35697 |
| Mat_22 | 4.458631 | 48.35697 |
| Mat_24 | 4.458631 | 48.35697 |

| | | |
|---|---|---|
| Mat_25 | 4.458631 | 48.35697 |
| Mat_4 | 4.458631 | 48.35697 |
| Mat_7 | 4.458631 | 48.35697 |
| Mon | -3.43597 | 55.37805 |
| Olg | 2.213749 | 46.22764 |
| Olg_1 | 2.213749 | 46.22764 |
| Ome | 2.213749 | 46.22764 |
| Pri | 12.56738 | 41.87194 |
| Rab | -7.09262 | 31.7917 |
| Sut_29 | -0.95927 | 53.35291 |
| Sut_33 | -0.95927 | 53.35291 |
| Sut_34 | -0.95927 | 53.35291 |
| Sut_4 | -0.95927 | 53.35291 |
| Sut_41 | -0.95927 | 53.35291 |
| Tal_10 | -3.45465 | 47.6997 |
| Tal_11 | -3.45465 | 47.6997 |
| Tal_14 | -3.45465 | 47.6997 |
| Tal_28 | -3.45465 | 47.6997 |
| Tal_3 | -3.45465 | 47.6997 |
| Tal_33 | -3.45465 | 47.6997 |
| Tal_34 | -3.45465 | 47.6997 |
| Tal_38 | -3.45465 | 47.6997 |
| Tal_40 | -3.45465 | 47.6997 |
| Vil_1 | -1.05034 | 45.09393 |
| Vil_15 | -1.05034 | 45.09393 |
| Vil_2 | -1.05034 | 45.09393 |
| Vil_21 | -1.05034 | 45.09393 |
| Vil_25p | -1.05034 | 45.09393 |
| Vil_27 | -1.05034 | 45.09393 |
| Vil_32 | -1.05034 | 45.09393 |
| Vil_35 | -1.05034 | 45.09393 |
| Vil_36g | -1.05034 | 45.09393 |
| Vil_36p | -1.05034 | 45.09393 |
| Vil_10 | -1.05034 | 45.09393 |
| Tym_16 | -3.55328 | 53.30307 |
| Tym_19 | -3.55328 | 53.30307 |
| Mat_17 | 4.458631 | 48.35697 |
| Tym_3 | -3.55328 | 53.30307 |
| Tym_26g | -3.55328 | 53.30307 |
| Tym_26p | -3.55328 | 53.30307 |
| Tym_41 | -3.55328 | 53.30307 |
| Sut_17 | -0.95927 | 53.35291 |
| Sut_36 | -0.95927 | 53.35291 |
| 6_29 | -5.71117 | 37.93551 |
| 6_1 | -5.71117 | 37.93551 |
| Mat_11 | 4.458631 | 48.35697 |

| | | |
|---|---|---|
| Dor_D | -2.01 | 50.6 |
| Dor_C | -2.01 | 50.6 |
| Dor_B | -2.01 | 50.6 |
| Dor_A | -2.01 | 50.6 |
| Tin | 5.291266 | 52.13263 |
| Tym_40 | -3.55328 | 53.30307 |
| Tym_5 | -3.55328 | 53.30307 |
| 3_27 | -5.55589 | 36.03633 |
| Tym_20 | -3.55328 | 53.30307 |
| 13_8 | -3.24035 | 43.02902 |
| Tym_30 | -3.55328 | 53.30307 |

## Appendix 23: Command line to filter for heterozygousity

```
#Prepare and format file in bcftools
#First, convert multiallelic sites to biallelic using bcftools:
bcftools norm -m - "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf"  >
allsamples_biallellic_convert.vcf
#Filter the alternative alleles under certain value
bcftools view -e "FORMAT/AD[:1]<2 && INFO/AD[1]<5"
"/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/allsamples_biallellic_convert.vcf" > biallelic-filtered.vcf
#Compress vcf output with bgzip (Samtools):
bgzip -@8 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/allsamples_biallellic_convert.vcf"
bgzip -@8 "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf"
#Produce vcf file statistics for files:
bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf.gz" > run3_stats.txt
bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/allsamples_biallellic_convert.vcf.gz" >
biallelic_convert_stats.txt
#het-filtered vcf:
bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/allpopulations.hetfiltered.vcf.gz" >
hetfiltered_stats.txt
bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/allsamples.hetfiltered.vcf.gz" > het_bigfile_stats.txt
#Extract Genotypes per sample:
```

```
bcftools query -f '%CHROM %POS[\t%GT]\n' "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/biallelic-filtered.vcf" > genotypesbysample_run3.txt
head -n 1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" | cut -c 1-1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" > genotypesbysample_cut_1000.txt
bcftools query -f '%CHROM %POS[\t%GT]\n' "/nobackup/mnkz72/new_refmap/vcftools/allele_filter/biallelic_filtered_populations_snps.vcf" > genotypesbysample_biallelic.txt
head -n 1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" | cut -c 1-1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" > genotypesbysample_run3_cut_1000.txt
#score hets as 1, homs as 0
sumx<-rep(NA,nrow(genosx))
sumy<-NULL
for (i in c(1:ncol(genosx))){
  for (j in c(1:nrow(genosx))){
    if(genosx[j,i]=="1/0" | genosx[j,i]=="0/1"){
      sumx[j]<-1
    } else if (genosx[j,i]=="0/0" | genosx[j,i]=="1/1"){
      sumx[j]<-0
    }
  }
  sumy<-cbind(sumy,sumx)
  sumx<-rep(NA,nrow(genosx))
}
sumy

plot(c(1:nrow(sumy)),sumy[,3])

#proportion of hets per ind moving average
```

```
sumz<-NULL
mod<-50 # number of SNPs for moving average
means<-NULL
sumzz<-NULL
inds<-NULL
for (k in c(1:ncol(sumy))){
  for (l in c(1:nrow(sumy))){
    if (l %% mod == 0){
      meanx<-sum(sumy[c((l-mod):l),k],na.rm=T)/mod #mean over window size including missing data
      means<-c(means,meanx)
    }
  }
  sumz<-cbind(sumz,means)
  ind<-rep(k,l %/% mod)
  inds<-c(inds,ind)
  sumzz<-c(sumzz,means)
  means<-NULL
}
sumz
sumzzz<-data.frame(inds,sumzz)
head(sumzzz)

boxplot(sumzzz$sumzz~sumzzz$inds,xlab="sample",ylab="prop het")
#view individuals of interest from boxplot to find regions of high het (replace number in sumz[,])
plot(c(1:nrow(sumz)),sumz[,15],type="l",xlab="SNP_position",ylab="prop het")
#view genos in region of interest (SNP_position*mod)
sumy[(13*mod):(14*mod),15]

#plot all inds
```

```
plot(c(1:nrow(sumz)),rep(max(sumz,na.rm=T),nrow(sumz)),ylim=c(-
0.05,max(sumz,na.rm=T)),type="l",col="white",xlab="SNP_position",ylab="prop het")
for (m in c(1:ncol(sumz))){
  lines(c(1:nrow(sumz)),sumz[,m],type="l")
}

#proportion of hets per location moving average
pop<-c(1:19)# list of inds to include (same population), numbers are order in the sample list
lenp<-length(pop)
sump<-sumy[,pop]
meanp<-NULL
for (m in c(1:nrow(sump))){
  meany<-sum(sump[m,],na.rm=T)/lenp #mean over population including missing data
  meanp<-c(meanp,meany)
}
meanp

plot(meanp,type="l",xlab="SNP_position",ylab="prop het")

#find het loci
hetlocs<-which(meanp>quantile(meanp,0.99))
hetvals<-meanp[meanp>quantile(meanp,0.99)]
hetgenos<-data.frame(hetlocs,hetvals,sump[hetlocs,])
hetgenos

#test with two pops
#proportion of hets per location moving average
pop<-list(c(1:8),c(9:19))# list of inds to include (same population), numbers are order in the sample list
meanp<-NULL
```

```
meanps<-NULL
for (n in 1:length(pop)){
  lenp<-length(pop[[n]])
  sump<-sumy[,pop[[n]]]
  for (m in c(1:nrow(sump))){
    meany<-sum(sump[m,],na.rm=T)/lenp #mean over population including missing data
    meanp<-c(meanp,meany)
  }
  meanps<-cbind(meanps,meanp)
  meanp<-NULL
}
meanps

plot(meanps[,1],type="p",pch=1,xlab="SNP_position",ylab="prop het")
for (p in 2:ncol(meanps)){
  points(meanps[,p],pch=p)
}

#find het loci (top 99% hets)
hetlocs<-NULL
for (q in 1:length(pop)){
  hetloc<-which(meanps[,q]>quantile(meanps,0.99))
  hetlocs<-sort(union(hetlocs, hetloc))
}
hetvals<-meanps[hetlocs,]
hetgenos<-data.frame(hetlocs,hetvals)
hetgenos
```

## Appendix 24: Command line to filter by loci location in vcftools

```
#vcftools:
#frequency analysis
vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --freq --out run3_populations_freqanalysis
#depth:
vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --depth --out run3_populations_depthanalysis
#allele counts:
vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --counts --out run3_populations_countsanalysis
#heterozygousity:
vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --het --out run3_populations_hetanalysis
#site-quality:
vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --site-quality --out run3_populations_sitequalityanalysis
#Allele-het-depth counts for raw vcg from ref_map:
vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --freq --out allpopulations_freqanalysis_freq
vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --depth --out allpopulations_freqanalysis_depth
vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --het --out allpopulations_freqanalysis_het
vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --counts --out allpopulations_freqanalysis_counts
vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --site-quality --out allpopulations_freqanalysis_site-quality
vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --missing-indv --out allpopulations_freqanalysis_missingindv
#Filter for bi-allelic alleles:
vcftools --vcf /nobackup/mnkz72/new_refmap/vcftools/allele_filter/allsamples_cp.vcf --min-alleles 2 --max-alleles 2 --out biallelic_filtered_populations_snps –recode
```

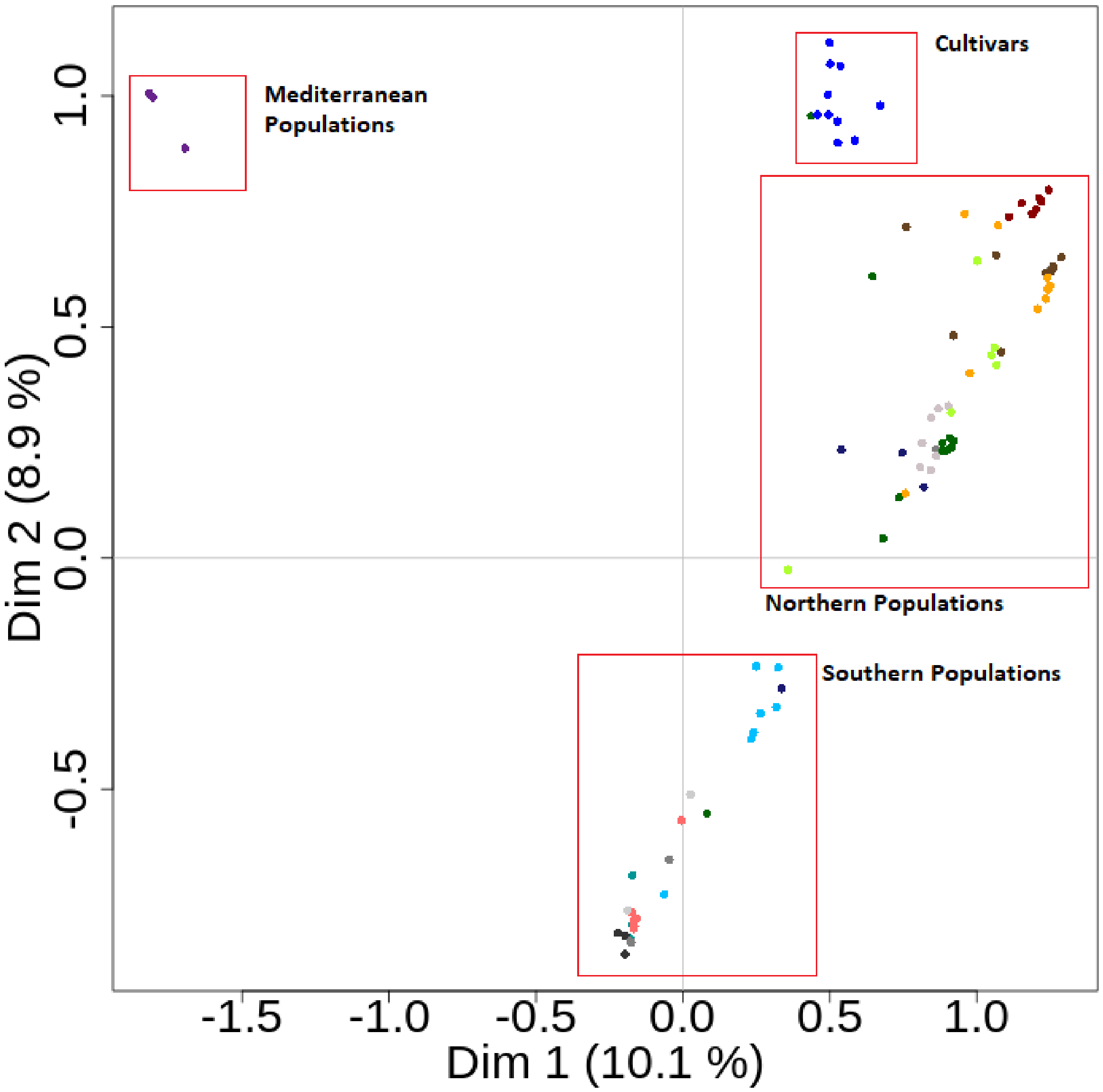
## Appendix 25: Full Individual and Population list for heterozygous filtered analysis

| name | pop | latitude | region | group |
|---|---|---|---|---|
| 3_11 | A-3 | 36.03633 | South_Spain | Southern |
| 3_13 | A-3 | 36.03633 | South_Spain | Southern |
| 3_14 | A-3 | 36.03633 | South_Spain | Southern |
| 3_15 | A-3 | 36.03633 | South_Spain | Southern |
| 3_27 | A-3 | 36.03633 | South_Spain | Southern |
| 1_1 | B-1 | 36.80044 | South_Spain | Southern |
| 1_26 | B-1 | 36.80044 | South_Spain | Southern |
| 1_6 | B-1 | 36.80044 | South_Spain | Southern |
| 10_15 | C-10 | 37.88211 | South_Spain | Southern |
| 10_26 | C-10 | 37.88211 | South_Spain | Southern |
| 10_27 | C-10 | 37.88211 | South_Spain | Southern |
| 10_30 | C-10 | 37.88211 | South_Spain | Southern |
| 6_1 | D-6 | 37.93551 | South_Spain | Southern |
| 6_29 | D-6 | 37.93551 | South_Spain | Southern |
| 12_10 | E-12 | 42.31008 | Mediterranean | Southern |
| 12_20 | E-12 | 42.31008 | Mediterranean | Southern |
| 12_8 | E-12 | 42.31008 | Mediterranean | Southern |
| 13_1 | F-13 | 43.02902 | North_Spain | Southern |
| 13_12 | F-13 | 43.02902 | North_Spain | Southern |
| 13_19 | F-13 | 43.02902 | North_Spain | Southern |
| 13_22 | F-13 | 43.02902 | North_Spain | Southern |
| 13_23 | F-13 | 43.02902 | North_Spain | Southern |
| 13_8 | F-13 | 43.02902 | North_Spain | Southern |
| Lla_17 | G-Lla | 43.40738 | North_Spain | Southern |
| Lla_20 | G-Lla | 43.40738 | North_Spain | Southern |
| Lla_23 | G-Lla | 43.40738 | North_Spain | Southern |
| Lla_33 | G-Lla | 43.40738 | North_Spain | Southern |
| Lla_A | G-Lla | 43.40738 | North_Spain | Southern |
| Lla_B | G-Lla | 43.40738 | North_Spain | Southern |
| Lla_C | G-Lla | 43.40738 | North_Spain | Southern |
| Vil_1 | H-Vil | 45.09393 | France | Northern |
| Vil_10 | H-Vil | 45.09393 | France | Northern |
| Vil_15 | H-Vil | 45.09393 | France | Northern |
| Vil_2 | H-Vil | 45.09393 | France | Northern |
| Vil_21 | H-Vil | 45.09393 | France | Northern |
| Vil_25p | H-Vil | 45.09393 | France | Northern |
| Vil_27 | H-Vil | 45.09393 | France | Northern |
| Vil_32 | H-Vil | 45.09393 | France | Northern |
| Vil_35 | H-Vil | 45.09393 | France | Northern |
| Vil_36g | H-Vil | 45.09393 | France | Northern |

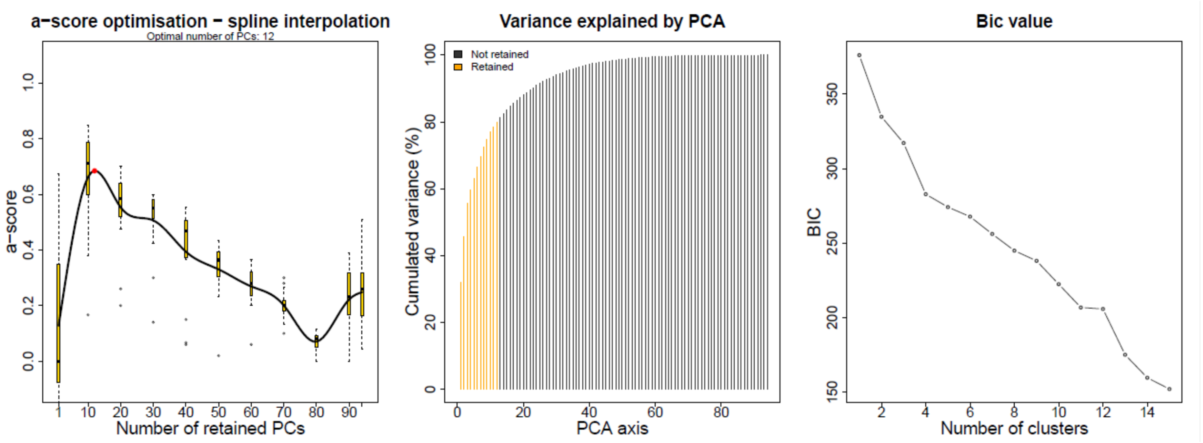
| | | | | |
|---|---|---|---|---|
| Vil_36p | H-Vil | 45.09393 | France | Northern |
| Tal_10 | I-Tal | 47.6997 | France | Northern |
| Tal_11 | I-Tal | 47.6997 | France | Northern |
| Tal_14 | I-Tal | 47.6997 | France | Northern |
| Tal_28 | I-Tal | 47.6997 | France | Northern |
| Tal_3 | I-Tal | 47.6997 | France | Northern |
| Tal_33 | I-Tal | 47.6997 | France | Northern |
| Tal_34 | I-Tal | 47.6997 | France | Northern |
| Tal_38 | I-Tal | 47.6997 | France | Northern |
| Tal_40 | I-Tal | 47.6997 | France | Northern |
| Mat_11 | J-Mat | 48.35697 | France | Northern |
| Mat_12 | J-Mat | 48.35697 | France | Northern |
| Mat_13 | J-Mat | 48.35697 | France | Northern |
| Mat_17 | J-Mat | 48.35697 | France | Northern |
| Mat_2 | J-Mat | 48.35697 | France | Northern |
| Mat_20 | J-Mat | 48.35697 | France | Northern |
| Mat_22 | J-Mat | 48.35697 | France | Northern |
| Mat_24 | J-Mat | 48.35697 | France | Northern |
| Mat_25 | J-Mat | 48.35697 | France | Northern |
| Mat_4 | J-Mat | 48.35697 | France | Northern |
| Mat_7 | J-Mat | 48.35697 | France | Northern |
| Dor_A | K-Dor | 50.6 | South_UK | Northern |
| Dor_B | K-Dor | 50.6 | South_UK | Northern |
| Dor_C | K-Dor | 50.6 | South_UK | Northern |
| Dor_D | K-Dor | 50.6 | South_UK | Northern |
| Iow2_10 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_13 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_14 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_2 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_25 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_26 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_7 | L-Iow2 | 50.68183 | South_UK | Northern |
| Iow2_9 | L-Iow2 | 50.68183 | South_UK | Northern |
| Tym_16 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_19 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_20 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_26g | M-Tym | 53.30307 | North_UK | Northern |
| Tym_26p | M-Tym | 53.30307 | North_UK | Northern |
| Tym_3 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_30 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_40 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_41 | M-Tym | 53.30307 | North_UK | Northern |
| Tym_5 | M-Tym | 53.30307 | North_UK | Northern |
| Sut_17 | N-Sut | 53.35291 | North_UK | Northern |
| Sut_29 | N-Sut | 53.35291 | North_UK | Northern |
| Sut_33 | N-Sut | 53.35291 | North_UK | Northern |

| | | | | |
|---|---|---|---|---|
| Sut_34 | N-Sut | 53.35291 | North_UK | Northern |
| Sut_36 | N-Sut | 53.35291 | North_UK | Northern |
| Sut_4 | N-Sut | 53.35291 | North_UK | Northern |
| Sut_41 | N-Sut | 53.35291 | North_UK | Northern |
| Gis | CUL | NA | Cultivar | Cultivar |
| Lir | CUL | NA | Cultivar | Cultivar |
| Mon | CUL | NA | Cultivar | Cultivar |
| Olg | CUL | NA | Cultivar | Cultivar |
| Olg_1 | CUL | NA | Cultivar | Cultivar |
| Ome | CUL | NA | Cultivar | Cultivar |
| Pri | CUL | NA | Cultivar | Cultivar |
| Rab | CUL | NA | Cultivar | Cultivar |
| Suz | CUL | NA | Cultivar | Cultivar |
| Tin | CUL | NA | Cultivar | Cultivar |

## Appendix 26: Correspondence analysis (CA) plot for chapter 4



## Appendix 27: DAPC Summary



## Appendix 28: Tajima D’s Summary

| pop | mean_pi | TajimaD | rare_alleles | demography | selection |
|---|---|---|---|---|---|
| A-3 | 82.3 | -0.42 | neutral | neutral | neutral |
| B-1 | 57.58 | -79.44 | many_rare | expansion_after_bottleneck | selective_sweep |
| C-10 | 18.3 | -71.09 | many_rare | expansion_after_bottleneck | selective_sweep |
| D-6 | 63.17 | -182.16 | many_rare | expansion_after_bottleneck | selective_sweep |
| E-12 | 39.25 | -125.65 | many_rare | expansion_after_bottleneck | selective_sweep |
| F-13 | 49.81 | -40.29 | many_rare | expansion_after_bottleneck | selective_sweep |
| G-Lla | 67.41 | -9.04 | many_rare | expansion_after_bottleneck | selective_sweep |
| H-Vil | 9.11 | -15.89 | many_rare | expansion_after_bottleneck | selective_sweep |
| I-Tal | 3.5 | -8.33 | many_rare | expansion_after_bottleneck | selective_sweep |
| J-Mat | 41.01 | -34.55 | many_rare | expansion_after_bottleneck | selective_sweep |
| K-Dor | 41.55 | -18.45 | many_rare | expansion_after_bottleneck | selective_sweep |
| L-low2 | 11.33 | -8.05 | many_rare | expansion_after_bottleneck | selective_sweep |
| M-Tym | 10.56 | -15.02 | many_rare | expansion_after_bottleneck | selective_sweep |
| N-Sut | 21.02 | -14.14 | many_rare | expansion_after_bottleneck | selective_sweep |
| | | | | | |
| P(T<=t) two-tail between North/South | 0.019298 | 0.023475 | | | |

## Appendix 29: Mean MAF for each population

| pop | mean_nsites | mean_pi | mean_maf |
|---|---|---|---|
| A-3 | 811.7 | 82.3 | 0.46 |
| B-1 | 833 | 57.58 | 0.66 |
| C-10 | 858 | 18.3 | 0.86 |
| D-6 | 692 | 63.17 | 0.57 |

| | | | |
|---|---:|---:|---:|
| E-12 | 862 | 39.25 | 0.91 |
| F-13 | 804.9 | 49.81 | 0.64 |
| G-Lla | 857.8 | 67.41 | 0.32 |
| H-Vil | 854.7 | 9.11 | 0.25 |
| I-Tal | 854.2 | 3.5 | 0.55 |
| J-Mat | 832.9 | 41.01 | 0.17 |
| K-Dor | 856 | 41.55 | 0.32 |
| L-Iow2 | 862 | 11.33 | 0.43 |
| M-Tym | 773.5 | 10.56 | 0.25 |
| N-Sut | 790 | 21.02 | 0.26 |

## Appendix 30: Chapter 5 R Commands

```
#Seed Measurements:
setwd("G:/Linum Project/rdir")
library(ggplot2)
library(ggpubr)
library(readr)
seed_measurements_wild_data_pop <- read_csv("seed_measurements_wild_data_pop.csv",
                                    col_types = cols(`Lat_(degrees_N)` = col_number(),
                                             Alt = col_number(), `Seed_length_(Cm)` = col_number(),
                                             `Seed_width_(Cm)` = col_number(),
                                             `Seed_Area_(Cm)` = col_number(),
                                             pc1 = col_number()))
View(seed_measurements_wild_data_pop)

#Modelling for seed area vs latitude:
g <- ggscatter(seed_measurements_wild_data_pop, x = "Lat_(degrees_N)", y = "Seed_Area_(Cm)", size = 1.0,
      rug = TRUE,                    # Add marginal rug
      color = "Experiment", palette = "uchicago", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Experiment), label.sep = ";", label.x.npc = "left", label.y.npc=0.20, method =
"pearson")+geom_point(aes(color = Experiment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Seed Area vs Latitude")

```

```
g +
  font("title", size = 22, color = "black", face = "bold")+
  font("xlab", size = 18, color = "black", face = "bold")+
  font("ylab", size = 18, color = "black", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 14, color = "black",)

#Modelling for seed area vs pc1:
h <- ggscatter(seed_measurements_wild_data_pop, x = "pc1", y = "Seed_Area_(Cm)", size = 1.0,
          rug = TRUE,                          # Add marginal rug
          color = "Experiment", palette = "uchicago", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Experiment), label.sep = ";", label.x.npc = "left", label.y.npc=0.20, method =
"pearson")+geom_point(aes(color = Experiment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Seed Area vs Climatic Variables (pc1)")

h +
  font("title", size = 22, color = "black", face = "bold")+
  font("xlab", size = 18, color = "black", face = "bold")+
  font("ylab", size = 18, color = "black", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 14, color = "black",)

#Vern/Non-vern on other traits:
library("car")
setwd("G:/Linum Project/rdir")
library(readr)
combined_vern_control_2018_only <- read_csv("combined_vern_control_2018_only.csv",
                                    col_types = cols(Lat = col_number(),
                                                Lon = col_number(), pc1 = col_number(),
```

```
                                        `Height_(Cm)` = col_number(), Stem_no = col_number(),
                                        Bud_no = col_number(), Days_to_fl = col_number()))
View(combined_vern_control_2018_only)
scatterplot(Days_to_fl ~ Height_(Cm) | Treatment, data = combined_vern_control_2018_only,
      smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Plant Height(Cm)", ylab = "Days to
Flowering")
abline(lm(Height ~ Days_to_fl data = combined_vern_control_2018_only),col="red")

ggscatter(combined_vern_control_2018_only, x = "Bud_no", y = "Days_to_fl", size = 2.0,
     rug = TRUE,                 # Add marginal rug
     color = "Treatment", palette = "jco", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
 stat_cor(aes(color = Treatment), method = "pearson")

#Model summary
summary1 <- summary(glm(Stem_no ~ Days_to_fl, data = combined_vern_control_2018_only))
adjRsq <- summary1$adj.r.squared
fStat <- summary1$statistic
pValue <- pf(fStat[summary1])
summary(summary1)

#Height - Days to fl test:
Heightdaysmod<-cbind(combined_vern_control_2018_only$Days_to_fl,
combined_vern_control_2018_only$'Height_(Cm)')
Heightaysmod2<-glm(Heightdaysmod~Treatment, family=quasibinomial,
data=combined_vern_control_2018_only)
summary(Heightaysmod2)
barcentres<-
barplot(tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatm
ent,mean),ylim=c(0,max(tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2
```

```
018_only$Treatment,mean)+5)), ylab = "Height (Cm)", xlab = "Treatment", main = "Plant Height vs
Treatments")
means<-
tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatment,mea
n)
ses<-
tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatment,sd)/
sqrt(tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatment
,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#Stem_no - Days to fl test:
Stemdaysmod<-cbind(combined_vern_control_2018_only$Stem_no,
combined_vern_control_2018_only$Days_to_fl)
Stemdaysmod2<-glm(Stemdaysmod~Treatment, family=quasibinomial,
data=combined_vern_control_2018_only)
summary(Stemdaysmod2)
barcentres<-
barplot(tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,
mean),ylim=c(0,max(tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only
$Treatment,mean)+2)), ylab = "Stem Number", xlab = "Treatment", main = "Stem Number vs Treatments")
means<-
tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,mean)
ses<-
tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,sd)/sqrt(
tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#Bud_no - Days to fl test:
buddaysmod<-cbind(combined_vern_control_2018_only$Days_to_fl,
combined_vern_control_2018_only$Bud_no)
buddaysmod2<-buddaysmodlm<-lm(buddaysmod~Treatment,data=combined_vern_control_2018_only)
```

```
summary(pc1daysmodlm)
barcentres<-
barplot(tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,m
ean),ylim=c(0,max(tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Tr
eatment,mean)+2)), ylab = "Bud Number", xlab = "Treatment", main = "Bud Number vs Treatments")
means<-
tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,mean)
ses<-
tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,sd)/sqrt(t
apply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,length))
arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

#PC1 - Height Correlation:
library(readr)
wildonly_vern18 <- read_csv("wildonly_vern18.csv",
                  col_types = cols(Lat = col_number(),
                             Lon = col_number(), pc1 = col_number(),
                             `Height_(Cm)` = col_number(), Stem_no = col_number(),
                             Bud_no = col_number(), Days_to_fl = col_number()))
View(wildonly_vern18)

#pc1 - height correlation:
a <- ggscatter(wildonly_vern18, x = "pc1", y = "Height_(Cm)", size = 1.0,
      rug = TRUE,                 # Add marginal rug
      color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.8, method =
"pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Plant Height (Cm) vs Climate Variable (pc1)")

a +
  font("title", size = 22, color = "black", face = "bold")+
```

```
  font("xlab", size = 18, color = "blue", face = "bold")+
  font("ylab", size = 18, color = "blue", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 18, color = "blue", face = "bold")


#Lat - height correlation:
b <- ggscatter(wildonly_vern18, x = "Lat", y = "Height_(Cm)", size = 1.0,
          rug = TRUE,                              # Add marginal rug
          color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.8, method =
"pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Plant Height (Cm) vs Latitude (degrees North)")

b +
  font("title", size = 20, color = "black", face = "bold")+
  font("xlab", size = 18, color = "blue", face = "bold")+
  font("ylab", size = 18, color = "blue", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 18, color = "blue", face = "bold")

#pc1 - Stem number correlation:
c <- ggscatter(wildonly_vern18, x = "pc1", y = "Stem_no", size = 1.0,
          rug = TRUE,                              # Add marginal rug
          color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.9, method =
"pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Stem Number vs Climate Variable (pc1)")

c +
```

```
  font("title", size = 22, color = "black", face = "bold")+
  font("xlab", size = 18, color = "blue", face = "bold")+
  font("ylab", size = 18, color = "blue", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 18, color = "blue", face = "bold")

#Lat - Stem number correlation:
d <- ggscatter(wildonly_vern18, x = "Lat", y = "Stem_no", size = 1.0,
          rug = TRUE,                              # Add marginal rug
          color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.85, method =
"pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Stem Number vs Latitude (degrees North)")

d +
  font("title", size = 22, color = "black", face = "bold")+
  font("xlab", size = 18, color = "blue", face = "bold")+
  font("ylab", size = 18, color = "blue", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 18, color = "blue", face = "bold")

#pc1 - Bud number correlation:
e <- ggscatter(wildonly_vern18, x = "pc1", y = "Bud_no", size = 1.0,
          rug = TRUE,                              # Add marginal rug
          color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.7, method =
"pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Bud Number vs Climate Variable (pc1)")

e +
```

```
  font("title", size = 22, color = "black", face = "bold")+
  font("xlab", size = 18, color = "blue", face = "bold")+
  font("ylab", size = 18, color = "blue", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 18, color = "blue", face = "bold")

#Lat - Bud number correlation:
f <- ggscatter(wildonly_vern18, x = "Lat", y = "Bud_no", size = 1.0,
          rug = TRUE,                           # Add marginal rug
          color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +
  stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.53, method =
"pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+
theme_light()+ggtitle("Bud Number vs Latitude (degrees North)")

f +
  font("title", size = 22, color = "black", face = "bold")+
  font("xlab", size = 18, color = "blue", face = "bold")+
  font("ylab", size = 18, color = "blue", face = "bold")+
  font("legend.title", size = 18, color = "black", face = "bold")+
  font("legend.text", size = 18, color = "blue", face = "bold")
```

## **Appendix 31: Plant Height, stem number, and bud number**

| Pop_ind | Height_(Cm) | Stem_no | Bud_no | Days_to_fl |
|---|---:|---:|---:|---:|
| Vil_25 | 76.5 | 10 | 2 | 292 |
| Vil_27 | 21.5 | 6 | 0 | 202 |
| Vil_27 | 21.5 | 6 | 0 | 202 |
| Vil_36 | 39.5 | 27 | 2 | 167 |
| Roc_12 | 48 | 16 | 2 | 247 |
| Roc_12 | 52.5 | 25 | 3 | 205 |
| Saf_16 | 29.5 | 10 | 0 | 212 |
| Saf_17 | 41 | 33 | 5 | 294 |
| Saf_19 | 34 | 48 | 2 | 299 |
| Saf_19 | 42 | 41 | 1 | 198 |
| Tal_10 | 34 | 6 | 1 | 149 |

| | | | | |
|---|---|---|---|---|
| Tal_10 | 26 | 3 | 0 | 69 |
| Mat_2 | 32 | 6 | 4 | 201 |
| Mat_2 | 39 | 33 | 13 | 191 |
| Mat_14 | 29 | 5 | 3 | 211 |
| Mat_14 | 24 | 12 | 2 | 177 |
| Mat_17 | 26.3 | 6 | 3 | 200 |
| Mat_17 | 49 | 20 | 16 | 198 |
| Mat_18 | 39 | 17 | 1 | 248 |
| Mat_18 | 36.5 | 21 | 12 | 190 |
| Mat_23 | 47 | 22 | 8 | 199 |
| Mat_23 | 31 | 5 | 5 | 196 |
| 19_26 | 43.5 | 11 | 0 | 184 |
| 19_26 | 43.5 | 11 | 0 | 184 |
| 19_30 | 39 | 9 | 3 | 213 |
| 14_21 | 40 | 36 | 3 | 214 |
| 13_1 | 34.5 | 12 | 1 | 191 |
| Lla_17 | 45 | 6 | 1 | 270 |
| Lla_20 | 55 | 33 | 10 | 300 |
| Lla_23 | 29 | 6 | 2 | 212 |
| Lla_43_A | 58 | 18 | 1 | 243 |
| Lla_B | 33 | 26 | 1 | 212 |
| Lla_C | 43 | 25 | 5 | 202 |
| Lla_C | 28 | 14 | 3 | 207 |
| 15_17 | 26.5 | 7 | 3 | 151 |
| 15_17 | 23 | 8 | 3 | 134 |
| 15_18 | 18 | 10 | 3 | 177 |
| 15_18 | 45 | 30 | 1 | 160 |
| 15_28 | 29.5 | 2 | 0 | 236 |
| 15_30 | 32 | 5 | 1 | 184 |
| 15_30 | 41 | 15 | 1 | 182 |
| 3_11 | 49.5 | 12 | 5 | 190 |
| 3_13 | 41.5 | 10 | 2 | 157 |
| 3_13 | 29.5 | 3 | 1 | 156 |
| 3_13 | 31 | 3 | 2 | 193 |
| 3_14 | 35.5 | 10 | 0 | 160 |
| 3_14 | 43 | 4 | 2 | 118 |
| 3_15 | 27 | 3 | 2 | 122 |
| 3_15 | 40.5 | 4 | 2 | 189 |
| 2_3 | 26.5 | 3 | 1 | 109 |
| 2_3 | 29 | 5 | 2 | 98 |
| 2_29 | 23.5 | 12 | 1 | 216 |
| 4_7 | 28 | 7 | 1 | 76 |
| 4_7 | 18.5 | 3 | 0 | 63 |
| 4_23 | 30 | 11 | 1 | 151 |

|  |  |  |  |  |
|---|---|---|---|---|
| 4_27 | 25 | 2 | 0 | 66 |
| 4_27 | 18.5 | 6 | 0 | 67 |
| 4_28 | 31 | 2 | 1 | 102 |
| 4_28 | 22 | 4 | 2 | 82 |
| 4_A | 33 | 12 | 4 | 186 |
| Tor_4 | 17.5 | 6 | 2 | 59 |
| Tor_4 | 16 | 5 | 1 | 71 |
| 1_1 | 25 | 13 | 0 | 138 |
| 1_26 | 30.5 | 5 | 1 | 152 |
| 1_26 | 46 | 5 | 0 | 178 |
| 1_A | 29.2 | 9 | 1 | 113 |
| 1_C | 39 | 12 | 1 | 128 |
| 1_C | 34.5 | 7 | 2 | 124 |
| 5_4 | 20 | 3 | 1 | 54 |
| 5_10 | 19 | 4 | 0 | 73 |
| 8_12 | 34.6 | 9 | 4 | 202 |
| 8_12 | 47 | 13 | 12 | 184 |
| 8_16 | 35.6 | 10 | 1 | 159 |
| 9_23 | 36.5 | 13 | 0 | 105 |
| 9_27 | 29.3 | 1 | 0 | 55 |
| 9_27 | 33.5 | 5 | 4 | 62 |
| 9_34 | 32 | 2 | 1 | 118 |
| 10_15 | 27.5 | 1 | 0 | 153 |
| 10_15 | 18.5 | 3 | 0 | 84 |
| 10_26 | 38 | 7 | 0 | 115 |
| 10_26 | 29.5 | 6 | 1 | 100 |
| 10_27 | 32 | 5 | 3 | 118 |
| 10_27 | 31 | 5 | 0 | 117 |
| 10_30 | 29 | 8 | 0 | 83 |
| 6_1 | 37.7 | 9 | 0 | 103 |
| 6_26 | 42.5 | 11 | 4 | 198 |
| 6_26 | 50 | 7 | 6 | 181 |
| 6_29 | 47.5 | 5 | 3 | 199 |
| 7_17 | 36 | 8 | 1 | 121 |
| 11_23 | 42 | 15 | 0 | 158 |
| Man_4 | 49 | 17 | 3 | 214 |
| Man_4 | 26 | 8 | 1 | 212 |
| Man_5 | 38 | 8 | 1 | 200 |
| Man_8 | 33 | 30 | 8 | 287 |
| Tym_5 | 33.5 | 22 | 0 | 199 |
| Tym_30 | 31 | 6 | 2 | 157 |
| IOW2_2 | 17.5 | 4 | 2 | 74 |
| IOW2_2 | 25.5 | 8 | 7 | 50 |
| IOW2_25 | 23 | 2 | 0 | 57 |

| | | | | |
|---|---|---|---|---|
| IOW2_25 | 26.5 | 5 | 1 | 66 |
| IOW2_26 | 23 | 18 | 1 | 212 |
| IOW1_11 | 36 | 15 | 0 | 307 |
| IOW1_17 | 36 | 18 | 1 | 272 |
| Ara | 78.5 | 1 | 2 | 73 |
| Ara | 93 | 3 | 4 | 77 |
| Ari | 63.5 | 4 | 8 | 57 |
| Ari | 88.5 | 2 | 4 | 45 |
| Ari | 91.5 | 3 | 3 | 66 |
| Ble | 62 | 1 | 2 | 49 |
| Ble | 45 | 3 | 8 | 38 |
| Ble | 58 | 1 | 4 | 47 |
| Ble | 40.5 | 1 | 9 | 50 |
| Bol | 54 | 1 | 2 | 68 |
| Bol | 64 | 1 | 0 | 52 |
| Ede | 73 | 1 | 2 | 65 |
| Ede | 76 | 1 | 0 | 63 |
| Ede | 73 | 1 | 2 | 65 |
| Gis | 53 | 3 | 3 | 83 |
| Gis | 52 | 1 | 1 | 58 |
| Gis | 52 | 3 | 2 | 61 |
| Gis | 52 | 5 | 2 | 63 |
| Lir | 57.5 | 2 | 4 | 49 |
| Lir | 86 | 3 | 8 | 49 |
| Lir | 45 | 3 | 6 | 56 |
| Mar | 56 | 2 | 5 | 49 |
| Mar | 44.5 | 3 | 6 | 63 |
| Mon | 47 | 3 | 6 | 54 |
| Mon | 78.5 | 1 | 3 | 40 |
| Mon | 77 | 1 | 3 | 53 |
| Olg | 16.5 | 1 | 0 | 90 |
| Olg | 58.5 | 1 | 0 | 111 |
| Ome | 36 | 1 | 5 | 49 |
| Ome | 44.5 | 1 | 5 | 40 |
| Ome | 47 | 3 | 8 | 49 |
| Ome | 55 | 1 | 10 | 46 |
| Pri | 49.5 | 3 | 3 | 65 |
| Pri | 38.5 | 1 | 2 | 48 |
| Pri | 37 | 1 | 3 | 76 |
| Pri | 32 | 7 | 1 | 129 |
| Rab | 48 | 8 | 14 | 118 |
| Rab | 48 | 3 | 11 | 38 |
| Rab | 57.5 | 1 | 10 | 45 |
| Rab | 63 | 3 | 3 | 57 |

| | | | | |
|---|---|---|---|---|
| Suz | 72 | 3 | 2 | 51 |
| Suz | 72 | 1 | 1 | 26 |
| Suz | 78 | 3 | 3 | 52 |
| Suz | 68.5 | 1 | 3 | 77 |
| Tin | 50 | 1 | 3 | 63 |
| Tin | 64 | 1 | 2 | 55 |
| Tin | 61 | 2 | 6 | 68 |
| Vol | 40 | 3 | 5 | 51 |
| Vol | 27.8 | 2 | 2 | 64 |
| Vil_21 | 27 | 3 | 3 | 98 |
| Vil_21 | 26 | 2 | 0 | 101 |
| Vil_25 | 19.5 | 3 | 0 | 105 |
| Vil_25 | 25.5 | 3 | 3 | 96 |
| Vil_27 | 24.5 | 3 | 2 | 93 |
| Vil_27 | 27.5 | 5 | 1 | 101 |
| Vil_36 | 30.5 | 3 | 1 | 92 |
| Vil_36 | 27.5 | 2 | 0 | 99 |
| Roc_12 | 38 | 4 | 2 | 93 |
| Roc_12 | 46 | 5 | 2 | 90 |
| Saf_9 | 21.5 | 4 | 1 | 95 |
| Saf_9 | 25 | 3 | 3 | 106 |
| Saf_10 | 31 | 3 | 1 | 100 |
| Saf_10 | 8 | 6 | 4 | 92 |
| Saf_16 | 25 | 3 | 2 | 98 |
| Saf_16 | 25.5 | 8 | 2 | 92 |
| Saf_19 | 27.5 | 5 | 0 | 110 |
| Saf_19 | 31 | 4 | 0 | 95 |
| Tal_4 | 26 | 4 | 5 | 95 |
| Tal_4 | 32 | 6 | 6 | 96 |
| Tal_23 | 21 | 3 | 2 | 93 |
| Tal_23 | 32.5 | 3 | 1 | 103 |
| Tal_25 | 31.5 | 3 | 0 | 131 |
| Tal_28 | 24.5 | 6 | 1 | 122 |
| Tal_28 | 30 | 4 | 1 | 117 |
| Mat_2 | 45 | 4 | 6 | 91 |
| Mat_2 | 23 | 7 | 5 | 119 |
| Mat_14 | 30.5 | 5 | 0 | 102 |
| Mat_14 | 34.5 | 2 | 2 | 97 |
| Mat_17 | 23.5 | 3 | 0 | 107 |
| Mat_17 | 19.5 | 4 | 0 | 95 |
| Mat_23 | 27.5 | 5 | 5 | 87 |
| Mat_23 | 34.5 | 1 | 1 | 103 |
| Mat_24 | 31 | 4 | 2 | 101 |
| Mat_24 | 45 | 6 | 5 | 98 |

|  |  |  |  |  |
|---|---|---|---|---|
| 12_20 | 36 | 3 | 1 | 85 |
| 19_30 | 25 | 9 | 0 | 99 |
| 19_30 | 30 | 3 | 2 | 106 |
| 14_6 | 17 | 5 | 0 | 105 |
| 14_21 | 31.5 | 6 | 1 | 96 |
| 14_21 | 42 | 2 | 0 | 99 |
| 13_1 | 39 | 1 | 3 | 98 |
| 13_1 | 22.5 | 6 | 2 | 89 |
| 13_5 | 24 | 8 | 0 | 134 |
| 13_12 | 21.5 | 4 | 0 | 106 |
| Lla_17 | 19 | 1 | 1 | 100 |
| Lla_20 | 32.5 | 3 | 1 | 99 |
| Lla_20 | 30 | 2 | 1 | 95 |
| Lla_23 | 25 | 3 | 2 | 131 |
| Lla_23 | 26 | 3 | 0 | 127 |
| Lla_33 | 30 | 3 | 2 | 106 |
| Lla_33 | 28 | 6 | 4 | 106 |
| Lla_A | 30.5 | 7 | 3 | 110 |
| Lla_A | 29.5 | 4 | 2 | 98 |
| Lla_B | 30 | 10 | 11 | 98 |
| Lla_C | 28.5 | 4 | 2 | 99 |
| 15_17 | 29.5 | 1 | 2 | 92 |
| 15_17 | 23 | 4 | 2 | 89 |
| 15_18 | 30 | 5 | 3 | 88 |
| 15_18 | 24.5 | 8 | 3 | 85 |
| 15_27 | 24.5 | 3 | 2 | 97 |
| 15_27 | 32 | 3 | 1 | 96 |
| 15_28 | 33 | 3 | 1 | 101 |
| 15_28 | 29 | 5 | 2 | 102 |
| 15_30 | 32 | 1 | 2 | 97 |
| 15_30 | 29 | 2 | 0 | 103 |
| 3_11 | 27 | 2 | 1 | 90 |
| 3_11 | 30.5 | 4 | 2 | 93 |
| 3_13 | 24.5 | 4 | 0 | 97 |
| 3_13 | 31 | 4 | 1 | 95 |
| 3_14 | 24.2 | 1 | 1 | 91 |
| 3_14 | 34 | 3 | 2 | 97 |
| 3_16 | 58 | 5 | 1 | 101 |
| 2_3 | 27 | 3 | 1 | 80 |
| 2_3 | 26.5 | 3 | 3 | 88 |
| 2_29 | 30 | 5 | 1 | 91 |
| 2_29 | 24 | 7 | 1 | 102 |
| 4_5 | 23 | 2 | 2 | 87 |
| 4_5 | 28.1 | 3 | 2 | 80 |

| | | | | |
|---|---|---|---|---|
| 4_7 | 26 | 3 | 1 | 89 |
| 4_7 | 25 | 4 | 3 | 92 |
| 4_23 | 19 | 1 | 1 | 88 |
| 4_23 | 28 | 1 | 1 | 86 |
| 4_27 | 24.3 | 3 | 2 | 84 |
| 4_28 | 25.3 | 1 | 1 | 82 |
| Tor_4 | 18.5 | 6 | 9 | 77 |
| Tor_4 | 17 | 4 | 3 | 79 |
| 1_26 | 24.5 | 3 | 0 | 90 |
| 1_26 | 25 | 3 | 3 | 88 |
| 1_A | 18.5 | 2 | 0 | 86 |
| 1_A | 28 | 3 | 2 | 91 |
| 5_4 | 28 | 4 | 3 | 79 |
| 5_4 | 24.5 | 3 | 3 | 78 |
| 5_4 | 21 | 3 | 2 | 78 |
| 5_10 | 19.5 | 1 | 2 | 83 |
| 5_10 | 18.7 | 2 | 2 | 84 |
| 8_16 | 32 | 5 | 7 | 87 |
| 8_16 | 21 | 7 | 5 | 85 |
| 8_27 | 24.2 | 3 | 0 | 82 |
| 9_24 | 24.5 | 5 | 5 | 82 |
| 9_27 | 23 | 5 | 2 | 84 |
| 10_15 | 24.5 | 5 | 2 | 74 |
| 10_26 | 28 | 5 | 0 | 77 |
| 10_26 | 19.5 | 1 | 0 | 84 |
| 10_27 | 20 | 5 | 0 | 78 |
| 10_27 | 26 | 5 | 7 | 78 |
| 10_30 | 19.3 | 5 | 1 | 76 |
| 6_26 | 21 | 5 | 0 | 92 |
| 6_26 | 24 | 5 | 1 | 95 |
| 6_29 | 30 | 1 | 0 | 98 |
| 6_29 | 27 | 7 | 3 | 98 |
| 11_23 | 24 | 9 | 4 | 86 |
| Man_4 | 21 | 5 | 3 | 94 |
| Man_4 | 24.5 | 3 | 2 | 99 |
| Man_5 | 29 | 5 | 2 | 98 |
| Man_5 | 25 | 3 | 1 | 98 |
| Man_6 | 27 | 4 | 2 | 98 |
| Man_6 | 29 | 5 | 1 | 93 |
| Man_8 | 27.5 | 5 | 0 | 92 |
| Man_8 | 19.5 | 5 | 1 | 105 |
| Tym_3 | 28.5 | 3 | 1 | 111 |
| Tym_3 | 33.5 | 7 | 1 | 105 |
| Tym_5 | 22.5 | 5 | 0 | 109 |

| | | | | |
|---|---|---|---|---|
| Tym_5 | 28.5 | 1 | 0 | 106 |
| Tym_26 | 25.5 | 3 | 0 | 106 |
| Tym_26 | 28.5 | 3 | 1 | 78 |
| Tym_30 | 28 | 1 | 0 | 108 |
| Tym_30 | 27.5 | 4 | 2 | 107 |
| Sut_4 | 27.5 | 3 | 0 | 103 |
| Sut_4 | 30.5 | 4 | 1 | 103 |
| Dor | 35 | 3 | 1 | 100 |
| Dor | 37.5 | 4 | 1 | 106 |
| Dor | 31 | 5 | 0 | 103 |
| Dor | 29.5 | 2 | 0 | 109 |
| Dor_C | 29 | 2 | 5 | 93 |
| Dor_C | 33 | 8 | 5 | 103 |
| IOW2_2 | 19 | 4 | 0 | 92 |
| IOW2_2 | 29 | 7 | 7 | 96 |
| IOW2_25 | 29 | 5 | 3 | 90 |
| IOW2_25 | 23.5 | 5 | 3 | 88 |
| IOW2_26 | 22.5 | 4 | 1 | 111 |
| IOW2_26 | 33 | 11 | 0 | 102 |
| IOW2_30 | 16.5 | 6 | 0 | 110 |
| IOW2_30 | 30 | 15 | 2 | 106 |
| IOW1_10 | 24.5 | 4 | 1 | 97 |
| IOW1_10 | 25 | 6 | 4 | 86 |
| IOW1_11 | 26.5 | 5 | 2 | 104 |
| IOW1_11 | 26.5 | 4 | 1 | 99 |
| IOW1_17 | 25.5 | 4 | 1 | 93 |
| IOW1_17 | 27 | 4 | 0 | 101 |
| Ara | 68 | 1 | 4 | 95 |
| Ara | 72.5 | 1 | 2 | 92 |
| Ari | 84 | 1 | 1 | 90 |
| Ari | 68 | 1 | 3 | 93 |
| Ari | 57.2 | 4 | 3 | 86 |
| Ari | 67 | 1 | 4 | 85 |
| Ble | 51 | 1 | 2 | 100 |
| Ble | 37.5 | 1 | 4 | 94 |
| Ble | 66 | 2 | 5 | 102 |
| Ble | 59 | 1 | 5 | 93 |
| Bol | 65.2 | 1 | 4 | 90 |
| Bol | 58 | 1 | 4 | 92 |
| Ede | 79 | 1 | 2 | 97 |
| Ede | 70.5 | 1 | 2 | 99 |
| Gis | 46 | 2 | 2 | 102 |
| Gis | 56 | 3 | 1 | 93 |
| Gis | 50 | 3 | 3 | 93 |

| | | | | |
|---|---|---|---|---|
| Gis | 66 | 3 | 4 | 106 |
| Lir | 18 | 1 | 0 | 77 |
| Lir | 52 | 3 | 5 | 86 |
| Mar | 41 | 1 | 3 | 83 |
| Mar | 43.5 | 1 | 6 | 93 |
| Mon | 52.5 | 1 | 4 | 96 |
| Mon | 81.5 | 2 | 6 | 90 |
| Mon | 71 | 1 | 6 | 82 |
| Mon | 56 | 1 | 3 | 97 |
| Olg | 74 | 1 | 2 | 100 |
| Olg | 74 | 1 | 1 | 92 |
| Ome | 25 | 1 | 2 | 82 |
| Ome | 28.5 | 1 | 1 | 83 |
| Pri | 47 | 3 | 6 | 87 |
| Pri | 41.5 | 3 | 4 | 94 |
| Pri | 42 | 2 | 5 | 97 |
| Pri | 23.5 | 9 | 2 | 123 |
| Rab | 62 | 1 | 2 | 94 |
| Rab | 47.5 | 1 | 4 | 90 |
| Rab | 42 | 3 | 6 | 97 |
| Rab | 49.3 | 3 | 11 | 93 |
| Suz | 56 | 1 | 4 | 97 |
| Suz | 56 | 1 | 3 | 94 |
| Tin | 45 | 2 | 1 | 97 |
| Tin | 46 | 1 | 2 | 92 |
| Tin | 53.5 | 1 | 1 | 104 |
| Tin | 63 | 3 | 3 | 93 |
| Vol | 36.5 | 2 | 3 | 95 |
| Vol | 29 | 1 | 2 | 96 |

## Appendix 32: Macro for ImageJ processing (Chapter 5)

```
#Opening file and setting known distance unit:
open("D:\\Linum Project\\Cap and seed photos\\Seed photos Aug 2018\\1_1 Seeds.tif");
makeLine(996, 210, 1014, 324);
run("Set Scale...", "distance=115.41 known=1 pixel=1 unit=mm");
run("Split Channels");
selectWindow("1_1 Seeds.tif (blue)");
run("Color Balance...");
run("Apply LUT");
```

```
run("Analyze Particles...", "size=25-Infinity pixel circularity=0.00-1.00 show=Outlines display");
setAutoThreshold("Default");
//run("Threshold...");
setThreshold(0, 254);
run("Convert to Mask");
setAutoThreshold("Default");
setThreshold(1, 255);
run("Convert to Mask");
run("Analyze Particles...", "size=25-Infinity pixel circularity=0.00-1.00 show=Outlines display");
saveAs("Results", "D:\\Linum Project\\Cap and seed photos\\Seed photos Aug 2018\\1_1 Seeds.xls");

#__________________________________________________________________________________
________________


#Aplying LUT and making binary:

run("Apply LUT");
run("Make Binary");
run("Analyze Particles...", "size=25-Infinity pixel circularity=0.00-1.00 show=Outlines display");
```

# BIBLIOGRAPHY

Adams, J., 2009. Vegetation-climate interaction: how plants make the global environment. Springer Science & Business Media.

Addicott, F.T. (1977) ‘Flower Behavior in Linum lewisii: Some Ecological and Physiological Factors in Opening and Abscission of Petals’, *American Midland Naturalist*, 97(2), p. 321. Available at: https://doi.org/10.2307/2425097.

Adhikari, K.N., Buirchell, B.J. and Sweetingham, M.W. (2012) ‘Length of vernalization period affects flowering time in three lupin species’, *Plant Breeding*, 131(5), pp. 631–636. Available at: https://doi.org/10.1111/J.1439-0523.2012.01996.X.

Adhikari, P.B., Liu, X. and Kasahara, R.D. (2020) ‘Mechanics of Pollen Tube Elongation: A Perspective’, *Frontiers in Plant Science*, 11, p. 1612. Available at: https://doi.org/10.3389/FPLS.2020.589712/BIBTEX.

Alcala, N. *et al.* (2013) ‘Peak and persistent excess of genetic diversity following an abrupt migration increase’, *Genetics*, 193(3), pp. 953–971. Available at: https://doi.org/10.1534/GENETICS.112.147785/-/DC1/GENETICS.112.147785-4.PDF.

Alexander, D.H., Novembre, J. and Lange, K. (2009) ‘Fast model-based estimation of ancestry in unrelated individuals’, *Genome Research*, 19(9), pp. 1655–1664. Available at: https://doi.org/10.1101/GR.094052.109.

Allaby, R.G. *et al.* (2005) ‘Evidence of the domestication history of flax (Linum usitatissimum L.) from genetic diversity of the sad2 locus’, *Theoretical and Applied Genetics*, 112(1), pp. 58–65. Available at: https://doi.org/10.1007/S00122-005-0103-3/TABLES/3.

Allard, A. *et al.* (2016) ‘Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population’, *Journal of Experimental Botany*, 67(9), pp. 2875–2888. Available at: https://doi.org/10.1093/JXB/ERW130.

Alonso, C. et al. (2013) ‘Among-species differences in pollen quality and quantity limitation: Implications for endemics in biodiverse hotspots’, *Annals of Botany*, 112(7), pp. 1461–1469. doi:10.1093/aob/mct213.

‘An Introduction to Generalized Linear Models’ (2018) *An Introduction to Generalized Linear Models, Fourth Edition* [Preprint]. Available at: https://doi.org/10.1201/9781315182780.

Analysis of relative gene expression using different real-time quantitative PCR - Abstract - Europe PMC (no date). Available at: https://europepmc.org/article/cba/641544 (Accessed: 10 June 2022).

Anderson, J.T., Willis, J.H. and Mitchell-Olds, T. (2011) ‘Evolutionary genetics of plant adaptation’, *Trends in Genetics*, 27(7), pp. 258–266. Available at: https://doi.org/10.1016/J.TIG.2011.04.001.

Andrews, K.R. *et al.* (no date) ‘Harnessing the power of RADseq for ecological and evolutionary genomics’. Available at: https://doi.org/10.1101/028837.

Arnold, B. *et al.* (2013) ‘RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling’, *Molecular Ecology*, 22(11), pp. 3179–3190. Available at: https://doi.org/10.1111/MEC.12276.

Asíns, M.J. (2002) ‘Present and future of quantitative trait locus analysis in plant breeding’, *Plant Breeding*, pp. 281–291. Available at: https://doi.org/10.1046/j.1439-0523.2002.730285.x.

Ataii, E. *et al.* (2021) ‘Genetic characterization of a diversity panel, selected from IPK linseed (Linum usitatissimum) world collection’, *Plant Breeding*, 140(5), pp. 919–932. Available at: https://doi.org/10.1111/pbr.12962.

ATES, E. (2012) ‘Some chemical and morphological properties of five clover species (Trifolium sp.) at different aspect of pasture in Belovets village (Razgrad), Bulgaria’, *International Journal of Plant Production*, 5(3), pp. 255–262. Available at: https://doi.org/10.22069/IJPP.2012.737.

Atkins, K.E. and Travis, J.M.J. (2010) ‘Local adaptation and the evolution of species’ ranges under climate change’, *Journal of Theoretical Biology*, 266(3), pp. 449–457. Available at: https://doi.org/10.1016/J.JTBI.2010.07.014.

Auge, G.A. *et al.* (2017) ‘Maternal vernalization and vernalization-pathway genes influence progeny seed germination’, *New Phytologist*, 216(2), pp. 388–400. Available at: https://doi.org/10.1111/nph.14520.

Bacilieri, R. *et al.* (2013) ‘Genetic structure in cultivated grapevines is linked to geography and human selection’, *BMC Plant Biology*, 13(1), pp. 1–14. Available at: https://doi.org/10.1186/1471-2229-13-25/TABLES/5.

Baiakhmetov, E. *et al.* (2021) ‘Evidence for extensive hybridisation and past introgression events in feather grasses using genome-wide SNP genotyping’, *BMC Plant Biology*, 21(1), pp. 1–20. Available at: https://doi.org/10.1186/S12870-021-03287-W/TABLES/3.

Ballerini, E.S. and Kramer, E.M. (2011) ‘In the light of evolution: A reevaluation of conservation in the CO-FT regulon and its role in photoperiodic regulation of flowering time’, *Frontiers in Plant Science*, 2(NOV), p. 81. Available at: https://doi.org/10.3389/FPLS.2011.00081/BIBTEX.

Barboni, D. et al. (2004) ‘Relationships between plant traits and climate in the Mediterranean region: A pollen data analysis’, *Journal of Vegetation Science*, 15(5), pp. 635–646. doi:10.1111/j.1654-1103.2004.tb02305.x.

Bay, R.A. *et al.* (2018) *Genomic signals of selection predict climate-driven population declines in a migratory bird*.

Behjati, S. and Tarpey, P.S. (2013) ‘What is next generation sequencing?’, *Archives of Disease in Childhood. Education and Practice Edition*, 98(6), p. 236. Available at: https://doi.org/10.1136/ARCHDISCHILD-2013-304340.

Beltrán, R. *et al.* (2019) 'Effect of temperature on pollen germination for several Rosaceae species: Influence of freezing conservation time on germination patterns', *PeerJ*, 2019(12). Available at: https://doi.org/10.7717/PEERJ.8195/SUPP-1.

Bigot, S. *et al.* (2018) 'Pivotal roles of environmental sensing and signaling mechanisms in plant responses to climate change', *Global Change Biology*, 24(12), pp. 5573–5589. Available at: https://doi.org/10.1111/GCB.14433.

Bilton, T.P. *et al.* (2018) 'Linkage Disequilibrium Estimation in Low Coverage High-Throughput Sequencing Data', *Genetics*, 209(2), p. 389. Available at: https://doi.org/10.1534/GENETICS.118.300831.

Bird, K.A. *et al.* (2017) 'Population structure and phylogenetic relationships in a diverse panel of Brassica rapa L.', *Frontiers in Plant Science*, 8. Available at: https://doi.org/10.3389/fpls.2017.00321.

Blackman, B.K., Michaels, S.D. and Rieseberg, L.H. (2011) 'Connecting the sun to flowering in sunflower adaptation', *Molecular ecology*, 20(17), pp. 3503–3512. Available at: https://doi.org/10.1111/J.1365-294X.2011.05166.X.

Blanquart, F. *et al.* (2013) 'A practical guide to measuring local adaptation', *Ecology Letters*, 16(9), pp. 1195–1205. Available at: https://doi.org/10.1111/ELE.12150.

Blázquez, M., Koornneef, M. and Putterill, J. (2001) 'Flowering on time: genes that regulate the floral transition: Workshop on the molecular basis of flowering time control', *EMBO Reports*, 2(12), p. 1078. Available at: https://doi.org/10.1093/EMBO-REPORTS/KVE254.

Blümel, M., Dally, N. and Jung, C. (2015) 'Flowering time regulation in crops — what did we learn from Arabidopsis?', *Current Opinion in Biotechnology*, 32, pp. 121–129. Available at: https://doi.org/10.1016/J.COPBIO.2014.11.023.

Boavida, L.C. and McCormick, S. (2007) 'TECHNICAL ADVANCE: Temperature as a determinant factor for increased and reproducible in vitro pollen germination in Arabidopsis thaliana', *The Plant Journal*, 52(3), pp. 570–582. Available at: https://doi.org/10.1111/J.1365-313X.2007.03248.X.

Bolsheva, N.L. *et al.* (2017) 'Evolution of blue-flowered species of genus Linum based on high-throughput sequencing of ribosomal RNA genes', *BMC Evolutionary Biology*, 17(2), pp. 23–36. Available at: https://doi.org/10.1186/S12862-017-1105-X/FIGURES/6.

Booker, T.R., Jackson, B.C. and Keightley, P.D. (2017) 'Detecting positive selection in the genome', *BMC Biology*, 15(1), pp. 1–10. Available at: https://doi.org/10.1186/S12915-017-0434-Y/FIGURES/2.

Boshier, D. et al. (2015) 'Is local best? examining the evidence for local adaptation in trees and its scale', Environmental Evidence, 4(1). doi:10.1186/s13750-015-0046-3.

Bougeard, S. and Dray, S. (2018) 'Supervised Multiblock Analysis in R with the ade4 Package', *Journal of Statistical Software*, 86, pp. 1–17. Available at: https://doi.org/10.18637/JSS.V086.I01.

BOX, G.E.P. (1953) 'NON-NORMALITY AND TESTS ON VARIANCES', *Biometrika*, 40(3–4), pp. 318–335. Available at: https://doi.org/10.1093/BIOMET/40.3-4.318.

Božič, A. and Šiber, A. (2022) 'Mechanics of inactive swelling and bursting of porate pollen grains', *Biophysical Journal*, 121(5), pp. 782–792. Available at: https://doi.org/10.1016/J.BPJ.2022.01.019.

Brandon, R.N., 2014. Adaptation and environment (Vol. 1040). Princeton University Press.

Brennan, A.C. *et al.* (2014) 'The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe', *BMC Plant Biology*, 14(1), pp. 1–14. Available at: https://doi.org/10.1186/1471-2229-14-17/FIGURES/4.

Breygina, M., Klimenko, E. and Schekaleva, O. (2021) 'Pollen Germination and Pollen Tube Growth in Gymnosperms', *Plants*, 10(7). Available at: https://doi.org/10.3390/PLANTS10071301.

Buira, A. et al. (2020) 'The contribution of the edaphic factor as a driver of recent plant diversification in a Mediterranean biodiversity hotspot', *Journal of Ecology*, 109(2), pp. 987–999. doi:10.1111/1365-2745.13527.

Bustin, S.A. *et al.* (2009) 'The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments', *Clinical chemistry*, 55(4), pp. 611–622. Available at: https://doi.org/10.1373/CLINCHEM.2008.112797.

Butte, A.J., Dzau, V.J. and Glueck, S.B. (2001) 'Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues"', *https://doi.org/10.1152/physiolgenomics.2001.7.2.95*, 7(2), pp. 95–96. Available at: https://doi.org/10.1152/PHYSIOLGENOMICS.2001.7.2.95.

Calvin, J.A. (1998) 'Regression Models for Categorical and Limited Dependent Variables', *Technometrics*, 40(1), pp. 80–81. Available at: https://doi.org/10.1080/00401706.1998.10485496.

Campetella, G. *et al.* (2020) 'Plant functional traits are correlated with species persistence in the herb layer of old-growth beech forests', *Scientific Reports*, 10(1). Available at: https://doi.org/10.1038/S41598-020-76289-7.

Carmona, R. *et al.* (2017) 'Automated identification of reference genes based on RNA-seq data', *BioMedical Engineering OnLine*, 16(Suppl 1). Available at: https://doi.org/10.1186/S12938-017-0356-5.

Catchen, J. *et al.* (2013) 'Stacks: an analysis tool set for population genomics', *Molecular ecology*, 22(11), pp. 3124–3140. Available at: https://doi.org/10.1111/MEC.12354.

Cave, R.L. and Johnston, M.E. (2010) "Vernalization promotes flowering of a heat tolerant Calandrinia while long days replace vernalization for early flowering of brunonia," Scientia Horticulturae, 123(3), pp. 379–384. Available at: https://doi.org/10.1016/j.scienta.2009.10.004.

Cave, R.L. *et al.* (2013) 'Modelling temperature, photoperiod and vernalization responses of Brunonia australis (Goodeniaceae) and Calandrinia sp. (Portulacaceae) to predict flowering time', *Annals of botany*, 111(4), pp. 629–639. Available at: https://doi.org/10.1093/AOB/MCT028.

Chambers, J. and Hastie, T., 1992. Statistical models in S. London: Chapman & Hall

Chandler, J., Wilson, A. and Dean, C. (1996) ‘Arabidopsis mutants showing an altered response to vernalization’, *The Plant Journal*, 10(4), pp. 637–644. Available at: https://doi.org/10.1046/J.1365-313X.1996.10040637.X.

Chapman, J.R. *et al.* (2009) ‘A quantitative review of heterozygosity–fitness correlations in animal populations’, *Molecular Ecology*, 18(13), pp. 2746–2765. Available at: https://doi.org/10.1111/J.1365-294X.2009.04247.X.

Chen, C.-C. and Gibson, P.B. (1973) ‘Effect of Temperature on Pollen-Tube Growth in Trifolium repense after Cross- and Self-Pollinations1’, *Crop Science*, 13(5), pp. 563–566. Available at: https://doi.org/10.2135/CROPSCI1973.0011183X001300050021X.

Chen, D. *et al.* (2018) ‘Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana’, *Nature Communications 2018 9:1*, 9(1), pp. 1–13. Available at: https://doi.org/10.1038/s41467-018-06772-3.

Chiluwal, A. *et al.* (2020) ‘Deterioration of ovary plays a key role in heat stress-induced spikelet sterility in sorghum’, *Plant Cell and Environment*, 43(2), pp. 448–462. Available at: https://doi.org/10.1111/pce.13673.

Cho, L.H., Yoon, J. and An, G. (2017) ‘The control of flowering time by environmental factors’, *The Plant Journal*, 90(4), pp. 708–719. Available at: https://doi.org/10.1111/TPJ.13461.

Chouard, P. (1960) ‘Vernalization and its Relations to Dormancy’, *Annual Review of Plant Physiology*, 11(1), pp. 191–238. Available at: https://doi.org/10.1146/ANNUREV.PP.11.060160.001203.

Clarke, J.D. (2009) ‘Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation’, *Cold Spring Harbor Protocols*, 2009(3), p. pdb.prot5177. Available at: https://doi.org/10.1101/PDB.PROT5177.

Cloutier, S. *et al.* (2009) ‘Development and analysis of EST-SSRs for flax (Linum usitatissimum L.)’, *Theoretical and Applied Genetics*, 119(1), pp. 53–63. Available at: https://doi.org/10.1007/S00122-009-1016-3/FIGURES/4.

Cloutier, S. *et al.* (2012) ‘Integrated consensus genetic and physical maps of flax (Linum usitatissimum L.)’, *Theoretical and Applied Genetics 2012 125:8*, 125(8), pp. 1783–1795. Available at: https://doi.org/10.1007/S00122-012-1953-0.

Clua, A.A. and Gimenez, D.O. (2003) ‘Environmental factors during seed development of narrow-leaved bird’s-foot-trefoil (Lotus tenuis) influences subsequent dormancy and germination’, *Grass and Forage Science*, 58(4), pp. 333–338. Available at: https://doi.org/10.1046/j.1365-2494.2003.00385.x.

Conner, J.K. (2002) ‘Genetic mechanisms of floral trait correlations in a natural population’, *Nature*, 420(6914), pp. 407–410. Available at: https://doi.org/10.1038/NATURE01105.

Corbesier, L. and Coupland, G. (2006) ‘The quest for florigen: a review of recent progress’, *Journal of Experimental Botany*, 57(13), pp. 3395–3403. Available at: https://doi.org/10.1093/JXB/ERL095.

Cox, M. and Doudna, J., 2015. Molecular Biology:2nd Revised edition. New York: W.H. FREEMAN & CO LTD, pp.221-238, 369-376, 592-593.

Çopur, O., Gur, M.A., Karakus, M. and Demirel, U., 2006. Determination of correlation and path analysis among yield components and seed yield in oil flax varieties (Linum usitatissimum L.). *Journal of Biological Sciences*, 6(4), pp.738-743.

Craufurd, P.Q. and Wheeler, T.R. (2009) 'Climate change and the flowering time of annual crops', *Journal of experimental botany*, 60(9), pp. 2529–2539. Available at: https://doi.org/10.1093/JXB/ERP196.

Crean, A.J. and Immler, S. (2021) 'Evolutionary consequences of environmental effects on gamete performance', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1826). Available at: https://doi.org/10.1098/RSTB.2020.0122.

Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*, 27(15), pp. 2156–2158. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR330.

Darapuneni, M.K. *et al.* (2014) 'Effect of vernalization and photoperiod on flax flowering time', *Euphytica*, 195(2), pp. 279–285. Available at: https://doi.org/10.1007/s10681-013-0996-x.

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9(5-6), 416–423. https://doi.org/10.1093/bfgp/elq031

Davey, J.W. *et al.* (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics 2011 12:7*, 12(7), pp. 499–510. Available at: https://doi.org/10.1038/nrg3012.

Debieu, M. *et al.* (2013) 'Co-Variation between Seed Dormancy, Growth Rate and Flowering Time Changes with Latitude in Arabidopsis thaliana', *PLoS ONE*, 8(5). Available at: https://doi.org/10.1371/journal.pone.0061075.

Delph, L. and Havens, K., 1998. Sperm Competition and Sexual Selection. Elsevier Inc.  Pp.149-173.

Dempewolf, H. *et al.* (2017) 'Past and Future Use of Wild Relatives in Crop Breeding', *Crop Science*, 57(3), pp. 1070–1082. Available at: https://doi.org/10.2135/CROPSCI2016.10.0885.

Descamps, C., Quinet, M. and Jacquemart, A.L. (2021) 'Climate Change–Induced Stress Reduce Quantity and Alter Composition of Nectar and Pollen From a Bee-Pollinated Species (Borago officinalis, Boraginaceae)', *Frontiers in Plant Science*, 12, p. 2264. Available at: https://doi.org/10.3389/FPLS.2021.755843/BIBTEX.

Devasirvatham, V. *et al.* (2012) 'Effect of high temperature on the reproductive development of chickpea genotypes under controlled environments', *Functional Plant Biology*, 39(12), pp. 1009–1018. Available at: https://doi.org/10.1071/FP12033.

Diamond, J. (2002) 'Evolution, consequences and future of plant and animal domestication', *Nature*, 418(6898), pp. 700–707. Available at: https://doi.org/10.1038/NATURE01019.

Dickinson, H., Rodriguez-Enriquez, J. and Grant-Downton, R. (2018) ‘Pollen Germination and Pollen Tube Growth of Arabidopsis thaliana: in vitro and Semi in vivo Methods’, *Bio-protocol*, 8(16). Available at: https://doi.org/10.21769/BIOPROTOC.2977.

Diederichsen, A. (2001) ‘Comparison of genetic diversity of flax (Linum usitatissimum L.) between Canadian cultivars and a world collection’, *Plant Breeding*, 120(4), pp. 360–362. Available at: https://doi.org/10.1046/J.1439-0523.2001.00616.X.

Diederichsen, A. and Raney, J.P. (2006) ‘Seed colour, seed weight and seed oil content in Linum usitatissimum accessions held by Plant Gene Resources of Canada’, *Plant Breeding*, 125(4), pp. 372–377. Available at: https://doi.org/10.1111/J.1439-0523.2006.01231.X.

Diederichsen, A. and Ulrich, A. (2009) ‘Variability in stem fibre content and its association with other characteristics in 1177 flax (Linum usitatissimum L.) genebank accessions’, *Industrial Crops and Products*, 30(1), pp. 33–39. Available at: https://doi.org/10.1016/j.indcrop.2009.01.002.

Dmitriev, A.A. *et al.* (2021) ‘Genome Sequencing of Fiber Flax Cultivar Atlant Using Oxford Nanopore and Illumina Platforms’, *Frontiers in Genetics*, 11, p. 1487. Available at: https://doi.org/10.3389/FGENE.2020.590282/BIBTEX.

Dole, J.M. and Wilkins, H.F. (1994) “Interaction of bulb vernalization and shoot photoperiod on `Nellie White' Easter Lily,” HortScience, 29(3), pp. 143–145. Available at: https://doi.org/10.21273/hortsci.29.3.143.

Donohue, K. *et al.* (2005) ‘Environmental and genetic influences on the germination of Arabidopsis thaliana in the field’, *Evolution*, 59(4), pp. 740–757. Available at: https://doi.org/10.1111/J.0014-3820.2005.TB01750.X.

Donohue, K. *et al.* (2010) ‘Germination, postgermination adaptation, and species ecological ranges’, *Annual Review of Ecology, Evolution, and Systematics*, 41, pp. 293–319. Available at: https://doi.org/10.1146/annurev-ecolsys-102209-144715.

Dray, S. and Dufour, A.B. (2007) ‘The ade4 Package: Implementing the Duality Diagram for Ecologists’, *Journal of Statistical Software*, 22(4), pp. 1–20. Available at: https://doi.org/10.18637/JSS.V022.I04.

Dudley, S.A. (1996) ‘THE RESPONSE TO DIFFERING SELECTION ON PLANT PHYSIOLOGICAL TRAITS: EVIDENCE FOR LOCAL ADAPTATION’, *Evolution*, 50(1), pp. 103–110. Available at: https://doi.org/10.1111/J.1558-5646.1996.TB04476.X.

Ejsmond, M.J. *et al.* (2011) ‘Does climate affect pollen morphology? Optimal size and shape of pollen grains under various desiccation intensity’, *Ecosphere*, 2(10), pp. 1–15. Available at: https://doi.org/10.1890/ES11-00147.1.

El-Beltagi, H. and Salama, Z. (2007) *Evaluation of fatty acids profile and the content of some secondary metabolites in seeds of different flax cultivars (Linum usitatissimum L.) Cytotoxic and antioxidant properties of active principals isolated from water hyacinth against four cancer cells lines View project Ascorbic Acid*

*Enhances Growth and Yield of Sweet Peppers (Capsicum annum) by Mitigating Salinity Stress View project*. Available at: https://www.researchgate.net/publication/255606154.

Elfving, N. *et al.* (2011) 'The Arabidopsis thaliana Med25 mediator subunit integrates environmental cues to control plant development', *Proceedings of the National Academy of Sciences of the United States of America*, 108(20), pp. 8245–8250. Available at: https://doi.org/10.1073/PNAS.1002981108.

Ellstrand, N.C., Prentice, H.C. and Hancock, J.F. (2003) 'Gene Flow and Introgression from Domesticated Plants into their Wild Relatives', *https://doi.org/10.1146/annurev.ecolsys.30.1.539*, 30, pp. 539–563. Available at: https://doi.org/10.1146/ANNUREV.ECOLSYS.30.1.539.

Essemine, J., Ammar, S. and Bouzid, S. (2010) 'Impact of heat stress on germination and growth in higher plants: physiological, biochemical and molecular repercussions and mechanisms of defence.', *Journal of Biological Sciences*, 10(6), pp. 565–572.

Fábián, A. *et al.* (2019) 'Stigma functionality and fertility are reduced by heat and drought co-stress in wheat', *Frontiers in Plant Science*, 10. Available at: https://doi.org/10.3389/FPLS.2019.00244/FULL.

Falconer, D.S., 1996. Introduction to quantitative genetics. Pearson Education India.

FAOSTAT data (2006) Food and Agriculture Organiza-tion of the United Nations, FAO Statistical Data-bases, URL: http://faostat.fao.org/

Fedorenko, O.M., Savushkin, A.I. and Olimpienko, G.S. (2001) 'Genetic Diversity in Natural Populations of Arabidopsis thaliana (L.) Heynh. from Karelia', *Russian Journal of Genetics 2001 37:2*, 37(2), pp. 162–167. Available at: https://doi.org/10.1023/A:1009041922186.

Firon, N., Nepi, M. and Pacini, E. (2012) 'Water status and associated processes mark critical stages in pollen development and functioning', *Annals of Botany*, 109(7), pp. 1201–1214. Available at: https://doi.org/10.1093/AOB/MCS070.

Flowers, J.M. *et al.* (2009) 'Population Genomics of the Arabidopsis thaliana Flowering Time Gene Network', *Molecular Biology and Evolution*, 26(11), pp. 2475–2486. Available at: https://doi.org/10.1093/MOLBEV/MSP161.

Fragallah, S.A.D.A. *et al.* (2019) 'Effects of Sucrose, Boric Acid, pH, and Incubation Time on in Vitro Germination of Pollen and Tube Growth of Chinese fir (Cunnighamial lanceolata L.)', *Forests 2019, Vol. 10, Page 102*, 10(2), p. 102. Available at: https://doi.org/10.3390/F10020102.

Frichot, E. *et al.* (2014) 'Fast and efficient estimation of individual ancestry coefficients', *Genetics*, 196(4), pp. 973–983. Available at: https://doi.org/10.1534/GENETICS.113.160572/-/DC1.

Frichot, E. and François, O. (2015) 'LEA: An R package for landscape and ecological association studies', *Methods in Ecology and Evolution*, 6(8), pp. 925–929. Available at: https://doi.org/10.1111/2041-210X.12382.

Fu, Y.-B., Diederichsen, A. and Allaby, R.G. (2011) "Locus-specific view of flax domestication history," Ecology and Evolution, 2(1), pp. 139–152. Available at: https://doi.org/10.1002/ece3.57.

Fu, Y.B. (2012) ‘Population-based resequencing revealed an ancestral winter group of cultivated flax: implication for flax domestication processes’, *Ecology and Evolution*, 2(3), pp. 622–635. Available at: https://doi.org/10.1002/ECE3.101.

Fu, Y.X. (1995) ‘Statistical Properties of Segregating Sites’, *Theoretical Population Biology*, 48(2), pp. 172–197. Available at: https://doi.org/10.1006/TPBI.1995.1025.

Furness, C.A. and Rudall, P.J. (2004) ‘Pollen aperture evolution – a crucial factor for eudicot success?’, *Trends in Plant Science*, 9(3), pp. 154–158. Available at: https://doi.org/10.1016/J.TPLANTS.2004.01.001.

Galloway, L.F. and Burgess, K.S. (2009) ‘Manipulation of flowering time: phenological integration and maternal effects’, *Ecology*, 90(8), pp. 2139–2148. Available at: https://doi.org/10.1890/08-0948.1.

Galloway, L.F. and Burgess, K.S. (2012) ‘Artificial selection on flowering time: influence on reproductive phenology across natural light environments’, *Journal of Ecology*, 100(4), pp. 852–861. Available at: https://doi.org/10.1111/J.1365-2745.2012.01967.X.

Gao, Y.-B. et al. (2014) “Low temperature inhibits pollen tube growth by disruption of both tip-localized reactive oxygen species and endocytosis in Pyrus Bretschneideri Rehd.,” Plant Physiology and Biochemistry, 74, pp. 255–262. Available at: https://doi.org/10.1016/j.plaphy.2013.11.018.

Gargiulo, R., Kull, T. and Fay, M.F. (2021) ‘Effective double-digest RAD sequencing and genotyping despite large genome size’, *Molecular Ecology Resources*, 21(4), pp. 1037–1055. Available at: https://doi.org/10.1111/1755-0998.13314.

Gelman, A. and Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge university press.

Gerristen, H., 2018. CRAN - Package mapplots. [online] Cran.r-project.org. Available at: <https://cran.r-project.org/web/packages/mapplots/index.html> [Accessed 22 June 2022].

Gilad, Y. *et al.* (2002) ‘Evidence for positive selection and population structure at the human MAO-A gene’, *Proceedings of the National Academy of Sciences*, 99(2), pp. 862–867. Available at: https://doi.org/10.1073/PNAS.022614799.

Giles, B.E. (1990) ‘The effects of variation in seed size on growth and reproduction in the wild barley Hordeum vulgare ssp. spontaneum’, *Heredity 1990 64:2*, 64(2), pp. 239–250. Available at: https://doi.org/10.1038/hdy.1990.29.

Gill KS (1966) Evolutionary relationships among Linum species. Ph.D. Thesis, University of California, Riverside, USA

Giménez-Benavides, L., Escudero, A. and Iriondo, J.M. (2007) ‘Reproductive limits of a late-flowering high-mountain Mediterranean plant along an elevational climate gradient’, *New Phytologist*, 173(2), pp. 367–382. Available at: https://doi.org/10.1111/J.1469-8137.2006.01932.X.

Ginzinger, D.G. (2002) ‘Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream’, *Experimental Hematology*, 30(6), pp. 503–512. Available at: https://doi.org/10.1016/S0301-472X(02)00806-8.

Giorno, F. *et al.* (2013) ‘Ensuring Reproduction at High Temperatures: The Heat Stress Response during Anther and Pollen Development’, *Plants 2013, Vol. 2, Pages 489-506*, 2(3), pp. 489–506. Available at: https://doi.org/10.3390/PLANTS2030489.

Girden, E.R., 1992. ANOVA: Repeated measures (No. 84). sage.

Glantz, S., Slinker, B. and Neilands, T., n.d. Primer of applied regression et analysis of variance.

Godwin, I.D., Aitken, E.A.B. and Smith, L.W. (1997) ‘Application of inter simple sequence repeat (ISSR) markers to plant genetics’, *ELECTROPHORESIS*, 18(9), pp. 1524–1528. Available at: https://doi.org/10.1002/ELPS.1150180906.

Goudenhooft, C., Bourmaud, A. and Baley, C. (2019) ‘Flax (Linum usitatissimum L.) fibers for composite reinforcement: Exploring the link between plant growth, cell walls development, and fiber properties’, *Frontiers in Plant Science*. Frontiers Media S.A. Available at: https://doi.org/10.3389/fpls.2019.00411.

Grapin, A. *et al.* (1998) ‘Diploid Musa acuminata genetic diversity assayed with sequence-tagged microsatellite sites’, *ELECTROPHORESIS*, 19(8–9), pp. 1374–1380. Available at: https://doi.org/10.1002/ELPS.1150190829.

Grassein, F., Lavorel, S. and Till-Bottraud, I. (2014) ‘The importance of biotic interactions and local adaptation for plant response to environmental changes: field evidence along an elevational gradient’, *Global Change Biology*, 20(5), pp. 1452–1460. Available at: https://doi.org/10.1111/GCB.12445.

Greer, S. *et al.* (2010) ‘Housekeeping genes; expression levels may change with density of cultured cells’, *Journal of Immunological Methods*, 355(1–2), pp. 76–79. Available at: https://doi.org/10.1016/J.JIM.2010.02.006.

Gregis, V. *et al.* (2009) ‘The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes’, *The Plant Journal*, 60(4), pp. 626–637. Available at: https://doi.org/10.1111/J.1365-313X.2009.03985.X.

Guan, S. *et al.* (2021) ‘Genetic analysis and population structure of wild and cultivated wishbone flower (Torenia fournieri Lind.) lines related to specific floral color’, *PeerJ*, 7, p. e11702. Available at: https://doi.org/10.7717/PEERJ.11702/SUPP-6.

Günther, T. *et al.* (2016) ‘Genomic and phenotypic differentiation of Arabidopsis thaliana along altitudinal gradients in the North Italian Alps’, *Molecular Ecology*, 25(15), pp. 3574–3592. Available at: https://doi.org/10.1111/MEC.13705.

Guo, D. *et al.* (2020) ‘Resequencing 200 Flax Cultivated Accessions Identifies Candidate Genes Related to Seed Size and Weight and Reveals Signatures of Artificial Selection’, *Frontiers in Plant Science*, 10. Available at: https://doi.org/10.3389/fpls.2019.01682.

Guo, Q. (2012) ‘Incorporating latitudinal and central–marginal trends in assessing genetic variation across species ranges’, *Molecular Ecology*, 21(22), pp. 5396–5403. Available at: https://doi.org/10.1111/MEC.12012.

Gupta, S. *et al.* (2021) ‘Identification and characterization of a novel long juvenile resource AGS 25’, *Genetic Resources and Crop Evolution*, 68(3), pp. 1149–1163. Available at: https://doi.org/10.1007/S10722-020-01055-7/TABLES/2.

Gutaker, R.M. *et al.* (2019) ‘Flax latitudinal adaptation at LuTFL1 altered architecture and promoted fiber production’, *Scientific Reports 2019 9:1*, 9(1), pp. 1–12. Available at: https://doi.org/10.1038/s41598-018-37086-5.

Gutiérrez-Valencia, J. *et al.* (2022) ‘Genomic analyses of the Linum distyly supergene reveal convergent evolution at the molecular level’, *Current Biology*, 32(20), pp. 4360-4371.e6. Available at: https://doi.org/10.1016/J.CUB.2022.08.042.

El Haddad, N. *et al.* (2021) ‘Crop wild relatives in durum wheat breeding: Drift or thrift?’, *Crop Science*, 61(1), pp. 37–54. Available at: https://doi.org/10.1002/CSC2.20223.

Hall, M.C. and Willis, J.H. (2006) ‘DIVERGENT SELECTION ON FLOWERING TIME CONTRIBUTES TO LOCAL ADAPTATION IN MIMULUS GUTTATUS POPULATIONS’, *Evolution*, 60(12), pp. 2466–2477. Available at: https://doi.org/10.1111/J.0014-3820.2006.TB01882.X.

Hämälä, T. and Savolainen, O. (2019) ‘Genomic Patterns of Local Adaptation under Gene Flow in Arabidopsis lyrata’, *Molecular Biology and Evolution*, 36(11), pp. 2557–2571. Available at: https://doi.org/10.1093/MOLBEV/MSZ149.

Hampe, A. and Petit, R.J. (2005) ‘Conserving biodiversity under climate change: the rear edge matters’, *Ecology Letters*, 8(5), pp. 461–467. Available at: https://doi.org/10.1111/J.1461-0248.2005.00739.X.

Han, Q. *et al.* (2021) ‘Association between RsFT, RsFLC and RsCOL5 (A&B) expression and flowering regulation in Japanese wild radish’, *AoB PLANTS*, 13(4). Available at: https://doi.org/10.1093/aobpla/plab039.

Hanley, M.E. *et al.* (2008) ‘Breeding system, pollinator choice and variation in pollen quality in British herbaceous plants’, *Functional Ecology*, 22(4), pp. 592–598. Available at: https://doi.org/10.1111/J.1365-2435.2008.01415.X.

Harvey, M.G. et al. (2016) “Sequence capture versus restriction site associated DNA sequencing for Shallow Systematics,” Systematic Biology, 65(5), pp. 910–924. Available at: https://doi.org/10.1093/sysbio/syw036.

Hayat, M.J. and Higgins, M. (2014) ‘Understanding poisson regression’, *Journal of Nursing Education*, 53(4), pp. 207–215. Available at: https://doi.org/10.3928/01484834-20140325-04.

He, Y., Chen, T. and Zeng, X. (2020) ‘Genetic and Epigenetic Understanding of the Seasonal Timing of Flowering’, *Plant Communications*, 1(1). Available at: https://doi.org/10.1016/J.XPLC.2019.100008/GENETIC_AND_EPIGENETIC_UNDERSTANDING_OF_THE_SEASONAL_TIMING_OF_FLOWERING.PDF.

Hedhly, A., Hormaza, J.I. and Herrero, M. (2004) ‘Effect of temperature on pollen tube kinetics and dynamics in sweet cherry, Prunus avium (Rosaceae)’, *American Journal of Botany*, 91(4), pp. 558–564. Available at: https://doi.org/10.3732/ajb.91.4.558.

Hedhly, A., Hormaza, J.I. and Herrero, M. (2005) ‘The effect of temperature on pollen germination, pollen tube growth, and stigmatic receptivity in peach’, *Plant Biology*, 7(5), pp. 476–483. Available at: https://doi.org/10.1055/S-2005-865850.

Heisler, M.G. and Jönsson, H. (2007) ‘Modelling meristem development in plants’, *Current Opinion in Plant Biology*, 10(1), pp. 92–97. Available at: https://doi.org/10.1016/J.PBI.2006.11.005.

Herburger, K. and Holzinger, A. (2016) ‘Aniline blue and Calcofluor white staining of callose and cellulose in the streptophyte green algae Zygnema and Klebsormidium’, *Bio-protocol*, 6(20). Available at: https://doi.org/10.21769/BIOPROTOC.1969.

Hernández, F. *et al.* (2019) ‘Genetic Diversity and Population Structure of Wild Sunflower (Helianthus annuus L.) in Argentina: Reconstructing Its Invasion History’, *Journal of Heredity*, 110(6), pp. 746–759. Available at: https://doi.org/10.1093/JHERED/ESZ047.

Hernandez, R.D. *et al.* (2019) ‘Ultrarare variants drive substantial cis heritability of human gene expression’, *Nature Genetics 2019 51:9*, 51(9), pp. 1349–1355. Available at: https://doi.org/10.1038/s41588-019-0487-7.

Higgins, J.A., Bailey, P.C. and Laurie, D.A. (2010) ‘Comparative genomics of flowering time pathways using brachypodium distachyon as a model for the temperate Grasses’, *PLoS ONE*, 5(4). Available at: https://doi.org/10.1371/journal.pone.0010065.

Higuchi, R. *et al.* (1992) ‘Simultaneous amplification and detection of specific DNA sequences’, *Bio/technology (Nature Publishing Company)*, 10(4), pp. 413–417. Available at: https://doi.org/10.1038/NBT0492-413.

Hjelmqvist, H. (1950) ‘The flax weeds and the origin of cultivated flax.’, *Botaniska Notiser*, 2, pp. 257–98.

Hoban, S. *et al.* (2016) ‘Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions’, *American Naturalist*, 188(4), pp. 379–397. Available at: https://doi.org/10.1086/688018/ASSET/IMAGES/LARGE/FG4.JPEG.

Hopkins, R., Schmitt, J. and Stinchcombe, J.R. (2008) ‘A latitudinal cline and response to vernalization in leaf angle and morphology in Arabidopsis thaliana (Brassicaceae)’, *New Phytologist*, 179(1), pp. 155–164. Available at: https://doi.org/10.1111/J.1469-8137.2008.02447.X.

Hothorn, T., Bretz, F. and Westfall, P. (2008a) ‘Simultaneous Inference in General Parametric Models’, *Biometrical Journal*, 50(3), pp. 346–363. Available at: https://doi.org/10.1002/BIMJ.200810425.

Hothorn, T., Bretz, F. and Westfall, P. (2008b) ‘Simultaneous Inference in General Parametric Models’, *Biometrical Journal*, 50(3), pp. 346–363. Available at: https://doi.org/10.1002/BIMJ.200810425.

Hu, T. *et al.* (2021) ‘Next-generation sequencing technologies: An overview’, *Human Immunology*, 82(11), pp. 801–811. Available at: https://doi.org/10.1016/J.HUMIMM.2021.02.012.

Huggett, J. *et al.* (2005) ‘Real-time RT-PCR normalisation; strategies and considerations’, *Genes & Immunity 2005 6:4*, 6(4), pp. 279–284. Available at: https://doi.org/10.1038/sj.gene.6364190.

Huis, R., Hawkins, S. and Neutelings, G. (2010) ‘Selection of reference genes for quantitative gene expression normalization in flax (Linum usitatissimum L.)’, *BMC Plant Biology*, 10(1), pp. 1–14. Available at: https://doi.org/10.1186/1471-2229-10-71/FIGURES/6.

Ibrahim, A. *et al.* (2018) ‘A regulator of early flowering in barley (Hordeum vulgare L.)’, *PLOS ONE*, 13(7), p. e0200722. Available at: https://doi.org/10.1371/JOURNAL.PONE.0200722.

IDT. 2022. Tips for resuspending and diluting your oligonucleotides. [online] Available at: <https://eu.idtdna.com/pages/education/decoded/article/tips-for-resuspending-and-diluting-your-oligonucleotides> [Accessed 2 June 2022].

Iguchi, A. *et al.* (2020) ‘Deep-sea amphipods around cobalt-rich ferromanganese crusts: Taxonomic diversity and selection of candidate species for connectivity analysis’, *PLOS ONE*, 15(2), p. e0228483. Available at: https://doi.org/10.1371/JOURNAL.PONE.0228483.

Innan, H. *et al.* (1999) ‘A Method for Estimating Nucleotide Diversity From AFLP Data’, *Genetics*, 151(3), pp. 1157–1164. Available at: https://doi.org/10.1093/GENETICS/151.3.1157.

Innes, P. *et al.* (2022) ‘Assessment of biogeographic variation in traits of Lewis flax (Linum lewisii) for use in restoration and agriculture’, *AoB PLANTS*, 14(2). Available at: https://doi.org/10.1093/aobpla/plac005.

Ionkova, I. *et al.* (2013) ‘Linum narbonense: A new valuable tool for biotechnological production of a potent anticancer lignan Justicidine B’, *Pharmacognosy Magazine*, 9(33), p. 39. Available at: https://doi.org/10.4103/0973-1296.108138.

Iovane, M. *et al.* (2022) ‘High temperature and humidity affect pollen viability and longevity in olea europaea l.’, *Agronomy*, 12(1). Available at: https://doi.org/10.3390/agronomy12010001.

Iwaizumi, M.G. and Takahashi, M. (2012) ‘Effects of pollen supply and quality on seed formation and maturation in Pinus densiflora’, *Journal of plant research*, 125(4), pp. 517–525. Available at: https://doi.org/10.1007/S10265-011-0463-5.

Jaakola, L. and Hohtola, A. (2010) ‘Effect of latitude on flavonoid biosynthesis in plants’, *Plant, Cell & Environment*, 33(8), pp. 1239–1247. Available at: https://doi.org/10.1111/J.1365-3040.2010.02154.X.

Jahnke, M.R. and Etterson, J.R. (2019) ‘Autonomous self-fertilization in Linum sulcatum, a native annual with a previously unknown mating system’, *Journal of plant research*, 132(1), pp. 57–67. Available at: https://doi.org/10.1007/S10265-018-1076-Z.

Jarman, S.N., Ward, R.D. and Elliott, N.G. (2002) ‘Oligonucleotide primers for PCR amplification of coelomate introns’, *Marine Biotechnology*, 4(4), pp. 347–355. Available at: https://doi.org/10.1007/s10126-002-0029-6.

Jensen, H. *et al.* (2007) ‘Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation’, *Molecular ecology*, 16(19), pp. 4066–4078. Available at: https://doi.org/10.1111/J.1365-294X.2007.03452.X.

Jhala, A. and Hall, L.M. (2014) *Flax (Linum usitatissimum L.): Current uses and future applications*. Available at: https://www.researchgate.net/publication/228952435.

Jhala, A.J. and Hall, L.M. (2010) ‘Flax (Linum usitatissimum L.): Current Uses and Future Applications’, *Australian Journal of basic and Applied Sciences*, 4(9), pp. 4304–4312.

Jiang, F., Cadotte, M.W. and Jin, G. (2021) ‘Individual-level leaf trait variation and correlation across biological and spatial scales’, *Ecology and Evolution*, 11(10), p. 5344. Available at: https://doi.org/10.1002/ECE3.7425.

Johnson, M.A., Harper, J.F. and Palanivelu, R. (1AD) ‘A Fruitful Journey: Pollen Tube Navigation from Germination to Fertilization’. Available at: https://doi.org/10.1146/annurev-arplant-050718.

Johnstone, A., 2001. Biology. Oxford: Oxford University Press, pp.94-95.

Jombart, T. and Ahmed, I. (2011) ‘adegenet 1.3-1: new tools for the analysis of genome-wide SNP data’, *Bioinformatics*, 27(21), pp. 3070–3071. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR521.

Jombart, T. and Bateman, A. (2008) ‘adegenet: a R package for the multivariate analysis of genetic markers’, *Bioinformatics*, 24(11), pp. 1403–1405. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTN129.

Jones, B. and Gliddon, C. (1999) *Reproductive biology and genetic structure in Lloydia serotina*, *Plant Ecology*.

de Jong, M.A. *et al.* (2011) ‘Mitochondrial DNA Signature for Range-Wide Populations of Bicyclus anynana Suggests a Rapid Expansion from Recent Refugia’, *PLOS ONE*, 6(6), p. e21385. Available at: https://doi.org/10.1371/JOURNAL.PONE.0021385.

de Jong, M.J. *et al.* (2021) ‘SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets’, *Molecular Ecology Resources*, 21(4), pp. 1369–1379. Available at: https://doi.org/10.1111/1755-0998.13339.

Jordon-Thaden, I.E. *et al.* (2020) ‘A basic ddRADseq two-enzyme protocol performs well with herbarium and silica-dried tissues across four genera’, *Applications in Plant Sciences*, 8(4), p. e11344. Available at: https://doi.org/10.1002/APS3.11344.

Jung, J.H. *et al.* (2007) ‘The GIGANTEA-Regulated MicroRNA172 Mediates Photoperiodic Flowering Independent of CONSTANS in Arabidopsis’, *The Plant Cell*, 19(9), pp. 2736–2748. Available at: https://doi.org/10.1105/TPC.107.054528.

Kakani, V.G. *et al.* (2002) ‘Response of in vitro pollen germination and pollen tube growth of groundnut (Arachis hypogaea L.) genotypes to temperature’, *Plant, Cell & Environment*, 25(12), pp. 1651–1661. Available at: https://doi.org/10.1046/J.1365-3040.2002.00943.X.

Kakani, V.G. *et al.* (2005) ‘Differences in in vitro Pollen Germination and Pollen Tube Growth of Cotton Cultivars in Response to High Temperature’, *Annals of Botany*, 96(1), pp. 59–67. Available at: https://doi.org/10.1093/AOB/MCI149.

Kamvar, Z.N., Tabima, J.F. and Gřunwald, N.J. (2014a) ‘Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction’, *PeerJ*, 2014(1), pp. 1–14. Available at: https://doi.org/10.7717/PEERJ.281/TABLE-6.

Kamvar, Z.N., Tabima, J.F. and Gřunwald, N.J. (2014b) ‘Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction’, *PeerJ*, 2014(1), pp. 1–14. Available at: https://doi.org/10.7717/PEERJ.281/TABLE-6.

Kassambara, A., 2022. 'ggplot2' Based Publication Ready Plots [R package ggpubr version 0.4.0]. [online] Cran.r-project.org. Available at: <https://cran.r-project.org/web/packages/ggpubr/index.html> [Accessed 9 October 2022].

Kayes, J.M. and Clark, S.E. (1998) ‘CLAVATA2, a regulator of meristem and organ development in Arabidopsis’, *Development*, 125(19), pp. 3843–3851. Available at: https://doi.org/10.1242/DEV.125.19.3843.

Keddy, P.A. and Constabel, P. (1986) ‘Germination of Ten Shoreline Plants in Relation to Seed Size, Soil Particle Size and Water Level: An Experimental Study’, *The Journal of Ecology*, 74(1), p. 133. Available at: https://doi.org/10.2307/2260354.

Khan, M.R.G., Ai, X.Y. and Zhang, J.Z. (2014) ‘Genetic regulation of flowering time in annual and perennial plants’, *Wiley Interdisciplinary Reviews: RNA*. Blackwell Publishing Ltd, pp. 347–359. Available at: https://doi.org/10.1002/wrna.1215.

Khanday, I. and Sundaresan, V. (2021) ‘Plant zygote development: recent insights and applications to clonal seeds’, *Current opinion in plant biology*, 59. Available at: https://doi.org/10.1016/J.PBI.2020.101993.

Kim, D.H. *et al.* (2009) ‘Vernalization: Winter and the Timing of Flowering in Plants’, *http://dx.doi.org/10.1146/annurev.cellbio.042308.113411*, 25, pp. 277–299. Available at: https://doi.org/10.1146/ANNUREV.CELLBIO.042308.113411.

Kinmonth-Schultz, H. *et al.* (2021a) ‘Flowering Times of Wild Arabidopsis Accessions From Across Norway Correlate With Expression Levels of FT, CO, and FLC Genes’, *Frontiers in Plant Science*, 12, p. 2376. Available at: https://doi.org/10.3389/FPLS.2021.747740/BIBTEX.

Kinmonth-Schultz, H. *et al.* (2021b) ‘Flowering Times of Wild Arabidopsis Accessions From Across Norway Correlate With Expression Levels of FT, CO, and FLC Genes’, *Frontiers in Plant Science*, 12, p. 2376. Available at: https://doi.org/10.3389/FPLS.2021.747740/BIBTEX.

Koonin, E. V. (2005) ‘Orthologs, paralogs, and evolutionary genomics’, *Annual review of genetics*, 39, pp. 309–338. Available at: https://doi.org/10.1146/ANNUREV.GENET.39.073003.114725.

Kurokura, T. *et al.* (2017) 'Fragaria vesca CONSTANS controls photoperiodic flowering and vegetative development', *Journal of Experimental Botany*, 68(17), pp. 4839–4850. Available at: https://doi.org/10.1093/JXB/ERX301.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B. (2017) 'lmerTest Package: Tests in Linear Mixed Effects Models', *Journal of Statistical Software*, 82(13), pp. 1–26. Available at: https://doi.org/10.18637/JSS.V082.I13.

Kyrkjeeide, M.O. et al. (2016) 'Long-distance dispersal and barriers shape genetic structure of peatmosses (sphagnum) across the Northern Hemisphere', *Journal of Biogeography*, 43(6), pp. 1215–1226. doi:10.1111/jbi.12716.

Landoni, B. *et al.* (2022) 'Local climate and vernalization requirements explain the latitudinal patterns of flowering initiation in the crop wild relative Linum bienne', *bioRxiv*, p. 2022.01.02.474722. Available at: https://doi.org/10.1101/2022.01.02.474722.

Langin, K.M. *et al.* (2017) 'Partial support for the central-marginal hypothesis within a population: reduced genetic diversity but not increased differentiation at the range edge of an island endemic bird', *Heredity*, 119(1), pp. 8–15. Available at: https://doi.org/10.1038/HDY.2017.10.

Larios, E. *et al.* (2014) 'Natural selection on seed size through the life cycle of a desert annual plant', *Ecology*, 95(11), pp. 3213–3220. Available at: https://doi.org/10.1890/13-1965.1.

Lascoux, M., Glémin, S. and Savolainen, O. (2016) 'Local Adaptation in Plants', *eLS*, pp. 1–7. Available at: https://doi.org/10.1002/9780470015902.A0025270.

Lazaro, A., Obeng-Hinneh, E. and Albani, M.C. (2018) 'Extended Vernalization Regulates Inflorescence Fate in Arabis alpina by Stably Silencing PERPETUAL FLOWERING1', *Plant Physiology*, 176(4), pp. 2819–2833. Available at: https://doi.org/10.1104/PP.17.01754.

Lee, D.R. *et al.* (2014) 'Developing local adaptation strategies for climate change in agriculture: A priority-setting approach with application to Latin America', *Global Environmental Change*, 29, pp. 78–91. Available at: https://doi.org/10.1016/J.GLOENVCHA.2014.08.002.

Lee, J. and Lee, I. (2010) 'Regulation and function of SOC1, a flowering pathway integrator', *Journal of Experimental Botany*, 61(9), pp. 2247–2254. Available at: https://doi.org/10.1093/JXB/ERQ098.

Lee, J.H. *et al.* (2013) 'Regulation of temperature-responsive flowering by MADS-box transcription factor repressors', *Science*, 342(6158), pp. 628–632. Available at: https://doi.org/10.1126/SCIENCE.1241097/SUPPL_FILE/PAPV2.PDF.

Lee, J.H., Kim, J.J. and Ahn, J.H. (2012) 'Role of SEPALLATA3 (SEP3) as a downstream gene of miR156-SPL3-FT circuitry in ambient temperature-responsive flowering', *https://doi.org/10.4161/psb.21366*, 7(9), pp. 1151–1154. Available at: https://doi.org/10.4161/PSB.21366.

Leijten, W. *et al.* (2018) ‘Translating Flowering Time from Arabidopsis thaliana to Brassicaceae and Asteraceae Crop Species’, *Plants 2018, Vol. 7, Page 111*, 7(4), p. 111. Available at: https://doi.org/10.3390/PLANTS7040111.

Leimu, R. and Fischer, M. (2008a) ‘A meta-analysis of local adaptation in plants’, *PLoS ONE*, 3(12). Available at: https://doi.org/10.1371/journal.pone.0004010.

Leimu, R. and Fischer, M. (2008b) ‘A Meta-Analysis of Local Adaptation in Plants’, *PLOS ONE*, 3(12), p. e4010. Available at: https://doi.org/10.1371/JOURNAL.PONE.0004010.

Lemuth, K. and Rupp, S. (2015) ‘Microarrays as Research Tools and Diagnostic Devices’, *RNA Technologies*, pp. 259–280. Available at: https://doi.org/10.1007/978-3-319-17305-4_13/FIGURES/2.

Li, H. (2011) ‘A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data’, *Bioinformatics*, 27(21), p. 2987. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR509.

Li, H. and Durbin, R. (2009) ‘Fast and accurate short read alignment with Burrows-Wheeler transform’, *Bioinformatics (Oxford, England)*, 25(14), pp. 1754–1760. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTP324.

Li, L. *et al.* (2015a) ‘Flowering responses to light and temperature’, *Science China Life Sciences 2016 59:4*, 59(4), pp. 403–408. Available at: https://doi.org/10.1007/S11427-015-4910-8.

Li, L. *et al.* (2015b) ‘Flowering responses to light and temperature’, *Science China Life Sciences 2016 59:4*, 59(4), pp. 403–408. Available at: https://doi.org/10.1007/S11427-015-4910-8.

Li, Y.X. *et al.* (2016) ‘Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population’, *The Plant Journal*, 86(5), pp. 391–402. Available at: https://doi.org/10.1111/TPJ.13174.

Lian, H., Roy, E. and Zheng, H. (2016) ‘Microglial Phagocytosis Assay’, *BIO-PROTOCOL*, 6(21). Available at: https://doi.org/10.21769/BIOPROTOC.1988.

Liang, W. *et al.* (2018a) ‘Selection and evaluation of reference genes for qRT-PCR analysis in Euscaphis konishii Hayata based on transcriptome data’, *Plant Methods*, 14(1), pp. 1–9. Available at: https://doi.org/10.1186/S13007-018-0311-X/FIGURES/4.

Liang, W. *et al.* (2018b) ‘Selection and evaluation of reference genes for qRT-PCR analysis in Euscaphis konishii Hayata based on transcriptome data’, *Plant Methods*, 14(1), pp. 1–9. Available at: https://doi.org/10.1186/S13007-018-0311-X/FIGURES/4.

Linck, E. and Battey, C.J. (2019) ‘Minor allele frequency thresholds strongly affect population structure inference with genomic data sets’, *Molecular Ecology Resources*, 19(3), pp. 639–647. Available at: https://doi.org/10.1111/1755-0998.12995.

Lischer, H.E.L. and Excoffier, L. (2012) ‘PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs’, *Bioinformatics (Oxford, England)*, 28(2), pp. 298–299. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR642.

Liu, J. *et al.* (2015) ‘Genetic and epigenetic control of plant heat responses’, *Frontiers in Plant Science*, 6(APR), p. 267. Available at: https://doi.org/10.3389/FPLS.2015.00267/BIBTEX.

Liu, Y. *et al.* (2019) ‘Genome-Wide Single Nucleotide Polymorphisms are Robust in Resolving Fine-Scale Population Genetic Structure of the Small Brown Planthopper, Laodelphax striatellus (Fallén) (Hemiptera: Delphacidae)’, *Journal of Economic Entomology*, 112(5), pp. 2362–2368. Available at: https://doi.org/10.1093/JEE/TOZ145.

Livak, K.J. and Schmittgen, T.D. (2001a) ‘Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method’, *Methods*, 25(4), pp. 402–408. Available at: https://doi.org/10.1006/METH.2001.1262.

Livak, K.J. and Schmittgen, T.D. (2001b) ‘Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method’, *Methods*, 25(4), pp. 402–408. Available at: https://doi.org/10.1006/METH.2001.1262.

Loha, A., Tigabu, M. and Fries, A. (2008) ‘Genetic variation among and within populations of Cordia africana in seed size and germination responses to constant temperatures’, *Euphytica 2008 165:1*, 165(1), pp. 189–196. Available at: https://doi.org/10.1007/S10681-008-9806-2.

Lotz, L.A.P. (1990) ‘The Relation Between Age and Size at First Flowering of Plantago Major in Various Habitats’, *The Journal of Ecology*, 78(3), p. 757. Available at: https://doi.org/10.2307/2260897.

Loveless, M.D. and Hamrick, J.L. (1984a) *Ecological Determinants of Genetic Structure in Plant Populations*, *Source: Annual Review of Ecology and Systematics*. Available at: https://about.jstor.org/terms.

Loveless, M.D. and Hamrick, J.L. (1984b) *Ecological Determinants of Genetic Structure in Plant Populations*, *Source: Annual Review of Ecology and Systematics*. Available at: https://about.jstor.org/terms.

Lovell, J.T. *et al.* (2013) ‘Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation’, *Proceedings of the Royal Society B: Biological Sciences*, 280(1763). Available at: https://doi.org/10.1098/rspb.2013.1043.

Lu, P. *et al.* (2014) ‘The Arabidopsis CALLOSE DEFECTIVE MICROSPORE1 Gene Is Required for Male Fertility through Regulating Callose Metabolism during Microsporogenesis’, *Plant Physiology*, 164(4), pp. 1893–1904. Available at: https://doi.org/10.1104/PP.113.233387.

Luo, Z. *et al.* (2019) ‘Genetic diversity and population structure of a Camelina sativa spring panel’, *Frontiers in Plant Science*, 10, p. 184. Available at: https://doi.org/10.3389/FPLS.2019.00184/BIBTEX.

Lütfi, P. (2002) *The Effects of Temperature on Pollen Germination and Pollen Tube Growth of Apricot and Sweet Cherry*. Available at: https://www.pubhort.org/ejhs/2002/file_1877.pdf (Accessed: 6 September 2022).

Lyakh, V.A. and Soroka, A.I. (2021) ‘Artificial medium for in vitro pollen germination of some ornamental Linum species’, *Advances in Horticultural Science*, 35(3), pp. 329–332. Available at: https://doi.org/10.36253/AHSC-10854.

Maisonneuve, B. and Den Nijs, A.P.M. (1984) ‘In vitro pollen germination and tube growth of tomato (Lycopersicon esculentumMill.) and its relation with plant growth’, *Euphytica*, 33(3), pp. 833–840. Available at: https://doi.org/10.1007/BF00021910/METRICS.

Manoli, A. *et al.* (2012) ‘Evaluation of candidate reference genes for qPCR in maize’, *Journal of Plant Physiology*, 169(8), pp. 807–815. Available at: https://doi.org/10.1016/J.JPLPH.2012.01.019.

Marth, G.T. *et al.* (2004) ‘The Allele Frequency Spectrum in Genome-Wide Human Variation Data Reveals Signals of Differential Demographic History in Three Large World Populations’, *Genetics*, 166(1), pp. 351–372. Available at: https://doi.org/10.1534/GENETICS.166.1.351.

Maruki, T. and Lynch, M. (2014) ‘Genome-wide estimation of linkage disequilibrium from population-level high-throughput sequencing data’, *Genetics*, 197(4), pp. 1303–1313. Available at: https://doi.org/10.1534/GENETICS.114.165514.

Mastretta-Yanes, A. *et al.* (2014) ‘Gene Duplication, Population Genomics, and Species-Level Differentiation within a Tropical Mountain Shrub’, *Genome Biology and Evolution*, 6(10), p. 2611. Available at: https://doi.org/10.1093/GBE/EVU205.

McDill, J. *et al.* (2009) ‘The phylogeny of Linum and Linaceae subfamily Linoideae, with implications for their systematics, biogeography, and evolution of heterostyly’, *Systematic Botany*, 34(2), pp. 386–405. Available at: https://doi.org/10.1600/036364409788606244.

McKinney, G.J. *et al.* (2017) ‘Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations’, *Molecular Ecology Resources*, 17(4), pp. 656–669. Available at: https://doi.org/10.1111/1755-0998.12613.

McMaster, G.S. *et al.* (2008) ‘Simulating the influence of vernalization, photoperiod and optimum temperature on wheat developmental rates’, *Annals of botany*, 102(4), pp. 561–569. Available at: https://doi.org/10.1093/AOB/MCN115.

McNeil, K.A., Newman, I. and Kelly, F.J., 1996. Testing research hypotheses with the general linear model. SIU Press.

Meek, M.H. *et al.* (2023) ‘Understanding Local Adaptation to Prepare Populations for Climate Change’, *BioScience*, 73(1), pp. 36–47. Available at: https://doi.org/10.1093/biosci/biac101.

De Mendonça Vilela, M. *et al.* (2017) ‘Analysis of Three Sugarcane Homo/Homeologous Regions Suggests Independent Polyploidization Events of Saccharum officinarum and Saccharum spontaneum’, *Genome Biology and Evolution*, 9(2), pp. 266–278. Available at: https://doi.org/10.1093/GBE/EVW293.

Miller, J.M., Cullingham, C.I. and Peery, R.M. (2020) ‘The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method’, *Heredity 2020 125:5*, 125(5), pp. 269–280. Available at: https://doi.org/10.1038/s41437-020-0348-2.

Miranda, J.A. and Steward, G.F. (2017) ‘Variables influencing the efficiency and interpretation of reverse transcription quantitative PCR (RT-qPCR): An empirical study using Bacteriophage MS2’, *Journal of Virological Methods*, 241, pp. 1–10. Available at: https://doi.org/10.1016/J.JVIROMET.2016.12.002.

Mittell, E.A. *et al.* (2020) ‘Feral populations of Brassica oleracea along Atlantic coasts in western Europe’, *Ecology and Evolution*, 10(20), p. 11810. Available at: https://doi.org/10.1002/ECE3.6821.

Mobaxterm., 2022. MobaXterm free Xserver and tabbed SSH client for Windows. [online] Available at: <https://mobaxterm.mobatek.net/> [Accessed 21 September 2022].

MOLES, ANGELA T. et al. (2004) “Small-seeded species produce more seeds per square metre of canopy per year, but not per individual per lifetime,” Journal of Ecology, 92(3), pp. 384–396. Available at: https://doi.org/10.1111/j.0022-0477.2004.00880.x.

Montalt, R. *et al.* (2019) ‘Influence of temperature on the progamic phase in Citrus’, *Environmental and Experimental Botany*, 166, p. 103806. Available at: https://doi.org/10.1016/J.ENVEXPBOT.2019.103806.

Montesinos-Navarro, A., Picó, F. and Tonsor, S. (2012) ‘Clinal variation in seed traits influencing life cycle timing in Arabidopsis thaliana’, *Evolution*, 66.

Moran, E. V. (2020) ‘Simulating the effects of local adaptation and life history on the ability of plants to track climate shifts’, *AoB PLANTS*, 12(1). Available at: https://doi.org/10.1093/aobpla/plaa008.

Mori, A.S. *et al.* (2022) ‘Feedbacks Between Biodiversity and Climate Change’, *The Ecological and Societal Consequences of Biodiversity Loss*, pp. 281–304. Available at: https://doi.org/10.1002/9781119902911.CH13.

Muravenko, O. V. *et al.* (2010a) ‘Karyogenomics of species of the genus Linum L.’, *Russian Journal of Genetics 2010 46:10*, 46(10), pp. 1182–1185. Available at: https://doi.org/10.1134/S1022795410100091.

Muravenko, O. V. *et al.* (2010b) ‘Karyogenomics of species of the genus Linum L.’, *Russian Journal of Genetics 2010 46:10*, 46(10), pp. 1182–1185. Available at: https://doi.org/10.1134/S1022795410100091.

Murray, B.R. *et al.* (2004) ‘Geographical gradients in seed mass in relation to climate’, *Journal of Biogeography*, 31(3), pp. 379–388. Available at: https://doi.org/10.1046/J.0305-0270.2003.00993.X.

Muthamia, K. (1994) ‘Effect of bulb size and duration of vernalization on onion (allium cepa l.) seed yield’. Available at: http://erepository.uonbi.ac.ke/handle/11295/23061 (Accessed: 9 October 2022).

Mylne, J. *et al.* (2004) ‘Epigenetic Regulation in the Control of Flowering’, *Cold Spring Harbor Symposia on Quantitative Biology*, 69, pp. 457–464. Available at: https://doi.org/10.1101/SQB.2004.69.457.

Nag, S., Mitra, J. and Karmakar, P.G. (2015) ‘ An Overview on Flax ( Linum usitatissimum L.) and its Genetic Diversity ’, *International Journal of Agriculture, Environment and Biotechnology*, 8(4), p. 805. Available at: https://doi.org/10.5958/2230-732x.2015.00089.3.

Nazareno, A.G. et al. (2021) ‘By animal, water, or wind: Can dispersal mode predict genetic connectivity in riverine plant species?’, *Frontiers in Plant Science*, 12. doi:10.3389/fpls.2021.626405.

Nei, M. (1972a) *Genetic Distance between Populations*, *Source: The American Naturalist*.

Nei, M. (1972b) *Genetic Distance between Populations*, *Source: The American Naturalist*. Available at: https://about.jstor.org/terms.

Newbigin, E., Anderson, M.A. and Clarke’, A.E. (1993) *Gametophytic Self-Incompatibility Systems*, *The Plant Cell*.

Nievola, C.C. *et al.* (2017) ‘Rapid responses of plants to temperature changes’, *Temperature: Multidisciplinary Biomedical Journal*, 4(4), p. 371. Available at: https://doi.org/10.1080/23328940.2017.1377812.

Normand, S. et al. (2011) ‘Postglacial migration supplements climate in determining plant species ranges in Europe’, *Proceedings of the Royal Society B: Biological Sciences*, 278(1725), pp. 3644–3653. doi:10.1098/rspb.2010.2769.

Nôžková, J., Remeselníková, K. and Bjelková, M. (2014) ‘Characterization and evaluation of flax seeds (Linum usitatissimum L.) on selected genotypes’, *Journal of Central European Agriculture*, 15(1), pp. 193–207. Available at: https://doi.org/10.5513/JCEA01/15.1.1434.

Odong, T.L. *et al.* (2011) ‘Determination of genetic structure of germplasm collections: Are traditional hierarchical clustering methods appropriate for molecular marker data?’, *Theoretical and Applied Genetics*, 123(2), pp. 195–205. Available at: https://doi.org/10.1007/S00122-011-1576-X/FIGURES/6.

Öhlund, E., Zurek, K. and Hammer, M. (2015) ‘Towards Sustainable Agriculture? The EU framework and local adaptation in Sweden and Poland’, *Environmental Policy and Governance*, 25(4), pp. 270–287. Available at: https://doi.org/10.1002/EET.1687.

Olejniczak, P. *et al.* (2018) ‘Seed size in mountain herbaceous plants changes with elevation in a species-specific manner’, *PLoS ONE*, 13(6). Available at: https://doi.org/10.1371/journal.pone.0199224.

Osnato, M. *et al.* (2022) ‘Photoperiod Control of Plant Growth: Flowering Time Genes Beyond Flowering’, *Frontiers in Plant Science*, 12, p. 3359. Available at: https://doi.org/10.3389/FPLS.2021.805635/BIBTEX.

Ozhimkova, E., Uschapovsky, I. and Manaenkov, O. (2023) “Study of varietal differences in the composition of heteropolysaccharides of oil flax and fiber flax,” Polysaccharides, 4(1), pp. 78–87. Available at: https://doi.org/10.3390/polysaccharides4010006.

Pacini, E. (2008) ‘Pollination’, *Encyclopedia of Ecology*, pp. 562–565. Available at: https://doi.org/10.1016/B978-0-444-63768-0.00859-3.

Pacini, E. and Dolferus, R. (2016) ‘The Trials and Tribulations of the Plant Male Gametophyte — Understanding Reproductive Stage Stress Tolerance’, *Abiotic and Biotic Stress in Plants - Recent Advances and Future Perspectives* [Preprint]. Available at: https://doi.org/10.5772/61671.

Paina, C. *et al.* (2014) ‘Vernalization Mediated Changes in the Lolium perenne Transcriptome’, *PLOS ONE*, 9(9), p. e107365. Available at: https://doi.org/10.1371/JOURNAL.PONE.0107365.

Pali, V. *et al.* (2014) ‘Identification of microsatellite markers for fingerprinting popular Indian flax (Linum usitatissimum L.) cultivars and their utilization in seed genetic purity assessments’, *Australian Journal of Crop Sciences*, 8(1).

Pan, C. et al. (2018) ‘Tomato stigma exsertion induced by high temperature is associated with the jasmonate signalling pathway’, *Plant, Cell & Environment*, 42(4), pp. 1205–1221. Available at: doi:10.1111/pce.13444.

Paradis, E. and Schliep, K. (2019a) ‘ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R’, *Bioinformatics*, 35(3), pp. 526–528. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTY633.

Paradis, E. and Schliep, K. (2019b) ‘ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R’, *Bioinformatics*, 35(3), pp. 526–528. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTY633.

Park, M.J. *et al.* (2016) ‘LATE ELONGATED HYPOCOTYL regulates photoperiodic flowering via the circadian clock in Arabidopsis’, *BMC Plant Biology*, 16(1), pp. 1–12. Available at: https://doi.org/10.1186/S12870-016-0810-8/FIGURES/5.

Peet, M.M. and Bartholemew, M. (1996) *Effect of Night Temperature on Pollen Characteristics, Growth, and Fruit Set in Tomato*, *J. AMER. SOC. HORT. SCI*.

Pembleton, L.W., Cogan, N.O.I. and Forster, J.W. (2013a) ‘StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations’, *Molecular Ecology Resources*, 13(5), pp. 946–952. Available at: https://doi.org/10.1111/1755-0998.12129.

Pembleton, L.W., Cogan, N.O.I. and Forster, J.W. (2013b) ‘StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations’, *Molecular Ecology Resources*, 13(5), pp. 946–952. Available at: https://doi.org/10.1111/1755-0998.12129.

Peterson, B.K. *et al.* (2012a) ‘Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species’, *PLOS ONE*, 7(5), p. e37135. Available at: https://doi.org/10.1371/JOURNAL.PONE.0037135.

Peterson, B.K. *et al.* (2012b) ‘Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species’, *PLOS ONE*, 7(5), p. e37135. Available at: https://doi.org/10.1371/JOURNAL.PONE.0037135.

Peterson, B.K. *et al.* (2012c) ‘Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species’, *PLOS ONE*, 7(5), p. e37135. Available at: https://doi.org/10.1371/JOURNAL.PONE.0037135.

Pfaffl, M.W. (2001) ‘A new mathematical model for  relative quantification in real-time RT–PCR’, *Nucleic Acids Research*, 29(9), p. e45. Available at: https://doi.org/10.1093/NAR/29.9.E45.

Porebski, S., Bailey, L.G. and Baum, B.R. (1997) ‘Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components’, *Plant Molecular Biology Reporter 1997 15:1*, 15(1), pp. 8–15. Available at: https://doi.org/10.1007/BF02772108.

Posé, D. *et al.* (2013) ‘Temperature-dependent regulation of flowering by antagonistic FLM variants’, *Nature 2013 503:7476*, 503(7476), pp. 414–417. Available at: https://doi.org/10.1038/nature12633.

Promega.co.uk. 2022. GoTaq® qPCR and RT-qPCR Systems | qPCR Master Mix | RT-qPCR Master Mix. [online] Available at: <https://www.promega.co.uk/products/pcr/qpcr-and-rt-qpcr/qpcr-kits/?catNum=A6020#resources> [Accessed 30 May 2022].

Przeworski, M. (2002) ‘The Signature of Positive Selection at Randomly Chosen Loci’, *Genetics*, 160(3), pp. 1179–1189. Available at: https://doi.org/10.1093/GENETICS/160.3.1179.

Purcell, S. *et al.* (2007a) ‘PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses’, *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at: https://doi.org/10.1086/519795.

Purcell, S. *et al.* (2007b) ‘PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses’, *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at: https://doi.org/10.1086/519795.

Purcell, S. *et al.* (2007c) ‘PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses’, *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at: https://doi.org/10.1086/519795.

Purugganan, M.D. and Fuller, D.Q. (2009) ‘The nature of selection during plant domestication’, *Nature*, 457(7231), pp. 843–848. Available at: https://doi.org/10.1038/NATURE07895.

Putterill, J., Laurie, R. and Macknight, R. (2004) ‘It’s time to flower: the genetic control of flowering time’, *BioEssays*, 26(4), pp. 363–373. Available at: https://doi.org/10.1002/BIES.20021.

Qaderi, M.M., Martel, A.B. and Dixon, S.L. (2019) ‘Environmental Factors Influence Plant Vascular System and Water Regulation’, *Plants*, 8(3), p. 65. Available at: https://doi.org/10.3390/PLANTS8030065.

Qin, Y. *et al.* (2009) ‘Penetration of the Stigma and Style Elicits a Novel Transcriptome in Pollen Tubes, Pointing to Genes Critical for Growth in a Pistil’, *PLOS Genetics*, 5(8), p. e1000621. Available at: https://doi.org/10.1371/JOURNAL.PGEN.1000621.

*QuantiFluor dsDNA System for High Concentration DNA* (no date). Available at: https://www.promega.co.uk/resources/pubhub/quantifluor-dsdna-system-for-high-concentration-dna/ (Accessed: 3 May 2022).

Rahimi, M.M., Zarei, M.A. and Arminian, A. (2011) ‘Selection criteria of flax (Linum usitatissimum L.) for seed yield, yield components and biochemical compositions under various planting dates and nitrogen’, *African Journal of Agricultural Research*, 6(13), pp. 3167–3175. Available at: https://doi.org/10.5897/AJAR11.382.

Rao, X. *et al.* (2013a) ‘An improvement of the 2ˆ(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis’, *Biostatistics, bioinformatics and biomathematics*, 3(3), p. 71. Available at: /pmc/articles/PMC4280562/ (Accessed: 13 June 2022).

Rao, X. *et al.* (2013b) ‘An improvement of the 2ˆ(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis’, *Biostatistics, bioinformatics and biomathematics*, 3(3), p. 71. Available at: /pmc/articles/PMC4280562/ (Accessed: 15 August 2022).

Rastogi, A. *et al.* (2013) *Effect of auxin and gibberellic acid on growth and yield components of linseed (Linum usitatissimum L.)*, *Crop Breeding and Applied Biotechnology*.

Rauf, S. *et al.* (2010) *Consequences of Plant Breeding on genetic diversity Bt cotton in Pakistan: Performance and adoption with imperfect information View project Drought tolerant Sunflower breeding material View project International Journal of Plant Breeding Consequences of Plant Breeding on Genetic Diversity*. Available at: https://www.researchgate.net/publication/235910237.

Raven, P., 1982. Biology of plants. New York: Worth Publishers, pp.132-137.

Ray, C. (1944) ‘CYTOLOGICAL STUDIES ON THE FLAX GENUS, LINUM’, *American Journal of Botany*, 31(4), pp. 241–248. Available at: https://doi.org/10.1002/J.1537-2197.1944.TB08027.X.

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/.

Ream, T.S. *et al.* (2014) ‘Interaction of Photoperiod and Vernalization Determines Flowering Time of Brachypodium distachyon’, *Plant Physiology*, 164(2), p. 694. Available at: https://doi.org/10.1104/PP.113.232678.

Reeves, P.H. and Coupland, G. (2000) ‘Response of plant development to environment: control of flowering by daylength and temperature’, *Current Opinion in Plant Biology*, 3(1), pp. 37–42. Available at: https://doi.org/10.1016/S1369-5266(99)00041-2.

Rehman, H.U. *et al.* (2014) ‘Seed Priming Influence on Early Crop Growth, Phenological Development and Yield Performance of Linola (Linum usitatissimum L.)’, *Journal of Integrative Agriculture*, 13(5), pp. 990–996. Available at: https://doi.org/10.1016/S2095-3119(13)60521-3.

Remans, T. *et al.* (2008) 'Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis thaliana exposed to increased metal concentrations', *Planta*, 227(6), pp. 1343–1349. Available at: https://doi.org/10.1007/S00425-008-0706-4/FIGURES/1.

Reynolds, W.W. and Casterlin, M.E. (1980) 'The Role of Temperature in the Environmental Physiology of Fishes', *Environmental Physiology of Fishes*, pp. 497–518. Available at: https://doi.org/10.1007/978-1-4899-3659-2_19.

Riedel, G. *et al.* (2014a) 'An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells', *PLOS ONE*, 9(3), p. e93031. Available at: https://doi.org/10.1371/JOURNAL.PONE.0093031.

Riedel, G. *et al.* (2014b) 'An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells', *PLOS ONE*, 9(3), p. e93031. Available at: https://doi.org/10.1371/JOURNAL.PONE.0093031.

Riedel, G. *et al.* (2014c) 'An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells', *PLOS ONE*, 9(3), p. e93031. Available at: https://doi.org/10.1371/JOURNAL.PONE.0093031.

Del Río, J.C. *et al.* (2011) 'Structural characterization of guaiacyl-rich lignins in flax (Linum usitatissimum) fibers and shives', *Journal of Agricultural and Food Chemistry*, 59(20), pp. 11088–11099. Available at: https://doi.org/10.1021/JF201222R/ASSET/IMAGES/LARGE/JF-2011-01222R_0001.JPEG.

Roach, D.A. (1987) *Variation in Seed and Seedling Size in Anthoxanthum odoratum*, *Source: The American Midland Naturalist*.

Rodriguez-Enriquez, M.J. *et al.* (2013) 'A novel method for efficient in vitro germination and tube growth of Arabidopsis thaliana pollen', *New Phytologist*, 197(2), pp. 668–679. Available at: https://doi.org/10.1111/NPH.12037.

Royo, C. *et al.* (2020) 'Agronomic performance of durum wheat landraces and modern cultivars and its association with genotypic variation in vernalization response (Vrn-1) and photoperiod sensitivity (Ppd-1) genes', *European Journal of Agronomy*, 120, p. 126129. Available at: https://doi.org/10.1016/J.EJA.2020.126129.

Rozhmina, T. *et al.* (2021) 'A comprehensive dataset of flax (Linum uitatissimum L.) phenotypes', *Data in Brief*, 37. Available at: https://doi.org/10.1016/j.dib.2021.107224.

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

Rúa, M.A. *et al.* (2016) 'Home-field advantage? evidence of local adaptation among plants, soil, and arbuscular mycorrhizal fungi through meta-analysis', *BMC Evolutionary Biology*, 16(1). Available at: https://doi.org/10.1186/s12862-016-0698-9.

Ruane, L.G. and Donohue, K. (2007) ‘Environmental effects on pollen–pistil compatibility between Phlox cuspidata and P. drummondii (Polemoniaceae): implications for hybridization dynamics’, *American Journal of Botany*, 94(2), pp. 219–227. Available at: https://doi.org/10.3732/AJB.94.2.219.

Rychel, S. *et al.* (2019) ‘FLOWERING LOCUS T, GIGANTEA, SEPALLATA, and FRIGIDA homologs are candidate genes involved in white lupin (Lupinus albus L.) early flowering’, *Molecular Breeding*, 39(3), pp. 1–17. Available at: https://doi.org/10.1007/S11032-019-0952-0/FIGURES/2.

Sa, R. *et al.* (2021) ‘Chromosome-Level Genome Assembly and Annotation of the Fiber Flax (Linum usitatissimum) Genome’, *Frontiers in Genetics*, 12, p. 735690. Available at: https://doi.org/10.3389/FGENE.2021.735690/FULL.

Saini, H.S., Sedgley, M. and Aspinall, D. (1983) ‘Effect of Heat Stress During Floral Development on Pollen Tube Growth and Ovary Anatomy in Wheat (Triticum aestivum L.)’, *Functional Plant Biology*, 10(2), pp. 137–144. Available at: https://doi.org/10.1071/PP9830137.

Sallam, A.H. et al. (2021) “Cold conditioned: Discovery of novel alleles for low-temperature tolerance in the Vavilov Barley Collection,” Frontiers in Plant Science, 12. Available at: https://doi.org/10.3389/fpls.2021.800284.

Sand Korneliussen, T. *et al.* (2013) *Calculation of Tajima’s D and other neutrality test statistics from low depth next-generation sequencing data*. Available at: http://www.biomedcentral.com/1471-2105/14/289.

Sanderson, B.J. et al. (2020) “A targeted sequence capture array for phylogenetics and population genomics in the salicaceae,” Applications in Plant Sciences, 8(10). Available at: https://doi.org/10.1002/aps3.11394.

Santamaría, L. *et al.* (2003) ‘PLANT PERFORMANCE ACROSS LATITUDE: THE ROLE OF PLASTICITY AND LOCAL ADAPTATION IN AN AQUATIC PLANT’, *Ecology*, 84(9), pp. 2454–2461. Available at: https://doi.org/10.1890/02-0431.

Sasaki, E. *et al.* (2017) ‘The genetic architecture of the network underlying flowering time variation in Arabidopsis thaliana’, *bioRxiv*, p. 175430. Available at: https://doi.org/10.1101/175430.

Schena, M. *et al.* (1995) ‘Quantitative monitoring of gene expression patterns with a complementary DNA microarray’, *Science (New York, N.Y.)*, 270(5235), pp. 467–470. Available at: https://doi.org/10.1126/SCIENCE.270.5235.467.

Schliep, K.P. (2011) ‘phangorn: phylogenetic analysis in R’, *Bioinformatics*, 27(4), pp. 592–593. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTQ706.

Schmittgen, T.D. and Livak, K.J. (2008) ‘Analyzing real-time PCR data by the comparative CT method’, *Nature Protocols 2008 3:6*, 3(6), pp. 1101–1108. Available at: https://doi.org/10.1038/nprot.2008.73.

Schmitz, R.J. and Amasino, R.M. (2007) ‘Vernalization: A model for investigating epigenetics and eukaryotic gene regulation in plants’, *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1769(5–6), pp. 269–275. Available at: https://doi.org/10.1016/J.BBAEXP.2007.02.003.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012a) ‘NIH Image to ImageJ: 25 years of image analysis’, *Nature Methods*, pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012b) ‘NIH Image to ImageJ: 25 years of image analysis’, *Nature Methods 2012 9:7*, 9(7), pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.

Schopf, J.W. *et al.* (1996) ‘Activation of floral meristem identity genes in Arabidopsis’, *Nature 1996 384:6604*, 384(6604), pp. 59–62. Available at: https://doi.org/10.1038/384059a0.

Schueler, S., Schlünzen, K.H. and Scholz, F. (2005) ‘Viability and sunlight sensitivity of oak pollen and its implications for pollen-mediated gene flow’, *Trees - Structure and Function*, 19(2), pp. 154–161. Available at: https://doi.org/10.1007/S00468-004-0376-1/FIGURES/4.

Schuerger, A.C., Brown, C.S. and Stryjewski, E.C. (1997) *Anatomical Features of Pepper Plants (Capsicum annuum L.) Grown under Red Light-emitting Diodes Supplemented with Blue or Far-red Light*, *Sedgwick Rd., Bldg*.

Schwaegerle, K.E. and Levin, D.A. (1990) ‘Quantitative genetics of seed size variation inPhlox’, *Evolutionary Ecology 1990 4:2*, 4(2), pp. 143–148. Available at: https://doi.org/10.1007/BF02270911.

Seppänen, M.M. et al. (2010) “Vernalization response of Phleum pratense and its relationships to stem lignification and floral transition,” Annals of Botany, 106(5), pp. 697–707. Available at: https://doi.org/10.1093/aob/mcq174.

Servedio, M.R. *et al.* (2014) ‘Not Just a Theory—The Utility of Mathematical Models in Evolutionary Biology’, *PLoS Biology*, 12(12), p. 1002017. Available at: https://doi.org/10.1371/JOURNAL.PBIO.1002017.

Shachar-Hill, B. *et al.* (2013) ‘Mercury-sensitive water channels as possible sensors of water potentials in pollen’, *Journal of Experimental Botany*, 64(16), pp. 5195–5205. Available at: https://doi.org/10.1093/JXB/ERT311.

SHAPIRO, S.S. and WILK, M.B. (1965) ‘An analysis of variance test for normality (complete samples)’, *Biometrika*, 52(3–4), pp. 591–611. Available at: https://doi.org/10.1093/BIOMET/52.3-4.591.

Sharma, N. *et al.* (2020a) ‘The molecular mechanism of vernalization in Arabidopsis and cereals: role of Flowering Locus C and its homologs’, *Physiologia Plantarum*, 170(3), pp. 373–383. Available at: https://doi.org/10.1111/PPL.13163.

Sharma, N. *et al.* (2020b) ‘The molecular mechanism of vernalization in Arabidopsis and cereals: role of Flowering Locus C and its homologs’, *Physiologia Plantarum*, 170(3), pp. 373–383. Available at: https://doi.org/10.1111/PPL.13163.

Shavrukov, Y. *et al.* (2017) ‘Early flowering as a drought escape mechanism in plants: How can it aid wheat production?’, *Frontiers in Plant Science*, 8, p. 1950. Available at: https://doi.org/10.3389/FPLS.2017.01950/BIBTEX.

Sheldon, C.C. *et al.* (2009) ‘Mechanisms of gene repression by vernalization in Arabidopsis’, *Plant Journal*, 59(3), pp. 488–498. Available at: https://doi.org/10.1111/J.1365-313X.2009.03883.X.

Sherry, R.A. and Lord, E.M. (1996) ‘DEVELOPMENTAL STABILITY IN LEAVES OF CLARKIA TEMBLORIENSIS (ONAGRACEAE) AS RELATED TO POPULATION OUTCROSSING RATES AND HETEROZYGOSITY’, *Evolution*, 50(1), pp. 80–91. Available at: https://doi.org/10.1111/J.1558-5646.1996.TB04474.X.

Shindo, C. *et al.* (2005) ‘Role of FRIGIDA and FLOWERING LOCUS C in Determining Variation in Flowering Time of Arabidopsis’, *Plant Physiology*, 138(2), pp. 1163–1173. Available at: https://doi.org/10.1104/PP.105.061309.

Shirsekar, G. *et al.* (2021) ‘Fine-scale Population Structure of North American Arabidopsis thaliana Reveals Multiple Sources of Introduction from Across Eurasia’, *bioRxiv*, p. 2021.01.22.427575. Available at: https://doi.org/10.1101/2021.01.22.427575.

Singh, N. *et al.* (2013) ‘Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties’, *PLOS ONE*, 8(12), p. e84136. Available at: https://doi.org/10.1371/JOURNAL.PONE.0084136.

Singh, N.B. *et al.* (1995) *RELATIONSHIP OF PLANT HEIGHT, DAYS TO FLOWERING AND MATURITY TO GRAIN YIELD IN SHORT-DURATION DETERMINATE PIGEONPEA* •. Available at: www.IndianJournals.com.

Smýkal, P. *et al.* (2018) ‘The Impact of Genetic Changes during Crop Domestication’, *Agronomy 2018, Vol. 8, Page 119*, 8(7), p. 119. Available at: https://doi.org/10.3390/AGRONOMY8070119.

Song, Y.H. *et al.* (2014) ‘Distinct roles of FKF1, GIGANTEA, and ZEITLUPE proteins in the regulation of constans stability in Arabidopsis photoperiodic flowering’, *Proceedings of the National Academy of Sciences of the United States of America*, 111(49), pp. 17672–17677. Available at: https://doi.org/10.1073/PNAS.1415375111/SUPPL_FILE/PNAS.201415375SI.PDF.

Song, Y.H., Ito, S. and Imaizumi, T. (2013) ‘Flowering time regulation: photoperiod- and temperature-sensing in leaves’, *Trends in plant science*, 18(10), p. 575. Available at: https://doi.org/10.1016/J.TPLANTS.2013.05.003.

Sork, V.L. (2018) ‘Genomic Studies of Local Adaptation in Natural Plant Populations’, *Journal of Heredity*, 109(1), pp. 3–15. Available at: https://doi.org/10.1093/jhered/esx091.

Soto-Cerda, B.J. *et al.* (2014) ‘The potential of pale flax as a source of useful genetic variation for cultivated flax revealed through molecular diversity and association analyses’, *Molecular Breeding*, 34(4), pp. 2091–2107. Available at: https://doi.org/10.1007/S11032-014-0165-5/FIGURES/3.

South, A. (2011) *rworldmap: A New R package for Mapping Global Data*. Available at: http://www.un.org/millenniumgoals/bkgd.

Speck, T. and Burgert, I. (2011) ‘Plant Stems: Functional Design and Mechanics’, *http://dx.doi.org/10.1146/annurev-matsci-062910-100425*, 41, pp. 169–193. Available at: https://doi.org/10.1146/ANNUREV-MATSCI-062910-100425.

Sporemex, 2022. Spores and Pollens. [online] Sporomex.co.uk. Available at: <https://www.sporomex.co.uk/technology/spores-and-pollens> [Accessed 8 August 2022].

Sreedharan, S.P., Kumar, A. and Giridhar, P. (2018) ‘Primer design and amplification efficiencies are crucial for reliability of quantitative PCR studies of caffeine biosynthetic N-methyltransferases in coffee’, *3 Biotech*, 8(11). Available at: https://doi.org/10.1007/s13205-018-1487-5.

Starlabgroup.com. 2022. 96-Well PCR Plate, Skirted, Low Profile, White - STARLAB. [online] Available at: <https://www.starlabgroup.com/GB-en/product/96-well-pcr-plate-skirted-low-profile-white-e1403-5209.html> [Accessed 30 May 2022].

Steel, R.G.D. and Torrie, J.H., 1960. Principles and procedures of statistics. Principles and procedures of statistics.

Stinchcombe, J.R. *et al.* (2004) ‘A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA’, *Proceedings of the National Academy of Sciences*, 101(13), pp. 4712–4717. Available at: https://doi.org/10.1073/PNAS.0306401101.

Stöcklin, J., Kuss, P. and Pluess, A.R. (2009) ‘Genetische Vielfalt, phänotypische Variabilität und lokale Adaptation in der Alpenlandschaft: Fallstudien mit vier alpinen Pflanzenarten’, *Botanica Helvetica*, 119(2), pp. 125–133. Available at: https://doi.org/10.1007/s00035-009-0065-1.

Sultan, S.E. (2000) “Phenotypic plasticity for plant development, function and life history,” Trends in Plant Science, 5(12), pp. 537–542. Available at: https://doi.org/10.1016/s1360-1385(00)01797-0.

Sun, J. *et al.* (2019) ‘Photoperiod sensitivity of Canadian flax cultivars and 5-azacytidine treated early flowering derivative lines’, *BMC Plant Biology*, 19(1). Available at: https://doi.org/10.1186/s12870-019-1763-5.

Sun, M. and Wong, K.C. (2001) ‘Genetic structure of three orchid species with contrasting breeding systems using RAPD and allozyme markers’, *American Journal of Botany*, 88(12), pp. 2180–2188. Available at: https://doi.org/10.2307/3558379.

Sun, S. and Frelich, L.E. (2011) ‘Flowering phenology and height growth pattern are associated with maximum plant height, relative growth rate and stem tissue mass density in herbaceous grassland species’, *Journal of Ecology*, 99(4), pp. 991–1000. Available at: https://doi.org/10.1111/J.1365-2745.2011.01830.X.

Sundaresan, V. (2005) ‘Control of seed size in plants’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(50), p. 17887. Available at: https://doi.org/10.1073/PNAS.0509021102.

Suter, L. *et al.* (2014) ‘Gene Regulatory Variation Mediates Flowering Responses to Vernalization along an Altitudinal Gradient in Arabidopsis’, *Plant Physiology*, 166(4), pp. 1928–1942. Available at: https://doi.org/10.1104/PP.114.247346.

Sveinsson, S. *et al.* (2014a) ‘Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (Linum) using transcriptomics’, *Annals of Botany*, 113(5), pp. 753–761. Available at: https://doi.org/10.1093/AOB/MCT306.

Szakiel, A., Pączkowski, C. and Henry, M. (2010) “Influence of environmental abiotic factors on the content of saponins in plants,” Phytochemistry Reviews, 10(4), pp. 471–491. Available at: https://doi.org/10.1007/s11101-010-9177-x.

Szpiech, Z.A. and Rosenberg, N.A. (2011) ‘On the size distribution of private microsatellite alleles’, *Theoretical Population Biology*, 80(2), pp. 100–113. doi:10.1016/j.tpb.2011.03.006.

Tajima, F. (1989) ‘Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism’, *Genetics*, 123(3), p. 585. Available at: https://doi.org/10.1093/GENETICS/123.3.585.

Tatham, S., 2022. PuTTY User Manual. [online] PuTTY. Available at: <https://the.earth.li/~sgtatham/putty/0.77/htmldoc/> [Accessed 21 September 2022].

Teper-Bamnolker, P. and Samach, A. (2005) ‘The Flowering Integrator FT Regulates SEPALLATA3 and FRUITFULL Accumulation in Arabidopsis Leaves’, *The Plant Cell*, 17(10), p. 2661. Available at: https://doi.org/10.1105/TPC.105.035766.

*The GNU Bourne-Again Shell* (no date). Available at: https://tiswww.case.edu/php/chet/bash/bashtop.html (Accessed: 5 May 2022).

Thomashow, M.F. (1999) “Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms,” Annual Review of Plant Physiology and Plant Molecular Biology, 50(1), pp. 571–599. Available at: https://doi.org/10.1146/annurev.arplant.50.1.571.

Tigano, A. and Friesen, V.L. (2016) ‘Genomics of local adaptation with gene flow’, *Molecular Ecology*. Blackwell Publishing Ltd, pp. 2144–2164. Available at: https://doi.org/10.1111/mec.13606.

Toräng, P. *et al.* (2015) ‘Large-scale adaptive differentiation in the alpine perennial herb Arabis alpina’, *The New phytologist*, 206(1), pp. 459–470. Available at: https://doi.org/10.1111/NPH.13176.

Tork, D.G. *et al.* (2022) ‘Ideotype Selection of Perennial Flax (Linum spp.) for Herbaceous Plant Habit Traits’, *Agronomy*, 12(12). Available at: https://doi.org/10.3390/agronomy12123127.

Trelease, W., 1931. Winter botany. New York: Dover Publications.

Tukey, J., 1949. Comparing Individual Means in the Analysis of Variance. Biometrics, 5(2), p.99.

Unacademy, 2022. Draw a Labelled Diagram of the Pollen Grain. [online] Unacademy.com. Available at: <https://unacademy.com/content/question-answer/biology/draw-a-labelled-diagram-of-the-pollen-grain/> [Accessed 8 August 2022].

Uysal, H. *et al.* (2010) ‘Genetic diversity of cultivated flax (Linum usitatissimum L.) and its wild progenitor pale flax (Linum bienne Mill.) as revealed by ISSR markers’, *Genetic Resources and Crop Evolution*, 57(7), pp. 1109–1119. Available at: https://doi.org/10.1007/S10722-010-9551-Y/FIGURES/4.

Uysal, H. *et al.* (2012) 'Variation in phenotypic characters of pale flax (Linum bienne Mill.) from Turkey', *Genetic Resources and Crop Evolution*, 59(1), pp. 19–30. Available at: https://doi.org/10.1007/S10722-011-9663-Z/FIGURES/10.

Valentim, F.L. *et al.* (2015a) 'A quantitative and dynamic model of the Arabidopsis flowering time gene regulatory network', *PloS one*, 10(2). Available at: https://doi.org/10.1371/JOURNAL.PONE.0116973.

Vandesompele, J. *et al.* (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', *Genome biology*, 3(7). Available at: https://doi.org/10.1186/GB-2002-3-7-RESEARCH0034.

Venables, W.N. and Ripley, B.D. (2002a) 'Random and Mixed Effects', pp. 271–300. Available at: https://doi.org/10.1007/978-0-387-21706-2_10.

Verdu, C.F. *et al.* (2016) 'Dealing with paralogy in RADseq data: in silico detection and single nucleotide polymorphism validation in Robinia pseudoacacia L.', *Ecology and Evolution*, 6(20), pp. 7323–7333. Available at: https://doi.org/10.1002/ECE3.2466.

Vieira, M.L.C. *et al.* (2016) 'Microsatellite markers: what they mean and why they are so useful', *Genetics and Molecular Biology*, 39(3), p. 312. Available at: https://doi.org/10.1590/1678-4685-GMB-2016-0027.

Villanueva, R.A.M. and Chen, Z.J. (2019) 'ggplot2: Elegant Graphics for Data Analysis (2nd ed.)', *https://doi.org/10.1080/15366367.2019.1565254*, 17(3), pp. 160–167. Available at: https://doi.org/10.1080/15366367.2019.1565254.

Vuletin Selak, G. *et al.* (2013) 'The effect of temperature and genotype on pollen performance in olive (Olea europaea L.)', *Scientia Horticulturae*, 156, pp. 38–46. Available at: https://doi.org/10.1016/J.SCIENTA.2013.03.029.

Walker, J.M. (1969) 'One-Degree Increments in Soil Temperatures Affect Maize Seedling Behavior', *Soil Science Society of America Journal*, 33(5), pp. 729–736. Available at: https://doi.org/10.2136/SSSAJ1969.03615995003300050031X.

Walters, D., Keil, D. and Walters, B., 1996. Vascular plant taxonomy. Dubuque, Iowa: Kendall/Hunt, p.598.

Wang, J. *et al.* (2018) 'A major locus controls local adaptation and adaptive life history variation in a perennial plant', *Genome Biology*, 19(1). Available at: https://doi.org/10.1186/s13059-018-1444-y.

Wang, Z. *et al.* (2012) 'The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads', *The Plant journal : for cell and molecular biology*, 72(3), pp. 461–473. Available at: https://doi.org/10.1111/J.1365-313X.2012.05093.X.

Watanabe, S. *et al.* (2011) 'A Map-Based Cloning Strategy Employing a Residual Heterozygous Line Reveals that the GIGANTEA Gene Is Involved in Soybean Maturity and Flowering', *Genetics*, 188(2), pp. 395–407. Available at: https://doi.org/10.1534/GENETICS.110.125062.

Wee, A.K. et al. (2020) ‘Genetic structures across a biogeographical barrier reflect dispersal potential of four Southeast Asian mangrove plant species’, *Journal of Biogeography*, 47(6), pp. 1258–1271. doi:10.1111/jbi.13813.

Weigel, D. *et al.* (1992) ‘LEAFY controls floral meristem identity in Arabidopsis’, *Cell*, 69(5), pp. 843–859. Available at: https://doi.org/10.1016/0092-8674(92)90295-N.

Weir, B.S. (2012) ‘Estimating F-statistics: A historical view’, *Philosophy of science*, 79(5), p. 637. Available at: https://doi.org/10.1086/667904.

Welch, S.M., Dong, Z. and Roe, J.L. (2004) *Modelling gene networks controlling transition to flowering in Arabidopsis*. Available at: www.cropscience.org.au.

Werner, C.R. *et al.* (2020) ‘How Population Structure Impacts Genomic Selection Accuracy in Cross-Validation: Implications for Practical Breeding’, *Frontiers in Plant Science*, 11, p. 2028. Available at: https://doi.org/10.3389/FPLS.2020.592977/BIBTEX.

White, N.J. and Butlin, R.K. (2021) ‘Multidimensional divergent selection, local adaptation, and speciation’, *Evolution*, 75(9), pp. 2167–2178. Available at: https://doi.org/10.1111/evo.14312.

Whitlock, M.C. (2015) ‘Modern approaches to local adaptation’, *American Naturalist*, 186, pp. S1–S4. Available at: https://doi.org/10.1086/682933.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. springer.

Wilczek, A.M. *et al.* (no date) ‘Genetic and physiological bases for phenological responses to current and predicted climates’. Available at: https://doi.org/10.1098/rstb.2010.0128.

Williams, G. C. 1966. Adaptation and natural selection. Princeton University Press, Princeton, NJ.

Williams, I.J., Martin, A.P. and Clark, S.J. (1990) ‘Pollination requirements of linseed (Linum usitatissimum)’, *The Journal of Agricultural Science*, 115(3), pp. 347–352. Available at: https://doi.org/10.1017/S0021859600075778.

Williams, J.H. and Reese, J.B. (2019) ‘Evolution of development of pollen performance’, *Current Topics in Developmental Biology*, pp. 299–336. doi:10.1016/bs.ctdb.2018.11.012.

Wolfe, Lorne M and Wolfe, L M (1995) ‘The genetics and ecology of seed size variation in a biennial plant, Hydrophyllum appendiculatum (Hydrophyllaceae)’, *Oecologia 1995 101:3*, 101(3), pp. 343–352. Available at: https://doi.org/10.1007/BF00328821.

Wright, S. (1965) ‘The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating’, *Evolution*, 19(3), p. 395. Available at: https://doi.org/10.2307/2406450.

Wright, S. (1978) ‘The Relation of Livestock Breeding to Theories of Evolution’, *Journal of Animal Science*, 46(5), pp. 1192–1200. Available at: https://doi.org/10.2527/JAS1978.4651192X.

Wu, C. *et al.* (2016) ‘Effect of plant age and vernalization on bolting, plant growth and enzyme activity of garlic (Allium sativum L.)’, *Scientia Horticulturae*, 201, pp. 295–305. Available at: https://doi.org/10.1016/J.SCIENTA.2016.02.006.

Xie, D. *et al.* (2018) ‘Genome-wide association study identifying candidate genes influencing important agronomic traits of Flax (Linum usitatissimum L.) using SLAF-seq’, *Frontiers in Plant Science*, 8, p. 2232. Available at: https://doi.org/10.3389/FPLS.2017.02232/BIBTEX.

Yan, L., Chouw, N. and Jayaraman, K. (2014) ‘Flax fibre and its composites - A review’, *Composites Part B: Engineering*, 56, pp. 296–317. Available at: https://doi.org/10.1016/j.compositesb.2013.08.014.

Yan, X. and Su, X., 2009. Linear regression analysis. Singapore: World Scientific.

Yang, C.Y. *et al.* (2007) ‘Luminescence of aniline blue in hydrophobic cavity of BSA’, *Journal of Photochemistry and Photobiology A: Chemistry*, 188(1), pp. 51–55. Available at: https://doi.org/10.1016/J.JPHOTOCHEM.2006.11.017.

Yang, H. *et al.* (2016) ‘Genetic Divergence between Camellia sinensis and Its Wild Relatives Revealed via Genome-Wide SNPs from RAD Sequencing’, *PLOS ONE*, 11(3), p. e0151424. Available at: https://doi.org/10.1371/JOURNAL.PONE.0151424.

You, F.M. *et al.* (2017) ‘Genetic variability of 27 traits in a core collection of flax (Linum usitatissimum L.)’, *Frontiers in Plant Science*, 8. Available at: https://doi.org/10.3389/fpls.2017.01636.

You, F.M. *et al.* (2018) ‘Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax’, *The Plant Journal*, 95(2), pp. 371–384. Available at: https://doi.org/10.1111/TPJ.13944.

Zamir, D., Tanksley, S.D. and Jones, R.A. (1981) ‘Low temperature effect on selective fertilization by pollen mixtures of wild and cultivated tomato species’, *Theoretical and Applied Genetics*, 59(4), pp. 235–238. Available at: https://doi.org/10.1007/BF00265501/METRICS.

Zhang, J. *et al.* (2020) ‘Genomic Comparison and Population Diversity Analysis Provide Insights into the Domestication and Improvement of Flax’, *iScience*, 23(4). Available at: https://doi.org/10.1016/j.isci.2020.100967.

Zhao, P. *et al.* (2022) ‘Precipitation and local environment shape the geographic variation of seed size across natural populations of sand rice (Agriophyllum squarrosum)’, *Journal of Experimental Botany*, 73(16), pp. 5682–5697. Available at: https://doi.org/10.1093/JXB/ERAC231.

Zheng, X. *et al.* (2012) ‘A high-performance computing toolset for relatedness and principal component analysis of SNP data’, *Bioinformatics*, 28(24), pp. 3326–3328. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTS606.

Zimmerman, S.J., Aldridge, C.L. and Oyler-Mccance, S.J. (2020) ‘An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern’, *BMC Genomics*, 21(1), pp. 1–16. Available at: https://doi.org/10.1186/S12864-020-06783-9/FIGURES/6.

Zinn, K.E., Tunc-Ozdemir, M. and Harper, J.F. (2010) ‘Temperature stress and plant sexual reproduction: uncovering the weakest links’, *Journal of Experimental Botany*, 61(7), p. 1959. Available at: https://doi.org/10.1093/JXB/ERQ053.

Zipper, H. *et al.* (2004) ‘Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications’, *Nucleic Acids Research*, 32(12), p. e103. Available at: https://doi.org/10.1093/NAR/GNH101.

Zohary, D. and Hopf, M. (2000) ‘Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley.’, *Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley.* [Preprint], (Ed.3).

Zonia, L. and Munnik, T. (2004) ‘Osmotically Induced Cell Swelling versus Cell Shrinking Elicits Specific Changes in Phospholipid Signals in Tobacco Pollen Tubes’, *Plant Physiology*, 134(2), p. 813. Available at: https://doi.org/10.1104/PP.103.029454.

---