Durham E-Theses

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

HABEAHAN, HORASMAN,FEBRICO

How to cite:

Use policy

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:

162

163

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

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

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

281 temperature treatments.

282   In chapter 4 of this thesis, we examined population genetics of wild *L. bienne* samples

283 originating from different latitudes across western Europe. We examined Wild flax *(L. bienne)*

284 populations across western Europe to provide insights into their genetic structure and diversity. This

285 population analysis will develop our understanding of adaptation in wild *Linum* in response to their

286 environment. A double-digest RAD sequencing (DDRadSeq) protocol was utilized to look at variation

287 in SNPs across different populations. Both *L. bienne* and *L. usitatissimum* samples were sequenced

288 and aligned to an *L. usitatissimum* whole genome. Genetic structuring of our *Linum* samples were

289 revealed across Western Europe. Cultivars in our collection revealed to be more genetically related

290 to the Northern accessions of our wild samples.

291   In chapter 5 we summarized plant architectures. In this chapter traits and their relationship

292 to the latitude were summarized in relation to the requirement for vernalization. Four traits were

293 measured: overall height of the plants after first flowering, the number of stems of the plants after

294 first flowering, and the number of flower buds on the plants at first flowering. In addition to this,

295 seed size was examined in terms of its area. We also examined correlation between traits and

296 environmental variables. We found that there was a relationship between traits when no-

297 vernalization occurred. With vernalization, these relationships became less significant and, in some

298 cases, not significant. This illustrates that vernalization influences the relationship of traits beyond

299 flowering initiation. There were also suggestions that the traits measured correlate with latitude

300 under no-vernalization treatments. For vernalized individuals, the correlation for the traits measured

301 and latitude was not significant. Seed sizes were strongly correlated with both latitude and climatic

302 variables in all cases of the treatments. The findings suggest there are effect of environmental

303 variables in these measured traits, which suggests differentiation within wild *Linum* species.

304

305

306

307

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

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

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

399

400

1.2 INTRODUCTION TO STUDY SPECIES

401   In this thesis, two species of the flowering plant in the genus *Linum* were studied. *Linum* is a

403 genus of flowering plants which contains more than 200 identified species, distributed in the

404 subtropics and temperate zones (Öhlund et al., 2015; Muravenko et al., 2010). *Linum* is found in

405 temperate into the subtropical regions with species such as the blue flaxes *L. lewisii* and the *L.*

406 *narbonense* often cultivated as garden ornamentals (Ionkova et al., 2013; Addicott, 1977). In

407 horticulture, prolific flowering makes them of interest in ornamental uses (Tork D. et al., 2022.

408 Geographically, *Linum* is distributed across Northern Americas, and the European continents.

409 Another cultivated species which is important industrially includes the *L. usitatissimum* (Jhala A. et

410 al., 2010). The wild ancestor of this cultivar is thought to be the pale flax *L. bienne* (Uysal et al.,

411 2010). Allaby *et al* (2005), reveals that there is a single domestication event for *Linum,* and they

412 suggest that the *sad2* locus of *Linum* was subject to artificial selection in the cultivars. The wild

413 relative is significant because it represents a primary genepool for development of the cultivar.

414 Previous study reveals that there is a genetic differentiation between the two species and suggests

415 that 48% of *L. bienne* alleles were unique (Soto-Cerda et al*.,* 2014). With this suggestion, it is also

416 possible to map *L. bienne* sequences to the cultivars to identify alleles favourable to the cultivars

417 (Soto-Cerda *et al.,* 2014). In this thesis, both wild (*L. bienne*) and cultivar (*L. usitatissimum*)

418 individuals will be examined.

419   Morphologically, *L. usitatissimum* tends to have taller stems when compared to *L. bienne*.

420 They also tend to have larger seeds, capsules, and flowers. The pale flax (*L. bienne* Mill) was

421 previously shown to vary in flowering phenology (Uysal et al., 2011) and the cultivated flax, *L.*

422 *usitatissimum* has been extensively studied in terms of its phenology with studies suggesting

423 development of certain phenotypes to optimise yield (Rehman et al., 2014). Wild and cultivated flax

424 differ in their stem structure seeds (Figure 1) .

425
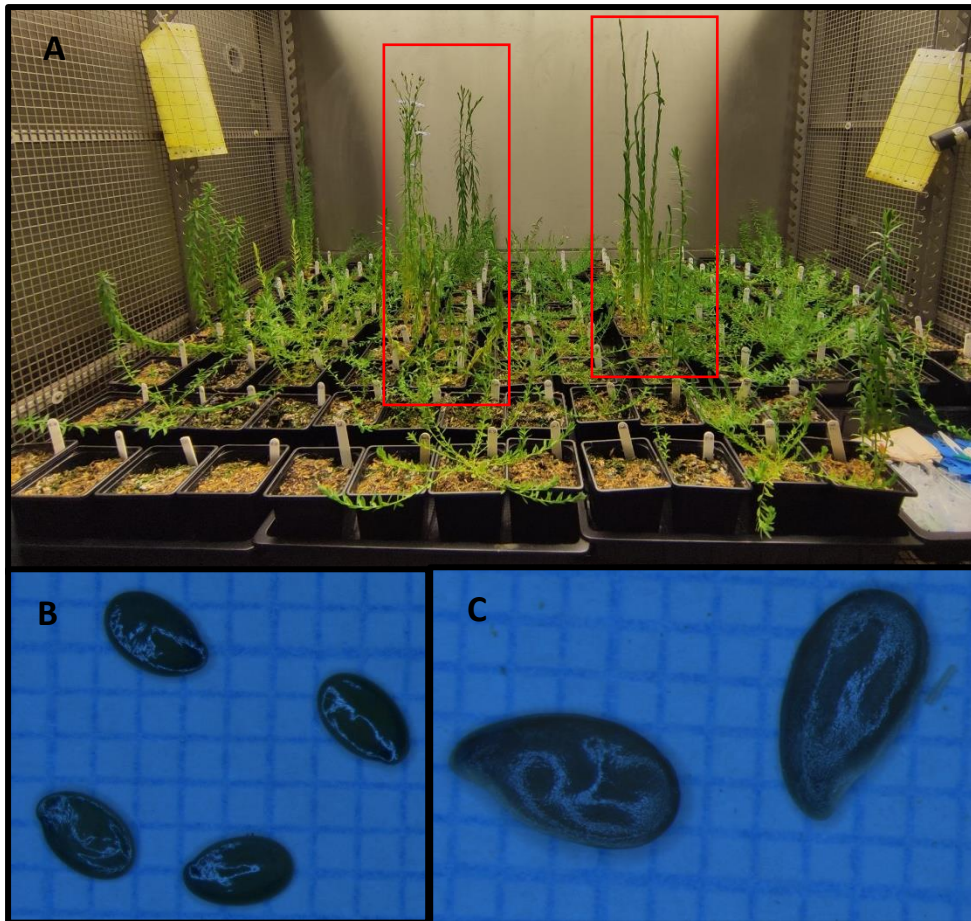
Figure 1. Collection of wild and cultivated S1 (selfing) *L. usitatissimum* and *L. bienne* grown for >3 months period. Cultivars (in red boxes) are growing taller, sturdier stems than wild relatives. There is variation in plant heights and number of stems in our *Linum* samples (1A). Wild (*L. bienne*) 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

442   fibre morphotypes have greater straw weight, plant height and protein content than the oil

443   morphotypes while the oil morphtypes have greater seed weight, seed oil content, and branching

444   capability (You F. et al., 2017). Measures of oil and stem fibre content of several oil/fibre

445   morphotypes also shows variability, suggesting variation within this species for industrial use

446   (Diederichsen and Ulrich., 2009; Rozhmina T. et al., 2021).

447   Historically, *Linum usitatissimum* has been cultivated for its fibres and seeds, with evidence

448   of domestication in the Middle East, revealed by the diversity of the *sad2* gene locus (Allaby et al.,

449   2005). Plant domestication is an important process which enables human civilizations to utilize plant

450   materials for advancements (Diamond, 2002). In terms of agriculture, traits for larger seed (size and

451   weight) and fibre properties have been selected for, when considering flax as cultivars with

452   development for said traits still of interest to this day (Rahimi M. et al., 2011; Yan L. et al., 2014; Guo

453   D. et al., 2020). As a result of domestication, genes for seed sizes, flowering time, and capsule

454   dehisence are revealed to be artificially selected (Zhang J. et al., 2020). More recently, qualities of

455   cell wall, stem strength and fibre properties are studied for using flax fibres as an environmentally

456   friendly composite alternative (Goudenhooft C. et al., 2019). Selective breeding in a trait may imply

457   better selection for that trait, however in long term, the consequences to those selected lines in

458   terms of survival and loss of functions in other genes have not been tested yet. It has been shown

459   that the temperate region of the world started to cultivate flax thousands of years ago.

460   Archaeological records illustrate that flax were cultivated during the Viking and Medieval age

461   (Ejstrud, et al., 2011) (Nag, et al., 2015). This suggests a process of long-term artificial selection in *L.*

462   *usitatissimum* with implications in several field of interests for agriculture. We are interested in

463   flowering time for this study. We suggest that selection for faster flowering in the cultivars may

464   result in loss of signalling in flowering time genes.

465   *Linum* seeds produces oleic compounds composed of triglycerides, and these are particularly

466   rich in α-linolenic acid, which gives linseed oil its ability to polymerize into solids. This is useful in

467   productions of industrial resins and solvents (Vereshchagin, & Novitskaya, 1965; McCullough, *et al*.,

468   2011), suggesting significance in the industry sector. In addition to this, recent research has found

469   that linseed oil has a positive effect on reducing the level of cholesterol in rabbits, illustrating

470   potential for linseed in the health food sector (Króliczewska, et al., 2018). Molecularly, it is suggested

471   that *Linum usitatissimum* seeds contains several biologically active components that are useful both

472   medically and industrially. This includes the presence of lignin in linseed, useful in resin production

473   (Del Rio *et al*., 2011; Touré & Xueming, 2010). These lignans have shown to be medically beneficial in

474   relation to their antioxidant properties (Hosseinian, et al., 2006). Fatty acid profiling has also

475   revealed that linseeds provide significant nutritional values (El-Beltagi, et al., 2007), suggesting

476 values in food security. Other potential uses of cultivated *Linum* species is their biogeographic trait

477 variation. This could be useful for ornamental purposes. In recent study of the ornamental *L. lewisii*,

478 traits such as flowering indeterminacy, seed mass and stem numbers suggest its potential use in field

479 restoration and agriculture (Innes, et al., 2022). These findings reveal that not only the common

480 cultivated flax was cultivated early on for their fibres, but also research into flax biproducts and

481 genetics is still ongoing with a strong potential in multiple fields such as industry, agriculture, the

482 health food sectors and horticulture. With these in mind, it is then of interest to optimise *Linum* crop

483 production. Examples of research in this area includes seed priming which influences crop growth

484 and development (Rehman H. *et al*., 2014), and experiments to employ plant growth regulators

485 (giberellic acid) for optimal yield and growth in *L. usitatissimum* (Rastogi A. *et al*., 2013). We propose

486 that, information into local adaptations and genetic implications of *Linum* wild crop relatives could

487 aid these efforts to further improve *Linum* as a crop plant. This may be done by exploring genetic

488 and phenotypic responses to environment changes to benefit agriculture. To this day, research for

489 implementation of wild relatives towards genetic engineering cultivated flax is relatively scarce. For

490 studies in this thesis, 18 varieties of *Linum usitatissimum* seeds were received from IPK World

491 Collection (https://www.ipk-gatersleben.de/en/). A list of cultivars and their morphotypes can be

492 found in appendix 1.

493

1.2.2 *Linum bienne:* Wild Flax

495   *Linum bienne* is a biennial plant and a wild relative of *Linum usitatissimum* (Gill KS, 1966). Its

496 common name is the "pale flax", considered the wild forebear of the cultivated *Linum usitatissimum*.

497 A study by Allaby *et al*, reveals this by exploring *sad2* loci from 30 accessions of wild and cultivated

498 flax. They found phylogenetic evidence that the wild type *L. bienne* was first domesticated for oil and

499 that there is an artificial selection of the loci, indicative of cultivation of the wild types *L. bienne*

500 (Allaby, et al., 2005). Further studies also revealed that *sad2* locus is a candidate domestication locus

501 associated with increased unsaturated fatty acid production in cultivated flax (Fu Y.B. et al., 2012).

502 This suggests implication for seed selection when *Linum* was first domesticated as opposed to fibre.

503 In the literature, there are suggestions that *L. bienne* are widely distributed around Western Europe,

504 the Mediterannean basins, North Africa and into Iran and the Caucasus. This represents its wide

505 distribution in sub-temperate into temperate zones (Zohary and Hopf, 1993). Just like the cultivar *L.

506 usitatissimum*, this wild progenitor has the same number of chromosome (*n=15*) and they're both

507 homostylous, in contrast to some distylous relatives such as the *Linum Tenue* (Gutiérrez-Valencia *et

508 al*., 2022).

509   Previously the wild *L. bienne* were considered as sub-species of *L. usitatissimum* (Uysal et al.,
510   2011), but now they are considered as the closest relative to *L. usitatissimum*. Phylogenetic studies
511   have illustrated that *L. bienne* is the closest relative of L. usitatissimum (Uysal, et al., 2010). Further
512   phylogenies from studies of *Linum* orthologs also suggests that *L. bienne* is the closest relatives of *L.*
513   *usitatissimum* (McDill J. et al., 2009; Sveinsson S. et al., 2011). This species represents a wider gene
514   pool, which can be of aid for *Linum usitatissimum* development in agriculture.

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

528

1.2.3 Production Values of *Linum* Biproducts:

530   With the above put into consideration, the current observation is that flax output is
531   economically significant. One of the biproducts of *Linum* is its fibre. Flax fibre production in 2020
532   stands at 976,113 tonnes (FAOSTAT data, 2022). In 2016 alone flax fibre output reached a
533   production value of 535 million USD. This value increased more in 2020, with flax world trade value
534   estimated at 726 million USD (Flax Fibers | OEC, 2022). The world's largest flax fibre exporter in 2020
535   is France, with an estimate of 51.4% in total world export value. The export share in the UK only
536   stands at 0.12% with an export value of 870 thousand USD. The world's top importer of flax fibres is
537   China with 47.8% total import value which stands at 347 million USD. The UK imports 0.14% of total
538   flax fibre import value, with an import value of 1.03 million USD (Flax Fibers | OEC, 2022). The export
539   value growth from the year 2000 – 2020 in the UK is -76.7%, with export values, massively reduced
540   from 3.74 million USD at the start of this period. This reveals, that at least in the UK, current export
541   value in flax fibre is decreasing. This being the case, in developing countries demands for flax fibre

542 are increasing. In India alone, the import value increased from 5.93 million USD in the year 2000 to

543 47.7 million USD in 2020, an increase of 703%. Other countries such as China reveals an import value

544 growth of 254% between the year 2000 and 2020 while Vietnam shows a >1000% increase (Flax

545 Fibers | OEC, 2022). This new trend in demand of flax fibre in developing countries could provide a

546 reason for temperate countries such as the UK to increase export of flax fibres.

547   Other major biproduct of *Linum* cultivars is linseed. In 2020, the total world production of

548 linseed is 3,367,331 tonnes, with Kazakhstan producing the most at 105,8247 tonnes (FAOSTAT data,

549 2022). In the same year the trade value of linseed worldwide is 980 million USD. Major exporters of

550 linseed include Canada (25.8%), Russia (23.5%) and Kazakhstan (21.7%). Major importers of linseed

551 include China (26.8%), Belgium (22.7%) and Germany (8.75%). In the UK, export of linseed grew

552 57.7% between the years 2000 – 2020. The total export value of linseed in the UK in 2020 is 7.28

553 million USD. The biggest increase is seen in Kazakhstan between the years 2000 – 2020 (Flax Fibers |

554 OEC, 2022). The increase in export of linseed illustrates the economical values linseed represents

555 worldwide in terms of industry and agriculture. Potential to improve cultivar types will be of huge

556 interest to the economic values represented in *Linum* biproducts derived from flax fibres and

557 linseeds.

558

559

1.3 LITERATURE REVIEW OF *LINUM* GENETIC DIVERSITY

560

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

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

594

1.3.1 Methods for studying population genomics in *Linum*

595

596   Within this study, we compare genetic diversity of wild *Linum bienne* samples, relative to the

597 cultivars. With regards to genetic analysis, it is worth noting Next-Gen sequencing (NGS) techniques.

598 NGS makes it possible to study local adaptation at a population level without a whole genome

599 sequence. This is of interest when looking at population genetics of wild relatives of agricultural

600 plants (Park & Kim, 2016). In relation to our study model, NGS is valuable in terms of gaining insights

601 into the genetic diversity of *L. bienne* and *L. usitatissimum.* Population analysis of wild relatives may

602 also be informative in the development of crop relatives. Several NGS techniques are available, and

603 the sequence data can be mapped de-novo or to a reference genome. This is relatively faster and

604 cheaper than producing a whole genome for the wild relatives.

605   With next generation sequencing technologies available, several *Linum* molecular markers

606 (microsattelites) have been developed (Fu & Peterson, 2010). It was shown that molecular markers

607 are transferable in many species of *Linum* (Soto-Cerda, et al., 2011). These transferable molecular

608 markers, often referred to as simple sequence repeats (SSRs), have been useful for data

609 enhancement for genetic and evolutionary studies. Because of this transferability, genetic and

610 evolutionary studies can compare between wild and cultivars of flax. Numerous species in the *Linum*

611 genus represent an interest to study their evolution, while the cultivar (*L. usitatissimum*) represents

612 an interest in the development of *Linum* as a crop through study of their genetic makeups and local

613 adaptation mechanisms such as self-incompatibility mechanisms (SI) and inbreeding depression

614 which their wild relatives may be able to reveal.

615   For whole genome sequencing, Illumina's Hi-C sequencing has become a mainstream

616 sequencing technique. It is preceded by the 3C method (Downes et al., 2021). The method is widely

617 used to examine organisation and conformation of chromosomes and secondary genetic structures

618 into chromosomal level organisation based on whole genome sequencing (Belton et al., 2012;

619 Lieberman-Aiden et al., 2009). The method involves the use of isoschizomer restriction enzymes

620 such as DpnII to digest cell nuclei isolates, recognising and cutting amino acid sites to generate an

621 overhang. These overhangs can then be used to to enable paired-end sequencing using adaptor

622 ligation (Belaghzal et al., 2017). However, to gain whole genome sequences often requires laborious

623 procedures that are relatively more expensive than reduced representation sequencing techniques

624 such as RAD Sequencing. This sequencing technique can help with identification of markers such as

625 single-nucleotide polymorphisms (SNP) that can be used to identify genetic signals such as

626 structuring between population and even identify novel loci for quantitative trait locus analyses.

627   The genome of *L. Usitatissimum* is estimated at 370Mb in size (Wang et al., 2012). Whole

628   genome sequences are publicly available down to chromosomal levels for the cultivated *L.*

629   *usitatissimum* having first been assembled using shotgun sequencing technologies in 2012 (Wang Z.

630   et al., 2012). The contig assembly contained 302 Mb of non-redundant sequence, representing an

631   estimated 81% genome coverage (Wang Z. et al., 2012). The initial genome is now referred to as CDC

632   Bethune v1 (Sa et al., 2021). The most widely used *L. usitatissimum* whole genome is the CDC

633   Bethune v2 (Sa et al., 2021; You et al., 2018). For this whole genome contigs contains a total N50 of

634   6.64Mb summing up 316 Mb of reads and a 97% coverage of annotated genes, considerably higher

635   than the previous CDC Bethune v1 (You et al., 2018). Another genome was assembled using long

636   reads for the Atlant variety of *L. usitatissimum*. Using a combined Oxford Nanopore and Illumina

637   sequencing technique, whole genome with a total length of 361.7 Mb, N50 of 350 kb, and 97.40%

638   completeness was recently achieved for the Atlant variant of the *L. usitatissimum* (Dmitriev et al.,

639   2021). The availability of this whole genome also suggests that continuity of these assemblies is

640   relatively poor as revealed by the relatively small contig N50. The largest contig N50 was found to be

641   365Kb (Sa et al., 2021; Dmitriev et al., 2021). They suggest that events such as recent whole-genome

642   duplication may result to the collapse of homologs and repeat sequences during the assembly

643   process, both under short and long reads sequencing (Sa et al., 2021; Dmitriev et al., 2021).

644   For the wild *L. bienne*, whole genome sequences are under development and not publicly

645   available. However, past studies using inter simple sequence repeats (ISSR) and RADSeq markers

646   have shown that the wild *L. bienne* illustrates ancestral relationship with the cultivated *L.*

647   *usitatissimum* and there were indications that suggests the contribution of *L. bienne* to *L.*

648   *usitatissimum* genome through gene flow (Fu, 2012; Gutaker RM., 2014). Molecular and cytogenic

649   studies of both cultivated and wild *Linum* were also undergone using high-throughput ribosomal

650   RNA. The studies suggests that 5S rDNA and ITS phylogeny is closely related between *L.*

651   *usitatissimum* and *L. bienne*. Both species have identical karyotypes and distribution of rDNA sites. In

652   both species, the coding sequences of 5S rDNA genes were similar (Bolsheva N. et al., 2017).

653   Comparison of ITS1 to 5.8S rDNA-ITS2 sequences showed that rDNA sequences were conservative in

654   all studied flax specimens (Bolsheva N. et al., 2017). In addition to this, Assessment of the number of

655   polymorphic sites in a genetic marker (*LuTFL* homologs) suggests mixed ancestry of the locus and

656   different copy of the locus suggests ancestry to the wild pale flax *L. bienne* (Gutaker RM., 2014).

657   More Northern populations are suggested to have high similarity to the *L. usitatissimum* (Gutaker

658   RM., 2014). This high similarity illustrates that whole genome for the cultivar is acceptable for use in

659   the wild *L. bienne* before the availability for its whole genome.

660   Other genetic events have also occurred as *Linum* species have undergone polyploidy

661   events. In plants, polyploidy is an event whereby whole genome duplication occurred increasing the

662   number of chromosomes. An example of this has also occurred in sugarcanes (Vilela et al., 2017).

663   Genetically, the genus *Linum* has experienced multiple polyploidy events. The last polyploidy event

664   occurring around 5-9 million years ago (mya) and a previous polyploidy event around 20-40 mya.

665   This reveals that every gene in the flax genome is potentially duplicated and multiplied through

666   polyploidy (Sveinsson et al., 2013). This needs to be accounted for when looking at their genetic

667   make-ups and for studying gene expression (Sveinsson et al., 2013). Polyploidy is important, as it is a

668   widely accepted to be a pervasive mechanism of plants and is often consequently followed by the

669   selective silencing of genes (Wendel & Adams, 2005). Studies have previously inferred that DNA

670   alterations in flax could induce changes that are heritable. These changes were also linked to

671   environmental factors (Schneeberger & Cullis, 1991). Polyploidy events can consequently rid of

672   genes, which were previously present in predecessors of cultivated flax. With the aid of cultivation,

673   breeding strategies of the genus, conditions may have changed and thus the genetic make-up of

674   cultivars is heritably different to those of the wild progenitor, representing a wider gene pool to be

675   studied.

676

1.3.2 Self-Incompatibility in *Linum*

678   In plants, control of outcrossing is often due to self-incompatibility (SI) genes at the S locus

679   (Newbigin E. et al., 1993). This locus is responsible for the prevention of self-fertilization and

680   therefore selfing in plants. Inactivation of the S-locus receptor kinase (SRK) and the S-locus cysteine-

681   rich protein (SCR) within this locus has led to the loss of self-incompatibility in some plants in the

682   *Brassicaceae* family (Sherman-Broyles, et al., 2007) and *Arabidopsis* (Suzuki, et al., 1999). Self-

683   Incompatibility genes is not yet fully understood in *Linum*, research focuses mainly on the formation

684   of hetero/homo styly. Differences in the S-genes expressed in the male and female reproductive

685   structure trigger self-rejection, controlling outcrossing, and leading to high heterozygosity

686   (Thompson & Kirch, 1992). Heterostyly in *Linum* means that there are morphological differences

687   between styles and anther filaments of *Linum* and it has been found as an outcrossing mechanism in

688   *Linum* (McDill, et al., 2009). Therefore, in *Linum,* polymorphisms of breeding organs contribute to

689   heterozygosity (Ebert, et al., 1989). In *Linum grandiflorum*, the S-locus controls the flower

690   morphology through regulations of transcriptional S-locus products (Ushijima, et al., 2011). This

691   reveals the connection of the S-locus regulation with the ability for *Linum* to promote heterozygosity

692   and therefore maintain genetic variation. The mating system of wild flax (*L. bienne*) is not yet fully

693   understood, although related species show distyly to ensure outcrossing, research regarding

694   polymorphisms in the wild relatives are still scarce to date. This thesis will explore aspects of *Linum*

695   reproduction by looking at pollen morphology, especially pollen viabilities measures such as the

696   availability of pollen to germinate under environmental stress, This will provide trait data to study

697   genetic mechanisms which may affect such morphology.

698

699   ### 1.3.3 Population Genetics Analysis

700   In modern genetics, there is an interest in determining genotypes which represent

701   underlying phenotypes. These are often crucial to a study model with potentially many populations,

702   representing a wide gene pool. To do this, it is crucial to gain insight into genetic sequences of many

703   individuals relatively quickly and cheaply. For the purposes of sequencing, high capacity and

704   relatively low costing technology for sequencing were first achieved using microarrays-based

705   sequencing technologies (Schena et al., 1995). Later, widely used genetic markers were introduced

706   as "microsatellites", also known as Single Sequence Repeats (SSRs) across the whole genome (Vieira

707   et al., 2016). This genotyping technology was a pre-cursor of Next-Generation Sequencing (NGS)

708   technologies such as RAD Sequencing. NGS provides sequence data for nucleic acids, DNA and RNA,

709   that can be analysed as biomarkers specific to regions of a genome. NGS also consists of several

710   techniques used today (Lemuth & Rupp, 2015). NGS genotyping technologies can often be an

711   alternative to the vastly more expensive and time consuming, whole genome strategies to look at a

712   study species. This is especially useful in studies involving species without publicly available genome

713   such as the wild relatives of *Linum* (Hu et al., 2021).

714   Peterson *et al* (2012) described one type of NGS genotyping technique as double-digest RAD

715   Sequencing or ddRADSeq. The protocol is a variation of the RAD sequencing protocol (Davey *et al*,

716   2010). DdRADSeq in comparison to RAD sequencing, uses a second restrictive digestion step to

717   improve the size selection step of the protocol (Peterson B. et al., 2012). This involves the use of a

718   second restriction enzyme and a second indexing step which allows for combinatorial indexing, more

719   specific to every individual (Peterson B. et al., 2012). When compared to a traditional RAD-

720   Sequencing technique, a double-digest RAD Sequencing can exclude regions which are flanked by a

721   very close or very distant recognition site. This then recovers libraries which consists of fragments

722   close to the specific target sequencing size (Peterson B. et al., 2012). The use of multiple enzymes

723   also contributes to a more diverse size selection availability in comparison to the traditional RAD

724   Sequencing technique (Peterson B. et al., 2012). In the study by Iguchi *et al* (2020), they were able to

725   identify diversity and selection mechanisms which leads to genetic variation in deep-sea amphipods

726 using a RAD sequencing protocol. This illustrates the significance of RAD sequencing for the use of

727 population diversity analysis, which is of interest when looking at our set of wild *L. bienne* samples.

728   ddRADSeq identifies Single Nucleotide Polymorphisms (SNPs) as genetic markers. SNPs

729 markers are simpler to genotype than microsatellites and are widely used in fine-scale population

730 genetic studies (Liu et al., 2019). The use of ddRADSeq have also been implied in herbarium samples

731 which have a relatively lower quality of preserved DNA, this protocol proves useful for obtaining

732 genetic datasets from a study species which is relatively fast to degrade such as plant leaf materials

733 (Jordon-Thaden et al., 2020). ddRADSeq have also been effective for genotyping organisms with a

734 larger genome size. A study using a non-model orchid species (genome size of 31.6Gbp) were viable

735 for ddRADSeq protocol, with inference of genetic diversity and differentiation (Gargiulo and Fay.,

736 2020). They also showed the requirement for a higher standard of quality (DNA extraction to library

737 preparation) when working with non-model organism of which their whole genome is not available

738 for reference (Gargiulo and Fay., 2020). This will prove as a challenge under the study model of this

739 thesis, as there is currently no whole genome publically available for the species *L. bienne*. This will

740 mean that any *de-novo* mapping of the extracted short sequences, will have to be of exceptionally

741 high quality. However, the use of *L. usitatissimum* whole genome is viable for reference to the later

742 mapping steps.

743   DdRADSeq uses indexing adaptors to "barcode" certain regions of a genome that have been

744 digested using two specific restriction enzymes. Subsequent bioinformatics analysis is then able to

745 highlight SNPs. The technique can be more cost effective than the previously developed SNP Chip

746 methods (Vieira et al., 2016). For any sequencing purposes, the first challenge is to extract a

747 sufficient amount of good quality DNA materials. There are multiple DNA extraction protocols

748 available to use. Nowadays, there are DNA extraction kits available to purchase such as Qiagen's

749 Dneasy kits. For ddRADSeq purposes, the most used protocol is a CTAB DNA extraction protocol

750 (Jordon-Thaden et al., 2020). CTAB DNA extractions can also be modified to target for high

751 polysaccharides, which may be the case in plant material extractions (Porebski et al., 1997). For

752 ddRADSeq purposes, other studies have used between 0.1 – 0.8 µg of total DNA per sample (Jordon-

753 Thaden, et al., 2020). This relatively high amount of DNA may represent a challenge in the extraction

754 of *Linum* DNA materials. During this study, optimisation of ddRADSeq protocols for *Linum* will be

755 performed and DNA extraction protocols will be explored in the interest of extracting relatively high-

756 quality DNA required to build ddRADSeq libraries. To build ddRADSeq libraries, one of the most

757 important step is to select restriction enzymes which are specified for digestion at associated length

758 (250-500bp for this study). This would enable short reads over many individuals which are

759 subsequently barcoded. For the barcoding process, a ligation process can be applied to the digested

760 DNA materials. Adapters (single indexing) can be ligated to the DNA materials at both ends of the

761 restrictedly cut DNA materials. Libraries can then be pooled and sequenced.

762 When successful, ddRADSeq outputs forward and reverse reads of the library. Reads can

763 then be mapped either *de-novo* or to a reference genome before further analysis could take place

764 (Peterson B. et al., 2012). In the interest of illustrating population structure and diversification within

765 our wild and cultivated samples, population genetic analysis is essential. Population genetic analysis

766 is often useful for genotyping studies involving samples from a wide range of localities. Population

767 genetic studies often utilise mathematical models to evaluate how a population varies in terms of

768 their genetics (Servedio et al., 2014). As part of an evolutionary biology study, protocols into

769 population genetics may infer adaptation, speciation, and structuring across populations which may

770 infer biological conclusion such as local adaptation, inbreeding depression, and potential transfer of

771 genetic material from one population to another in the sample set (Hoban et al., 2016).

772 Often local adaptation can lead to population genetic structures, with population forming

773 specific niche to the local environments (Leimu & Fischer, 2008). In plants this can be an important

774 mechanism in which biotic interactions can often lead to responses to environmental changes such

775 as elevational changes (Grassein et al., 2014). For *Linum,* observations of population genetics for the

776 wild relatives are scarce compared to wild relatives of other species such as the wild crop relatives of

777 cabbage, *Brassica oleracea*. Population genetics studies in *Brassica oleracea* have revealed

778 population structuring along Western Europe for their wild relatives (Mittell et al., 2020). The

779 structuring can suggest diversity in wild relatives of cultivars, which represent a wider genepool for

780 the improvement of cultivar relatives. While this is explored in the *Brassicaceae*, it hasn't been

781 explored in the wild relatives of flax, *L. bienne*. In another example, population genetic analyses

782 suggests gene flow and hybridization in 12 plant species suggesting extinction due to intogression of

783 cultivar types to the wild (Ellstrand N. et al., 2003). This adds to the relevance of population genetic

784 analysis, especially in illustrating genetic relation in wild and cultivar relatives of plants. In this study,

785 initial population structure analysis will look to answer the question of "Can genetic distance infer

786 population structuring within the Western wild samples of *L. bienne* and its cultivar relatives?" Other

787 relationships in population genetics will also be investigated for evidence of genetic diversity within

788 our samples.

789

790 ### 1.3.4 Genetic Expression in Flowering Time:

791 Within the model *Arabidopsis,* various environmental cues are known to affect the floral

792 regulatory network genes with light and temperature being the most significant (Li et al., 2015).

793   Timing of flowering in plants is often determined by environmental and endogeneous factors.

794   Environmental factors that have been shown to influence flowering time includes availability of

795   nutrients, ambient temperatures, drought, salinity, and the presence of pathogenic microbes (Cho L.

796   et al., 2017). Further, genetic expression of certain genes in the network affects other genes. As an

797   example, expression in genes such as *SOC1* have been shown to affect further genes in the pathway

798   such as *AP1* and *LFY*. There is an importance in quantitative input to each gene in the network for

799   expression in flowering times (Leal Valentim et al., 2015)*.*

800   To look at specific gene expression within plants, specific genes can be targeted using

801   primers and real-time qPCR studies (Higuchi et al., 1992). Primers are relatively short stretches of

802   nucleic acid sequences which can uniquely identify a specific region of the whole genome (Cox and

803   Doudna, 2015). They complement and amplify specific nucleic acids using apolymerase enzyme.

804   Therefore, these primers are useful for studies looking at specific gene expressions.

805   With additional information, and access to sequencing technologies; this thesis will aim to

806   better understand flowering responses in *Linum* by generating genetic expression data. In the

807   studies performed in this thesis, both *L. usitatissimum* and *L. bienne* genetic materials were

808   collected, tested, and analysed regarding differences in species and population localities. It is of

809   interest to find out whether there are differences in genetic expression of flowering time related

810   genes in response to environmental factors such as temperature. How would these changes in

811   genetic expression measure up against population localities, and climate within those localities?

812   Within this thesis, we hypothesize that differences in relative gene expression are present in our

813   *Linum* collection and that this is correlated to population localities and climatic variables, reflective

814   of local adaptation.

815

816   <u>1.3.5 Observation of Pollen Viabilities</u>

817   In addition to observing variation in flowering time due to response to the environment, we

818   wanted to measure other adaptation in flowering for *Linum*. Since most studies already looked at

819   hetero/homo styly, we approached this using the viability or *Linum* pollen. In addition to flowering

820   time, we investigated if temperature affects the number of pollen, pollen tube formation and its

821   ability to germinate the ovule. Pollens are an important part of a plant's breeding strategy and act as

822   a male counterpart to the ovule. Plants often show differences in pollen response related to

823   differences in breeding systems of said plants (Hanley et al., 2008). In addition, climate has been

824   observed to affect pollen performance in plants (Jan Ejsmond et al., 2011; Iovane M. et al., 2022).

825   Temperature acts as a major climatic factor which may limit the germination of pollen in plants. It is

826   deemed that temperature changes affects distribution of angiosperm species as the reproductive

827   stage is proven to be vulnerable to temperature. In the cultivated peach, it is observed that

828   increases of temperature to 20°C from their original 5.7°C in the field results in reduction in

829   receptivity of stigmas to pollens (Hedhly A. et al., 2005). They suggested that effects of temperature

830   on male and female organs of a flowering plants may be species-specific and could provide plants

831   with a level of plasticity to withstand environmental effects, such as ambient temperature, to make

832   sure they achieve fertilization (Hedhly A. et al., 2005). An older study in *Trifolium repense*, suggests

833   that growth of pollen tubes is affected by temperature and that pollen tube penetration into the

834   ovary is also observed to be negatively affected by colder and warmer temperatures (Chen and

835   Gibson, 1973). This effect in pollen viability due to temperatures could even be observed in plants

836   originating in warmer climates. In a species of groundnut (*Arachis hypogea*) it is observed that

837   warmer air temperatures (>35°C) results in failure of the setting of pods and pegs due to its lower

838   pollen viability. They found that for pollen germination and tube growth rate were less predictive in

839   discriminating genotypes for higher temperature than other parameters such as pollen tube length

840   and pollen germination rates (Kakani V. et al., 2002). This suggests plasticity in pollen viability

841   parameters for discriminating genotypes against temperatures in plants. In tomatoes (*Lycopersicon*

842   *esculentum*), it is suggested that temperatures affect *in-vitro* pollen germination and hybridization of

843   populations originating from different altitudes (Zamir D. et al., 1981; Maisonneuve and Den Nijs,

844   1984). Studies regarding pollen germination sensitivity to temperature stimulus can be achieved *in-*

845   *vivo* or *in-vitro*, with *in-vitro* procedures proving to be more elusive in plants (Boavida and

846   McCormick, 2007).

847   In several *Linum* species, it has been recently suggested that *in-vitro* germination is possible to

848   achieve up to 50% pollen germination rates (Lyakh and Soroka, 2021). However, they didn't include

849   the species *L. usitatissimum* or *L. bienne* within their study. Using the various sample populations

850   across western Europe that were in our collection, it is with interest to look at whether pollen

851   viability as pollen performance measure were affected under different treatments and whether

852   populations correlate with geographic and climatic variables, such as latitude, temperature, and

853   precipitation. We hypothesize that geographic and climatic variables affect pollen performance and

854   viability in our *Linum* collection and therefore its potentially affecting plasticity within population

855   fertilization strategy.

856   ## CHAPTER 2: FLOWERING TIME IN LINUM DUE TO VERNALIZATION AND

857   ## EXPRESSION OF FIVE FLOWERING TIME GENES.

858

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

874   Several studies have identified key flowering time genes that act as primary determinants of
875   the flowering-time network in the model *Arabidopsis thaliana* (Ballerini and Kramer., 2011; Welch S.
876   et al., 2004; Figure 1). Amongst these genes, the most rigorously studied are, the floral pathway
877   integrators such as FLOWERING LOCUS T (*FT*) and the SUPPRESSOR OF OVEREXPRESSION OF
878   CONSTANS (*SOC*), with FLOWERING LOCUS C (*FLC*) also studied (Sasaki *et al.*, 2017; Li *et al.*, 2015;
879   Valentim *et al.*, 2015). In the model *Arabidopsis,* it has been illustrated that, *FT*, and *SOC* loci
880   expression activates downstream flowering meristem identity genes such as LEAFY(*LFY)*,
881   APETALA1(*AP1*), SEPALLATA3 (*SEP3*) and FRUITFULL(*FUL*). This is because both *FT* and *SOC* are
882   considered transcription factor loci, and in the *Arabidopsis* flowering network, they are considered
883   as floral integrators (Welch S. et al., 2004; Lee and Lee, 2010). Furthermore, Ballerini et al (2011)
884   suggested that there is an element of conservation of *FT* homologs within flowering plants. Although
885   this is the case, they suggest that there is complexity in their regulation and evolution that is still of
886   interest in different angiosperm species to this day (Ballerini and Kramer., 2011). This suggests the
887   importance of *FT* as a floral regulator and the potential expression variation it may have in different
888   species, potentially due to environmental stimulus. In addition, there are two pathways that are

889   worth looking at in response to light and temperature regulation of *FT*. These are the Photoperiod

890   and Vernalization pathway. The expression of photoperiod genes CONSTANS (*CO*) and GIGANTEA

891   (*GI*) can activate *FT* due to photoperiodicity (Kurokura et al., 2017; Song Y. et al., 2014) and the

892   vernalization gene FRIGIDA (*FRI*) is responsible in determining variation of flowering time due to

893   vernalization requirements, defined as the requirement for seeds to sense cold to flower faster

894   (Shindo et al., 2005). Finally, the downstream meristem genes function as an initiator for floral organ

895   developments. Upregulation in these flowering meristem genes signals groups of meristem cells to

896   develop into flowering cells, instead of shoot cells (Teper-Bamnolker and Samach, 2005), thus

897   resulting in floral intitation after complex regulations of several genes in response to the

898   environment.

899
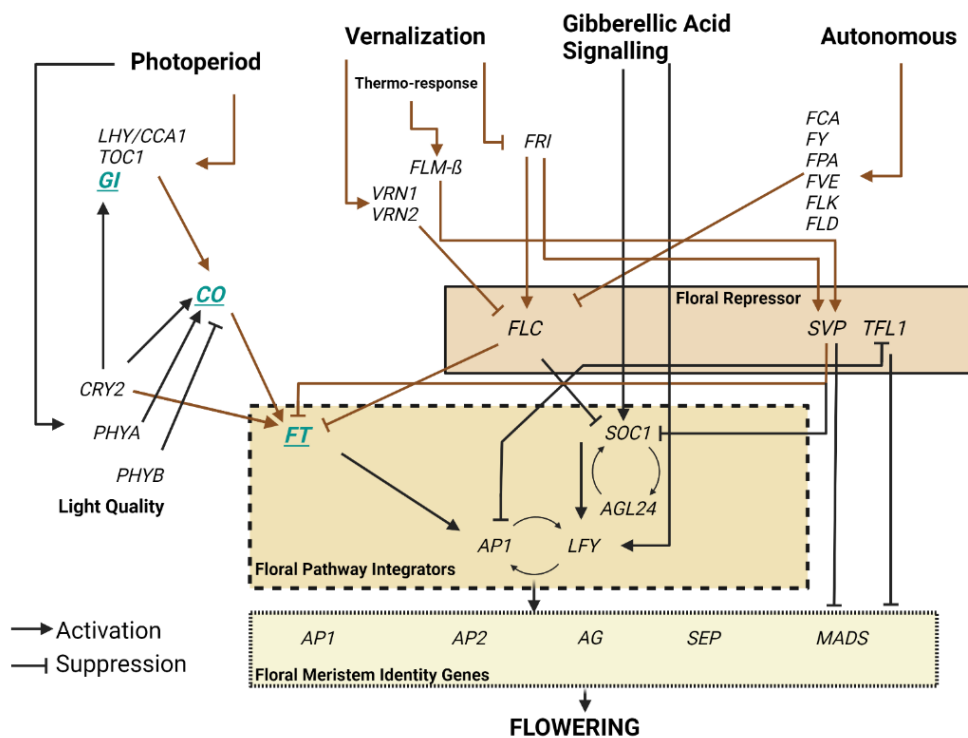
900   Figure 1. An overview of the flowering time network in *Arabidopsis thaliana* adapted from the literature (Sharma N. et al.,

901   2020; Leitjen W. et al., 2018; Chen et al., 2018; Ballerini and Kramer., 2011; Welch S. et al., 2004) Underlined are the genes

902   tested. We also highlighted pathways that affects *FT* (pathways in brown).

903

904   Comparative genomics have also been used to model flowering time pathways in other

905   species such as the temperate grasses (*Brachypodium distachyon*) (Higgins et al., 2010). They

906   suggest that flowering time pathways in *Brachypodium distachyon* are highly like the model

907   *Arabidopsis*. Reviews from Leitjen et al (2018) suggests that despite the overall conservation of these

908   flowering time network genes, there is evidence of divergence of flowering time regulation in both

909   the model *Arabidopsis thaliana* and several crop species due to environmental factors such as

910   temperature. With this, it can be assumed that different plant populations express varied responses

911   to flowering due to the environment. To what extent and which direction each gene is affecting each

912   plant population is still of research interest to this day, especially in non-model, cultivated plants. As

913   an addition, this variation due to flowering responses doesn't just occur in the model *Arabidopsis*.

914   Flowering initiation in non-model organisms such as the common sunflower (*Helianthus annuus)*,

915   have been suggested to be "exploited" by natural selection through genetic expression and tissue-

916   specific expressions of the *SOC1* genes affected by environmental factors. One of these factors is

917   photoperiodicity. Sunflower (*Helianthus annuus*) has shown diversity in photoperiodic responses,

918   due to changes in genetic expression modulated through natural selection (Blackman, et al., 2011).

919   Blackman et al, 2011, also suggests the occurrence of convergent evolution of photoperiod

920   responses in cultivated and wild types of sunflowers, which suggests phenotypic plasticity is weakly

921   constrained by genetic regulation (Blackman et al., 2011). In Japanese wild radish (*Raphanus*

922   *sativus*), recent research suggest that northern accessions were more sensitive to exposure to

923   prolonged cold conditions, in the sense that they require colder conditions to flower. This was

924   suggested by FLC expression, which is part of the vernalization pathway (see below). The Southern

925   wild radish population seem to be more sensitive to photoperiodicity, suggested by the expression

926   of photoperiodic genes (Han Q. et al., 2021). This is observed to correlate with multiple flowering

927   time gene expressions (Han Q, et al., 2021).

928   Vernalization is an adaptation, often found in temperate plants, which sense cold to allow

929   optimal timing of flowering initiation in spring (Kim *et al.*, 2009). Vernalization is the requirement for

930   seeds to be cooled from germination for individuals to flower earlier. In the model *Arabidopsis*

931   vernalization is widely studied. The vernalization network has been shown to progress around the

932   MADS-domain protein complex comprising of several flowering time genes forming complexes with

933   the MADS-protein SVP. One of these genes is the *FLC* (Flowering Locus C) which acts as a repressor

934   of *SOC1* (Chen et al., 2018). The closely related *FLM* (Flowering Locus M) also have been shown to

935   form complexes with the MADS-protein *SVP*. These proteins have shown to mediate response to

936   changes in ambient temperature by forming a complex with more specifically the β form of *FLM* at

937   lower temperatures. This in turn represses flowering at colder temperatures (Lee J. et al., 2013;

938   Chen et al., 2018; Posé et al., 2013). At relatively higher temperature, the MADS-protein *SVP*

939   degrades and the complex with *FLM-β* is not formed, thus producing reduced repressive complex,

940   allowing flowering initiation to proceed. Several vernalization related loci play an important part in

941   vernalization sensitivity and interact with other pathways leading to flowering initiation (Blázquez,

942   Koornneef and Putterill, 2001) (Chandler, Wilson and Dean, 1996). In this chapter, we will investigate

943 natural variation in flowering time responses for the temperate plant genus *Linum*. Both wild (*L.*

944 *bienne*) and cultivar types (*L. usitatissimum*) will be observed in terms of their phenotypic and

945 genetic responses to flowering initiation due to temperature changes in relation to vernalization.

946 This will imply potential adaptation in vernalization responses for both wild and cultivar types. The

947 main environmental variable tested will be temperature, as this corresponds more towards

948 vernalization requirements. However, expression of flowering time genes that may regulate

949 photoperiodicity were also observed.

950   Studies of *Arabidopsis thaliana* have also established that the latitude differences in

951 populations is associated with co-variation in growth and flowering time (Debieu et al., 2013a). In

952 respect of the wide distribution of *L. bienne* (as discussed in chapter 1), it is of interest to look at

953 differences in genetic expression from a flowering time perspective, based on phenotypic and

954 genotypic results. In *Linum*, vernalization requirement has been little studied, with studies mainly

955 focused on the cultivar, *L. usitatissimum* (Darapuneni *et al.*, 2014a). The wild relative (*L. bienne*) is of

956 interest, especially for questions regarding climates, as the wild population are likely to be locally

957 adapted to their environments (Landoni et al., 2022). This adaptation can be observed as clines of

958 flowering initiation times within the wild species. Several genes are of interest when looking at the

959 development of *Linum* against latitude which are part of the flowering genes network. *Gutaker et al.*

960 have previously identified the alteration and expression of the flowering gene *LuTFL1,* which could

961 reflect latitudinal adaptation and crop selection for fibre production in northern populations

962 (Gutaker *et al.*, 2019). However, other flowering time genes were not explored in this case. Other

963 flowering time genes could involve genes that are in the photoperiodicity network. Several flowering

964 time genes in cultivated *Linum* such as *LuCO* (CONSTANS) and *LuGI* (GIGANTEA) were previously

965 explored by Sun et al (2019) where they found variation in the expression of these genes under

966 several experiments with long days (22 °C/16 °C (12 h/12 h) for a 16 h/8 h photoperiod at a light

967 intensity of 300 µmol photons m− 2 s− 1) and short days (Similar to long days treatment but with

968 10 h/14 h photoperiod instead) which revolves around photoperiodicity (Sun *et al.*, 2019).

969   Under temperate wild *Arabidopsis* accessions, it has been recently observed that different

970 expression of genes of interest in this thesis correlates with the flowering time of the temperate

971 accessions (Kinmonth-Schultz *et al.*, 2021). The main aim for this chapter is to find out whether

972 expression of important flowering time genes, such as Flowering Locus C (FLC) and Constans (CO)

973 influence the flowering initiation of different, vernalized, and non-vernalized *Linum* populations

974 reflecting varying environmental conditions across the range. The hypothesis was that flowering

975 initiation requirements based on vernalization differ within the *Linum* collection dependent on

976 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
START			**FINISH**
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.		

1007   Figure 2. A workflow summary of the study methods in this chapter from the glasshouse to analyses in
1008   computing facilities.

1009

1010   To measure phenotypic and genetic differences between representative *Linum* populations;

1011   controlled growth of plant materials was conducted. Samples for this experiment were collected as

1012   described in sections "1.2.1 Linum usitatissimum: Cultivated flax" and 1.2.2 "Linum bienne: Wild

1013   Flax" of this thesis. Samples were collected from across Portugal, Spain, France, and the UK (Fig 3). A

1014   full list of the individuals used is available under appendix 2.

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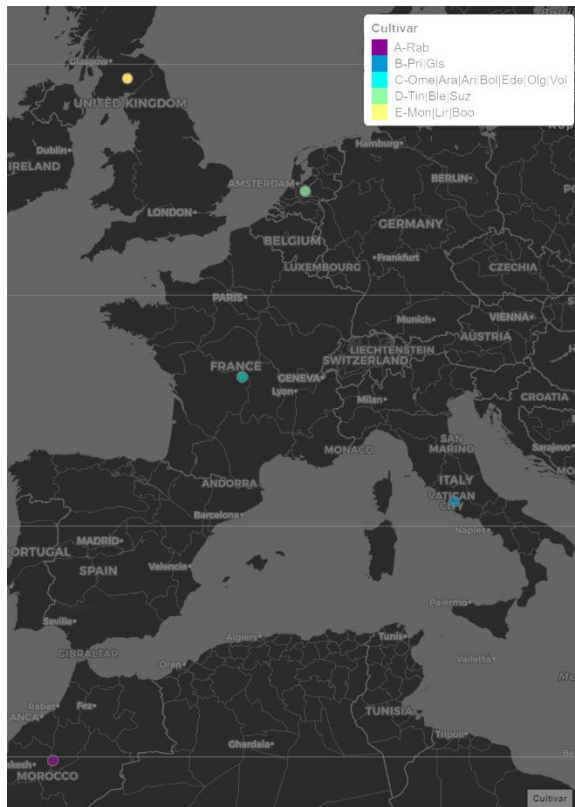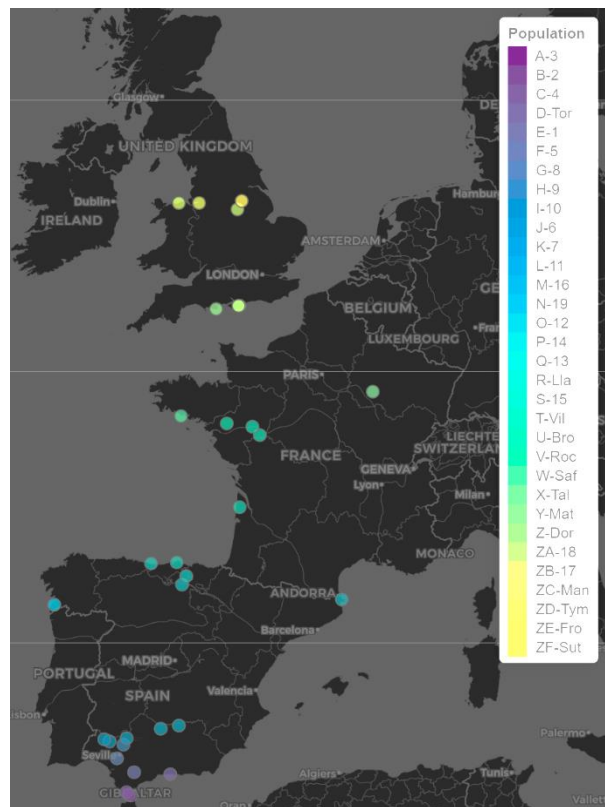
Figure 3. A map of the origin of collection of cultivars (A) and wild (B) *Linum*. Each population is alphabetically ranked by
their latitude, with the most Southern population ranked A. For cultivars, since latitude was only recorded by country level,
numerous populations belonging to the same latitude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

1034   from a collection of S1 and S2 seeds previously grown under experimental conditions based on no

1035   vernalization or under controlled glasshouses available in the department of Bioscience of Durham

1036   University. In total, there were 47 germinating populations from both vernalized and non-vernalized

1037   experiments. After 40 days of treatment, plants were morphologically examined by their flowering

1038   time, overall height, and number of stems when first flowering. Morphological measures between

1039   these traits will be examined in chapter 5 of this thesis. The vernalization conditions (see below)

1040   were setup in two Weiss Gallenkamp growth chambers model numbers A3655 and A3658. The

1041   conditions were kept the same throughout the experiment. The number of days to flower (flower

1042   initiation) was a point of phenotypic interest and genetic expression (RNA materials) as a point of

1043   interest for gene expression studies. Therefore, collection of RNA materials also took place at the

1044   corresponding first flowering of every individual.

1045   Preliminary measures for the S0 population grown in the controlled glasshouse were made

1046   before sowing of S1 generation for measures regarding vernalization which are plant height, stem

1047   numbers and bud numbers. The plants were exposed to vernalization conditions as follows; after

1048   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*

1049   *with 16:8 h light ratio*. After vernalization plants were kept at 24°C to 16:8 h light ratio indefinitely.

1050   The non-vernalized conditions were as follows; after sowing, *4°C (0-hour lights) for 72 hours, 22°C for*

1051   *10 days with 16:8 light ratio, 24°C for 40 days with 16:8 light ratio indefinitely*. Note the difference

1052   between the vernalization and the non-vernalized being the *4°C for 40 days cold treatment*. The

1053   lightings consist of numerous fluorescent tube lighting units, which were "Philips Master TL5 HE".

1054   These light tubes are 14 Watts in power requirements for each tube and emits 4000 Kelvin light

1055   temperature in colour by specifications. Plants from these S1 generations under vernalization

1056   treatments were also collected for RNA materials, for the genetic analysis purposes of this study. The

1057   phenotypic traits measured were divided into two categories: vegetative (stem number and plant

1058   heights) and reproductive (first day to achieve first flowering (flowering initation) and seed size),

1059   with particular interest in reproductive measures, due to the interest in flowering initiation. The

1060   phenotypic measures were measured at the time of first flowering of every individual (±1 day).

1061   Leaf materials were collected 14 weeks after sowing, and no more than three days after

1062   vernalization treatments have occurred. Leaf samples were collected from each treatment during a

1063   2-day window period from 11am to 1pm to avoid expression differences due to time of collection.

1064   Ten to twelve leaves from the top 2cm of the longest shoot on each plant were collected inside a

1065   1.5ml Eppendorf tube for each sample and were each labelled, and flash frozen in liquid N^2 before

1066   storing in the -80°C freezer. Two vernalization experiments were conducted during this study under

1067   different years (2019/20 and 2020/21). However due to restrictions on caused by the Covid-19

1068   pandemic, the 2020/21 vernalization experiment was cut short. Due to this, differences between the

1069   two experiments were unavoidable. These differences were quantified by analysing comparative

1070   analysis of the two experiments. This can be found under the "Results" part of this chapter.

1071   When considering locality of a wild population, its environment is the climate of the local

1072   area in terms of longitude, latitude, and altitude of the places where these wild populations were

1073   collected. Local climate was analysed based on different climate variables such as precipitation,

1074   temperatures, and sunlight availability that were extracted from over a 30-year period. The data was

1075   retrieved from WorldClim database. Data were available for precipitation (mm), solar radiation

1076   (kJ/day*m2), average temperature (°C), minimum and maximum temperature (°C), vapour pressure

1077   (kPa), and wind speed (m/s). Climate variables for principle component analysis (PCA) were used at

1078   30arcsec resolution over seasonal months (June July August for summer, September October

1079   November for autumn, December January February for winter and March April May for spring). To

1080   include all 6 dimensions of climatic variables, they were loaded into a principle component analysis

1081   with a principle component one (PC1) level explaining 63.8% of climatic variables (data available in

1082   appendix 3). Loaded value also suggests that lower PC1 values are associated with colder

1083   temperature associated with more Northern latitude. These data were used in a linked study of

1084   flowering time study (Landoni B. et al., 2022). All the climate variables were loaded in a collective

1085   PCA, whereby climatic PC1 values for each population were processed as the final representative

1086   values for all climatic variables. This would make further downstream analysis, regarding climate

1087   variables easier to process.

1088   When plotting PC1 and latitude of origin for each of our population, there is a visible

1089   relationship between the two variables, suggesting that lower PC1 values reflects higher latitudes.

1090   This can be seen in the plot below:
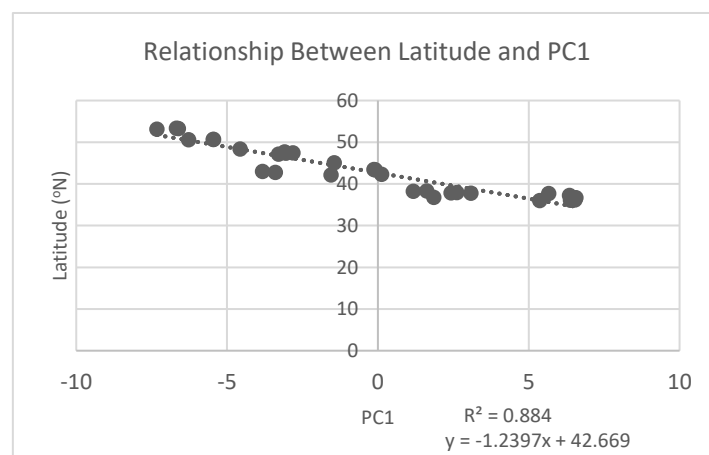
1091   Figure 4. Scatterplot to show negative relationship between climate variable representative (PC1) and latitude.

1092   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°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

1126   detectable RNA were diluted down to the same concentration as the lowest concentrated sample.

1127   This dilution was done using nuclease-free water accordingly. For the no template control, RNA was

1128   replaced with nuclease free water. 10µL was added per NTC. For the NRT, 5µL of random RNA

1129   sample with 6µL of nuclease free water were added.

1130   For the rest of the reactants, a master-mix were prepared in an Eppendorf tube per

1131   manufacturer's instructions. Master-mixes were vortexed for 10 seconds. After the addition of the

1132   master-mix ,1µL of Multiscribe Reverse Transcriptase were added into each sample and the NTC

1133   control. Samples were then centrifuged briefly and vortexed subsequently. After the addition of

1134   both master mix and Transcriptase, a thermocycling step can take place. The thermocycling steps for

1135   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	∞

1136   Table 1. Thermocycling steps for cDNA synthesis.

1137   After the thermocycling steps were finished, samples were stored in a 4°C fridge for short

1138   term (<72 Hours) and a -21°C freezer for long term storage(>72 Hours). The final Ct-value from the

1139   later real time qPCR protocol would serve as quality control. If quality of RNA from the extractions

1140   are satisfactory there was no need to measure cDNA output (Schmitz and Amasino, 2007).

1141

1142   <u>2.2.4 Expression of Genes of interest and House Keeping genes:</u>

1143   Housekeeping genes (HKGs) are required for maintenance of cells (Butte, Dzau and Glueck,

1144   2001). These genes are usually expressed relatively constantly under different conditions, which

1145   makes them suitable for comparison against the expression of other genes. Generally, HKGs

1146   constitute an important component for rt-qPCR procedures. However, a study suggested that

1147   expression of HKGs may change with cellular density of samples and that they could be affected by

1148   experimental conditions (Greer *et al.*, 2010a). Therefore, use of multiple HKGs within the same

1149   experiment can guard against these effects. *GAPDH* (Glyceraldehyde-3-phosphate dehydrogenase)

1150   and *UBI* (Ubiquitin) represent a couple of the commonly used HKGs in plants and therefore used for

1151   assessing different abiotic stress (Liang *et al.*, 2018a) (Carmona *et al.*, 2017). For *Linum,* the HKGs

1152   which will be used for this experiment are *LuGAPDH* and *LuUBI2.* This reference genes for *Linum*

1153   were developed by *Huis et al.* and are specifically useful for real time PCR (qRT-PCR) protocols (Huis,

1154   Hawkins and Neutelings, 2010). Our study will be incorporating these HKGs, as were also found on

1155   the study investigating photoperiodicity in *Linum* by *Sun et al* (Sun *et al.*, 2019)*.*

1156   We hypothesized that there would be variation in gene expression for flowering time genes

1157   according to latitude and due to vernalization. We tested flowering time genes already studied in

1158   *Linum usitatissimum* by Sun et al (Sun *et al.*, 2019). In our study however, we will be testing the gene

1159   expression based on vernalization stimulus and will be controlling for other environmental stimulus

1160   such as light availability. Whilst the study by Sun *et al* (2019) had found significant differences in

1161   expression of flowering time genes for *L. usitatissimum* under different light-length treatments, they

1162   have not tested vernalization as an environmental factor. In addition to this, they have not tested

1163   the wild relatives. It would be of interest to look at the expression of floral integrator gene, *LuFT,*

1164   along with the other photoperiodicity genes when differences in vernalization treatments is applied.

1165   This is because *Linum* is a temperate plant found across the sub-temperate into the temperate

1166   regions, thus different population may have adapted differently to vernalization as a stimulus to

1167   flowering initiation.

1168   In addition, the lineage leading to *L. usitatissimum and L. bienne* as a genus have been

1169   understood to have undergone a polyploidy event 20-40*mya* (Sveinsson *et al.*, 2014a). This has

1170   repercussions as there potentially are multiple copies (paralogues) of the same genes within *Linum*.

1171   This would mean there may be a need to test multiple copies of the same gene to see an effect

1172   which could be from one or multiple set of these copies. Wild and cultivars were tested for seven

1173   paralogue-specific flowering time genes using a modified real-time qPCR protocol. These genes were

1174   *LuFT1*(*Flowering Locus T 1*), *LuFT2*(*Flowering Locus T 2*), *LuGI1.1* (*GIGANTEA 1.1*), *LuGI1.2* (*GIGANTEA*

1175   *1.2*), *LuGI2* (*GIGANTEA 2*), *LuCO1*(*CONSTANS 1*), and *LuCO2* (*CONSTANS2*). A full list of nucleobases

1176   codes used for each primer can be found in appendix 4 (Sun *et al.*, 2019).

1177

1178   ## 2.2.5 Primer Testing

1179   The primers for the genes of interest were initially tested by performing PCR and analysing

1180   the products on an agarose gel (Jarman, Ward, and Elliott, 2002). Primers was tested using

1181   Promega's Go-Taq green master-mix; 2.5µl of each Forward and Reverse primers along with 2.5µl of

1182   *Linum* cDNA template for a 25µl total reaction volume. This would mean that the concentration of

1183   each primer is 1µM with < 250ng of cDNA template. We added 12.5µl of the Go-Taq green

1184   mastermix for a 1x concentration solution. The rest of the solution is made up with nuclease-free

1185   water up to 25µl. PCR mixes were treated in a thermal cycler using the following programme:

1186   Denaturation at 95°C for 3 minutes; Annealing with 35 cycles of 95°C for 45 seconds, 55°C for 30

1187   seconds, 72°C for 1 minute; and finally, an extension stage at 73°C for 5 minutes. Products were

1188   examined on a 2% TAE gel for PCR products. Products at <200bps with minimal amount of smearing

1189 observed were deemed acceptable. A decision was made to not use LuGI1.2 and LuGI2 primers

1190 based on inconsistencies with the gel such as smearing, and product sizes observed as > 200bps,

1191 leaving 5 primers (*LuFT1, LuFT2, LuCO1, LuCO2*, and *LuGI1.1* with the 2 housekeeping genes

1192 (*LuGAPDH* and *LuUBI2*).

1193

1194 ## 2.2.6 Quantitative Real-Time Procedure

1195   For real-time quantitative PCR reactions, an Applied Biosystem model 7300 real-time

1196 thermal-cycler was used with StarLab's 96-Well PCR Plates (96-Well PCR Plate, Skirted, Low Profile,

1197 White – STARLAB, 2022).

1198 Diluted cDNA was prepared for each sample. Primers were ordered from Integrated DNA

1199 Technologies (IDT). Each primer was aliquoted and diluted to 10µM working solution. To perform a

1200 real time qPCR, GoTaq qPCR SYBR Green were used (Promega Corporation). A master mix of the

1201 reaction components were then prepared as follows:

1202

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%

1205 *Table 2. The components required for the real-time qPCR reaction.*

1206

1207 SYBR Green Mix (included in the Promega qPCR GoTaq RT-qPCR system), and Nuclease free water

1208 were mixed by considering three repetitions per sample and 2 controls (No reverse transcriptase

1209 (NRT) and No template control (NTC)). For 14 samples we had (3×(14+2)) = 48 × 2 = 96. As there are

1210 possibilities for pipetting errors, 100 samples were considered for the SYBR and the water. SYBR and

1211 water were mixed in a 1.5ml Eppendorf tube and vortexed at 2000RPM for 30 seconds.

1212   The small volume of primers was often easier to be dispensed onto the sides of each well,

1213 this way loaded wells can be marked by the presence of the primer on the side to avoid

1214 contamination by other primers. After the primers were fully loaded, the plate was centrifuged

1215 down for 1 minute at 3000RPM at room temperature using the F2096 rotor in an Allegra X22

1216 refrigerated bench centrifuge. Making sure all the liquid has reached the bottom of the well, 8.2µL of

1217 the SYBR Green mix were loaded in with 5µL of each respective diluted cDNA samples following.

1218 Plates were sealed using an appropriate 96-well plate plastic sealer (we used Starlab's Self-Adhesive

1219 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:

1249
$$R - squared = \frac{Sum\ of\ squares\ due\ to\ regression}{Total\ sum\ of\ squares}$$

1250

1251   In this case the dilutions of primer tested were as follows; undiluted, 1/20, 1/200, 1/2000.

1252 After applying *Log* to the sample quantity, the coefficient of determination (slope) was calculated.

1253 The efficiency (%) was then calculated using the following equation described by Ginzinger

1254 (Ginzinger D, 2002):

1255
$$Efficiency\ (\%) = \left(10^{\frac{-1}{Slope}} - 1\right) \times 100$$

1256

1257   Efficiency values from the above equation are represented in percentage. However, for the

1258 relative gene expression calculations, the percentage values were converted into decimal values

1259 whereby a value of 2.00 will indicate 100% efficiency and 1.00 will indicate 0% efficiency. This

1260 conversion is essential for the final efficiency value input calculation in relation to ΔCt of each gene

1261 further downstream. Efficiency values were converted from % using the following equation

1262 (Ginzinger D, 2002):

1263
$$Efficiency\ Value = \frac{Effiency\ \%}{100} + 1$$

1264   Three primer efficiency tests were performed using real-time qPCR procedures for each

1265 respective primer but only the third test values used further for down-stream calculations as these

1266 seemed the best results from several tests, based on the efficiency (%) values closer to 90-110%. The

1267 results from the primer tests are noted in appendix 5.

1268 These primer efficiency values represent values which were inputted as a correction when

1269 calculating relative gene expression (RGE) of each gene tested. The efficiency value was calculated

1270 for each of the tested primers as well as the tested HKGs according to the separate ΔCt (avg Ct

1271 values – actual) for each gene and for each separate real time run (i.e repeats). The final Primer

1272 efficiency correction were calculated using this formula:

1273   *Final efficiency input = Efficiency value*^ΔCt for each gene

1274 The measured efficiencies values were not optimal as efficiencies of some primers are <95% as

1275 recommended (Miranda and Steward, 2017). Ideally these primers would be re-designed and re-

1276 ordered, but with the RNA materials in some cases already extracted and time limit considerations, a

1277 decision was made to lower the threshold efficiency to 85% to include all the tested primers were

1278 above the threshold.

1279

2.2.8 Real time qPCR of Vernalised and Non-vernalised Samples

1280

1281   Regarding the real time qPCR experiment described above, a set of samples that had

1282 experienced both vernalization treatments were chosen to compare flowering time gene expression

1283 differences. The final set of samples included 14 wild and 14 cultivar pairs each with samples from

1284 the non-vernalization and vernalization experiment. Some additional samples could not be utilized

1285 due to low quality RNA. Subsequently, cDNA from 2 samples at a time was added to 96 well plates

1286 designed to test the 5 Flowering time genes along with the 2 HKGs.

1287   Primer-sample combinations tested totalled up to 14 per plate with 3 repetitions each. The

1288 primer sequences are available under appendix 4. These, with the addition of No Reverse

1289 Transcriptase (NRT) and No Template Control (NTC) made up a 96-well plate.

1290

2.2.9 Clean-ups and quality controls:

1291

1292   As a pre-cursor to calculating relative gene expression (RGE), the Ct-values of the

1293 Housekeeping (HK) genes were checked for quality control. This was of importance because the

1294 observed primer-efficiency values were not ideal (87.273% for *LuGAPDH* and 51.464% for *LuUBI2*).

1295 Individual Ct values were removed if either; Ct-values were either too high (>29 cycle-threshold

1296 values) or too far away from the other values (i.e >1 Ct variation between the triplicates) for

1297 LuGADHP; 51/144 Ct-values were either too far away from other Ct-values or too high. These steps

1298 addressed potential experimental errors such as poor qualities and low quantities of cDNA and

1299 pipetting errors, which will impact the real-time qPCR quantification process.

1300   When applying the delta-delta Ct ($2^{-\Delta\Delta Ct}$) method to the final relative gene expression values,

1301 all initial samples were included, except for one wild sample (Saf_10) for both Flowering Time genes.

1302 This sample was excluded because of the much bigger relative gene expression values that make it

1303 an extreme outlier. This could be due to experimental errors such as the quality of RNA.

1304

2.2.10 Relative Gene Expression: Quantification

1305

1306   RGE data were processed in Microsoft Excel. For quantification purposes, a template was

1307 built for primer efficiency using formula stated under section 2.2.7 "Primer Efficiency" part of this

1308 chapter. In addition to these, further calculations are required to calculate relative gene expression.

1309 The Ct results for each gene of interest (GOI) was added to a Microsoft Excel template and are

1310 specified to the same House Keeping Genes (*GAPDH*) Ct-values for each respective samples. To

1311   calculate relative gene expression, a difference (delta) of Ct values were calculated between a

1312   reference house-keeping gene and the gene of interest based on previous method described by

1313   *Livak et al* (Livak and Schmittgen, 2001a). There are two methods which are available to explore

1314   using one HKG or using two HKGs. Both are described below.

1315   The default formula for calculating relative gene expression differences is the delta-delta Ct

1316   ($2^{-\Delta\Delta Ct}$) method. This method uses only one HKG. This is where, for each gene and sample, average

1317   Ct-values for the gene of interest (GOI) and the housekeeping gene (HKG) are calculated. Differences

1318   (delta-Ct) between the gene of interest and the housekeeping gene Ct-values were calculated and

1319   then the delta-delta Ct (ΔΔCt) were calculated using a calibrator value (i.e., *delta-Ct – calibrator*

1320   *value*). The choice calibrator value differs depending on experimental design but in this experiment,

1321   the average value of all the non-vernalized treatment ΔCt for the gene of interest was used as the

1322   calibrator value for that specific gene. This is an important factor to the quantification of RGE, as the

1323   choice of calibrator will impact the final gene-fold value (Schmittgen and Livak, 2008).

1324   ΔCt for a specific gene, using one HKG was calculated using (Rao X *et al*, 2013):

1325 $\Delta Ct = Ct(a\ Gene\ of\ interest\ (GOI)) - Ct(a\ reference\ gene\ (HKG))$

1326   ΔΔCt are then able to be calculated using (Rao X *et al*, 2013):

1327 $\Delta\Delta Ct = \Delta Ct(a\ Gene\ of\ interest\ (GOI)) - \Delta Ct(a\ reference\ gene\ (HKG))$

1328   The formula above can be applied to a specific GOI with a specific HKG which are constant for the

1329   entire experiment (i.e GAPDH only). Using this method, a fold-change of the target gene (either

1330   controlled (non-vernalized) or treated (vernalized) can be calculated, normalized, and related to a

1331   HKG using calibrator values as described above. The method was fully described and updated in a

1332   data analysis by Rao *et al* (Rao X et al., 2013).

1333   Although the delta-delta Ct method is the most used, the use of multiple house-keeping

1334   genes is regarded as more reliable in giving a background HKG expression against which to calculate

1335   relative gene expression (Riedel *et al.*, 2014a) (Manoli *et al.*, 2012). The variation that may occur

1336   between the multiple different house-keeping genes is resolved by normalization steps (Huggett *et*

1337   *al.*, 2005). According to *Vandesompele et al,* 2002*,* the normalization is done by geometric averaging

1338   of the multiple house-keeping genes involved as controls (Vandesompele et al., 2002; Riedel et al.,

1339   2014). The equation for using multiple reference genes is as follows:

1340 $$Relative\ Gene\ Expression = \frac{(E_{GOI})^{\Delta CtGOI}}{GeoMean[(E_{REF})^{\Delta CtREF}]}$$

1341

1342   The delta-Ct (ΔCt) values for each of the tested samples were determined using the following

1343   equation:

1344 $$\Delta Ct = Control\ Ct - Treated\ Ct$$

1345   The treated Ct values in this case will be the Ct values from the vernalized samples and the control

1346   from the non-vernalized. The PCR performed consisted of 2 reference genes (*LuGAPDH* and *LuUBI2*)

1347   with a gene of interest from one of our 5 chosen loci. Using the equations above and the primer

1348   efficiency values gained previously as formulated using methods described under section 2.2.5

1349   "Primer Testing", it is possible to compare relative genetic expression between treated (vernalized)

1350   and control (non-vernalized) individuals in our *Linum* sample set.

1351

1352

1353   2.3 STATISTICAL ANALYSES FOR RELATIVE GENE EXPRESSION

1354   Relative Gene Expression (RGE) is calculated using the $2^{-\Delta\Delta Ct}$ formula as described in section

1355   2.2.10 "Relative Gene Expression: Quantification" above. For follow-up analysis, various packages

1356   were used in R (R Core Team, 2022) ran under the R graphical interface of Rstudio (Rstudio Team,

1357   2020). ΔΔCt can be calculated by creating a template using the formula to calculate ΔCt (see section

1358   2.2.10) for each of the genes that were tested, including the HKG. The template was created in

1359   Microsoft Excel. the raw data were further analysed in R.

1360   For comparison between the different experiments that took place during different years

1361   (2018 and 2021), an F-test (Box G, 1953) was applied in R. For genetic expression analyses, templates

1362   for the calculation of Ct-values (see "Relative Gene Expression: Quantification") were made and bar

1363   charts can be plotted in Microsoft Excel. After values for Relative Gene Expresisons (RGE) were

1364   gathered for each respective locus, species and treatment, data could be modelled against

1365   environmental variables.

1366   RGE differences of vernalized and non-vernalized treatments, and wild or cultivar

1367   combinations was tested for each gene using a 2-sample, 2-tailed unequal variance t-test. This test

1368   compared fold-gene expression of the gene of interest between the two treatments (vernalized and

1369   non-vernalized). The tests were performed in Microsoft Excel. The formula used for the t-test is

1370   *"=TTEST(array1,array2,2,3)"*. RGE differences were visualised using R package ggplot including the

1371   command: *"+stat_compare_means(label = "p.signif", method = "t.test",ref.group = ".all.")"*

1372   (appendix 6).

1373   The differences of flowering initation between vernalized and non-vernalized individuals

1374   were correlated against latitude of origin (^{o}N) using Pearson's coefficient of determination (r^2) in R.

1375   This analysis measured how correlated flowering initiation is to latitude in each data sets (wild and

1376   cultivated; Steel R. *et al.,* 1960). P-values and r^2 values explained by the models were visualised

1377   within scatter plots of the data, showing lines of regression using the R-package "ggplot2" (Wickham

1378   H.*,* 2016). The full command used for drawing correlation coefficients for this analysis can be found

1379   under appendix 6. To further analyse these relationships using modelling, a Linear model analysis

1380   was able to run using R. The command can also be found under appendix 7. For modelling against

1381   environmental variables such as "latitude" and "climate" (described under the "Study and Methods"

1382   part of this chapter), a GLM (General Linear Modelling) approach was used. Various commands used

1383   for the GLM analysis in R can be found under appendix 7.

2.4 RESULTS: FLOWERING TIME BETWEEN EXPERIMENTS FROM DIFFERENT

1385 YEARS

1386

1387   The vernalization results were divided into three different categories: "non-Vernalized
1388 2018", "Vernalized 2018" and "Vernalized 2021". Due to COVID-19 pandemic, vernalization
1389 experiment in 2021 had to be terminated early in March 2020. Consistencies between the different
1390 years couldn't be kept the same and a bias towards earlier flowering plants can be observed in the
1391 2021 experiment as a result.

1392   Differences between the 2018 and the 2021 vernalization experiments was observed when
1393 looking at variation in number of days to flower. This was due to bias towards earlier flowering
1394 plants in the 2020-2021 replicate as the experiment was stopped prematurely (due to Covid-19
1395 restrictions), resulting on limited data collection for late flowering plants. Statistical comparison of
1396 the number of days to flower (flowering initiation) between the two vernalization experiments,
1397 under the different years found significant differences between "vernalized 2018" and "vernalized
1398 2021" experiments (*F-test, F=2.523, df=196, p=<0.005)* (Figure 5A). When data were filtered to the
1399 same individuals present in both experiments, the difference between the two datasets remained,
1400 with biases for earlier flowering individuals expressed *(2-sample, 2 tailed t-test p=<0.001)* (Figure
1401 5B). The significant variation between flowering initiation under the 2018 and 2021 vernalization
1402 experiment suggests that at least for phenotypic measures, data from the 2021 vernalization
1403 experiment should be excluded to avoid biases towards earlier flowering plants.
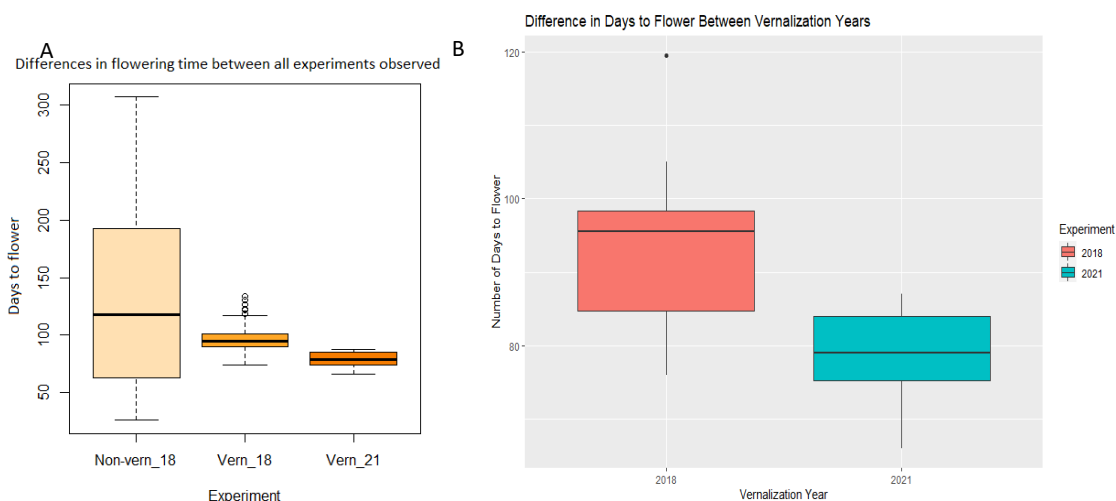
Figure 5A. A boxplot showing flowering time differences between all the individuals tested across the 2018 non-vernalization, the 2018 vernalization and the 2021 vernalization experiment (F-test, F=2.523, =196, p=<0.005). Figure 5B. A boxplot representing flowering time of a subset of all the mutual 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.
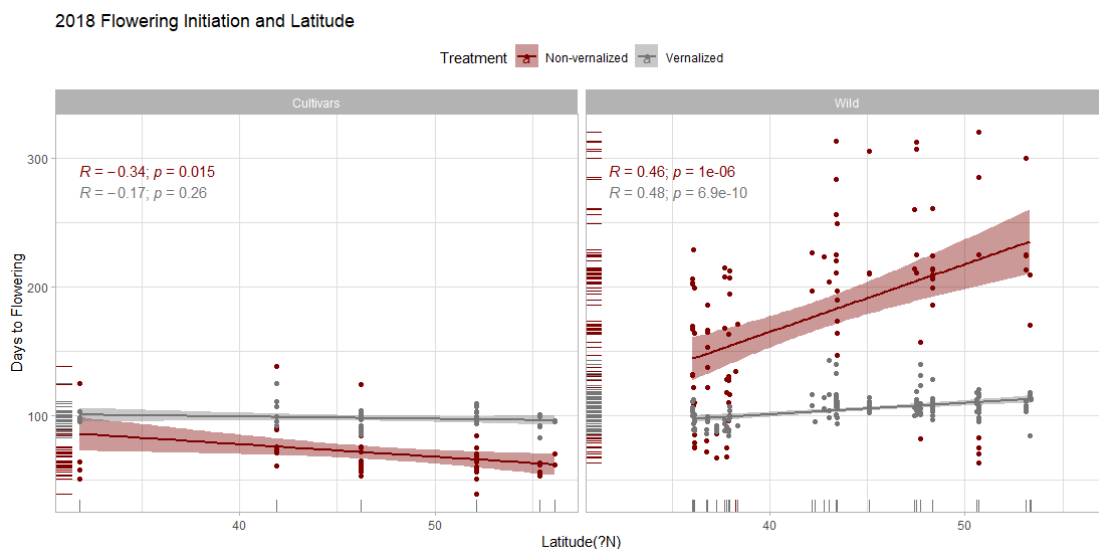
Figure 6. Scatterplots with regression line and linear models for flowering initiation to latitude. Note the differences because of latitude and vernalization in the number of days to flowering between wild and cultivated and samples. Reduction of more than 100 days to flowering initiation is observed 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.

1430

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).

1441

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).

1443

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

1456   delta-delta ct ($2^{-\Delta\Delta Ct}$) method would be further applied for the rest of the analysis regarding relative
1457   gene expression.

1458

1460
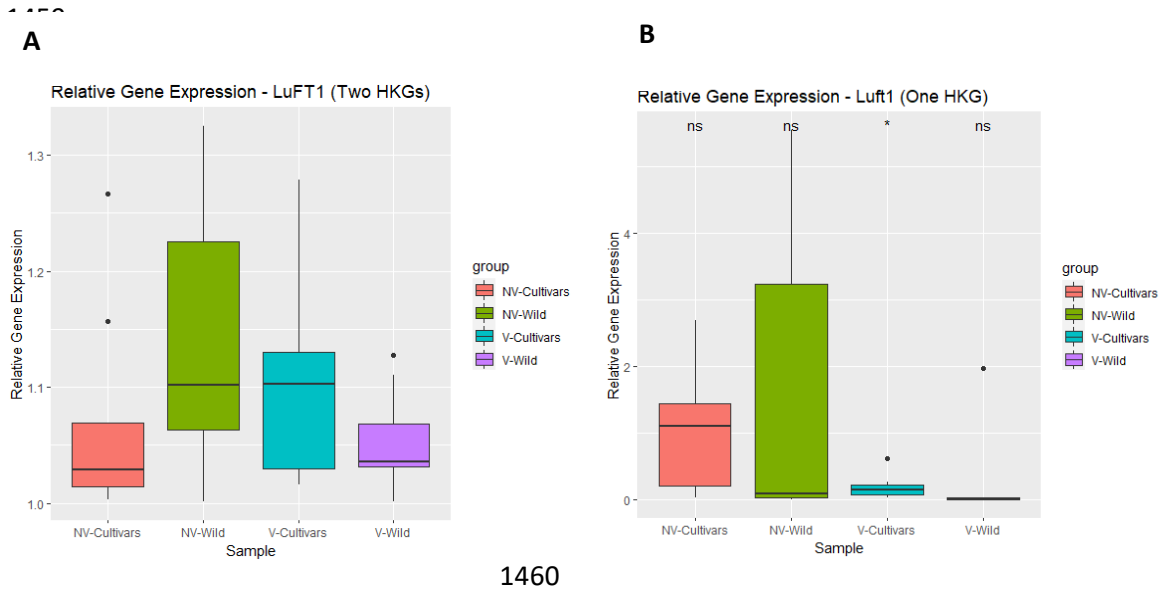
1461   Figure 7. Boxplots comparing Relative Gene Expression for *LuFT1* using two (Figure 6A) or just one (Figure 6B) house-
1462   keeping genes. In Figure 5B the relative gene expression of *LuFT1* is found to be significantly different (marked by *) to the
1463   other treatments (p = <0.05, multiple pairwide t-test).

1464

2.4.4 Results: Relative Gene Expression in Response to Vernalization

1465

1466   Relative gene expression results were initially sorted into groups reflecting treatments and
1467   species (non-vernalized-cultivated, vernalized-cultivated, non-vernalized-wild, and vernalized-wild).
1468   Relative gene expression ($2^{-\Delta\Delta Ct}$) was compared across each of the treatments to see whether any
1469   species/treatments show significant differences in relative gene expression respective of each gene
1470   for the treatment. Boxplots can be illustrated with a Bonferroni-corrected multiple pairwise t-test to
1471   reveal significance between treatments (Figure 8).

1472   Variation in expression was mostly seen in cultivars with a potential down-regulation in all
1473   the tested genes when cultivars were vernalized (Figures 7 A-E). Less strong RGE differences were
1474   seen in the wild samples. In the case of the comparisons between treatments, most of the
1475   differences in the data were observed when treating cultivars to vernalization. This is interesting as
1476   phenotypically; this difference wasn't to be expected. This is because flowering initiation in the
1477   cultivars did not show faster flowering initiation in response to vernalization and instead showed
1478   slightly slower flowering initiation (Figure 4). The multiple pairwise t-test however, was comparing
1479   differences of relative gene expression when vernalized, against all wild and cultivar individuals

1480   which are perhaps behaving very differently in terms of expressing each of these genes and maybe

1481   independent of each other. Therefore, variation among individuals was explored in more detail.

1482   Comparisons between expression responses (relative gene expression) against individuals with

1483   different treatment were undergone using a paired t-test as in the next part of this chapter.

1   A
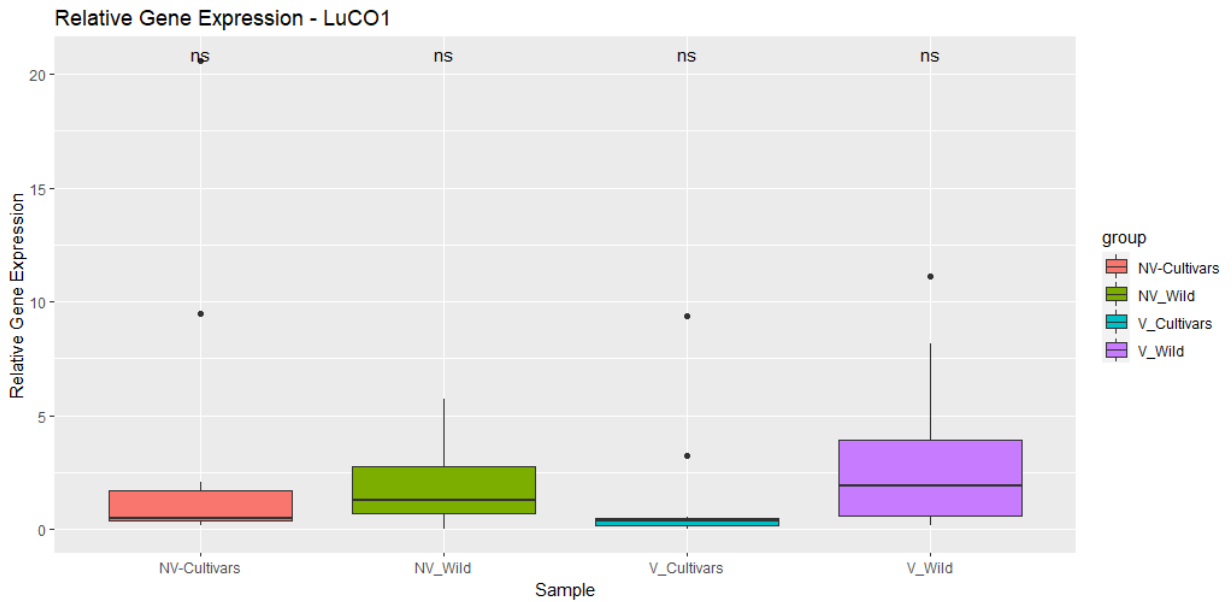
1485   Figure 8A. Relative Gene Expression (RGE) of samples and treatments for the locus *LuCO1*. The significance of multiple

1486   pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot. The labels "ns" means no-

1487   significant differences were observed between 4 observed treatments.

B
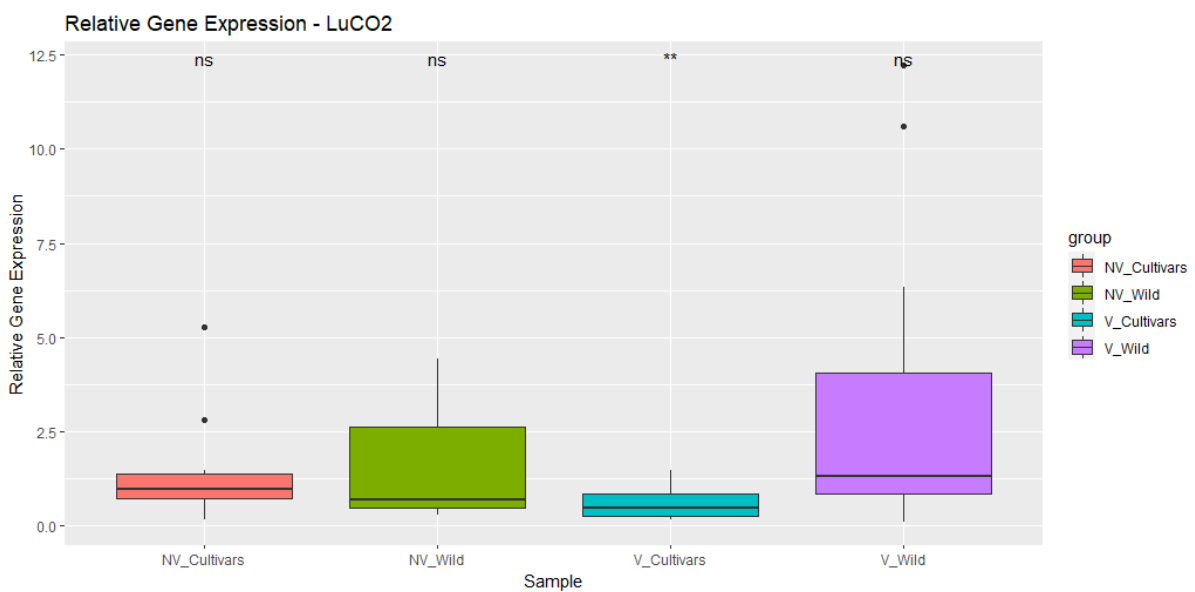
1488

1489

1490   Figure 8B. Relative Gene Expression (RGE) of samples and treatments for the locus *LuCO2*. The significance of multiple

1491   pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot. "Vernalized cultivated"

1492   ("NV_Cultivars") show a significant RGE variation when compared to other treatments ("**" means p= <0.01).

1493
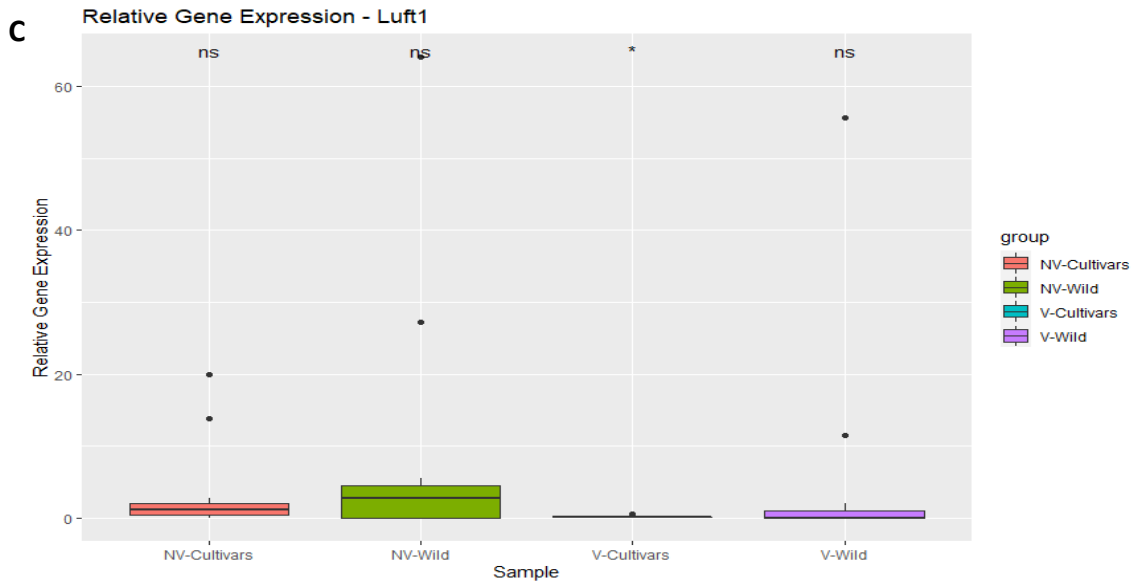
1494   Figure 8C. Relative Gene Expression (RGE) of samples and treatments for the locus *LuFT1*. The significance of multiple

1495   pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot . "Vernalized cultivated" (V-

1496   Cultivars) shows a significant RGE reduction ("*" means p= <0.05).

1497
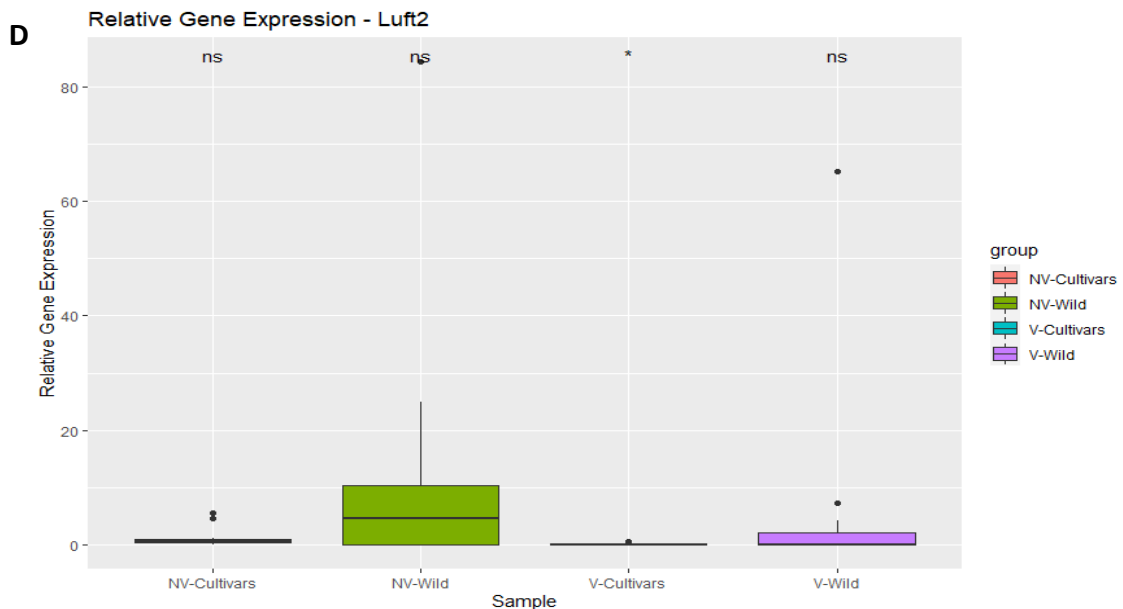
1498   Figure 8D. Relative Gene Expression (RGE) of samples and treatments for the locus *LuFT2*. The significance of multiple

1499   pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot . "Vernalized cultivated" (V-

1500   Cultivars) shows a significant RGE reduction ("*" means p= <0.05).

1501

1502

E Relative Gene Expression - LuGI1.1
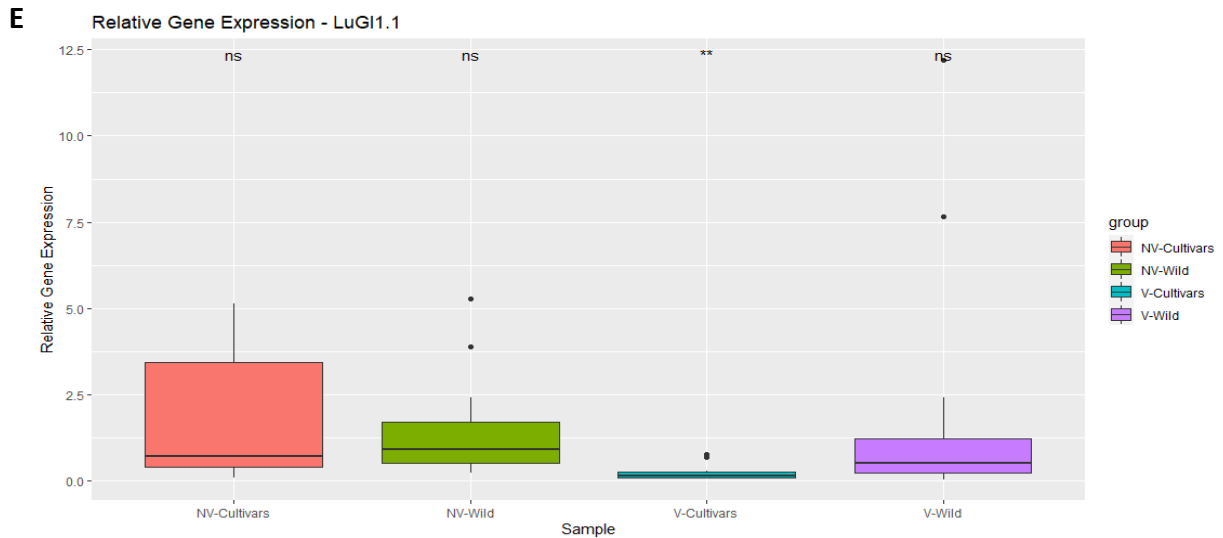
1503 Figure 8E. Relative Gene Expression (RGE) of samples and treatments for the locus *LuGI1.1*. The significance of multiple

1504 pairwise t-tests, adjusted with Bonferroni correction, are shown at the top of each boxplot . "Vernalized cultivated" (V-

1505 Cultivars) shows a significant RGE reduction ("**" means p= <0.01).

1506

1507   Variation in expression was mostly seen in cultivars with a potential downregulation in all

1508 the tested genes when cultivars were vernalized. Less strong RGE differences were seen in the wild

1509 samples. In the case of the comparisons between treatments, most of the differences in the data

1510 were observed when treating cultivars to vernalization. This is interesting as phenotypically; this

1511 difference wasn't to be expected. This is because flowering initiation in the cultivars did not show

1512 faster flowering initiation in response to vernalization and instead showed slightly slower flowering

1513 initiation (Figure 6). The multiple pairwise t-test, however, was comparing differences of relative

1514 gene expression when vernalized, against all wild and cultivar individuals which are perhaps

1515 behaving very differently in terms of expressing each of these genes and maybe independent of each

1516 other. Therefore, variation among individuals was explored in more detail. Comparisons between

1517 expression responses (relative gene expression) against individuals with different treatment were

1518 undergone using a paired t-test as in the next part of this chapter.

1519

1520 2.4.5 *LuGI1.1 (GIGANTEA)* Expression Response

1521   We tested fold-change differences (delta-delta ct) as a relative gene expression (RGE)

1522 measure to investigate whether differences in gene expression under vernalization treatment can be

1523 observed. To be more conservative, since we have two treatments compared with 5 genes, our

1524 alpha needs to be adjusted using Bonferroni's correction to 0.005 (0.05/10). Overall differences in

1525 *LuGI1.1* expression in cultivars was the closest to approaching significance, however, are not

1526 significant when considering Bonferroni's correction *(paired, 2-tailed t-test, p = 0.012 (p > α), t-stat =*

1527 *3.043, t-Critical two-tail= 2.228)*. The greatest differences were observed in cultivars EDE, BLE, PRI,

1528 GIS, and MON (Figure 9).

1529   There was no significant difference expression of *LuGI1.1* between treatments in wild

1530 samples *(paired, 2-tailed t-test, p = 0.407, t-stat = -0.864, t-Critical two-tail= 2.228)* (Figure 10).

1531 Expression varied more among samples with observations of both up and down-expression of

1532 *LuGI1.1*. Samples Lla_A and Tor_4 increased *LuGI1.1* expression the most and vernalized sample

1533 Tal_4 decreased expression the most.
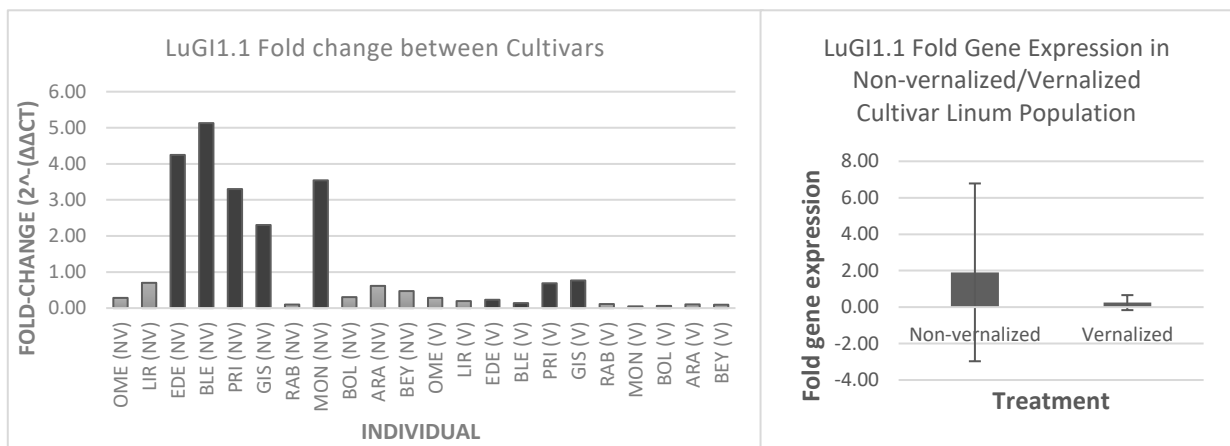
1534 Figure 9. Bar charts showing *LuGI1.1* fold change between cultivars under non-vernalized and vernalized treatments

1535 *(paired, 2-tailed t-test, p = 0.012 (p=>0.005))*. The cultivars showing the largest downregulation differences are shaded in

1536 darker grey.

1537
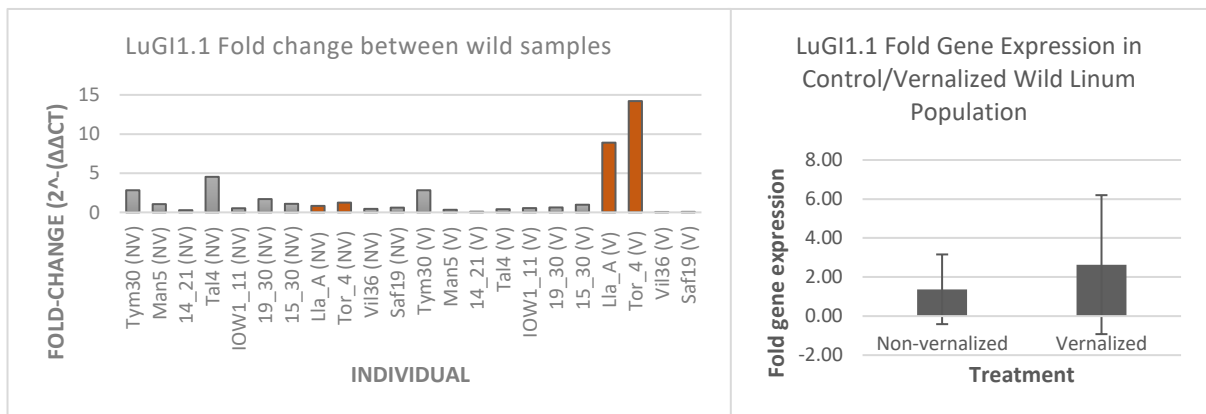
1538 Figure 10. Bar charts showing *LuGI1.1* fold change between wild samples under non-vernalized and vernalized treatments.

1539 Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression of

1540 *LuGI1.1* when vernalized, respectively.

1541

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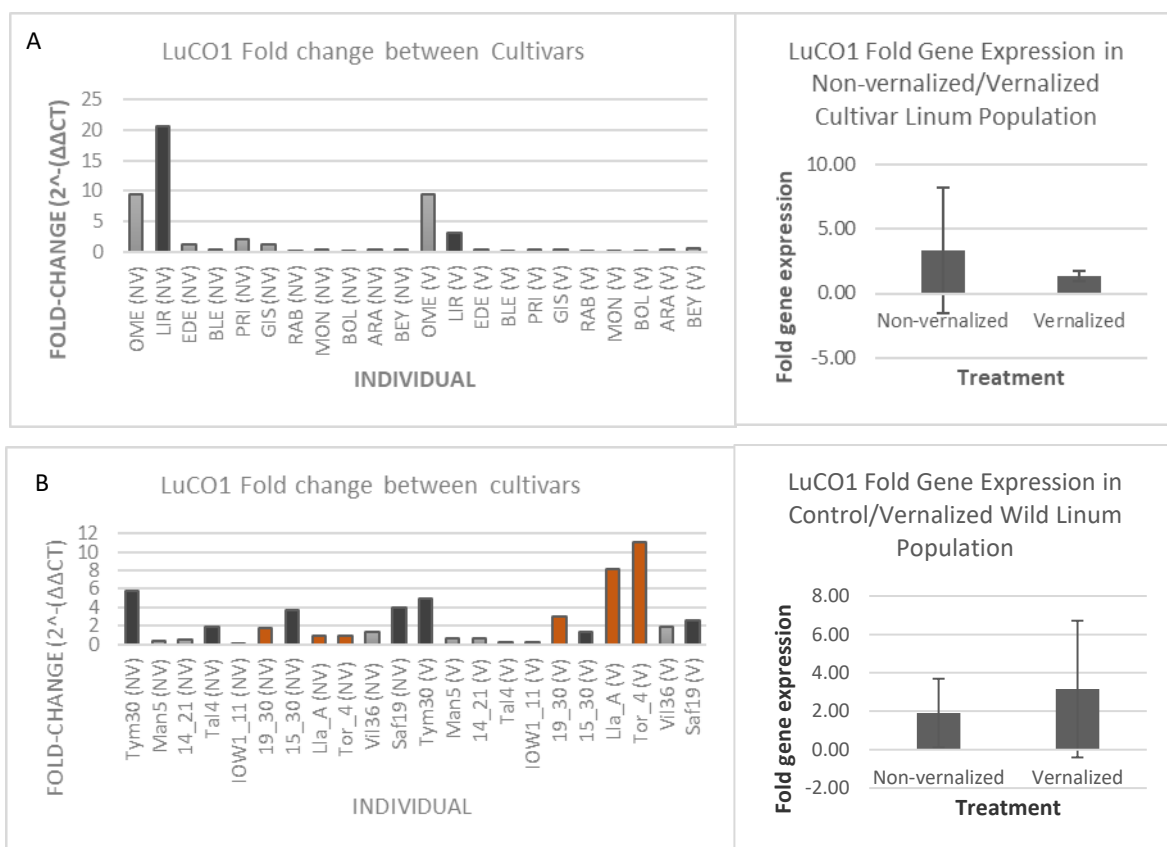
Figure 11. Bar charts showing LuCO1 fold change between cultivars (A) and wild types (B) under non-vernalized and vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and increase expression when vernalized, respectively.

1559   There were non-significant differences in *LuCO2* expression between the non-vernalized and

1560   the vernalized treatments for cultivars *(paired, 2-tailed t-test, p = 0.132, t-stat = 1.640, t-Critical two-*

1561   *tail= 2.228)*, and wild samples *(paired, 2-tailed t-test, p = 0.222, t-stat = -1.301, t-Critical two-tail=*

1562   *2.228)* (Figure 12). The trend was for *LuCO2* expression to decrease in cultivars during vernalization,

1563   while the trend for wild samples was to increase *LuCO2* expression. This revealed that although

1564   there were overall significant differences when comparing vernalized cultivated samples gene

1565   expression for *LuCO2*, when testing between vernalized and non-vernalized cultivated samples only,

1566   the expression of *LuCO2* shows no significant difference and results to vernalization having no

1567   significant effect in the relative gene expression of *LuCO2*, for both cultivar and wild samples.

1568
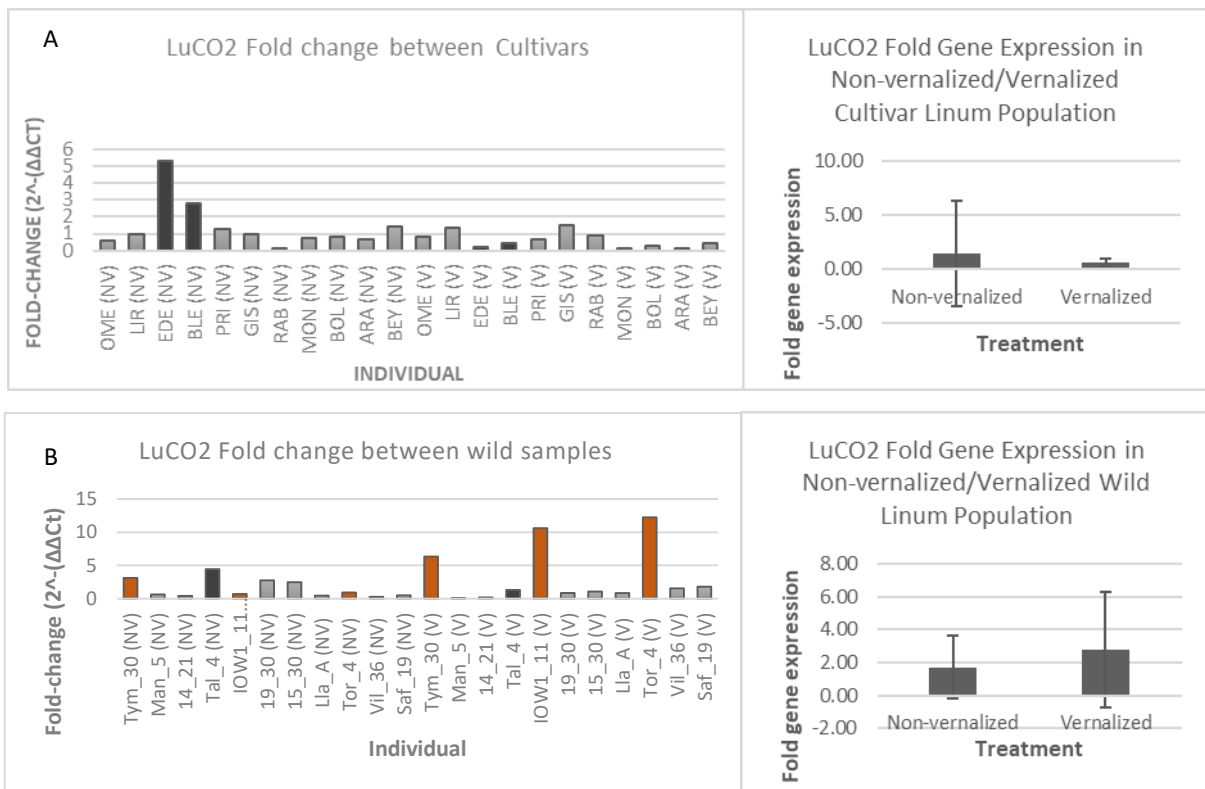
1569   Figure 12. Bar charts showing *LuCO2* fold change between cultivars (A) and wild types (B) under non-vernalized and

1570   vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and

1571   increase expression when vernalized, respectively.

1572

1573

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

1575   The cultivars showed a trend of decreased expression of *LuFT1* when vernalized. This was

1576   seen more in cultivar samples EDE and BLE, alike the *LuCO2* results. The overall change in gene

1577   expression was not significant *(paired, 2-tailed t-test, p = 0.093, t-stat = 1.856, t-Critical two-tail=*

1578   *2.228)* (Figure 12A). Wild samples showed a mixed pattern with one potential anomaly for sample

1579   Saf_10 (not shown) with much greater *LuFT1* expression than other wild samples across both

1580   treatments. There was no significant expression change in *LuFT1* expression when wild samples were

1581   vernalized *(paired, 2-tailed t-test, p = 0.669, t-stat = 0.440, t-Critical two-tail= 2.228)* (Figure 13B).

1582
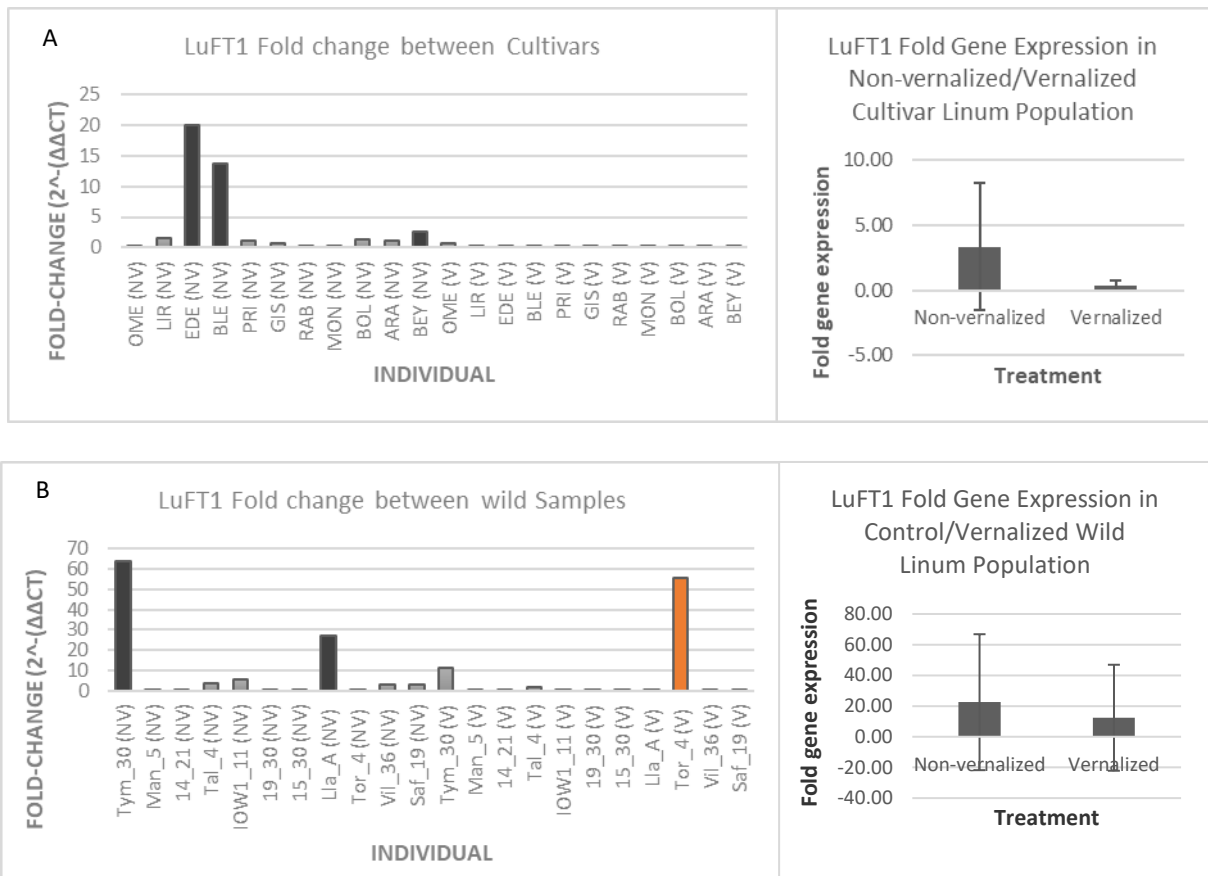
1583   Figure 13. Bar charts showing *LuFT1* fold change between cultivars (A) and wild types (B) under non-vernalized and

1584   vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and

1585   increase expression when vernalized, respectively.

1586

1587   As was observed for *LuFT1*, cultivar samples EDE and BLE showed the most reduced *LuFT2*

1588   expression after vernalization with non-significance on the fold change between the two treatments

1589   for the entire cultivar sample set *(paired, 2-tailed t-test, p = 0.087, t-stat = 1.916, t-Critical two-tail=*

1590   *2.262)* (Figure 14A). Among the wild samples, an exceptionally high downregulation in expression of

1591   *LuFT2* (>100 fold difference, not shown) was observed sample Saf_10. This sample was treated as an

1592   anomaly. Nearly the same trends as *LuFT1* were seen, with sample Tor_4 showing the most increase

1593   in expression of *LuFT2*. However, the results between the two treatments for the wild sample set

1594   were still not statistically significant altogether *(paired, 2-tailed t-test, p = 0.575, t-stat = 0.579, t-*

1595 *Critical two-tail= 2.228)* (Figure 14B). Relative gene expression was also tested between species in

1596 only the vernalized samples using a paired, 2-tailed t-test. None of the tested flowering time genes

1597 have shown significant expression difference between species when vernalized (α = >0.05).

1598
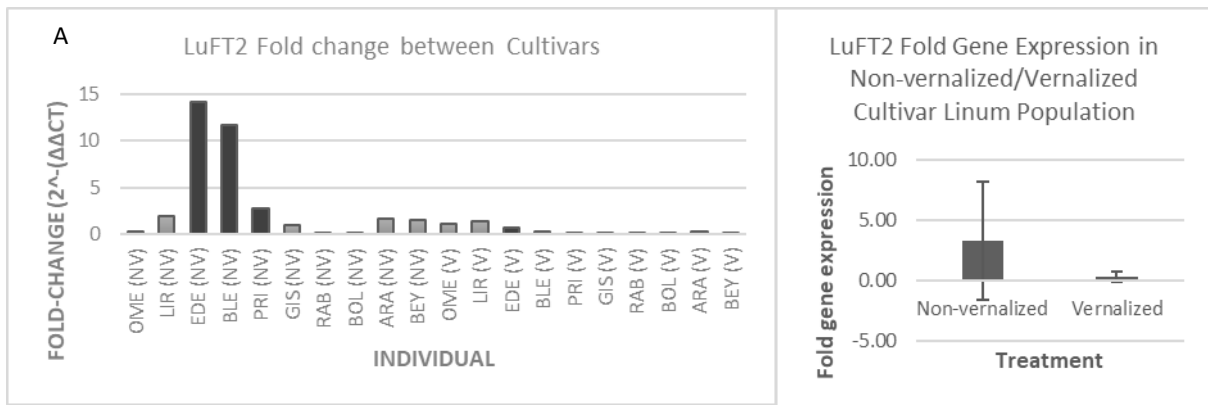
1599
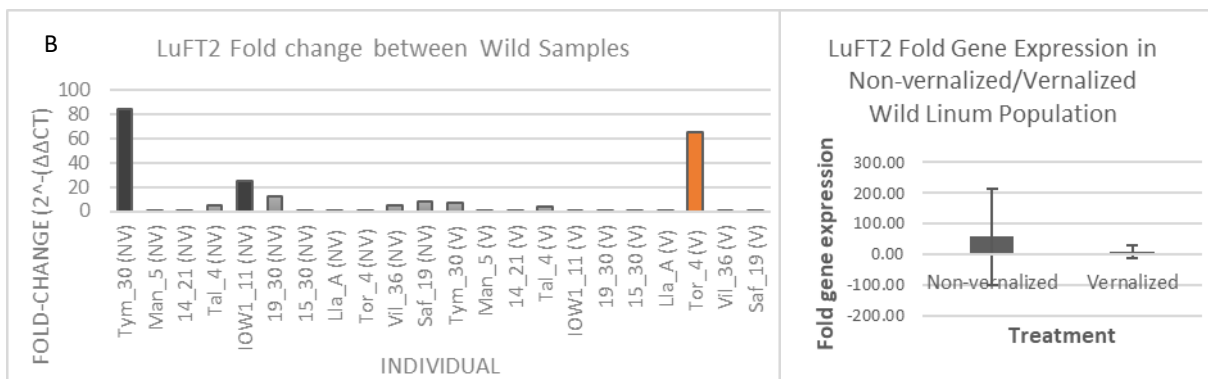
1600

1601 Figure 14. Bar charts showing *LuFT2* fold change between cultivars (A) and wild types (B) under non-vernalized and

1602 vernalized treatments. Highlighted in the bar charts with darker grey and orange are samples which seem to decrease and

1603 increase expression when vernalized, respectively.

1604

1605 <u>2.4.8 Gene Expression and Flowering Time</u>

1606 General linear models (GLM) were used to test whether the difference in RGE between

1607 vernalized and non-vernalized wild and cultivated *Linum* were associated with the flowering time

1608 initiation (Figure 15). In wild samples, the relative gene expression (RGE) differences of *LuFT1* and

1609 *LuFT2* were significantly positively associated with days to flower after vernalization (*LuFT1 t=6.359*

1610 *p=<0.001; LuFT2 t=5.128 p=<0.001*). The other tested loci, *LuCO1*, *LuCO2*, and *LuGI1.1* showed no

1611 significance when tested for association with flowering time (*LuCO1 t=1.510 p=0.162; LuCO2 t=1.324*

1612 *p=0.215; LuGI1.1 t=1.766 p=0.108*).

1613   For the cultivars, it was found that *LuFT1* and *LuFT2* expression differences were not

1614   significantly associated with flowering initiation (*LuFT1 t=1.226 p=0.237; LuFT2 t=1.732 p=0.122*).

1615   Neither did *LuCO1* and *LuCO2* expression differences show significant associations with flowering

1616   initiation (*LuCO1 t=-1.575 p=0.150; LuCO2 t=1.287 p=0.230*). Interestingly, *LuGI1.1* expression

1617   differences showed a significant positive association with flowering initiation (*LuGI1.1 t=2.932*

1618   *p=0.016*).

1619
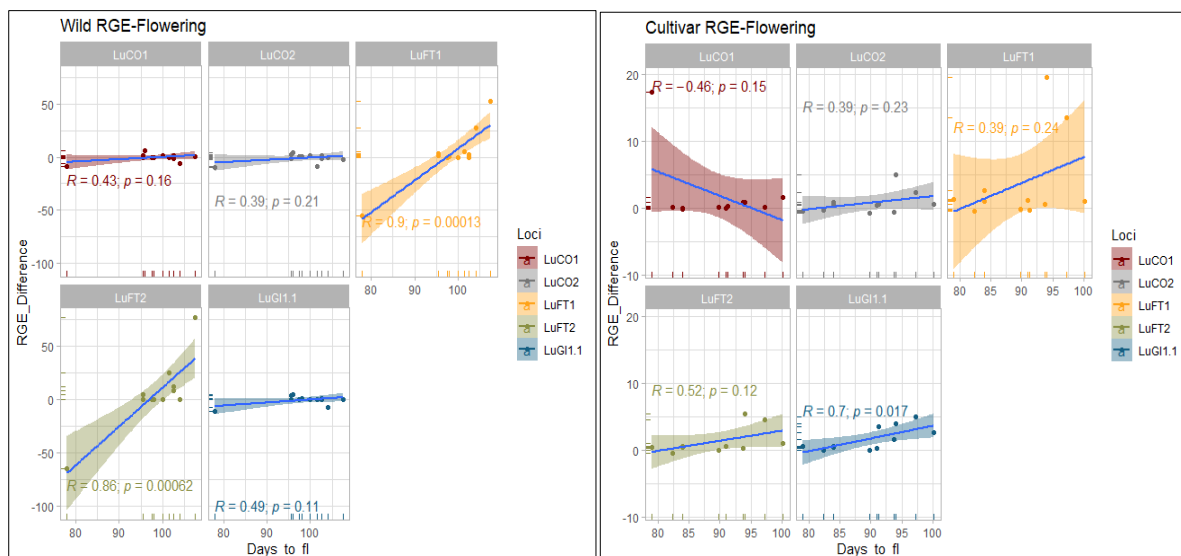
1620   Figure 15. Scatterplots showing number of days to flower after vernalisation labelled as "Days_to_fl" in relation to relative

1621   gene expression difference in wild samples and cultivars separately.

1622

2.4.9 Relative Gene Expression and Latitude

1624   The earlier phenotypic analysis found that vernalized wild individuals dramatically reduce

1625 the number of days required to flower when compared to non-vernalized wild plants, with

1626 populations from more northerly latitude of origin showing vernalization sensitivity. Here we tested

1627 latitude against relative gene expression for the cultivars and wild samples (Figure 16).

1628   Latitude was positively associated with relative gene expression difference in the genes; *LuFT1*

1629 *(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

1630 was no correlation between any gene expression tested in this study and latitude.

1631
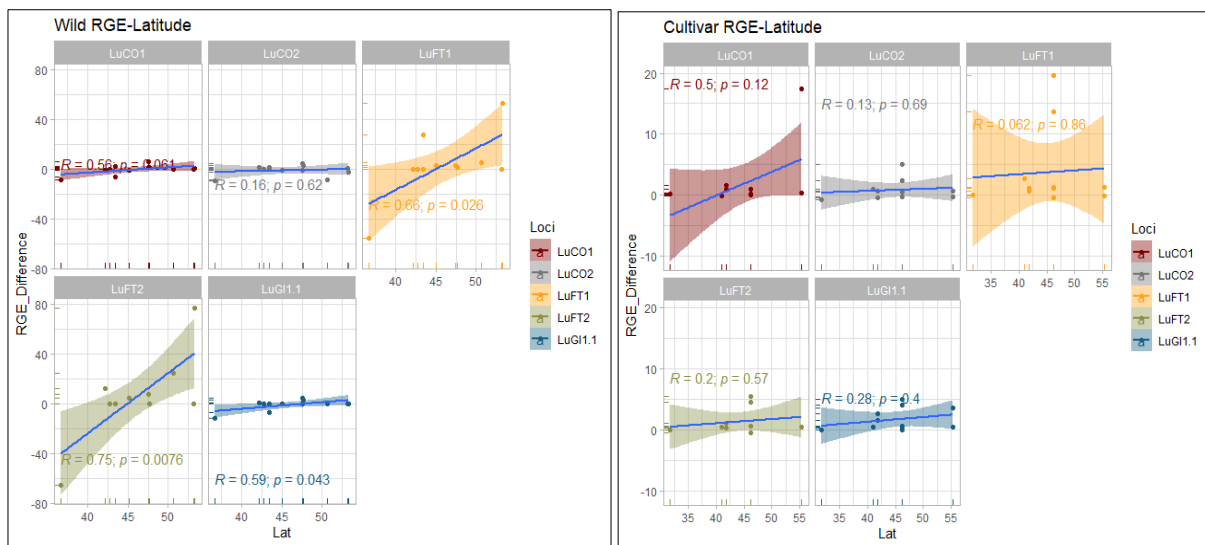
1632 Figure 16. Scatterplots showing latitude of origin (Lat) in relation to relative gene expression difference in wild samples and

1633 cultivars separately.

1634

2.4.10 Results: Relative Gene Expression and Climate

1635

1636 More Northerly location (higher latitude) suggests of colder, more wet, and windier

1637 climates. Wild samples were analysed for the relationship of flowering regulations genes expression

1638 differences and several climatic variables under that latitude. Climate of origin was summarised into

1639 principal components, and princliple component 1 (PC1) was then tested against relative gene

1640 expression differences as mentioned under section 2.2.1 "Samples and Experiment" of this chapter.
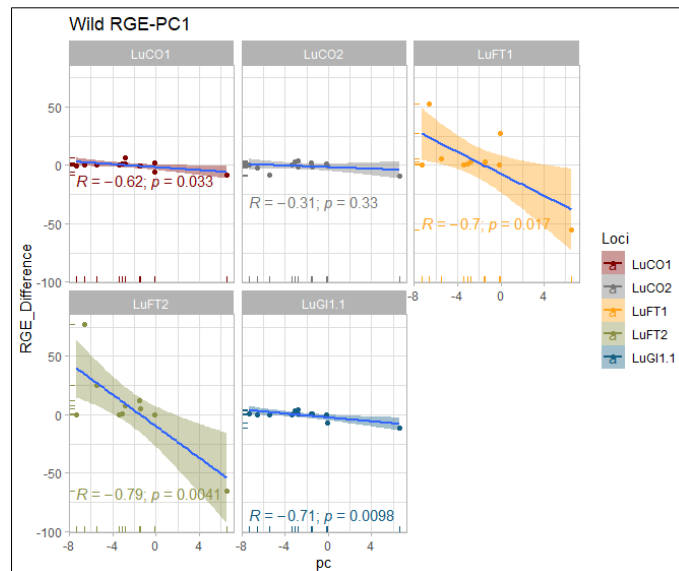
1641 Figure 17. Scatterplots showing climate variable PC1 (pc) in relation to relative gene expression difference of wild samples.

1642 Expression differences of *LuFT1* and *LuFT2* were significantly negatively associated with PC1.

1643 Lower PC1 values suggests colder climate (appendix 3) a summary of all the climatic variables (*LuFT1*

1644 *t=-2.905, p=0.017; LuFT2 t=-3.812 p=0.004*). There was also a significant negative association

1645 between climate and *LuGI1.1* (*t=-3.180 p=0.009*). The correlation observed in the wild for *LuGI1.1*

1646 against PC1 is more significant than it is observed under latitude alone. This suggests that climatic

1647 variables represent more significant correlation to the RGE observed under *LuGI1.1* and that that

1648 locus may be more varied in the wild according to the local climate than the latitude alone.

1649

1650

2.5 DISCUSSION AND CONCLUSION

1651

1652

1653   In the 2018 vernalization experiment, vernalization reduced the number of days to flowering

1654   in the wild *L. bienne*, especially in more Northern populations. This suggests that vernalization is an

1655   important mechanism in *L. bienne* and there was variation in regulation of flowering time in the wild

1656   relatives based on their latitude of origin. In addition to this, the results for *L. usitatissimum* suggests

1657   that requirements for vernalization were different between wild and cultivated *Linum*. This

1658   difference could be due to artificial selection for faster flowering of cultivars as part of domestication

1659   in *L. usitatissimum*. In the literature, genetic and association mapping have identified hundreds of

1660   genes as targets of divergence due to domestication (Smýkal P. et al., 2018). Evidence of this has

1661   been previously observed through resequencing of candidate genes related to seed sizes and weight

1662   (Guo *et al.*, 2020a). Vernalization insensitive flowering time could also be linked to productivity of

1663   *Linum* in temperate climates, and thus mechanism to vernalization is of interest in *Linum* breeding

1664   (Gutaker *et al.*, 2019).

1665   Other studies concluded that temperature, photoperiod (day length), and light availability

1666   (amount of sun light during the day) can influence developmental rates in crop plants such as wheat

1667   and other annual crops (Craufurd and Wheeler, 2009; Cave *et al.*, 2013; McMaster *et al.*, 2008). In

1668   wild *L. bienne*, variation in environmental conditions at different latitudinal localities have selected

1669   for differences in flowering initiation due to vernalization. This variation was not observed in our

1670   cultivated *L. usitatissimum* collection, suggesting no sensitivity to vernalization. However, in the

1671   literature there are suggestion that winter type *L. usitatissimum* found in Texas, USA, are sensitive to

1672   vernalization (Darapuneni M. et al., 2014). They suggest that these winter types are different to

1673   varieties grown in more Northern areas because they are grown in Autumn due to higher spring and

1674   summer temperatures in Texas. We saw no evidence of sensitivity to vernalization in our Northern

1675   European Spring and Winter varieties of *L. usitatissimum*. We suggest that there is wider variation

1676   within *L. usitatissimum* for requirement to vernalize. Perhaps growing seasons of each variety can

1677   affect the sensitivity of *L. usitatissimum* variant to vernalization.

1678   Another result illustrated in this study is the difference of gene expression due to

1679   vernalization conditions. After considering correction for multiple testing and variation among

1680   individuals, there was no significant difference due to vernalization in the fold change of any of the

1681   tested genes in both wild and cultivar types. Sun et al., 2019 also did not find an expression

1682   difference for *LuGI1.1*, between their cultivar lines (Sun et al. 2019). In addition to this, vernalization

1683   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,

1718 2016) may influence the non-significant relative expression differences of *LuFT* to flowering initiation

1719 and may have a major role to play in terms of this loss of vernalization requirement in the cultivars.

1720 Determining how vernalization affects the flowering time gene network, especially in *Linum* cultivars

1721 would be useful for agriculture to find genes which are important for specific growing conditions and

1722 exploring the potential loss of certain genetic functions due to selective breeding and human

1723 cultivation.

1724   For most of the tested genes, vernalized, and non-vernalized expression changes were not

1725 significantly different in wild and cultivated samples. In the literature however, there are some

1726 studies in other non-model organisms, which reveal that *GIGANTEA* homologous genes were

1727 differently expressed under vernalization treatments. This was true in a study using Ryegrasses

1728 (Paina *et al.*, 2014), and white lupins (Rychel *et al.*, 2019). However, studies regard *GIGANTEA* to be a

1729 photoperiodic response gene. This comes with implication that *GIGANTEA* interacts with another

1730 photoperiodic gene, the *CONSTAN* (CO). However, some studies have illustrated that *GIGANTEA* may

1731 be independently mediating photoperiodic control of flowering (Jung *et al.*, 2007). Although, there is

1732 evidence the *GIGANTEA* may be variably expressed when vernalized in other non-model organisms,

1733 the experiment in this case was limited in the number of paired samples that experienced both

1734 vernalization and non-vernalization treatments in 2021 (9 wild and 8 cultivated). This is due to

1735 availability of DNA materials (only available for 2021 experiments) and restrictions caused by the

1736 Covid-19 pandemic. There were 62 germinating individuals in the vernalization 2021 chamber.

1737 However, only around 20 individuals germinated in the non-vernalization 2021 chamber. Some of

1738 these have not had enough time to produce leaf materials for RNA extraction. In this study, there

1739 were only 38 individuals which experienced vernalization in both 2018 and 2021 experiments for

1740 which there was both expression and phenotypic data. In the 2018 experiment (pre-Covid-19) there

1741 were 197 vernalized individuals. This gives an idea about the limitation of data available for genetic

1742 study under the 2021 experiment due to restrictions caused by the Covid-19 pandemic. However, as

1743 in Figure 4, flowering time results from the 2018 experiment revealed correlation in flowering time

1744 and latitude in wild relatives and in addition suggests the loss of vernalization in cultivars. In the

1745 future, time permitting, an experiment involving vernalization, looking at a larger set of samples,

1746 may in turn reveal a stronger difference between the gene fold-change in vernalized and non-

1747 vernalized individuals.

1748   Another aspect that could be of interest from the genetic point of view is epigenetic

1749 regulation. Epigenetics is related to the idea that genetic regulation may be affected by different

1750 environments, which can include temperature changes. This was studied in the model *Arabidopsis,*

1751 looking at genetic regulation of flowering time (Khan, Ai and Zhang, 2014) and epigenetic responses

1752 to heat (Liu *et al.*, 2015). In the current study, certain genes in the wild samples showed a mix of

1753 increased and decreased expression across treatments. For example, the Southern samples Lla_A

1754 and Tor_4 tended to show increased expression (upregulation) when vernalized, and some Northern

1755 samples such as Tym_30 showed decreased expression (downregulation) of multiple flowering

1756 initiation genes when vernalized. Across wild sample populations, there were differences in genetic

1757 regulation based on their locality, or in this case latitude for *LuFT1*, *LuFT2*, and *LuGI1.1*. This further

1758 adds to the point that different population may regulate flowering time genes differently based on

1759 adaptation to their local environment which often affect downstream processes such as biosynthesis

1760 of plant hormones leading to developmental differences (Jaakola and Hohtola, 2010). A study of the

1761 model *Arabidopsis* comparing differences in altitude reveals that there was variation in genetic

1762 regulation of flowering time regulators, for populations under different altitude. These include a

1763 tested flowering regulator in this study, the Flowering Locus T (*FT*), with addition of vernalization

1764 genes such as VERNILIZATION INSENSITIVE 3 (*VI3*). Regulations of flowering regulator have been

1765 revealed to regulate MADS protein downstream and this differed between low and high altitudes

1766 genotypes (Suter *et al.*, 2014). Test of more *Linum* samples in consideration of altitude of different

1767 populations could help verify these results for *Linum* in future studies.

1768   Regarding the cultivars, most of the expression differences were reduced expression

1769 following vernalization. Reduced expression of flowering time genes such as the *LuCO1* and *LuCO2*

1770 (CONSTANS 1 and 2) could be a response to the expression of downstream flowering meristem

1771 identity such as SUPPRESSOR OF OVEREXPRESSION OF CONSTANS (*SOC*) to initiate flowering

1772 following vernalization as has been found in the model *Arabidopsis* (Sasaki *et al.*, 2017)(Li *et al.*,

1773 2015a)(Valentim *et al.*, 2015a). Therefore, further studies relating expression of homologues to *SOC*

1774 genes in *Linum* to expression changes of *CO* genes in response to vernalization could help confirm

1775 this vernalization pathway in *Linum*.

1776   Flowering time gene expression of *LuFT1, LuFT2, and LuGI1.1* were found to be associated

1777 with environmental variables related to latitude and climate of origin of wild samples. This result

1778 suggests that in the wild, phenological adaptation to the local environments could be due to

1779 different gene expression responses of these three flowering time genes. This is supported by the

1780 suggestion that *FT* was found to be a major locus contributing to local adaptation in flowering of

1781 perennial plants (Wang J. et al., 2018). The strongest associations between genetic expression as

1782 well as phenotypic responses (flowering time) to the local environments was shown by both *LuFT*

1783 genes. This contrasts with the result for wild flowering time that was not significantly associated

1784 with *LuGI1.1* relative gene expression. However, latitude and environmental variables were

1785 associated with *LuGI1.1* relative gene expression. In addition, in the cultivar set, *LuGi1.1* was the only

1786   gene that showed an association with flowering time. It is of interest to look at genes that maybe

1787   influencing the expression of *LuGI1.1*. It may be that expression of those genes were more related to

1788   flowering time than *LuGI1.1* itself. For example, there are multiple variants of the gene *GI* within

1789   *Linum* that were not tested in this study. These genes could be tested as part of future to further

1790   explore the flowering time network in *Linum* and implications of environmental variables. In

1791   addition, other vernalization related genes have also been observed to be related to local

1792   adaptation. The *FRI* (*FRIGIDA*) gene are related to local adaptation to drought tolerance by

1793   controlling flowering time in *Arabidopsis thaliana* (Lovell J. et al., 2013; Tigano and Friesen., 2016). In

1794   future studies, *FRI* can be another gene of interest when suggesting local adaptation to vernalization

1795   responses in *Linum*. In conclusion, evidence of local adaptation can be inferred from the phenotypic

1796   variation observed in the wild *L. bienne* tested here. This is further supported by the variation in

1797   expression of *FT*, linked to correlation towards environmental variables.

1798   One issue in this study was the decision to use of only one HKG. The use of multiple HKGs is

1799   regarded as superior to using only one HKG for quantification purposes as it gives a better picture of

1800   background expression (Remans *et al.*, 2008) (Manoli *et al.*, 2012). However, in the case of this

1801   experiment, the efficiency of *LuUBI2*, one of the tested HKG primers was very low (51%). This means

1802   that using both genes may skew the results due to the less efficient HKG so the Ct-values for *LuUBI2*

1803   were dropped from the analysis to conserve for this. In addition, the efficiency of some of the genes

1804   of interest (GOI) could be improved as part of future research. Quantification of relative gene

1805   expression using two HKGs was also explored as part of the analysis, with the less efficient *LuUBI2*.

1806   The results showed a drastic difference in relative gene expression when using two HKGs, with no

1807   significant differences between the treatments, potentially skewing relative expression. Using two-

1808   HKG indicate that the genes *LuFT1* and *LuCO2* showed the greatest difference in expression between

1809   vernalized and non-vernalized (non-vernalized) samples, but these differences were not statistically

1810   significant (appendix 8). This result suggests that using the less efficient primer may skew our results.

1811   We suggest improving the efficiency of *LuUBI2* for future research looking on the relationship of

1812   flowering time gene expressions with vernalization, using multiple HKGs. This would strengthen our

1813   results suggesting *LuFT1, LuFT2, and LuGI1.1* associations with environmental variables as found

1814   using one HKG in this study. Until such a time, our results are to be treated cautiously as a signal for

1815   the effects of environmental variables in relative gene expressions of *LuFT1, LuFT2, and LuGI1.1*.

1816   The relationship between relative gene expressions (RGE) difference calculated with two

1817   HKGs was tested against latitude of origin of wild samples (appendix 9) and no significant

1818   relationships were found. These results contrast to the results found above, using one HKG and

1819   suggest that the low efficiency value of *LuUBI2* could have contributed to masking some significance

1820   seen with only one HKG. In the future, using two HKGs with a higher efficiency value are

1821   recommended to test the relationships between gene expression to latitude, and climatic variables.

1822   We reflected that using one HKG is not ideal as most recent relative gene expression studies

1823   recommends using multiple HKGs. However, limited results on the expression of the floral integrator

1824   *FT* using one HKG are suggestive. To robustly test the expression of *FT* in *Linum* under vernalization,

1825   observation of HKG with higher efficiencies are recommended for future research.

1826   To further this research, other important regulatory flowering time genes in the network

1827   could be tested such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* (*SOC*) and as well as

1828   homologues of further downstream meristem identity genes found in the model Arabidopsis such as

1829   *LEAFY* (*LFY*), *APETALA1* (*AP1*), *SEPALLATA3* (*SEP3*) and *FRUITFULL* (*FUL*) (Schopf *et al.*, 1996) (Gregis

1830   *et al.*, 2009). Meristem development in *Linum* could be measured directly as a phenotypic

1831   comparison as has been done in other studies (Kayes and Clark, 1998) (Heisler and Jönsson, 2007).

1832   The complex network of flowering time genes could also imply that there is no single gene

1833   responsible for the phenotypic differences seen when both wild and cultivar type are vernalized. It

1834   may occur that several genetic regulations contribute, including the five major genes tested here.

1835   However, real-time primers that are specific to the two paralogues of each gene present in *Linum*

1836   would first need to be designed and tested, which is not a trivial task (Sun et al.). More genes which

1837   are specific to the vernalization pathway should be tested. Outside of the main floral pathway

1838   integrators, which *FT* (Flowering locus T) is a part of, this study looked only at previously designed

1839   flowering time genes which were related to photoperiodicity, such as *CO* and *GI*. It is interesting

1840   therefore, that the two *FT* loci showed significant relationships with latitude and climate of origin

1841   when observing the wild types. Vernalization-specific genes such as *FLC* (Flowering locus C) and *FRI*

1842   (FRIGIDA) have been shown to also play a role in the regulation network to suppress flowering locus

1843   T in the model *Arabidopsis* (Flowers *et al.*, 2009a). It would be of interest then to look at these

1844   vernalization pathway genes with appropriately designed primers specific for those genes in *Linum*

1845   as part of future studies.

1846   For further studies, it would also be of interest to measure traits such as seed yield. *Gutaker*

1847   *et al.*, have talked about potential link of flowering time to fruiting. Expression data for genes which

1848   may affect flowering time could be linked to whether seed yield from these measured individuals are

1849   affected by the changes of expression. Furthermore, variation in seed germination due to

1850   vernalization has been illustrated in the model *Arabidopsis* (Auge *et al.*, 2017). This may potentially

1851   link to productivity in *Linum* that may vary due to the requirement of vernalization and therefore

1852   affecting flowering time.

CHAPTER 3: THE QUANTIFICATION OF POLLEN AND POLLEN TUBES TO REVEAL

1854 POLLEN GERMINATION VIABILITY IN *LINUM* UNDER DIFFERENT TEMPERATURE

1855 TREATMENTS.

1856

1857   Pollen is the vehicle in which the male gamete is found. It is considered as a

1858 microgametophyte (sperm-producing gametophyte) in plants, a powder-like substance where the

1859 male gamete is produced and transferred (Johnstone, 2001). Pollen grains needs to disperse and

1860 travel to the stigma to fertilize egg cells which are found in the ovule and are non-dispersal. Pollen

1861 will often encounter many ecological and environmental pressures as they disperse (Loveless &

1862 Hamrick., 1984). They are considered as one of the major contributaries to the ability of flower to

1863 turn into seeds and produce progenies, a transformation from plant haploid gametophytes into

1864 diploid sporophytes (Pacini & Dolferus, 2016). Pollen is essential for seed production and thus

1865 agriculturally significant in crop plants such as *Linum*. As well, wild flowering plants often deploy

1866 different flowering strategies to optimise germination by pollen, and along with this, have evolved

1867 different germination mechanisms useful for artificial selections of male genotypes (Williams, 2012)

1868 (Tel Zur, et al., 2020). In this chapter we quantified pollen and pollen tube counts in *Linum*

1869 individuals treated under different temperature stress.

1870   Pollen must fertilize the ovary through the stigma-style-ovary system to form diploid

1871 sporophytes and form seeds (Pacini & Dolferus, 2016). The receptive ability of the stigma-style-ovary

1872 system is another aspect that could be observed with regards to the success of pollen germination.

1873 In this study however, we will focus on the pollen's ability to germinate and fertalize ovaries.

1874 Assessment of *in vivo* pollen germination success can be done by counting the number of pollen

1875 tubes in the style (Alonso et al., 2013; Williams & Reese, 2019). We decided to observe pollen

1876 initially because pollen is relatively easier to stain and study under confocal microscopy than the

1877 stigma. In a book published in 1998, Delph F, *et al,* reviewed variation in pollen growth, where there

1878 is an observation that pollen growth varies between species due to pollen/sperm competition

1879 whereby pollen traits such as, how quickly pollen reach the ovules, or how many pollen grains are

1880 transferred influence the success of germination. In addition, pollen competition in plants is

1881 described as gametophytic selection, potentially affected by environmental cues such as ambient

1882 temperature. This competition may cause pollen traits to differ in different flowering plants (Delph

1883 and Havens, 1998). In our *Linum* samples, we observed the viability of pollen and pollen tubes for

1884 the germination into seeds. We are particularly interested in the effect of temperature towards

1885 pollen viability under both wild and cultivated species.

3.1. Pollen Structure

1887

1888   In terms of their physical appearance, pollen grains varies in size across species (2-200μm)

1889 and contains parts where the generative cells are stored, often with a tube nucleus, within an intine,

1890 exine, and an aperture pore whereby the gametes can be omitted from using a structure widely

1891 known as pollen tubes (Johnstone, 2001; Pacini E, 2008). Pollen grains are microscopic. The smallest

1892 pollen size in diameter is found in *Myosotis*, with a diameter of 2.4-5μM (Sporemex, 2022), this

1893 often makes microscopy an ideal tool for observing pollen structures. The structure of a typical

1894 pollen is illustrated as below (Unacademy, 2022):
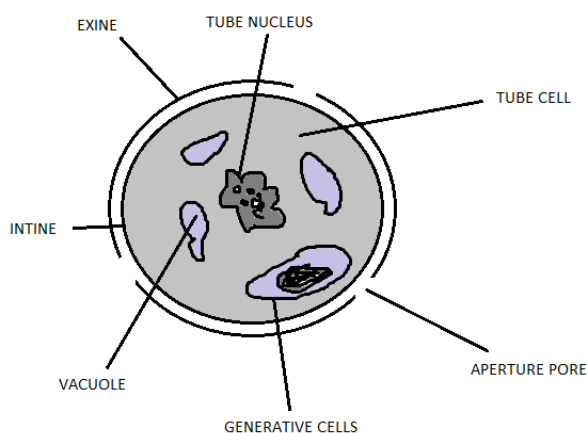
1895 Figure 1. A diagram depicting the well-known structure of a mature pollen.

1896

1897 During germination, the tube nucleus part of the pollen expands into a pollen tube whereby the

1898 male gamete of the pollen can be transferred to the ovary. Pollen tube expands out of the pollen

1899 through the "aperture pore". The number of aperture pores can vary between species. It is thought

1900 that increased number of apertures in angiosperm pollen grains can offer the species a selective

1901 advantage as it increases the number of germination sites and facilitates the chances of contact

1902 between the pollen tubes and the female ovule (Furness & Rudall, 2004).

1903   For germination, pollen often require active RNA synthesis. This was evident in the conifers,

1904 where RNA synthesis for pollen germination was studied for its dependence on transcription and

1905 translation (Breygina et al., 2021). When pollen is germinating, the tube nucleus will extend towards

1906 the aperture pore and will then form a pollen tube. These pollen tubes are organelles which can

1907 elongate and are responsible for the transfer of pollen male gametes into the stigma and down to

1908 reach the ovule (Adhikari et al., 2020). When pollen tubes reach ovules, male gametes within the

1909 tube nucleus enter the ovaries to form zygotes. The zygotes are the progenitor stem cells which can

1910   form embryos that later form into seeds (Kanday and Sundaresan, 2021). With this knowledge, it is

1911   then important to consider pollen viabilities in terms of their ability to form zygotes. This includes,

1912   amount of pollen, amount of pollen tube , and the rate which pollen tube reaches ovary. This can

1913   lead to factors influencing seed formation success (Iwazumi and Takahashi, 2012).

1914   Pollen viability is therefore of interest related to the breeding strategies of *Linum* as both

1915   wild and cultivated types are present under different environmental conditions. In *L. usitatissimum*,

1916   pollen can be used as a tool to measure geneflow using pollen as a mediator, which may in turn can

1917   detect different strategies and local adaptations related to the rate of pollen germination under

1918   different conditions (Jhala, et al., 2010). It is therefore of interest, to look at pollen germination rate

1919   and pollen viabilities when looking at local adaptations as well as strategies of breeding systems. In

1920   the literature, study species such as cotton (*Gossypium hirsutum*) shows a high tolerance to heat.

1921   Optimum temperature for pollen growth and pollen tube length in cotton have been observed to be

1922   above 32°C (KAKANI et al., 2005). Other observations in temperate cultivars such as apricots and

1923   sweet cherries (*Prunus spp.*) suggests that pollen germination and pollen tube length are higher in

1924   some cultivars treated with colder temperature (5°C), although varying within different lines (Pirlak.,

1925   2002). Previous observations of pollen tubes were made in another agriculturally significant plant,

1926   peach (*Prunus persica*), by Hedhly A. *et al* (Hedhly & Herrero., 2008). This study found that drastic

1927   increase in temperature had a negative effect on stigmatic reception of the pollen tubes, thus

1928   significantly slowing pollen tube growth. These results suggest that stigmatic cells' ability to sustain

1929   and adhere the pollen tube cells were reduced because of drastic temperature effect and the

1930   kinetics of pollen tube growth was also affected. This behaviour was also observed in the field

1931   (Hedhly & Herrero., 2008). A similar negative effect of temperature on pollen kinetics and stigmatic

1932   cell adhesion was also previously observed by the same group in cherries (Hedhly et al., 2004).

1933   It has been determined that pollen germination and pollen tube growth could be observed

1934   *in vitro* and semi *in vivo* for the model *Arabidopsis thaliana* (Dickinson et al., 2018), making them

1935   versatile for observations in the laboratory or the wild. In angiosperms, the dynamics between the

1936   pollen and its pollen tube formation right into the ovaries represents some of the most

1937   environmentally sensitive part of their sexual reproduction. Temperature is considered an important

1938   variable, currently studied under various environmental research because it is a "fragile" climatic

1939   variable, and temperature represents one of the main environmental variables which could affect

1940   the health of living organisms, especially in plants (Reynolds & Casterlin, 1980) (Nievola et al., 2017)

1941   (Qaderi et al., 2019).Pollen tube growth dynamics have been highlighted to potentially be controlled

1942   by $Ca2^+$ signalling and sensitive to heat stress, making it vulnerable to changes to the environment

1943   (Johnson et al., 2019). It has been suggested that $Ca2^+$ signalling acts as a signal transduction factor

1944 under heat stress conditions in the model *Arabidopsis thaliana* (Xu et al., 2022). This may affect
1945 pollen tube dynamics in *Linum*.

1946 This part of the chapter will test several growing temperatures to establish whether pollen
1947 germination is affected by using quantification and analysis of pollen structures. One of the interests
1948 in pollen observation is whether different temperature conditions impact the performance of pollen
1949 tubes to reach the ovule and form a gamete. To do this, pollen tube would need to be able to travel
1950 down the style into the ovule which contains the female egg cells. Whether temperature presents a
1951 significant stress for pollen tubes to be able to reach the ovule is of important agricultural interest in
1952 crops including *Linum*. A further interest is if there would be a difference in species and population
1953 level across cultivated *Linum* (*L. usitatissimum*) and its wild relatives (*L. bienne*). Another interest in
1954 terms of pollen viability is whether different climate of origin variables is associated with pollen
1955 viability. For the purposes of the study of this chapter, the vulnerability to changes in temperature
1956 are hypothesized to be of key importance for seed set in *Linum*. Pollen viabilities were previously
1957 tested in *L. usitatissimum* and were found to be little affected by heat stress. However, findings
1958 suggest the formation of bolls and seeding were negatively affected by heat stress (CROSS et al.,
1959 2003). Temperature effectson pollen have not been tested in *L. bienne*. Other cultivated species
1960 showed a negative effect on pollen viability and germination post high temperature exposure (Aloni
1961 et al., 2001).

1962 In this study, temperature as an environment variable was evaluated against pollen viability
1963 in the wild and cultivated *Linum* samples which were also used across other studies in this thesis.
1964 Observations were made of pollen viability at various stages. These were pollen count, pollen tubes
1965 count, and whether pollen tubes were able to reach the ovaries as a sign of pollen gamete being
1966 "successfully" transferred. Different temperatures were tested *in vivo* for pollen germination
1967 strategies comprising heat and cold treatments as well as a control typical temperature treatment.
1968 The main hypothesis assessed in this chapter was that pollen viability is affected by temperature as
1969 an environmental variable in wild and cultivated *Linum*. We also expect wild and cultivated flax to
1970 show variability within their pollen viabilities under different treatments. With local adaptation in
1971 mind, we expect differences in pollen responses to cold and heat treatments between populations
1972 originating in warmer and colder climates. This may be presented as correlations between pollen
1973 viability and local climate variables. More Northern plants are expected to perform better under
1974 colder temperatures in terms of pollen viability and vice versa for Southern plants.

1975

1976

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°C daytime temperature and 13°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°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°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

2011 weeks of flowering, using a combination of pollen staining and confocal microscopy techniques

2012 (described below). Between the different temperature treatments, samples were randomly swapped

2013 over a 72 hour rotation period with varying number of flowers at "open stage" able to be harvested

2014 per-observation.

2015

2016 3.2.1 Pollen Viability *In Vitro*

2017 Studies have previously shown several ways to test pollen viability. These often consist of

2018 staining by a fluorescent dye and assessing the presence of pollen and pollen tubes under a confocal

2019 microscope, as performed previously in *Panicum* (Ge et al., 2011). Optimization of pollen

2020 fluorescence microscopy is also of interest in the model *Arabidopsis* (Bou Daher et al., 2008) and

2021 thus this study involving the non-model *Linum* will explore staining methods for fluorescent

2022 microscopy.

2023 As a precursor to quantification, tests were done using *in vitro* germination under heat stress

2024 which uses different germination media to *in vitro* germinate pollen from different temperature

2025 treatments. At the aim was to find out whether *in vitro* germination using germination medium was

2026 possible in *Linum* to determine heat stress (Rodriguez-Enriquez et al., 2012). In the model

2027 *Arabidopsis thaliana*, it is suggested that *in vitro* germination is possible, with a varied germination

2028 levels at different treatments, however pollen germination and tube growth were dependent on

2029 pollen density in both liquid and solid medium (Boavida & McCormick, 2007). Recent past studies

2030 have tested the effect of germination media with extra sucrose, and boric acid for *in vitro* pollen

2031 tube germination in the non-model Chinese fir (Fragallah et al., 2019). Then, *in vitro* methods were

2032 used to germinate different pollens under different environmental conditions. We assessed different

2033 solutions in *Linum in vitro* germination. Different germination solutions were made using Sucrose as

2034 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%

2035 Table 1. A table of the three different *in vitro* germination solution which were tested in the laboratory.

2036 The pollens were firstly place within the different solutions for at least 3 hours before observation.

2037 To assess germination success, a control was also provided with just nuclease free water as the

2038 solution. 20% Toluidine blue was used to dye the pollens after the treatments and pollens were

2039 observed under a Leica DM500 microscope with a camera model MC190HD connected to a

2040 computer running Leica's LAS software. Results for the *in vitro* germination were not very convincing

2041 (see the results part of this chapter). This result would mean that *in vitro* germination procedures in

2042 *Linum* pollen needed progress. Therefore, it was decided that pollen would need to germinate *in*

2043 *vivo* on flower pistils, with measures at different pollen maturity stages to measure heat stress under

2044 different temerpature conditions.

2045 For *in vitro* germination attempts, three solutions were prepared according to Table 1 and

2046 was adjusted to pH of between seven and eight. Three to five drops were put onto a glass

2047 microscope slide with a Pasteur pipette. Wild *Linum* flowers from control 25 C conditions that were

2048 still not opened (to avoid *in vivo* germination) were collected by cutting from the stem and all petals

2049 were carefully removed using forceps to expose the stigmas, and anthers. These parts were then

2050 dipped onto the solution on the microscope slides several times to release pollen into the solution.

2051 At least 3 Hours have passed to allow germination before the staining procedure was undertaken.

2052 For the stain, 20% Toluidine blue were prepared within a mixture of 50% nuclease free water and

2053 50% glacial acetic acid. Two drops of the dye solution were added to the slide and the pollen

2054 samples. Observation took place using a Leica DM500 microscope fitted with MC19HD camera,

2055 connected to a computer able to run Leica's LAS software. In this study 30×10 and 50×10

2056 magnifications were used for the pollen *in vitro* germination observations.

2057

2058 ## 3.2.2 Pollen Treatments

2059 For the different temperature treatments, individuals were randomised under a controlled

2060 glass-house conditions before measurements under confocal microscopy. Plants were initially grown

2061 at 25°C during daytime. Plants were then treated with either a heat (+5°C) or a cold (-5°C) treatment

2062 under the corresponding tropical or arctic chambers as the first flower were observed. Plants were

2063 treated for at least 72 hours before further progress to allow for acclimatation. Post acclimatation,

2064 flowers were collected at an "open stage.

2065 Aniline blue staining protocol was deployed along with a confocal microscope for visual aid.

2066 Full protocol and information of the Aniline blue are available under the "3.2.3 Pollen Observation"

2067 part found below in this chapter.

2068 As described in *Schewe et al,* flower at an "open stage" are categorised as flower at anthesis,

2069 showing an opened 5-part whorl where the stamens are found with the stigma (stage 11) (Schewe,

2070 et al., 2011). Whole flowers were harvested and preserved with 70% EtOH inside separate 1.5ml

2071 Eppendorf tubes for each line. Two solutions were evaluated for pollen preservation. One solution is

2072 Formaldehyde-acetic acid alcohol (50ml of >95% ethyl alcohol, 2.5 ml of glacial acetic acid, 5.5 ml of

2073 formaldehyde, and 42 ml of dH_2O) (FAA), and the other solution was 70% ethanol with dH_2O

2074 (70%EtOH). Observations found both solutions to be adequate for floral and pollen preservation

2075 after 24-hour periods (appendix 12). After collection, flowers were submerged in at least 1ml of 70%

2076 EtOH inside a 1.5ml Eppendorf tube for preservation. Petals were removed as well as anthers which

2077 results in the boll and the stigma exposure. Some errors in pollen measures are expected as pollens

2078 may or may not be washed out during preservation stages, but the same treatments were applied to

2079 all samples under all treatments. Measures of pollen and pollen tube counts have been accumulated

2080 over several weeks. Between the different temperature treatments, individuals were randomly

2081 swapped, and different flowers were also measured per treatment.

2082 ### 3.2.3 Pollen and Pollen Tube Observation

2083 For the purposes of pollen and tubes observations, a modified protocol based on the

2084 protocol by Lu Yongxian (Lu Y, 2011) were used. The fluorescent staining dye applied into stigmas to

2085 observe pollen number and observe pollen tubes was 0.5% (w/v) Aniline blue in 0.1M KH_2PO_4.

2086 Glycerol was recommended in some fluorescent dye protocols, but it was not used in this study as

2087 fluorescence dye were observed better in non-glycerol solutions. Optimisation of aniline blue

2088 dilution was of interest, as too dilute will limit penetration into the pollen cells and too concentrated

2089 will act as a background noise, which may interfere with the images surveyed (Herburger &

2090 Holzinger, 2016). Preliminary optimisation found that 0.5% (w/v) Aniline blue in Potassium

2091 Phosphate (KH_2PO_4) provided images that were useful for pollen and pollen tubes observations

2092 (Appendix 13).

2093 Bolls and stigma attached were examined onto a 25mm x 75mm glass microscope slide with

2094 a thickness of ≤1.2mm. 2-3 drops of 0.5% aniline blue were sufficient for dying purposes. In theory

2095 aniline blue can penetrate callose better than other cells (Herburger & Holzinger, 2016), and

2096 therefore more sensitive towards pollens and pollen tubes. This made it easier to distinguish the two

2097 pollen tissues from other cell types and to quantify them. Samples were covered using 22mm x

2098 22mm cover slip with a thickness of 0.1mm to allow for optimal viewing under a confocal

2099 microscope. Samples were stored at room temperature for >3 hours prior to viewing to allow for the

2100 dye to penetrate the callous cells. The observations were also done in a dark room with as little light

2101 as possible.

2102 A Zeiss model 880 Confocal Microscope was used to observe fluorescence and presence of

2103 pollen and pollen tube structures with fluorescence absorbance between 400-500nm for aniline blue

2104 absorption (Yang et al., 2007) and images were further processed by Zeiss' own software (ZEN Black

2105 edition) to optimise image settings. Zeiss' 10× magnification and Zeiss' 20× magnification eye pieces

2106 all combined with a 10× eyepiece was used to observe pollen and pollen tubes. Different focal layers

2107 were examined and in cases where pollens could not be observed under one layer of focus, a 3D

2108 image with appropriate number of Z-stacks was acquired to scan through all axes.

2109

2110 ### 3.2.4 Pollen Counts

2111 *Linum* pollen structures were circular vessel like grains and have high callose content and

2112 therefore would be highlighted more intensively by the aniline blue dye than surrounding cells. We

2113 observed no differences in structure for the wild and cultivar species of *Linum* observed for this

2114 study. The pollen was expected to be found adhered around the stigma. Pollen tube structure were

2115 identified as thread like structures coming out of the pollen tube and penetrating the stigma, often

2116 in a straight line down to the ovary (Figure 2, next page).
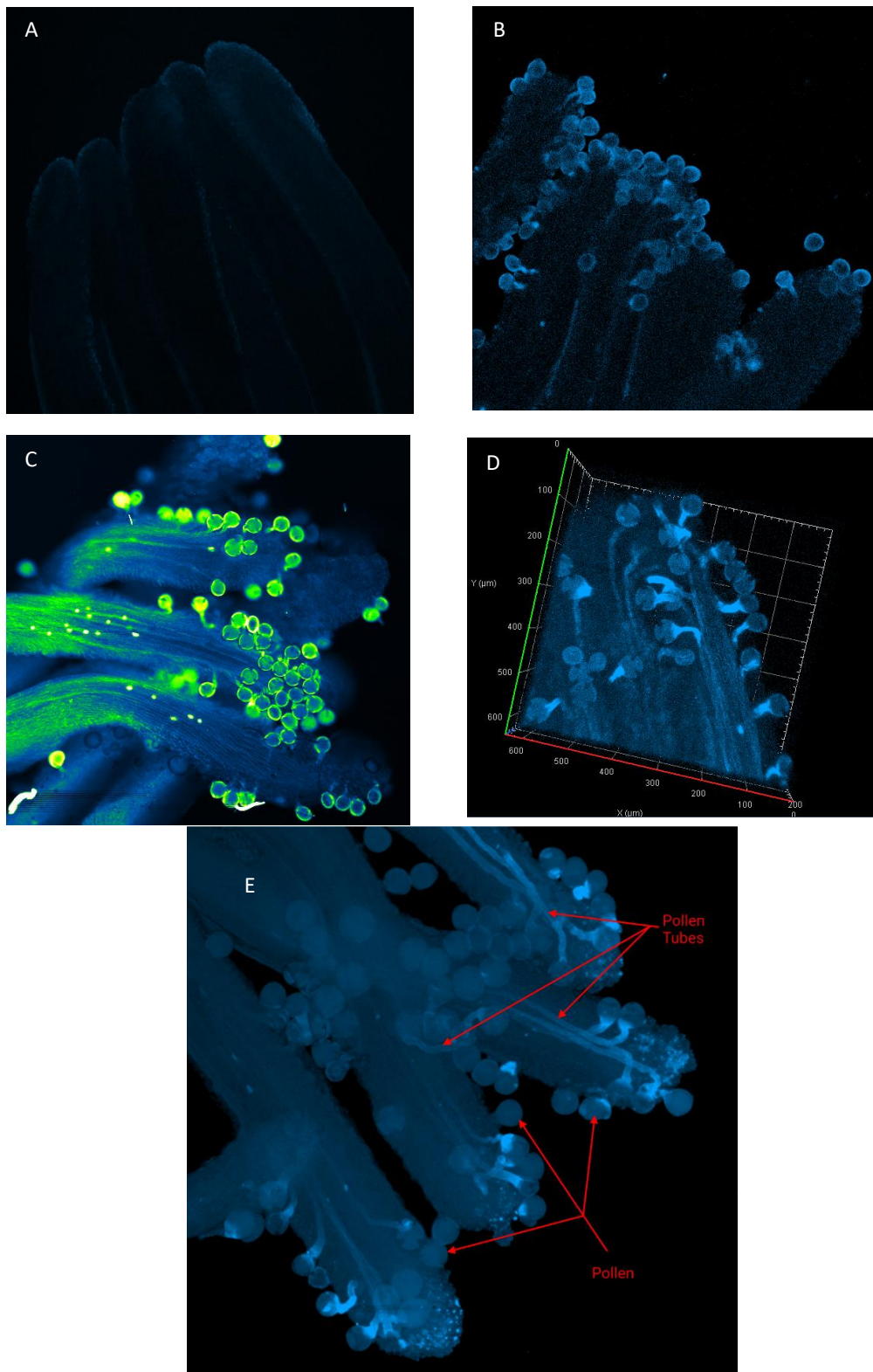
2117

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.

2122   Absence of pollen and pollen tube structures can be observed in some cases under fluorescent

2123   microscopy (Figure 2A). In other cases, prescence of pollen and pollen tubes at the stigma were also

2124   be observed (Figure 2B-E). Pollen were rounded in shape and fluoresce brighter than the stigma

2125   cells. The pollen tubes were string-like structures growing out of pollen into the stigma (Figure 2E).

2126   Images were further processed in Zeiss' Zen Black software to apply colouration to highlight

2127   fluorescence absorption at 400-500nm wavelength and increase contrasts between the pollen and

2128   pollen tube cells and surrounding stigma/style cells. Callose in pollen and pollen tubes can be seen

2129   as more fluorescent (Figure 2C). A 3-Dimensional observation can also be made using the Z-stacks

2130   function of Zeiss' Zen Black software (Figure 2D). This feature enabled layers to be taken stack by

2131   stack under a new Z dimension, thus building a 3-dimensional image. This tool represents an aid, as

2132   pollen and pollen tubes may have not been fully observed on a 2-dimensional image. After

2133   optimisation for each slide, sample images were captured, and pollen and pollen tubes were

2134   manually counted to build a dataset.

2135

3.2.5 Pollen Tube Counts

2136

2137   Observation was made for three different temperature treatments to observe whether

2138   pollen tubes were able to reach the ovule at the open flower stage. Aniline blue staining and

2139   confocal microscopy methods were used as before, but for this observation, we observed pollen

2140   tubes inside pistils and stigma cells. Colouration of the aniline blue were referred to "cyan" and

2141   contrast were increased with using Zeiss' Zen Black software. For stigma and styles with little pollen

2142   tube growth, wavelength was contrasted from fluorescent green to blue, highlighting dyed cells

2143   fluorescing green. The processed images could then be observed to see whether pollen tubes were

2144   able to reach down to the bottom of the style and reach the ovule. This was observed for each of the

2145   temperature treatments and recorded in binary, 0 representing pollen tube not reaching ovary and 1

2146   representing pollen tube reaching ovary for every individual tested. The ovary as well as the pollen

2147   tube fluoresce when dyed with the Aniline blue. An observation of a pollen tube structure going into

2148   the ovary at the bottom of the pistils would suffice the question of whether the pollen tube have

2149   managed to reach the ovary or not. The resulting difference in images with pollen tubes and without

2150   pollen tubes could be observed on the Figure below:

2151

2152
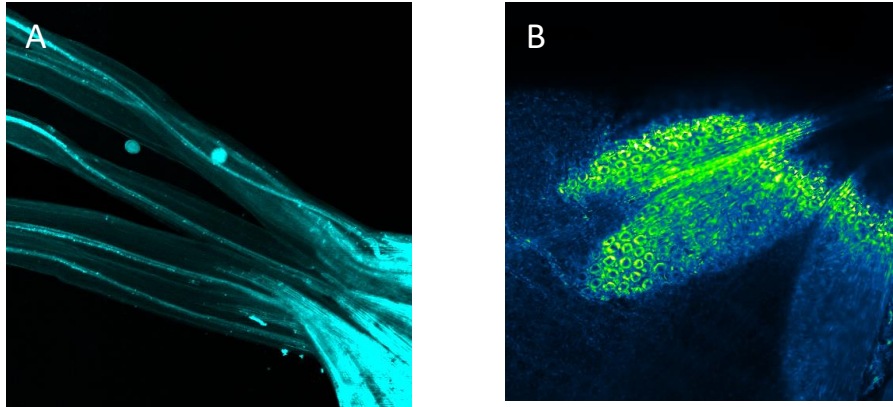
Figure 3. A. 10×20 magnification of a flower pistil with observed pollen tube growth fluorescence highlighted in cyan. B. 10×20 magnification of fluoresced ovary cells highlighted in "fluorescent green", depicting no pollen tubes going into the ovary.

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

2200 using the R package "lmerTest" (Kuznetsova et al., 2017). A 'logit' link was added using the command

2201 "(link = 'logit')" as an addition to the quasibinomial family specification which attempts to describe

2202 additional variance in the data that cannot be explained by a Binomial distribution alone. After the

2203 linear modelling, analysis of variance (ANOVA) for the different variables (Girden ER, 1992) was

2204 applied to the model in R using the command "anova(model,test=F)". Interaction between

2205 treatments and species was measured using the "interaction()" command of R. The Tukey's contrast

2206 treatments were added to the model using the R package multcomp (Hothorn et al., 2008), using the

2207 command: "model<-glht('modelprefix',mcp(interact="Tukey"))". The full model commands can be

2208 found under appendix 14.

2209   To evaluate whether pollen tube was able to reach the ovary, a binomial "yes or no" test

2210 was utilized. To test for this, a general linear model with a binomial adjustment was run in R with the

2211 function "*family=binomial(link='logit')*" added to the general linear model command (Gelman, A. and

2212 Hill, J., 2006).

2213

2214 ### 3.3.3 Analysis: Population of Origin Variables

2215   To consider environmental effect, we gathered geographical data for the wild *L. bienne*.

2216 Climatic variables from the origin of the wild individuals were gathered from WorldClim database in

2217 *Chapter 2* of this thesis. Latitudinal and climatic data was only reliably available for the wild *L. bienne*

2218 samples. This is because location data for our cultivars was not precise. Therefore, for the analysis

2219 against local variables, only wild *L. bienne* individuals were considered. Climatic variables were

2220 processed into summary principle components (PCs) as was described previously in chapter 2 (see

2221 section 2.2.1 "Samples and Experiment"). Latitude and climatic variables were used to further

2222 describe temperature variables of the local areas which the wild *L. bienne* were collected from. More

2223 Northern origin plants are expected to perform better under colder temperatures in terms of pollen

2224 viability and vice versa for Southern origin plants. The loadings of the principal component analysis

2225 are as in appendix 3. The temperature loadings of climate PCs will be of particular interest in this

2226 chapter. Since the temperature loadings were positive, a higher PC1 value indicated a higher

2227 temperature. For example, the PC1 value for the Southern population '3' is *5.367* while the PC1

2228 value for the Northern population 'Tym' is *-6.616*. The average temperature loading for population

2229 '3' collected from the WorldClim database is at 12.07^oC in winter (December, January, February)

2230 while for population 'Tym' it is 5.02^oC. For the summer (June, July August) the average temperature

2231 for population '3' is 22.6^oC and for population 'Tym' it is 15.05^oC. A linear model with a Pearson's

2232 correlation coefficient can be used to test whether latitude of locality and climatic variables affects

2233   either pollen count, pollen tube counts and or the proportion of pollen able to reach the ovary. A

2234   scatterplot of relationships was generated using the "ggscatter" command of the package ggplot2

2235   built for R (Villanueva & Chen, 2019).

2236

2237

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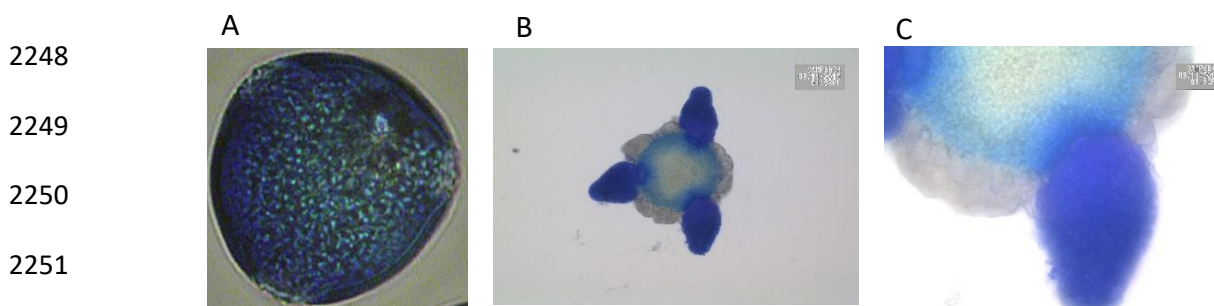
Figure 4. A. 10×50 cropped magnification of a mature pollen dyes with 20% Touludine blue, observed in nuclease free water and no added germination solutions. B. A 10×20 magnified image of a 20% Touludine blue dyed pollen cell which is undergoing bursting from all the aperture pore sites due to the germination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

2269

2270
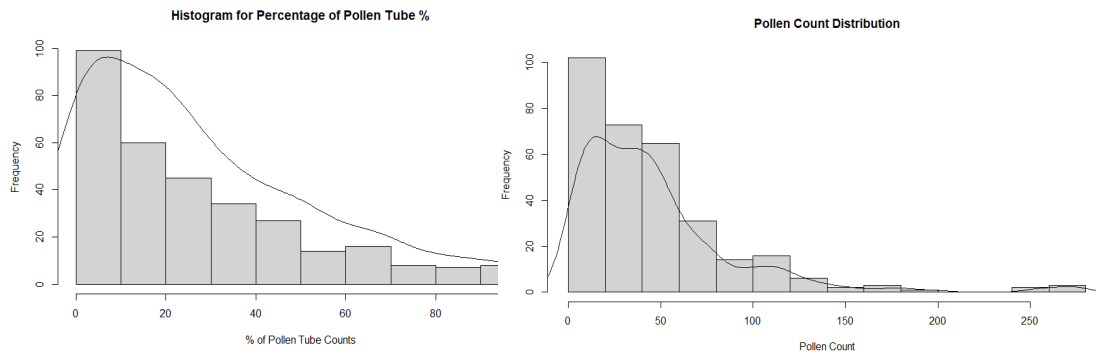
Figure 5. Histograms and density curves for the data distribution of both pollen tube % and pollen count showing a non-normal distribution.

2278

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).
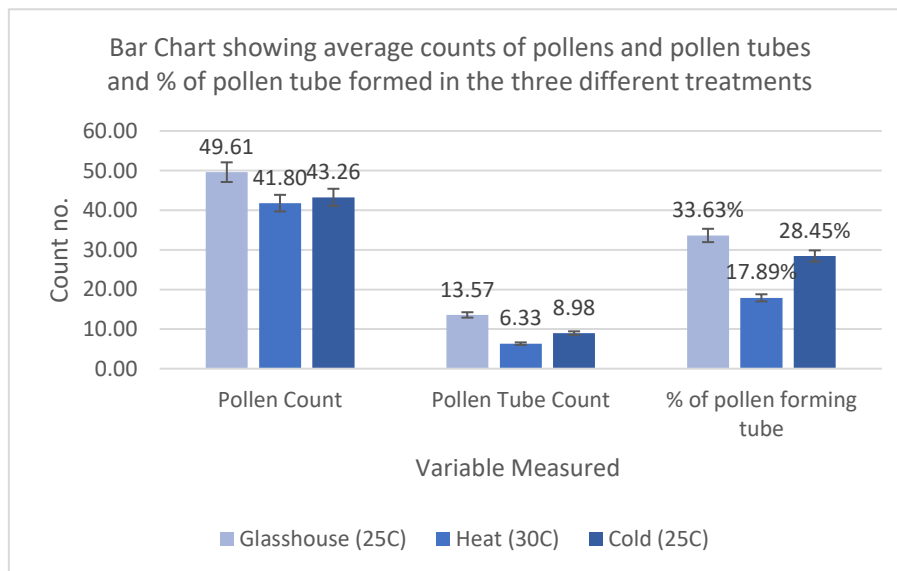
2285

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.

2288

2289　　　　　Pollen count of each treatment were statistically compared to check for effects on the

2290　number of pollens due to temperatures according to the hypothesis that the temperature affects the

2291　number of viable pollens within both wild and cultivated *Linum*. There was variation in the average

2292　pollen count when comparing the three treatments (Figure 6). To statistically test for this variation,

2293　we ran an ANOVA with a Tukey's post-hoc for multiple comparisons and we illustrated a model using

2294　a Poisson's regression model formula of "Pollen$Count ~ Pollen$Treatment" for the data comprising

2295　all the samples within this study. The analyses suggested that there was significant decrease

2296　between the pollen counts in the glasshouse (25°C) compared with both the cold (20°C) and heat

2297　(30°C) treatments (Table 2 below). This sugests that the treatments have a significant effect on the

2298　number of pollens within the samples tested in this study.

2299

2300

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	*

2301

2302

2303

2304　Table 2. A linear model with Tukey's correction for pollen count summary between the different temperature treatments.
2305　Significance (*p=<0.05*) was showing for a linear hypothesis between the 25°C – 20°C and the 30°C – 25°C treatments.

2306

2307　　　　　To see whether the same could be seen in the number of pollen tubes, average pollen tube

2308　count was also tested for each treatment (Figure 6). When statistically compared, the number of

2309　pollen tube on average between the treatments decreased with both cold and heat treatments. Just

2310　as were seen in the pollen count dataset. In addition, the number of pollen tubes seem to be

2311　significantly reduced (*p=<0.05*) for the heat (30°C) and cold (20°C) treatments (see Table 3).

2312

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	*

2313　Table 3. A linear hypothesis with a Tukey's adjustment for the number of pollen tubes observed between the different
2314　treatments. The test shows significant differences between all the treatments (*p=<0.05*).

2315

2316   We observed variation in pollen tube count and pollen count when looking at different

2317   temperature treatments. However, in relation to pollen count we wanted to see if there is variation

2318   in the proportion of pollen tubes able to be formed under our temperature treatments. To observe

2319   whether the proportion of pollen tubes to pollen count is substantially affected by the temperature

2320   treatments, the average proportions of pollen tubes formed against pollen for each temperature

2321   treatments were calculated and modelled. The results suggests that against the glasshouse (25°C)

2322   treatments, the cold (20°C) treatments showed a decrease in the proportion of pollen forming pollen

2323   tubes but that the difference was not significant (Table 4). Just as the cold (20°C) treatments, the

2324   heat (30°C) treatments showed a decrease in the proportion of pollen tube forming against the

2325   pollen count (Table 4). This decrease in proportion, however, was shown to be significantly different

2326   to the glasshouse (25°C) treatments when modelled against each other ($p = <0.001$) (Table 4). This

2327   suggest that warmer treatments could potentially affect proportion of pollen tube able to be formed

2328   in relation to the pollen count. Below is a linear hypothesis summary (with Tukey's modification) for

2329   the proportion of pollen tube formed against the number of pollens for each treatment.

2330

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**	*

2331   Table 4. Linear hypothesis with a Tukey's adjustment for the proportion of pollen tube forming, showing significance

2332   between 30°C and 25°C treatments (*p=<0.05*).

2333   To summarize the linear hypothesis significance for section 3.4.3 "Results: Pollen and Pollen

2334   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

2335   Table 5. Summary of linear hypothesis tests significance between different temperature treatments and counts as per

2336   section 3.4.3 "Results: Pollen and Pollen Tubes Count".

2337

83

3.4.4 Results: Pollen Tube Reaching the Ovary

2339   To test a model of whether any of the three treatments were significantly different in the

2340   number of pollen tubes able to reach the ovary, a binomial general linear model (GLM) was

2341   performed for "yes or no", binomial data under the different treatments. The model was

2342   summarised using a Tukey's adjustment for abnormally distributed data (see section "3.3.2 Analysis:

2343   Proportion of Pollen Tubes"). The results reveal that the percentage of pollen reaching the ovary is

2344   significantly different in the 30°C treatments (see Figure 7 and table 6).

2345   The results from the general linear model suggests that between the glasshouse (25°C) and

2346   the cold (20°C) the difference in the proportion of pollen reaching the ovary was not significantly

2347   different (*p = 0.85)*. This being the case, significant difference in the number of pollen tube reaching

2348   ovaries could be observed between the heated (30°C) treatment and both the glasshouse (25°C) and

2349   cold (20°C) treatments (*p=<0.001 for both*) (see table 6).

2350
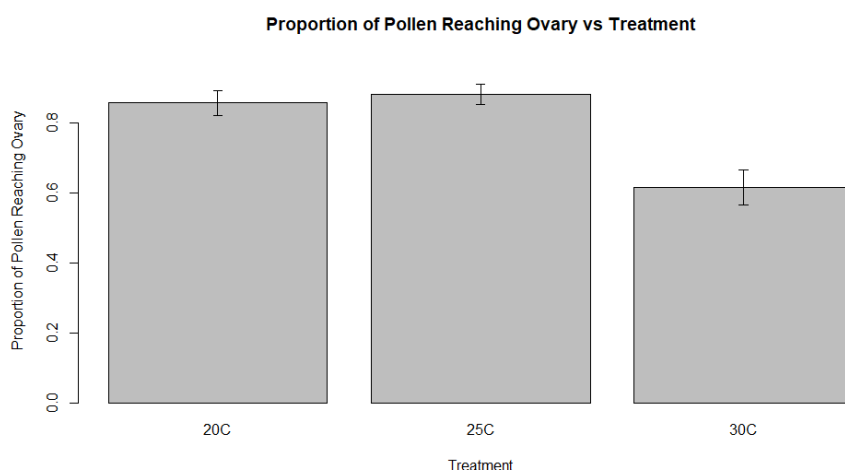
2351   Figure 7. Bar chart showing the proportion of pollen reaching ovary according to the different temperature treatments.

2352

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**	*

2353

2354   Table 6. General linear model with Tukey's adjustment for the proportion of pollen tubes reaching ovaries. The test shows

2355   significant differences between the temperature treatments of 30°C – 20°C and 30°C – 25°C (*p=<0.05*).

2356

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*

Proportion of Pollen Tube vs Treatment for L.bienne

B – *L. usitatissimum*

2382

Proportion of Pollen Tube vs Treatment for L.usitatissimum
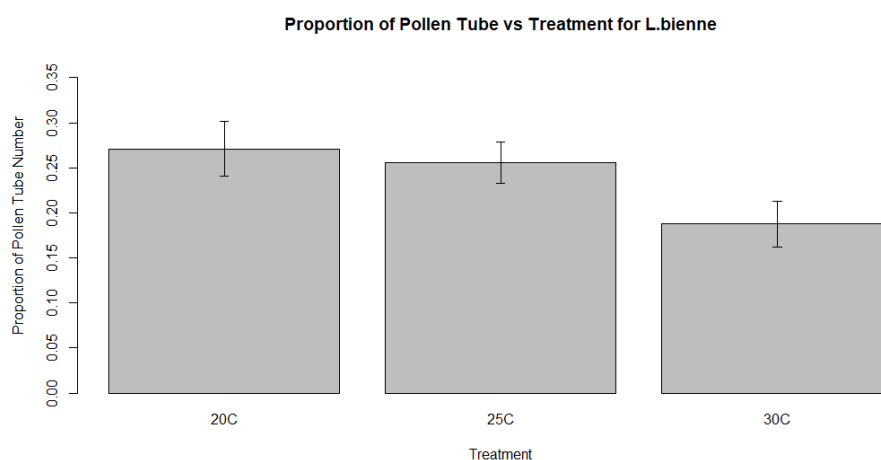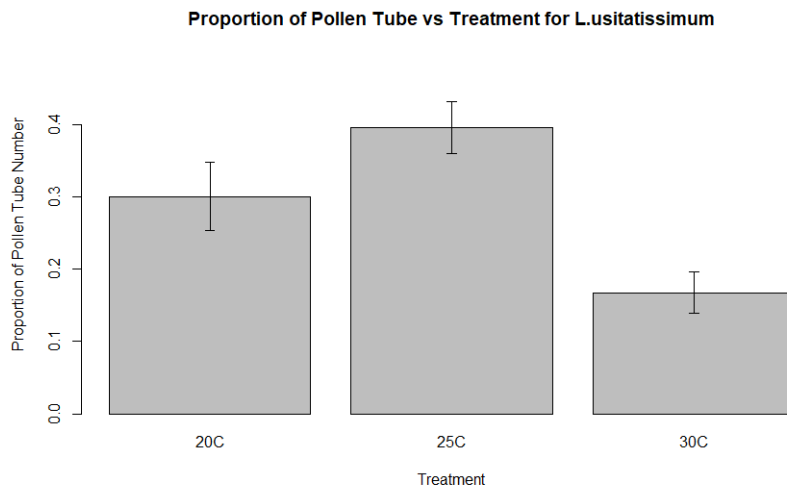
2383 Figure 8. Bar chart to represent the proportion of pollen tube forming against temperature treatment for the wild relatives
2384 *L. bienne* (8A) and the cultivar *L. usitatissimum* (8B).

2385

2386 **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

2387 **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**	*

2388

2389 Table 7 A and B. Summary tables of the linear model with Tukey's adjustments for the proportion of pollen tube formation
2390 between each *L bienne* (A) and *L. usitatissimum* (B). No significant differences were observed between the different
2391 treatments for *L. bienne*. There were significant differences between the 25°C – 20°C and the 30°C – 25°C treatments for the
2392 *L. usitatissimum* ($p=<0.05$).

2393

2394　　　　　To summarize the linear hypothesis for section 3.4.4 Results: "Pollen Tube Reaching the

2395　Ovary" and section 3.4.5 "Results: Pollen Tube Proportion Over Pollen Count for Different *Linum*

2396　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

2397　Table 8. Summary of linear hypothesis tests significance between different temperature treatments, ability of pollen

2398　reaching ovary and proportion of pollen tube forming for each species.

2399

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

2401　　　　　We evaluate the ability of pollen tube reaching the ovary for the different species. Evaluation

2402　of whether pollen tube was able to reach the ovary was applied as described under section "3.3.2

2403　Analysis: Proportion of Pollen Tubes" of this chapter. An initial Anova, revealed that the interaction

2404　between species and the treatments together were not significant (Treatment/Species *p = 0.2684*).

2405　The bar plot below (see Figures 9A) reveals a reduction in the proportion of *L. bienne* pollen able to

2406　reach the ovaries, under the heat (30°C) treatment.

2407　　　　　Summaries of the general linear models reveals that for both wild and cultivated species, the

2408　proportion of individuals with pollen tube able to reach the ovary is significantly affected by the heat

2409　treatments (30°C) (Tables 9A (*L. bienne*) and 9B (*L. usitatissimum*)). This was significant when

2410　compared against the cold (20°C) treatment (for *L. bienne p=<0.00179*, for *L usitatissimum*

2411　*p=0.00195*), and the glasshouse (25°C) treatments (for both species *p=<0.001* (<α of 0.05)). The

2412　result also shows some reduction on the proportion of individuals with pollen tube able to reach the

2413　ovary when observing the glasshouse and the cold treatments. However, this difference isn't

2414　statistically significant (*p=>0.05* in both cases).

2415

2416   **A – *L. bienne***
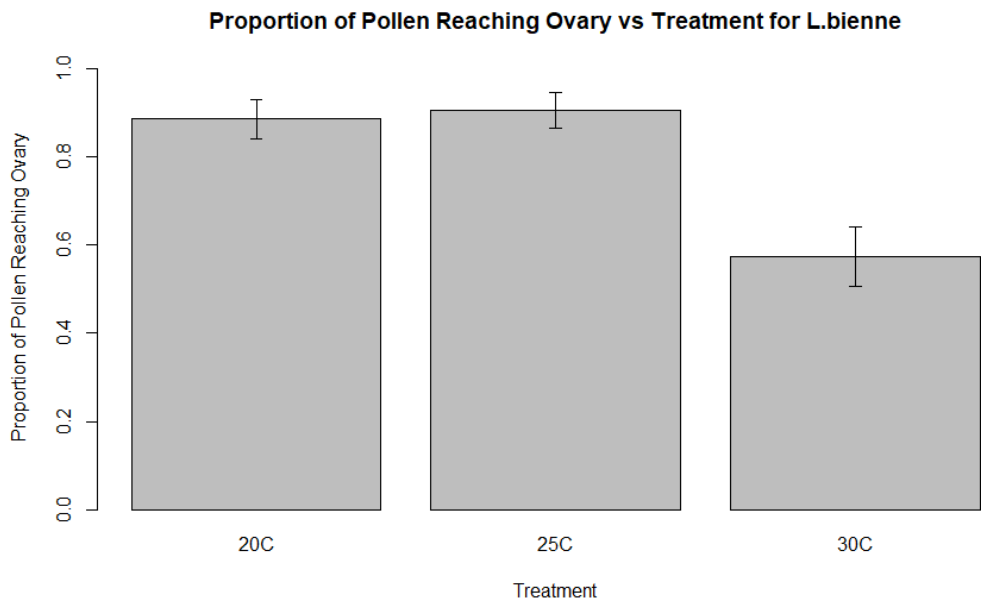
2417

2418   **B – *L. usitatissimum***
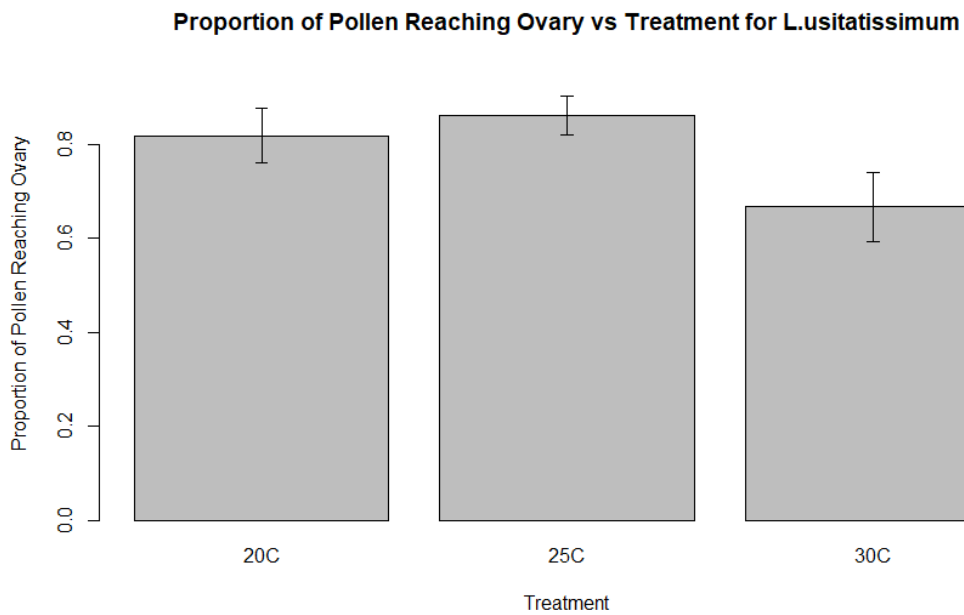
2419   Figure 9 . Bar chart for the proportion of pollen tube reaching the ovaries between different treatment for *L. bienne* (A) and

2420   *L. usitatissimum* (B).

2421

2422

2423

2424

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**	*

2426

2427 **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**	*

2428 Table 9A and 9B. Summary tables of the general linear model with Tukey's adjustment for the proportion of pollen tubes

2429 reaching the ovary between the different treatments for the different species. In both species the test shows a significantly

2430 reduced proportion of pollen reaching the ovary *(p=<0.05)*.

2431

2432 To summarize our results for section 3.4.6 "Results: Propotion of Pollen Tubes Reaching

2433 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

2434 Table 10. Summary of linear hypothesis tests significance between different temperature treatments and proportion of

2435 pollen tube reaching the ovary for the two species.

2436

2437 3.4.7 Results: Pollen and Latitude of Origin

2438 For the wild *L. bienne*, there was interest in looking at whether locality of origin of the wild

2439 plant had any roles to play on either the number of pollen count, the number of pollen tube observed

2440 and or the proportion of pollen tubes reaching the ovary. This is of interest, as this would add to the

2441 argument that local adaptation is present within our samples. Using Person's correlation and a linear

2442   modelling (represented by the blue line in Figure 10A), pollen count had no correlation with the

2443   latitude of locality of the different wild *L. bienne* samples within the study.

2444   The number of pollen count in the glasshouse and heat-treated plants seems to increase, the

2445   more northern an individual is localized (For glasshouse (*R=0.024, p0.86*), for heat (*R=0.1 p=0.45*)), but

2446   none of the correlations were significant (*all p values = >0.05*).

2447   Both glasshouse and cold treated individuals showed an increase in the number of observed

2448   pollen tubes, the more northern they are localized, but only the cold treatment was significantly

2449   correlated (*R=0.24, p=0.085* (glasshouse), *R=0.44 p=0.001* (*p=<0.05*)) (Figure 10B). This illustrates

2450   that, for our *L. bienne* individuals, northern individuals are more likely to form more pollen tube

2451   under cooler temperatures. When observing the heat treatment, it can be observed that correlation

2452   is opposite to that of the glasshouse and cold treatments (*R=-0.16 p=0.23*), all be it not significantly

2453   correlated. This change in correlation trend, shows some signs of the Northern individuals in this

2454   sample test, being more sensitive in their pollen tube formation to increases in temperature. This

2455   illustrates that the more Northern an individual within our *L. bienne* sample is localized, the more

2456   likely that they would favour cooler temperature for pollen tube formation.

2457   In Figure 10C, we plotted the proportion of pollen tube observed over the pollen count (in

2458   percentage) and against latitude. The trend illustrates the positive correlation in the cold treatment,

2459   of pollen tube amount against latitude is even more significant when considering the proportion of

2460   pollen tube able to form against the pollen count per individuals (*R=0.58, p=<0.0001*). This

2461   strengthens the point that the more Northern an individual is, the more able they are to form pollen

2462   tubes under the cold treatments.

2463

Pollen Count vs Latitude

Pollen Tube Count vs Latitude

Pollen Tube To Pollen Count Proportion (%) vs Latitude
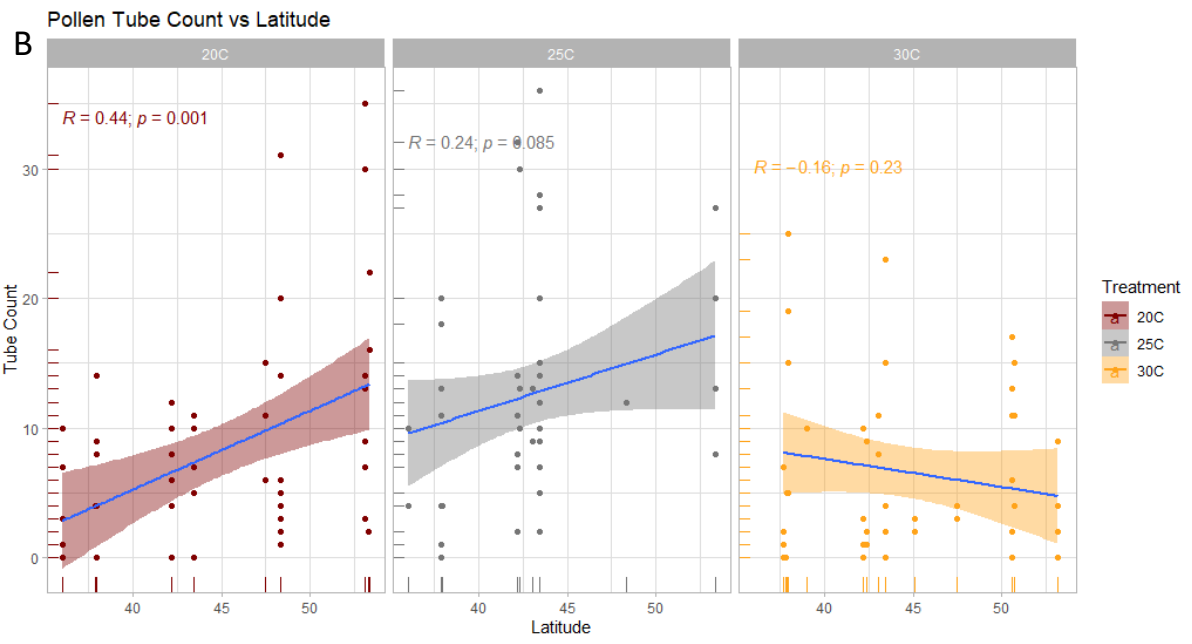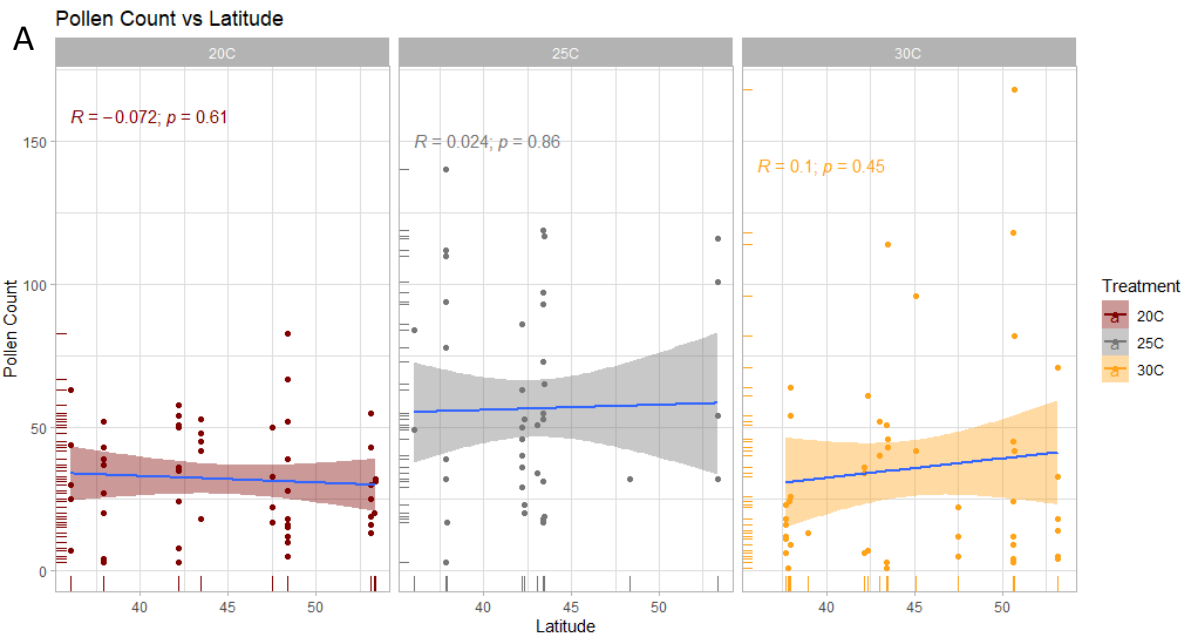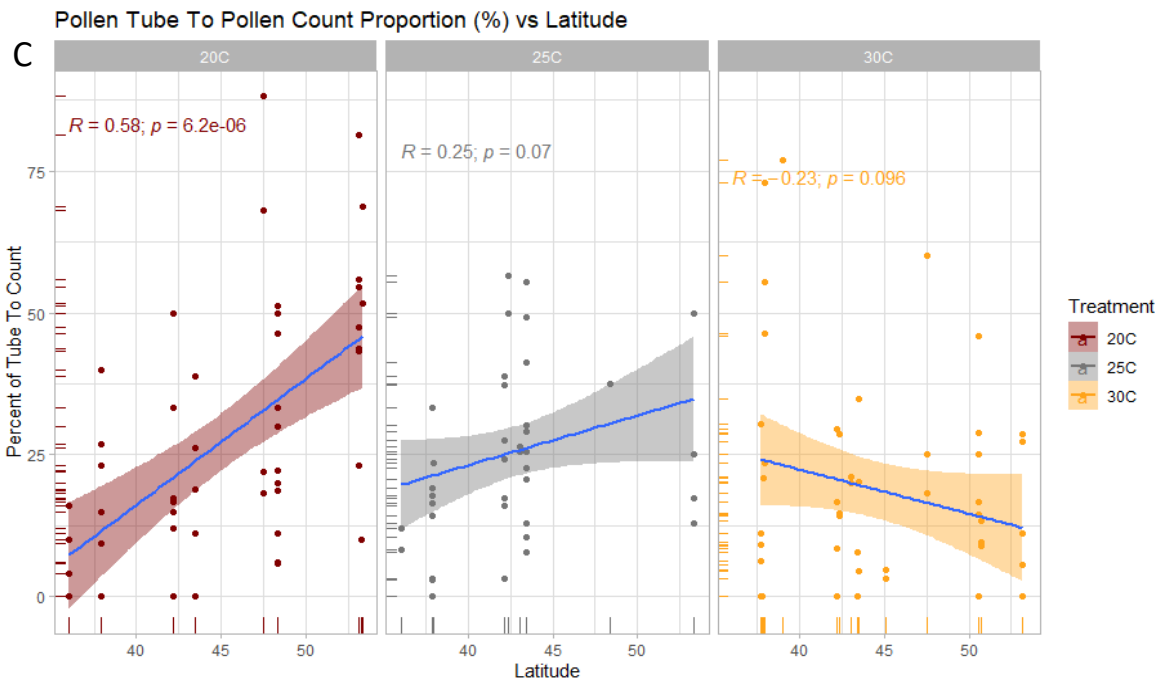
2466

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

2472

2473   With the latitude of localities showing significance on the ability of pollen to form pollen

2474 tubes under the cold treatment, it is possible to link pollen viability with climatic variables, especially

2475 variables such as average temperature. We suggest that more Northern individuals are varying

2476 sinigicantly in terms of pollen tube formations to the Southern individuals. We used summary PC for

2477 climate variables from chapter 2 to compare against pollen variables and climate on our wild

2478 samples.

2479   The number of pollen counts observed (Figure 11A) illustrates that there was no significant

2480 correlation between the pollen count and the principle component 1 (pc1) of the climatic variables

2481 (*p=>0.05* for all temperature treatments). This agrees with the findings for pollen count against

2482 latitude (see above section) whereby pollen count doesn't correlate with latitude of locality. The

2483 number of pollen tubes, reveals a difference in trend against the cold and glasshouse treatments

2484 compared with the heat treatment. With higher levels of PC1 observed, the less pollen tube is

2485 observed within the cold and the glasshouse treatment while under the heat treatment, the higher

2486 the level of PC1 for climatic variables the more pollen tube observations were made. However, just

2487 like under the findings against the latitude model, only the cold treatment showed a significant

2488 decrease in the number of pollen tubes (*R=-0.41 p=0.0026*).

2489   More significant correlations between pollen tube proportions and climatic variables can

2490 also be summarised by PC1 (Figure 11C). The correlations reveal the proportion of pollen tube

2491 decreases more significantly with an increase in the PC1 value under the cold treatment (*R=-0.54,*

2492 *p=<0.0005*). However, in contrast with the pollen tube count alone; when considering pollen tube

2493 formation in relation to the pollen count, the glasshouse treatment also shows a significant decrease

2494 in proportion in the event of an increase in PC1 value, a trend which was not significant under pollen

2495 tube count only. This reveals that proportion (in percentage) of pollen tube able to form were

2496 significantly correlated with climatic variables in the form of PC1 under both cold and glasshouse

2497 treatments. There was still a positive trend when considering the heat treatment. However, this

2498 remained unsignificant (*R=0.12, p=0.42*).

2499   A higher level of PC1 reveals a loading of warmer average temperatures for the local areas

2500 where the population were collected from in the wild (see section "3.3.3 Analysis: Population of

2501 Origin Variables"). Therefore, the significant decrease in the number of pollen tube count against a

2502 higher PC1 value reveals that at least in the cold treatment, the number of pollen tube able to be

2503 formed in our wild samples were significantly correlated with the higher temperature of their local

2504 climate. This result illustrates that under cold temperature treatments, individuals that are localised

2505   to a warmer climate are less able to form pollen tubes than that of individuals originating from

2506   colder climate.

2507
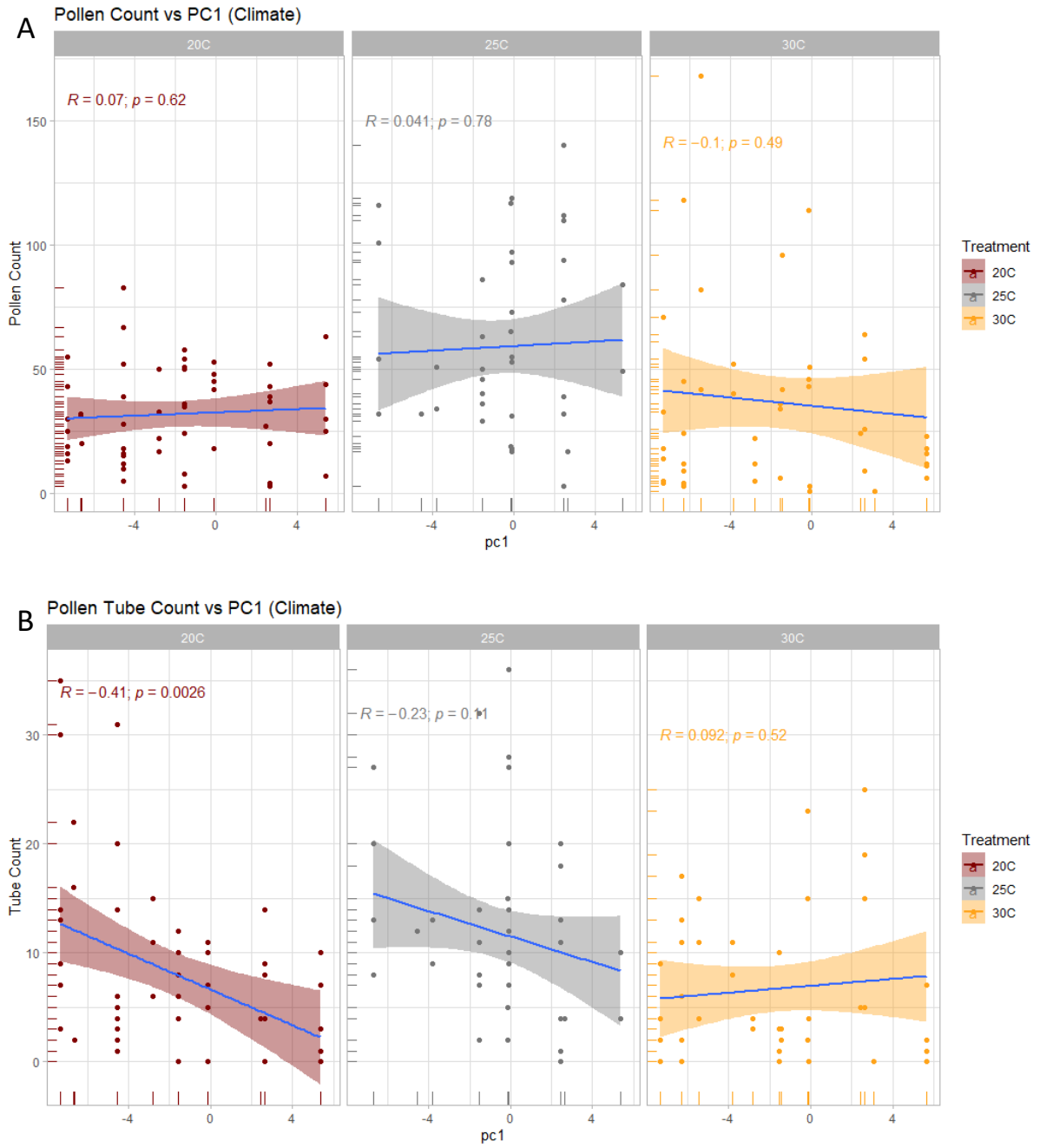
2508
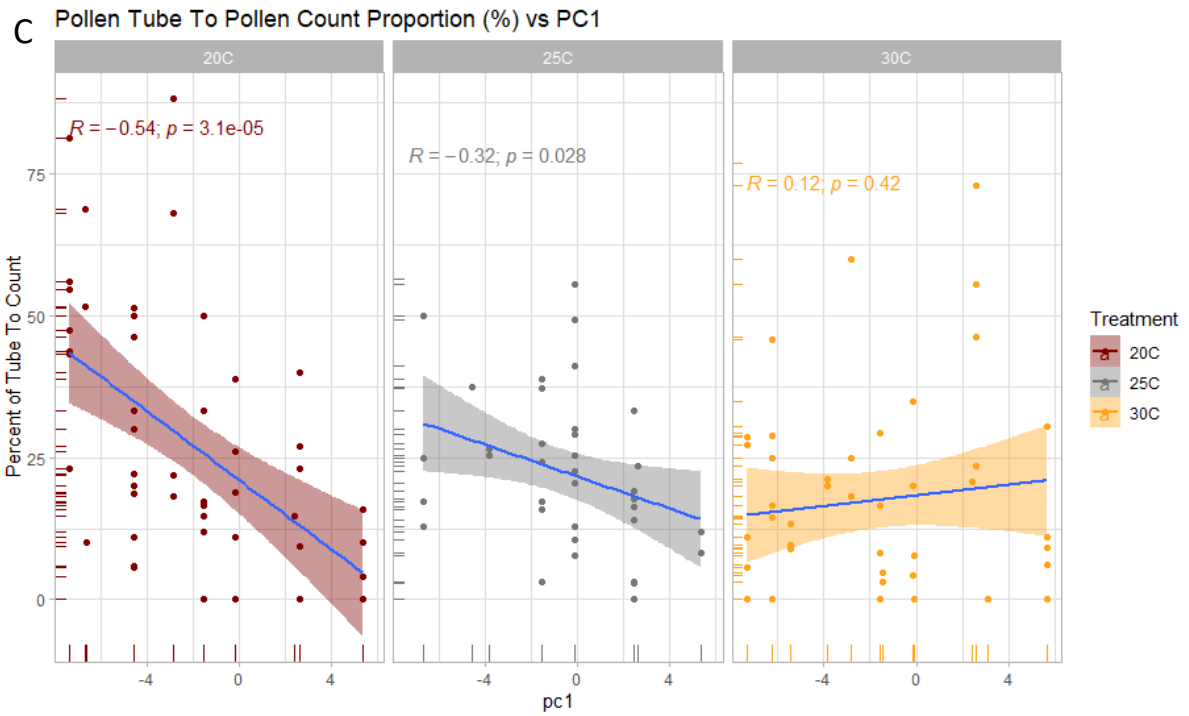
Figure 11. Pearson's correlations and linear modelling between pollen counts (A) and pollen tube counts (B) and climate PC1 values. The only significant correlation seen was between pollen tube count and PC1 values at cold temperature as seen in Fig 17 B (R=-0.41, p=0.0026, <α of 0.05). Pearson's correlations 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°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°C to 30°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°C daytime and 13°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

2584 because under the different temperature treatments, pollen tube proportions were significantly

2585 smaller in both cold and heat treatments with the cultivars. There was a more significant reduction in

2586 the proportion of pollen tubes than their wild ancestors. The effects of higher temperature stress

2587 negatively affected pollen tube formation more so than lower temperatures, especially in the

2588 cultivars. This effect of higher temperature stress on pollen viability is not new in plants. In chickpea

2589 (*Cicer arietinum*) for example, the pollen grains have been shown to be more sensitive to heat stress

2590 than their stigma counterparts (Devasirvatham et al., 2012). The fact that the crop wild relative *L.*

2591 *bienne* showed a potentially more resilient trait to rising temperature could be of interest in crop

2592 development. This importance has been highlighted in durum wheat (*Triticum*) before, which shows

2593 wild relatives yielding more grain under temperature stress when compared to the cultivar relatives

2594 (El Haddad et al., 2020). The fact that lower temperature affects pollen tube formation has been

2595 observed in other plants. Studies using pear also suggests disruption in pollen tube formation. They

2596 suggest that this is mediated by mitochondrial metabolic dysfunctions (Gao et al., 2014). The

2597 CBF/DREB1 proteins have also been identified to control expression of cold-induced genes

2598 (Thomashow, 1999). In Barley, novel alleles to frost resistance have also been identified as *FR-H2*

2599 amongst others (Sallam et al., 2021). There is potential that these cold resistance genes identified in

2600 other study models may play a role in higher tolerance to cold temperature observed in *L. bienne*

2601 pollen tube formation ability. In *Linum* this phenotypic difference in temperature stress pollen tube

2602 sensitivity may represent more diverse genetic resources for the improvement of crop heat stress

2603 resilience, as was reviewed in various wild relatives for crop improvements (Dempewolf et al., 2017).

2604 In terms of the ability of pollen tube to reach the ovary, the heat treatments revealed a

2605 significant reduction of pollen tubes that were able to reach the ovary. This was not the case in the

2606 cold temperature treatments. This suggests that an increase in temperature had a significantly

2607 negative effect on the ability of *Linum* pollen tubes to reach the ovary (Table 5). This result illustrated

2608 that *Linum* pollen's abilities to reach ovaries are more sensitive to an increase in temperature rather

2609 than a decrease. In a study with wheat (*Triticum aestivum* L.) pollen tube growth into the ovary was

2610 reduced due to an increase in temperature. It was suggested that this was because heat causes

2611 abnormal conditions for the ovary (Saini et al., 1983). In peach (*Prunus persica* L.), stigmatic receptivity

2612 to pollen tube was negatively affected by a rise in temperature (Hedhly et al., 2005). Suggestions in

2613 the literature often reveal negative effects of temperature increase on the female parts such as stigma

2614 and ovary that results in the observed reduced ability for pollen to reach the ovary. In this study

2615 however, we did not look at stigmatic and/or ovary 98arlierpment. In line with what is seen with the

2616 proportion of pollen tube reaching ovaries in this study, it would be of further interest to see how this

2617   negative response to the increase in temperature correlates to stigmatic and/or ovary development

2618   in our *Linum* samples.

2619   Furthermore, when analysed separately, wild *L. bienne* showed a more significant reduction

2620   in the number of pollen tube reaching the ovaries than those of their cultivar relatives. This trend was

2621   not expected since the proportion of pollen tubes to pollen count is significantly more reduced in the

2622   cultivar *L. usitatissimum* compared to the wild *L. bienne*. In the proportion of pollen tube able to reach

2623   ovary/species the wild *L. bienne* seems to have responded more strongly than their cultivar relatives.

2624   There is a suggestion here that even though wild pollen tube to pollen count proportion were higher

2625   than the cultivars, their ability to reach the pollen was more sensitive towards warmer temperature.

2626   Perhaps the pollen's female counterparts such as the stigma and the ovary were more sensitive to

2627   temperature changes in wild relatives (Fábián et al., 2019). A smiliar observation was also made in

2628   peach whereby stigmas tend to lose their capacity to support pollen tubes (Hedhly A. et al., 2005).

2629   Hedhly A. et al., 2005 also suggests that there are contrasting effect of temperature on the male and

2630   female parts of the flower. It maybe the case that, the amount of pollen tubes is affected by the ability

2631   of female counterparts, such as the stigma, to sustain pollen tubes at different temperatures. This

2632   aspect is yet to be explored in *Linum.* In this study, we showed that the heat (30°C) treatments cause

2633   a significant reduction in the number of *Linum* pollen tubes, and pollen tubes reaching the ovaries.

2634   However, this can be due to the sensitivity of female counterparts to rising temperature, as is evident

2635   in other species (Pan et al., 2018; Fábián et al., 2019). In future studies, it would be interesting to

2636   observe the effects of temperature on the growth rate of *Linum* pollen tube inside the stigma. This

2637   can be done using live images from confocal fluorescence microscopy on *in vitro* germination studies.

2638   A major aim in this thesis was to look at local adaptations, that was done here by a comparison

2639   between pollen measures and latitude of locality and climatic variables. We were interested in

2640   whether there was any correlation between the pollen count, pollen tube count and/or the proportion

2641   of pollen tube forming against local geographic and climatic variables. None of the different

2642   treatments shows a correlation in latitudinal and climatic variables when considering pollen counts.

2643   This suggests that the number of pollen in our sample set is neither positively nor negatively correlated

2644   with local climates. Other study observing percentages of pollen across the Mediterranean region

2645   found ecological trends in not only pollen data but also other plant traits to climatic gradients across

2646   the region (Barboni et al., 2004). We expected to see a trend in *Linum* pollen count with climatic

2647   variables, however, we didn't observe a significant trend in our sample set. Although this was the case

2648   for our study, there was some evidence of a change in direction of correlation trend in the number of

2649   pollens observed and local climatic variables, when looking at the different temperature treatments.

2650   This correlation was not statistically significant but perhaps with a larger sample set and from a more

2651   varied local environment, the trend would be more significant and more obvious, and a larger sample

2652   set would help in statistical power as well. Another study suggests that 'fresh' pollen should be used

2653   for observation studies as pollen conservation may impact germination abilities of pollen (Beltrán et

2654   al., 2019). Although laborious, it would be of interest to use 'fresh' pollen instead of a preservation

2655   method, as was done in this study. The number of pollen observed under different climatic variables

2656   is important as this may influence the possibility of gene flow between different populations and

2657   species. As *Linum* is a self-pollinating species, it is perhaps plausible that environmental segregation

2658   of populations by pollen production could influence gene flow dynamics as well as pollen/pistil

2659   incompatibility as was shown post-pollination in *Polemoniaceae* (Ruane & Donohue, 2007).

2660   When observing pollen tube count however, there was a significant association when looking

2661   at the cold treatments. This observation was true for both the latitudinal and climatic variables. Even

2662   within our limited sample sets, there was a latitudinal trend that, the more northern our sample was

2663   collected from, the more pollen tube it can produce under colder temperatures. The correlation with

2664   climatic variables also agrees with this observation. Local climate may, in the case of wild *L. bienne,*

2665   *influence* pollen viability, not just in terms of temperature, but also other climatic viables which may

2666   have played a role. In oak, pollen viability and sunlight availability in the local areas possibly influence

2667   pollen mediated gene flow between populations (Schueler et al., 2004).

2668   In addition, more northern wild population showed potential local adaptation with higher proportions

2669   of pollen tube forming observed in the cold treatments (Table 6A). This suggests that wild *L. bienne*

2670   are perhaps more adapted to colder temperature. There is potential for GWAS study here for Northern

2671   wild *L. bienne* to identify novel genes for cold tolerance. This can have implication towards

2672   improvement of cultivars. There was also a difference in the trend with heat treatments. When

2673   heated, the correlation changed direction between the latitude of localities and climatic variables and

2674   pollen and pollen tube counts. This was not significant but could be tested with a larger wild individual

2675   from a wider range of local variables in the future.

2676   In conclusion, different temperatures affected male reproductive function in terms of pollen

2677   number, pollen tube number, proportion of pollen tubes forming against pollen count, and proportion

2678   of pollen tubes reaching the ovary. In all cases there seemed to be an adverse effect of increases of

2679   temperature, be it pollen, and pollen counts and the amount of pollen tubes able to reach the ovary.

2680   Therefore, in our samples, an increase in temperature lessened pollen viabilities for *Linum*. There was

2681   also evidence that in the colder temperature treatments, pollen viability is correlated with latitude of

2682   localities and climatic variable. Local adaptation in pollen viability may have resulted in the differences

2683   seen. Additionally, we would like to observe the female counterpart of these *Linum* samples do in

2684   relation to the pollen viability trend that was seen in this study. Further study on the reception of

2685   female structures, such as the the ovary may reveal that, in line with the trend seen in pollen viabilities
2686   in *Linum*, the female counterparts also play a major part. Illustration in Sorghum have observed
2687   deterioration of ovary under heat stress (Chiluwal A. et al., 2020). There is also interest in looking at
2688   further colder treatments for the wild population to see whether there is optimal temperature for
2689   pollen viability and to generate results for more samples from more diverse latitudes. The differences
2690   seen here are evidence of local adaptation in the wild *Linum* species. More northern individuals seem
2691   to be more locally adapted to colder temperatures and could be more adversely affected by an
2692   increase in temperature in terms of their ability to form pollen tube and the ability for pollen tube to
2693   reach ovaries.

2694

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

2728 genetic bottlenecks events which revealed consequences of cultivation of the species (Guan et al.,

2729 2021). This is also of interest in terms of molecular ecology. Genetic analysis has previously

2730 identified localization and invasion histories of non-model plants (Hernández et al., 2019). To look at

2731 differences across population, it is of interest to establish a structure and diversity between the local

2732 population.

2733 In *Linum,* previous research used Single Sequence Repeats (SSR) techniques. In the research,

2734 SSRs were developed for 34 Turkish wild flax (*Linum bienne*) accessions and accessions from

2735 different cultivated flax (*Linum usitatissimum*) lines. The research found variation in polymorphisms

2736 between the accessions and clustering of 493 individuals according to their respective types,

2737 suggesting wild flax are more closely related to the deshiscent type. Along with this, the research

2738 also suggests genetic distancing among the wild types are significantly related to their geographical

2739 distances as well as their elevation (Uysal et al., 2010). The research revealed potential for marker

2740 studies using *Linum* types to understand genetic structuring, domestication, and genetic diversity of

2741 *Linum*. For this purpose, short reads within the whole sequence were used as a genetic marker for

2742 different individuals. This is substantially easier than obtaining a whole genome, as a whole genome

2743 for the wild *Linum* species is not necessary for observation of population structure based on these

2744 markers. This being the case for the wild species, a chromosomal-level whole genome is readily

2745 available for the cultivar relatives *L. usitatissimum* (Sa et al., 2021). In this chapter, we use ddRAD

2746 sequencing to investigate whether there are structuring, diversity and or divergence within our

2747 *Linum* samples.

2748 We have a collection of wild *L. bienne* and several oil and fibre cultivar variety of *L.*

2749 *usitatissimum* originating from Western Europe. Within the *Linum* samples available under this

2750 project, there were a variety of individual plant populations that originated from different places

2751 across the latitude of Western Europe. We hypothesize that there is genetic variation due to

2752 localization within different environments. This could be implied by structuring and genetic diversity

2753 measures which may have taken place across time as different population form and localize to their

2754 local environment. When looking at the collection of *Linum* samples from across western Europe,

2755 genetic structure would be revealed as the genetic distance between the different *Linum*

2756 populations as well as between the wild and cultivar types. This would help in explaining their

2757 population genetics and as well as their breeding systems. There will be an interest in answering the

2758 question of "can we use genetics to interpret population structures and breeding systems in

2759 *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

2794 in our samples for the implication of genetic distinction within our *Linum* populations.The

2795 expectation in the *Linum* sample, is that individuals would be grouped based on their population and

2796 location, and that grouping between the northern and southern populations would match what was

2797 previously observed in previous chapters which have examined the Linum samples morphologically.

2798 This would reveal that, structurally, northern, and southern populations in the Linum collection are

2799 genetically distant to each other. Further analysis could explore genetic traits such as heterozygosity

2800 and whether any population is genetically more diverse than other populations in the sample set.

2801 This would highlight potential breeding system differences in our Linum samples.

2802

2803

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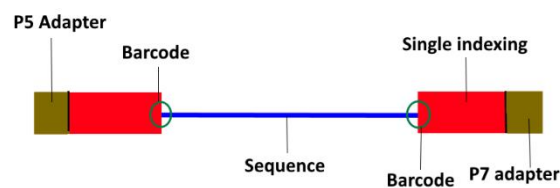
Figure 1. An illustration of the ligated DNA material per sample which are both adapted and barcoded.

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

2831   wild species collected directly from the field. A modified CTAB DNA extraction protocol was used

2832   from methods described by Soltis lab (Doyle, 1987).

2833   We modified the main CTAB washing buffer for DNA extractions. The presence of salt and

2834   polysaccharides could potentially interfere with downstream processes. To reduce this interference,

2835   suggestions were adapted from Clarke *et al* (Clarke J., 2009). The final modifications to the samples

2836   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%

2837   Table 1. Reagents for the CTAB extraction buffer

2838   With the CTAB buffer made, the protocol were carried out as described by Doyle (1987).

4.2.3 The Modified CTAB protocol

2839

2840   In the laboratory, dry baths were pre-heated to 60°C and RNAse were removed and ice-

2841   thawed from -21°C storage. In the glasshouse, 10-12 young, 'green' leaves were collected per

2842   individual. Samples were individually homogenised. Alternative homogenising methods were also

2843   initially used, using 3mM metal beads, a pinch of sand and a Tissue-Lyser II. However, access to

2844   Tissue-Lyser II were limited due to Covid-19 restrictions and most samples were ground using a

2845   tissue homogeniser. Samples were ground as finely as possible inside Eppendorf tubes and flash

2846   frozen in liquid Nitrogen until all samples were ready.

2847   Per sample, DNA was cleaned as per Soltis lab protocol (Doyle, 1987). Samples were

2848   homogenised further with a vortexer until ground tissues are thoroughly mixed with the CTAB and

2849   RNAse A solution. Samples were then placed on a 60°C pre-heated dry bath for at least 60 minutes

2850   (modified to allow time for optimal DNA lysis). Samples were then thawed to chill on ice for 3

2851   minutes before further processes.

2852   Samples were further washed using phenol-chloroform solutions as per protocol. To wash

2853   away any potential phenol after the wash, 200μL of chloroform isoamyl were added. Samples were

2854   centrifuged at 14,000RPM for 10 minutes before carefully removing the aqueous layer into newly

2855   labelled Eppendorf tubes. Na Acetate were then added as per protocol. Salt was washed away at

2856   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°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°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

2889   Biolabs). Two different combinations were tested: *Mse1/Pst1* and *Mse1/Sbf1* with the digestion

2890   protocol using NEB's buffer 2.1.500ng genomic DNA and 5µL of buffer 2.1 was used as

2891   recommended by the supplier for optimum enzyme efficiency. We prepared a 30µL reaction. A

2892   master-mix of the buffer and enzymes were prepared for every sample. For incubation, and

2893   inactivation purposes, a Prime Techne thermal cycler was used. A 30µl digestion reaction were

2894   prepared as per manufacturer's protocol.

2895   ### 4.2.5ii RAD – Anneal adapter and Ligation

2896   Digested DNA was subsequently ligated by using a ligase enzyme (manufactured by New

2897   England Biolabs). During ligation, digested DNA was adapted and barcoded specifically. To do this,

2898   adapter oligos were first annealed together per manufacturer's protocol.

2899   After mixing briefly adapters were annealed using a protocol recommended by Eurofin

2900   Operon. In a thermo-cycler, adapters were heated to 95°C for 2 minutes and ramp cooled to 25°C

2901   over a period of 45 minutes. Annealed adapters were then stored in a -21°C freezer for long term

2902   and in a 4°C fridge for short term (<72 Hours). The process was repeated for every adapter available.

2903   Samples and adapters were now ready for ligation stage. In the experiment, one forward and one

2904   reverse working stock adapters were combined with fragmented sample DNA and ligated to barcode

2905   the samples. New England Biolab's T4 ligase enzyme at 400U/µL concentration and a 10× T4 ligase

2906   buffer were used for this. The samples were loaded according to the 30µL final reaction volume from

2907   the digestion steps. The following tables describes each component for ligation. Samples were

2908   incubated at 23°C for 30 minutes and then heat inactivated at 65°C for 10 minutes. Samples were

2909   then cooled down at 2°C per 90 seconds until they reached 23°C.

2910

2911   ### 4.2.5iii RAD – Pooling and Cleaning Ligated products

2912   To pool and clean ligation products from primers, SeraMag magnetic cleaning beads were

2913   used. This step required all samples to be pooled into a single container. Total volume of the pooled

2914   samples was calculated as the volume of leftover ligation product after PCR (PCR product) × total

2915   amount of samples. 1.5mL Eppendorfs were used for pooling and cleaning. For each tube, the

2916   volume of pooled ligated product would be no more than 300µL, to be cleaned separately.

2917   Before using the SeraMag beads, working solutions were prepared per manufacturer's

2918   instructions. 1mL of the working solution were then transferred into a 1.5mL Eppendorf tube and

2919   further processed as per manufacturer's instructions.

2920   After the final magnetic stage, tube with the magnetic beads and 1mL TE were put set in a

2921   rack (non-magnetic) and PEG-8000 and Tween 20 solution was added per manufacturer's

2922   instructions. When the solution is mixed, 1mL of the previously prepared 1mL TE + SeraMag beads

2923   were pipetted into the conical flask. Conical flask was then filled up to 50mL volume with MilliQ

2924   water and mixed gently until the beads are evenly spread across the solution. The solution were

2925   then transferred into a 50mL Falcon tube, wrapped in aluminium foil, and stored in a 4°C fridge

2926   ready for use.

2927   Before using on ligated products, the cleaning beads were tested against a 100-1Kb DNA

2928   ladder (BioLine's HyperLadder). After confirmation against DNA Ladder (different ratio of beads

2929   should cut at different fragment sizes), the prepared Sera-Mag beads were then ready to be used for

2930   the ligated products. To do this, clarification needed to be made about the ratio of beads to be used

2931   for the fragment size which are of interest. In the case of this experiment, 250-500bp regions were

2932   of interest. A 1.5× SeraMag to ligated product volume ratio were used.

2933   After the addition of the beads working solution, solution was incubated at room

2934   temperature for 5 minutes. After this, samples were placed back into the magnetic stand and

2935   processed as per manufacturer's protocol. The produced DNA materials were then washed and dried

2936   per manufacturer's instructions. This wash step was repeated twice. When the second wash

2937   finished, beads were placed on a 37°C heat block until beads were dry. Rehydration took place with

2938   20μL of nuclease free water as elutants. When the beads were mixed in the water, samples were

2939   placed back in the magnetic stand until all beads were pulled towards the magnet. When all beads

2940   were pulled towards the magnet, the supernatant were then extracted and transferred to a new

2941   1.5mL Eppendorf tubes or into the next 300μL pooled tube until all were cleaned. Cleaned samples

2942   were then quantified using a NanoDrop ND1000 spectrometer.

2943

2944   <u>4.2.5iv RAD – PCR Amplification to Generate Illumina Sequencing Libraries</u>

2945   High-fidelity PCRs were undergone with all the samples for ligation verification and

2946   generation of sequencing libraries for Illumina. ThermoFisher's Phusion High Fidelity DNA

2947   Polymerase were used for this protocol with P1 and P2 adapter oligos as the primers. A total volume

2948   of 20μL reactions were prepared as per manufacturer's instructions. For this protocol, 20ng of input

2949   DNA were used. The input DNA into the digestion product was 500ng. 20ng in volume of DNA were

2950   calculated as follows:

2951 $$Input\ DNA = \frac{500\ (digestion\ DNA\ input)}{20(DNA\ required\ for\ PCR)} = 25$$

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

2952

2953 The input DNA from ligation were calculated as 1.5μL. The samples were then run on the 111arlier-

2954 cycler for 12 cycles of the following programme: *98°C for 2 minutes, 98°C for 10 seconds, 65°C for 30*

2955 *seconds, 72°C for 30 seconds and a final 72°C for 10 minutes*. After the cycles were done samples

2956 were held at 4°C. A 1×TBE gel electrophoresis was performed to check the presence of products

2957 against a DNA ladder.

2958

2959 4.2.5v RAD – Pippin size selection Preparation

2960 To construct a ddradseq library of 250-500bp fragments, a Pippin Prep facility was used

2961 (Sage science). A 1.5% Agarose gel cassette (with marker L) was used for Pippin preparation using

2962 manufacturer's protocol. The instrument was calibrated prior to the run as per manufacturer's

2963 protocol.

2964 To run the Pippin size selection, the cassette was sealed with the provided seal. The

2965 automatic test was run and the current measured in each elution channel at room temperature. A

2966 size selection protocol was then manually inputted or selected from the "Protocol editor" tab with

2967 sample ID provided. Under the "Cassette" input, an appropriate cassette file was loaded and under

2968 the "Reference Lane" input, the marker well was inputted and applied to all lanes. The pippin was

2969 subsequently run for each sample well to select at 250-500bp for the purpose of this experiment.

2970 When the Pippin prep finished running, 30μL of sample was collected from the elution module. An

2971 Agilent Tapestation 2200 facility was used to identify and quantify the region of interest, with a High

2972 SensitivityD1000 ScreenTape.

2973

2974 4.2.5vi RAD – qPCR Quantification

2975 For a final quantification, a quantitative Polymerase Chain Reaction (qPCR) procedure was

2976 explored to quantify DNA materials. In this part of the quantification a BioRad CFX96 Real-Time PCR

2977 System was used. For this purpose, 4 dilutions of library materials were explored. The dilutions were

2978 made using previous pooling from section "4.2.5iii RAD – Pooling and Cleaning Ligated products" of

2979 the RAD methods, each with the appropriate amount for the pooled DNA in 10mM Tris-HCL solution

2980 as required per dilution. All dilutions were repeated three times. No template controls (NTC) were

2981 used as a control for contamination in the qPCR runs. There were 4 standard dilutions used as a

2982 standard in the Bio-Rad system for quantification purposes for each of the dilutions. In total there

2983 was 48 wells for the samples and dilutions, 3 NTCs, and 4×3 wells for the standards. This makes up to

2984 63 wells for the run. For the standards, we used 4 concentrations (10pMol, 1pMol, 0.1pMol, and
2985 0.01pMol).

2986 For qPCR purposes, a KAPA Library Quantification Kit (Roche Molecular Systems) were used. When
2987 reagents were fully made, a 96-well qPCR plate were designed. Standards and samples were loaded
2988 as per manufacturer's instructions. An appropriate cycle programme was then set up and ran.

2989   After the protocol had finished running, a "Quantification CQ result" can be seen. CQ stands
2990 for quantification cycles, in other qPCR machines these are also specified as cycle-threshold. We
2991 used starting quantity (SQ) values to calculate the final concentration for each dilution factor. For
2992 example, for a dilution factor of 1:2000, the SQ-value were multiplied by 2000. Size corrections were
2993 able to be calculated by multiplying the average fragment length (250-500bp) by the standards
2994 (452bp). Units were able to be converted from picomolar to nanomolar by multiplying by 1000 and
2995 subsequently nanomolarity to a concentration unit (ng/μL). Using the delta-CQ values, consistencies
2996 between the triplicates were checked before calculations of concentrations.

2997

2998 ### 4.2.5vii RAD – Sequencing
2999   When satisfied witih quality checks, an Illumina 2500 Hi-Seq sequencer was used at Durham
3000 University's DBS Genomics facility to sequence the ddradseq library. An appropriate amount of
3001 pooled library was calculated for sequencing. All the DNA library poolings from the previous steps
3002 were combined into 1.5mL Eppendorf tube. Illumina Hi-Seq 2500 requires certain amount of
3003 genomic DNA to be read successfully. DBS Genomics at Durham University's Biosciences department
3004 recommends at least 10ng/10μl concentration of processed DNA library to process the samples
3005 through the sequencer. After quality check measures were made and 2×10μL tubes of sample were
3006 sent through the Illumina Hi-seq 2500 sequencer measuring at >10ng/μL each. An Eppendorf
3007 centrifugal vacuum concentrator were used as required to combine and concentrate pooled libraries
3008 to achieve a 10ng/μL threshold required for sequencing. After the sequencing process, two raw
3009 library files which needed further demultiplexing and mapping were outputted. The raw sequences
3010 consist of a forward and a reverse sequence for the pooled library which needed to be processed to
3011 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

3044   default parameter filters, 115/121 individuals were kept, 6 individuals were filtered out due to either

3045   excessive number of uncalled bases or low read qualities.

3046

3047   4.3.3 Analysis: Mapping to a Reference Genome

3048   Post-demultiplexing, samples were able to be mapped either without a reference genome

3049   (*de_novo*) or with a reference genome. We were interested in GWAS analysis, whereby a reference

3050   genome was essential. The sample set in this case was inclusive of cultivars (*L. usitatissimum*), and

3051   the wild pale flax (*L. bienne*) individuals which are widely regarded this wild species whereby the

3052   cultivars were domesticated from (Allaby et al., 2005). These are amongst the justification for opting

3053   to map to a whole genome which was already available for the cultivars of flax but not for the wild

3054   type.

3055   The Genome for the cultivar is publicly available and are assembled down to chromosomal

3056   levels, making it suitable to be mapped to (Sa et al., 2021). Mapping to a reference genome were

3057   undergone using the ref_map.pl programme of the STACKS v2.61 software (Catchen et al., 2013).

3058   The programme calls on each of the STACKS components noting "gstacks" as the component which

3059   align the positions for the RADSeq reads and calls SNPs in each of the sample based on the

3060   alignments. The programme expects the data to have been previously aligned with the genome

3061   using a separate aligner. For this purpose, BWA_MEM algorithm of BWA v0.7 was used to align the

3062   RADSeq reads to the cultivar genome (Li & Durbin., 2009). The outputs were pipelined to Samtools

3063   v1.15 to be converted to .BAM files which are the input file type required for ref_map.pl to work.

3064   BWA and Samtools were piped under one command. The command can be found under (appendix

3065   18). After obtaining .BAM files for each of the samples, ref_map.pl can be executed using the

3066   "ref_map.pl" command. The optional command to execute the "populations" programme of STACKS

3067   were also piped into the ref_map.pl command. This is a filter to accept a minimum percentage of

3068   individuals in a population to process a locus. This was set to 80% by using the command "-r 0.8".

3069   The populations command also allows further file output options such as the PLINK and VCF formats

3070   to be processed further downstream with a population genetic specific software. A population map

3071   was also specified to map which samples belongs to which population as well as to map a

3072   geographical region specific to that population. This is useful for further analysis downstream. A

3073   population map file for the *Linum* samples is listed as in appendix 19. The full command for the

3074   ref_map.pl pipeline can be found under (appendix 20). Post ref_map.pl run, several outputs will be

3075   of interest. The first if the VCF output which can be used for analysis quality of the reads such as the

3076   read depths, allele frequency and PGD-Spider (Lischer and Excoffier, 2012), PLINK (Purcell et al.,

3077   2007) or vcftools (Danecek et al., 2011) was used to convert the SNP data into PED and MAP.

3078

4.3.4 Analysis: Processing Mapped Reads

3079

3080   To process the mapped reads an initial step was to convert .PED and .MAP output files into

3081   .RAW and additional .BED files for read inputs into further downstream processes. To do this, PLINK

3082   v1.90 was able to be used (Purcell et al., 2007). PLINK was able to specify .PED and .MAP files in a

3083   directory given the same name. The additional command "--recodeA" is a data management options

3084   whereby both .PED and .MAP files were able to be processed into a single .RAW data file which

3085   includes formatting that are useful for further population genetics analysis. For further analysis a

3086   .BED file format were also required. This is obtained through the "--make-bed" addition to the PLINK

3087   command. There were numerous (>1) chromosomes in the .MAP files. This means that PLINK needs

3088   to acknowledge and allow for this in the output .RAW file. To do this the "--allow-extra-chr" option

3089   for PLINK were inputted into the command line for both make-bed and .RAW file options. The full

3090   command for PLINK conversions can be found under appendix 21. After both .RAW and .BED files

3091   were constructed, further population analysis could take place. In this case downstream analysis

3092   would be carried out using a programme written in "R". R was run in Durham University's HPC server

3093   for the purposes of this analysis.

3094

4.3.5 Analysis: Processing Data in SambaR

3095

3096   To further the analysis of the mapped reads, output .RAW and .BED files were further

3097   processed through using a programme constructed under "R". The programme is called "SambaR"

3098   (Snp 115arl Management and Basic Analyses in R) and is available for open use with no requirement

3099   for a licence. SambaR functions to integrate numerous R packages such as "Adegenet", "poppr",

3100   "FactoMineR", et cetra. SambaR functions as "collections of functions which increase the power of

3101   existing R tools for population-genetic analyses" (De Jong et al., 2021). This convenience makes

3102   SambaR an ideal tool to save time with running population genetics analysis, which can include

3103   population structure, diversity, and selection analysis. SambaR were run in the "R" environment;

3104   therefore, the statistical tool R is essential for further downstream processing. For the analysis of the

3105   data within this thesis, R (version 4.1.3) were loaded from the HPC server into the environment.

3106   SambaR could be manually downloaded using from the official Github page

3107   (https://github.com/mennodejong1986/SambaR) and loaded into R manually. Alternatively, it could

3108   be loaded into R using the command

3109 "source("https://raw.githubusercontent.com/mennodejong1986/SambaR/master/SAMBAR_v1.07.tx

3110 t")". This command executes a source code from the Github online server where the source code for

3111 SambaR can be found. Further downstream SambaR commands worked after the source codes were

3112 executed. R packages were able to be automatically checked and added through the "getpackages()"

3113 command. This command outputted a .txt file whereby each packages required could be checked as

3114 been successfully loaded or not. In the case of this analysis, all recommended packages were able to

3115 be loaded in addition to the essential packages to run the SambaR process. Data can be manipulated

3116 easily in SambaR using several pre-loaded R packages.

3117 SambaR also accepts a geographical input file. This is useful for implications of the

3118 geographical range within our samples. From collection, the samples within the pooled library

3119 should have a latitudinal range across Western Europe. This provides latitudinal cline for structure

3120 analysis. SambaR uses the function 'getMap' of the R package rworldmap-1.3.6 (South, 2011). The

3121 addition of pie charts was included for mapping in SambaR, using the function 'add.pie' of the R

3122 package mapplots-1.5.1 (Gerritsen, 2018). An input .txt file, consisting for 3 tab-separated columns

3123 were created using the geographical information of the different population observed within the

3124 sample. The information for this file is available in appendix 22. As an addition to SambaR's

3125 "importdata()" command, the option "geofile=fileprefix" were included.

3126 For different analysis, data were processed differently. For genetic structure analysis it is

3127 better to have as many individuals and populations as possible to infer better structuring between

3128 the populations in the collections. However, for population diversity and differentiation analysis, we

3129 determined that it was better to include as many SNPs as possible. Therefore, filters that enables to

3130 retain as many SNPs as possible were desirable during population diversity and population

3131 differentiation analysis. This may in turn reduce the number of individuals in the samples, thus,

3132 number of populations may be reduced under the filters for population diversity and population

3133 differentiation. For recommendations, SambaR recommends not using snpmiss=>0.05 for population

3134 structure analysis while for diversity and selection analysis, SambaR recommends higher snpmiss

3135 parameters.

3136 4.3.6 Analysis: Data Management in SambaR

3137 SNP data management and analyses were performed in R-4.1.2 (R Core Team, 2022) using

3138 wrapper functions of the R package SambaR (Github page:

3139 https://github.com/mennodejong1986/SambaR). The data was then imported into R and stored in a

3140 Genlight object using the function 'read.PLINK' of the R package adegenet-2.1.5 (Jombart, 2008;

3141 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.

3174
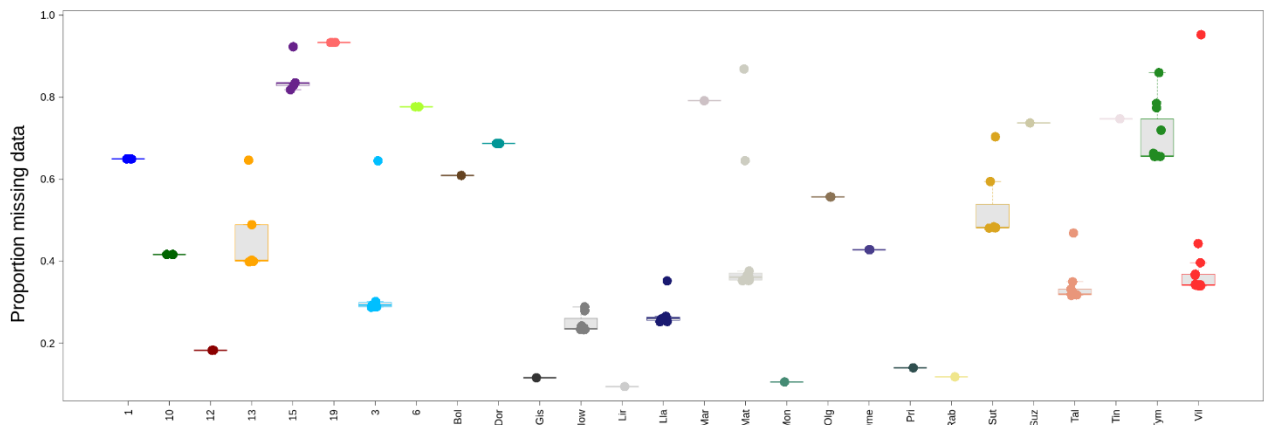
3175   Figure 2. Boxplot to show the proportion of missing data after the filtering procedure over populations.

3176

3177   The "*dohefilter*" option in the "*filterdata()*" command refers to filtering of function which

3178   remove SNPs with heterozygosity levels which are potentially indicative of paralogs. Paralogs are

3179   genes that are present in a particular organism which are related to each other through gene

3180   duplication events (Koonin., 2005). Paralogous genes are often exacerbated in plant genetics

3181   because the events of gene duplications such as polyploidy is more prevalent in plants (Mastretta-

3182   Yanes et al., 2014). Paralogous genes have also been found to bias population genetic estimates,

3183   which will affect downstream population genetic type of analyses (Verdu et al., 2016). Since *Linum*

3184   was found to have undergone polyploidy, the identification of paralogs due to heterozygosity may

3185   be an issue for further population genetic type analysis (Sveinsson et al., 2014). The identification of

3186   paralogs was able to be inferred from locus specific heterozygosity against the locus specific minor

3187   allele frequency. The plot below (Figure 3) indicates these locus specific frequencies and reveals the

3188   amount of data which SambaR were able to filter out, due to the identification of paralogs: (The SNP

3189   dataset in red were subsequently removed after the "dohefilter" option was set to "TRUE").

3190

3191

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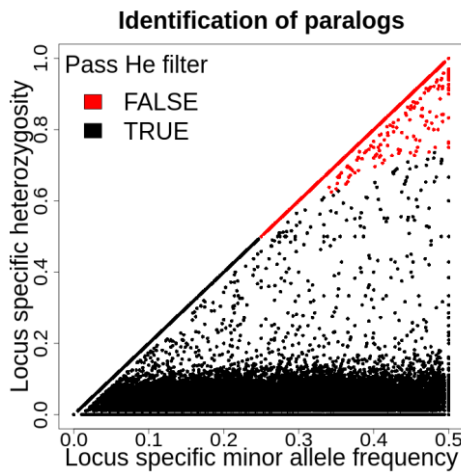
Figure 3. A scatterplot to reveal Sambar's identification and filtering of paralogs.

3198

3199

3200   In this case, SambaR's default parameters for removing the paralogs were only removing a

3201 small proportion of heterozygosity (highlighted in red in Figure 3). This reveals that not all paralogs

3202 were potentially removed with SambaR's "*dohefilter*" option. This was perhaps due to SambaR's

3203 conservative use of Hardy Weinberg's Equilibrium (HWE). HWE makes assumptions that in the

3204 populations, there are: no mutation, random mating, no gene flow, infinite population size, and no

3205 selection. With the knowledge that *Linum* are highly selfing (see previous chapters), the "*dohefilter*"

3206 option of SambaR may not be optimised for removing paralogs in *Linum* sequences (Jahnke &

3207 Etterson., 2019). Heterozygousity needs to be manually addressed from the input files before

3208 processing in SambaR. SambaR's "*filterdata*" function filters individuals based on their proportion of

3209 missing datapoints considering all SNPs. As an addition to heterozygosity filter, SambaR were also

3210 instructed to filter SNPs based on SNPs depth ("*snpdepthfilter=TRUE*" command), This command

3211 filters SNP which has a high read depth. Subsequently, the function filters SNPs based on their

3212 proportion of missing datapoints considering retained individuals only. The filtering of individuals

3213 could return different estimates of missing data and, as a result, different numbers of retained SNPs

3214 and individuals occurs. In population structure filtering (i.e. more individuals and less SNPs), the GC-

3215 content of the retained dataset equalled to 0.53 and the 'transversion vs transition'-ratio equalled

3216 0.59. For genetic diversity and selection analysis (i.e. less individuals and more SNPs), the GC-

3217 content of the retained dataset equalled 0.53 and the 'transversion vs transition'-ratio equalled 0.57.

3218 In the filtering for diversity and selection, linkage disequilibrium (LD) estimates were able to be

3219 calculated using *PLINK (-genome —r2 —ld-window-kb 1000000 —ld-window -r2 0)*. LD can infer trait-

3220 associated region(s) of a genome which may be of interest to further studies. However, for short-

3221 read sequences LD estimates may be redundant. This is because most of the reads do not cover all

3222   sites and consequently restricts LD to loci which are potentially very close (Maruki & Lynch., 2014)

3223   (Bilton et al., 2018).

3224

3225   <u>4.3.7 Analysis: Genetic Analysis in SambaR</u>

3226   For the purposes of Structure analysis, several analyses took place within SambaR. The

3227   analysis is often based on *Nei's genetic distance* between populations (Nei, 1972). This includes

3228   Correspondence analyses (CA), Principal coordinate analyses (PcoA), Principal component analyses

3229   (PCA), DAPC analyses using principles such as Landscape and Ecological Association (LEA) tests

3230   (Frichot & Francois., 2015). Other genetic distance related analyses were also illustrated in SambaR.

3231   For Correspondence Analyses (CA), analysis was performed using the function "*dudi.coa*" of the R

3232   package ade4-1.7.19 (Dray and Dufour, 2007; Bougeard and Dray, 2018). Data was imputed per

3233   SNP/individual by calculating genotype probabilities from population specific minor allele

3234   frequencies. Principal coordinate analyses (PcoA) were performed using the function "*pcoa*" of the R

3235   package ape-5.6.2 (Paradis and Schliep, 2018). This occurred on distance matrices containing 3

3236   different measures of genetic distance, with Nei's genetic distance, calculated with the function

3237   "*stamppNeisD*" of the R package StAMPP-1.6.3 (Pembleton et al., 2013), Hamming's genetic

3238   distance, calculated with the function "*bitwise.dist*" of the R package poppr-2.9.3 (Kamvar et al.,

3239   2014), and pi (pairwise sequence dissimilarity), calculated with the function "*calcpi*" of the R package

3240   SambaR. The principal component analyses (PCA) were performed using the function "*snpgdsPCA*"

3241   of the R package SNPRelate-1.28.0 (Zheng et al., 2012). DAPC analyses were performed using the

3242   function "*dapc*" of the R package adegenet-2.1.5 (Jombart, 2008; Jombart and Ahmed, 2011), both

3243   with and without prior population assignment. Multi-dimensional scaling (MDS) was performed

3244   using the function "*cmdscale*" (metric MDS) of the R package stats-4.1.3 (R Core Team, 2022) and

3245   the function "*isoMDS*" (non-metric MDS) of the R package MASS-7.3.57 (Venables and Ripley, 2002),

3246   on a Euclidean distance matrix generated with the function "*dist*" of the R package stats-4.1.3 (R

3247   Core Team, 2022). Neighbourhood joining (NJ) clustering was performed using the function "*NJ*" of

3248   the R package phangorn-2.8.1 (Schliep, 2011), using as input a Hamming's genetic distance matrix

3249   between individuals, calculated with the function "*bitwise.dist*" of the R package poppr-2.9.3

3250   (Kamvar et al. 2014). Bayesian population assignment (BPA) probabilities were calculated and

3251   plotted using the functions 'assign2pop' and 'plotassign2pop' of the R package SambaR. The optimal

3252   number of clusters (K) was determined using the elbow method on cross-entropy scores generated

3253   by the 'snmf' function, with the assumption that the startpoint of a plateau represents the optimal

3254   K.

3255   Divergence measures were measured using locus specific Fst estimates (according to Wright

3256   (1943), Nei (1977), and Cockerham and Weir (1987) (for all pairwise population comparisons) were

3257   subsequently calculated with the functions 'runWrightFst', 'locusNeiFst', and 'locusWCFst' of the R

3258   package SambaR. HWE, (2D) folded site frequency spectra (SFS), Tajima's D and genome wide

3259   heterozygosity analyses were able to be executed using the function 'calcdiversity' of the R package

3260   SambaR.

3261   Genome wide heterozygosity (genomeHe) was calculated in SambaR for each sample using

3262   the formula: *genome He = (He_seg * N_seg)/N_total*, in which: N_seg = the quantity of sites

3263   segregating within the population to which the tested individual belonged. He_seg = the proportion

3264   of Heterozygous sites within the investigated individual for those segregating sites. N_total = the

3265   total length of sequenced sites (polymorphic as well as monomorphic) which were able to pass the

3266   filter settings.

3267   Geographical maps were generated with the function 'getMap' of the R package rworldmap-

3268   1.3.6 (South, 2011). Piecharts were added using the function 'add.pie' of the R package mapplots-

3269   1.5.1 (Gerritsen, 2018). Admixture coefficients were implied with the functions 'snmf' and 'Q' of the

3270   R package LEA-3.8.0 (Frichot and Francois, 2014). Alpha was set to 10, tolerance to 0.00001, and

3271   number of iterations to 200. Ancestry coefficients were calculated with the software Admixture-1.3

3272   (Alexander et al., 2009) and illustrated using the 'plotstructure'-function of SambaR.

3273

3274   4.3.8 Analysis: Ancestry Coefficients

3275   As part of our population structure analysis, individual ancestry coefficients were able to be

3276   inferred using Landscape and Ecological Association (LEA) test. This was done using the package LEA

3277   under R (Frichot & Francois., 2015). These tests analysed population structure based on selection on

3278   a whole genome level. LEA applies landscape genomic data and identification of allele frequencies

3279   that illustrates genetic association with ecological associations. The LEA package derives adaptive

3280   alleles from large data sets, often referring to previous ecological association which is implied as

3281   ancestry coefficients (Frichot & Francois., 2015).

3282

3283   4.3.9 Analysis: Using different Genome reference

3284   There are multiple whole genomes now publicly available for *L. usitatissimum* and none for

3285   *L. bienne*. For this study we mapped to the *L. usitatissimum* whole genomes. They utilized different

3286   sequencing tools and have different read depths. For the purposes of this analysis, these differences

3287   between the reference genome could translate into differences in *snpmiss* and *indmiss* parameters

3288   in the SambaR input. These differences translate into the number of SNPs and individuals able to be

3289   analysed. For the purposes of comparison, we used another *L. usitatissimum* (Atlant variant) genome

3290   sequenced by Nanopore and Illumina sequencing and contain another cultivar type when compared

3291   to the CDC Bethune genome used for previous analyses in this thesis. The cultivar type used for this

3292   whole-genome sequencing was found to have low variability of morphological and anatomical

3293   characteristics under stress conditions, suggesting a variant that is adept under stress conditions.

3294   After sequencing they found 8.4 Gb of sequence data with N50 of 12kbp and a read coverage of 23×,

3295   and Illumina read coverage of 30× (Dmitriev et al., 2021). The CDC Bethune (v2) contained a

3296   summary of 21.80 Gb HiFi reads generated with N50 of ≥12kbp (Sa et al., 2021). The difference in

3297   initial data quality can be summarised in the below data quality plot when compared against our

3298   short reads:

3299

3300
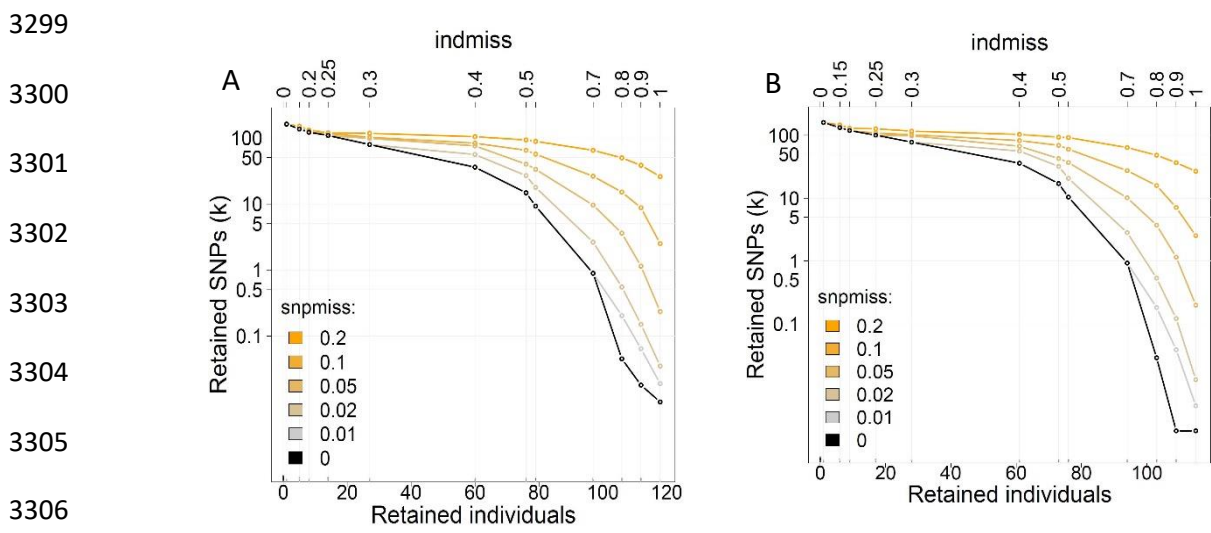
3307   Figure 4. Sambar's data quality plots for the CDC Bethune genome (4A) and the Atlant genome (4B).

3308

3309   The plot above illustrates the data quality when short reads are aligned to either the CDC

3310   Bethune genome (Figure 4A) or the Atlant genome (Figure 4B). This quality plot revealed that

3311   retained individuals for CDC Bethune genome was around the same as the Atlant whole genome

3312   (120 individuals). A *snpmiss* parameter of 0.05 and *indmiss* parameter of 0.7 for example, will return

3313   ~100 individuals with ~10k retained SNPs (before thinning). When aligning to the Atlant genome, this

3314   number was reduced by 10. This reveals that at least for these two different genomes (both

3315   different in terms of sequencing methods and cultivar type), that alignment of the short read

3316   sequences to different reference genome have some effects in terms of the quality of retained SNPs

3317   and retained individuals after filtering in SambaR. The CDC Bethune genome reveals a higher

3318   preference based on the number of individual samples kept.

3319   One of the most important qualities when looking at the quality after filtering, is the amount

3320   of heterozygosity in the data after filtering. For this, two plots can be shown to illustrate linear

3321   relationship between the proportion of Heterozygosity in all sites and segregating sites. Using the

3322   same filtering parameters, two plots were illustrated with to compare between the different

3323   reference the short reads were aligned to.
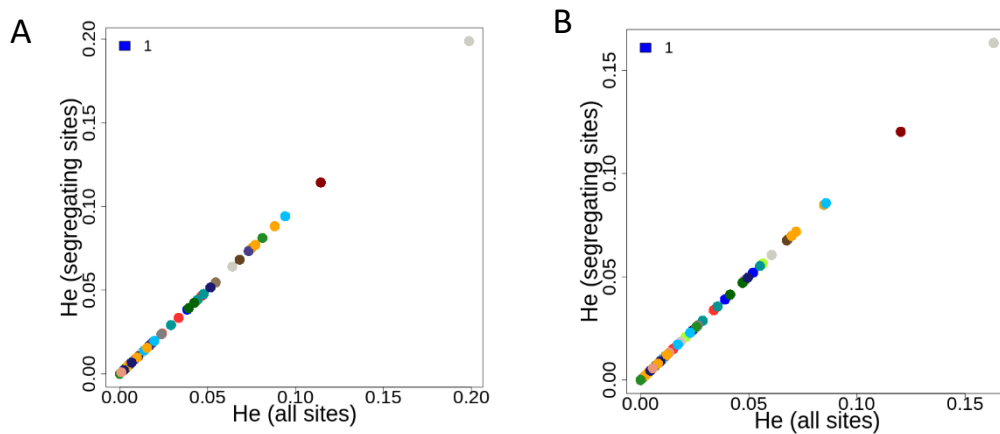
3331   Figure 5. Plots for heterozygosity after filtering short reads with the CDC Bethune genome (5A) and the Atlant

3332   genome (5B). Each dots illustrates different individuals differentiated by population (colours).

3334   Figure 5A is the linear plot for Heterozygosity (He) proportion on all sites and segregating

3335   sites for the data aligned to the CDC Bethune reference genome. The Figure 5B is the same plot for

3336   the data aligned to the Atlant reference genome. This revealed that aligning the short reads to either

3337   genome will result in significant amounts of Heterozygosity. There were differences in the rankings

3338   of the individuals within the population, noted by the differences in colour labelling. This reveals that

3339   proportion of heterozygosity remains mostly the same even using different reference genome. This

3340   potentially convey those heterozygous reads in our short reads are real heterozygous reads or

3341   perhaps reads in the genome due to the polyploidic nature of *Linum*. However, it is interesting how

3342   the rankings of these heterozygosity have changed within populations in our short reads. For the

3343   purposes of structure, diversity and differentiation analysis in the populations sampled here, we

3344   aligned to the more commonly used CDC Bethune genome.

4.3.10 Analysis: Removing Heterozygosity

3346

3347   As was seen in Figure 5A, all sites heterozygosity ranges from 0.000 to 0.20 within the

3348   samples and under different reference genomes. This was potentially an issue as the presence of

3349   heterozygosity might present an issue with paralogs which is especially of high importance to

3350   organisms which have previously underwent whole-genome duplication events such as polyploidy

3351   (McKinney et al., 2016). This is also revealed in the "locus specific heterozygosity" in the plot used

3352   for identification of paralogs in SambaR's "dohefilter" function. There were potential paralogs due to

3353   these heterozygousities. SambaR's filtering option may not remove all paralogs (Figure 6). In

3354   addition, we expected *Linum* to be highly selfing and polyploidy. As a result of this, explored options

3355   to filter for heterozygosity.

3356
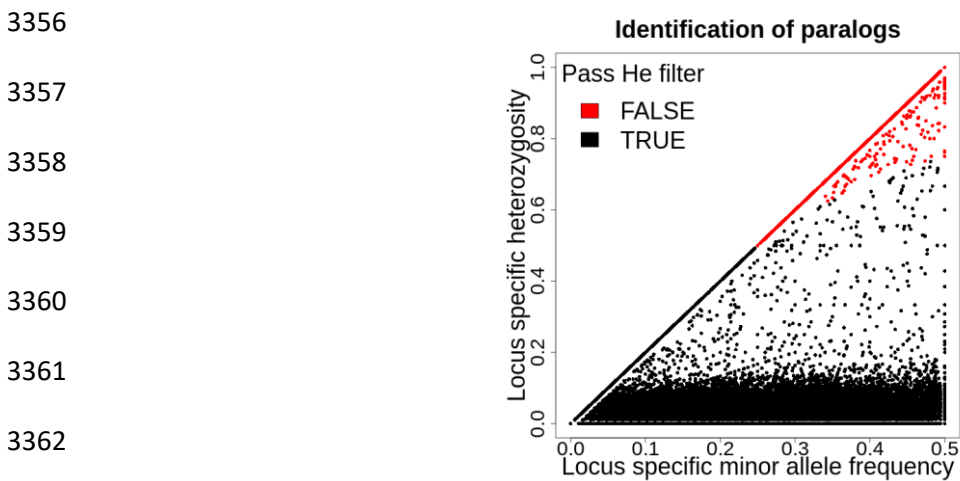
3363   Figure 6. Scatterplots to reveal Sambar's identification and filtering of paralogs for the default "dohefilter"

3364   option, revealing potential paralogs not filtered by SambaR (in black).

3365

3366   An option to avoid paralogs is to remove heterozygous alleles altogether, however the

3367   consequence of removing 100% heterozygosity is that further analysis such as the population

3368   diversity analysis will be made redundant due to the lack of heterozygous alleles (Chapman et al.,

3369   2009). In SambaR this will return an error whereby SambaR will specify that all samples are

3370   homozygous and therefore a population diversity measure is made redundant. Measures of

3371   heterozygosity, such as multi-locus heterozygosity (MLH) is useful for predicting whether

3372   populations are inbreeding or outcrossing in terms of their breeding strategies (Jensen et al., 2007).

3373   It is then preferable if the heterozygosity in the data were not entirely removed. To do this there is a

3374   requirement to specifically filter out a certain number of heterozygous reads. This can be done by

3375   removing SNPs and/or loci locations which have a higher than the threshold amount of

3376   heterozygosity, manually.

3377

4.3.11 Analysis: Filtering for Heterozygous Alleles

3378

3379   To avoid the redundancy of a population diversity analysis, a specific genotype filter

3380 measure to exclude locations with exceptionally higher heterozygosity were applied to the original

3381 population output of STACKS. To do this, the original VCF output file were first converted from muti-

3382 allelic to biallelic genotype/allele reads. This was done using the function "*bcftools norm -m*" under

3383 bcftools V1.10. Formatting of alternative alleles were done using the command "*bcftools view -e*

3384 *"FORMAT/AD[:1]<2 && INFO/AD[1]<5""* and extraction of genotype per sample is done by using the

3385 command "*bcftools query -f '%CHROM %POS[\t%GT]\n'*". Formatting of alleles and extraction of

3386 genotypes were done in bcftools V1.10 (Li., 2011). Once genotyped and formatted to biallelic reads,

3387 the file were processed in R using genotyping to mark heterozygous allele as "1" and homozygous

3388 allele as "0". Once heterozygous allele is marked as "1", a filter was applied whereby loci location

3389 with more than extreme heterozygosity can be identified and outputted. We decided to filter only

3390 for the top 1% heterozygous alleles to avoid missing important information from these potentially

3391 true heterozygous alleles. Filtering for heterozygousity can be done using the addition of

3392 "which(meanps[,q]>quantile(meanps,0.99))" to the initial allele output in R. This will remove top 1%

3393 of loci with excessive "1s" or heterozygous alleles. R was then asked to output a "*.txt*" file whereby it

3394 listed the loci position which can then be filtered out in vcftools. The full command lines for this

3395 process can be found under appendix 23.

3396   We used *vcftools' "--exclude-positions"* command in Vcftools for filtering against the output

3397 heterozygousity file (Danecek et al., 2011). First, the ".txt" file was made sure to be tab-delimited

3398 and was converted into a "Unix .txt" format using the Linux command "*awk '{ sub("\r$", ""); print }'*

3399 *winfileinput.txt > unixfileoutput.txt*". These "*.txt*" formatting options were essential so that vcftools

3400 could read the loci positions without errors under *Linux* commands. In the same command, the file

3401 can be recoded using the "*—recode*" command of vcftools and can be pipelined to output a .VCF file

3402 output whereby exceptionally heterozygous loci location was filtered out. The full command for this

3403 process can be found under appendix 24. This would keep out loci with exceptionally high

3404 heterozygosity which may represent paralogous alleles while preserving an amount of

3405 heterozygosity which potentially represent outcrossing populations within our sample sets.

3406

4.3.12 De_novo mapping

3407

3408   Mapping to L. usitatissimum was preferred in this analysis, to conserve the number of

3409 individuals and populations able to be analysed. However, using L. usitatissimum whole genome as a

3410   reference for population samples of L. bienne may not be optimal for SNP calling during further

3411   bioinformatic analysis, this may have implication on retained SNPs. We investigated how mapping

3412   "de novo" will impact the results gathered from mapping to the L. usitatissimum whole genome.

3413   After filtering for missing SNPs and individuals in SambaR, we found only 68 individuals were viable

3414   for further analysis when mapped de novo. After filtering and thinning, we found a total of 3029

3415   kept SNPs. The number of individuals kept for de novo mapping were significantly less than that

3416   from the option of using L. usitatissimum whole genome as a reference, where we kept 100

3417   individuals. We also lost potentially significant results because population 12 (Mediterranean

3418   population) were not kept from de novo mapping. Despite the loss of individuals using de novo

3419   mapping, we kept a significantly higher number of SNPs than mapping to the L. usitatissimum whole

3420   genome (kept only 800 SNPs with L. usitatissimum reference). This suggests that potential SNPs may

3421   have been lost due to mapping to L. usitatissimum whole genome. However, in the interest of

3422   keeping more individuals for population structure inference, we mapped to the L. usitatissimum

3423   reference. For future repeats of this population analysis, we suggest the use of L. bienne whole

3424   genome when they become publically available and/or including many more individual/population

3425   samples in the ddRADSeq protocol before mapping de novo. This may result in increases in the

3426   number of individuals retained after de novo mapping. We also suggest using more population from

3427   the Mediterranean region of Europe to observe for more variation from this region.

3428

3429

3430

4.4 RESULTS: IMPLICATIONS OF POPULATION ANALYSIS

3431

3432

3433   The final heterozygous-filtered data contained 866 SNPs (after filtering and thinning). After

3434   the filtering procedure, SambaR retained 100 individuals that makes up to 15 populations with

3435   retained grouping as Northern (> 45° North) and Southern (≤45° South) (Table 2), as described in

3436   Landoni et al (2022). We expected to see structuring between Northern and Southern populations,

3437   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
low2	50.68183	Northern	L-low2
Tym	53.30307	Northern	M-Tym
Sut	53.35291	Northern	N-Sut

3438   Table 2. List of output populations and their details of locality after the final filtering and thinning options of

3439   SambaR. Populations which are found ≤45° South were grouped as a more Southern population than those

3440   >45° in latitude. Apart from the cultivars (CUL), the populations were alphabetically ordered by latitudes with

3441   the most Southern population first. Populations were specified and grouped as per previous reciprocal results

3442   under Landoni et al (Landoni et al., 2022).

3443   In the following results the prefix of the population would be in the format of "Latitude (in

3444   ascending order)-Population". For example, for individuals from the most Southern population 3 the

3445   label would be "A-3" and the most Northern population Sut the labels would be "N-Sut". For Figures

3446   depicting individual comparison in the sample (such as genetic distance trees), the data will be

3447   formatted by "population_individual", for example individual 10 belonging to the population 12 will

3448   be formatted as "12_10". The full individual and population details is available in appendix 25.

3449   Our samples represent a wide range of latitude within Western Europe, representing Northern and

3450   Southern groups as was observed in Figure 7. In this results section we analysed the *Linum* samples

3451   as stated above in terms of their genetic diversity, structure, and divergence.
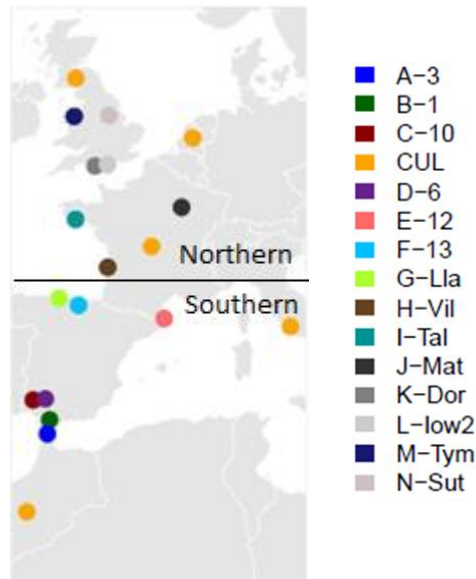
3452

3453   Figure 7. A colour map revealing the geography of the samples after passing filters and thinning with reference

3454   to their colours, population, and regions on the labels. For our analyses, Southern population are the wild

3455   Spanish population and Northern population are wild population found in France and the UK. This grouping

3456   aligns with previous reciprocal study groupings (Landoni et al., 2022)

3457

3458   ## 4.4.1 Results: Genetic Diversity

3459   ## 4.4.1 i Genome-wide Diversity

3460   We aimed to see population differentiation as well as to observe potential differences in

3461   breeding strategies. We looked at genome wide diversity to observe these. We expected to see a

3462   difference in diversity between _Northern_ (> 45° North) and _Southern_ (≤45° South) wild populations in

3463   our samples.

3464   The bar plot reveals a lower proportion of segregating sites in most of the Northern

3465   populations (Figure 8A). French populations such as "I-Tal" and "H-Vil" seems to reveal the lowest

3466   proportion of segregating sites (0.0462 and 0.1039) along with the Northern UK populations of "L-

3467   Iow2" and "M-Tym" (0.0727 and 0.1016). In exception of this is the French population of "J-Mat",

3468   which seems to have a high proportion of segregating sites. In contrast to this low segregating site

3469   proportion in most of the Northern populations, the more Southern populations have a higher

3470   proportion of segregating sites (>0.2). The highest proportion of segregating sites is observed in the

3471   Mediterranean population "E-12" (0.3915). These observations may suggest that the Southern

3472   populations are more likely to contain genes which are not conserved and are therefore potentially

3473   more outcrossing in terms of their breeding strategy when compared to their Northern relatives.

3474   In addition to this, the proportion of heterozygous sites can also suggest potential diversity

3475   within populations. When looking at heterozygosity within segregating sites only (Figure 8B), there

3476   wasn't much of a pattern to be observed. The proportion of heterozygote alleles for all sites,

3477   however, suggests that most of the latter, Northern population are less heterozygous. This suggests

3478   that perhaps the Southern populations are employing a strategically different breeding system. We

3479   expected the level of heterozygousity to not differentiate as much between the Southern and

3480   Northern population should they all be "selfing" populations. It may be the case that the more

3481   Southern populations are more outcrossing in breeding strategy than those of the more Northern

3482   populations in our sample lists.

3483

3484
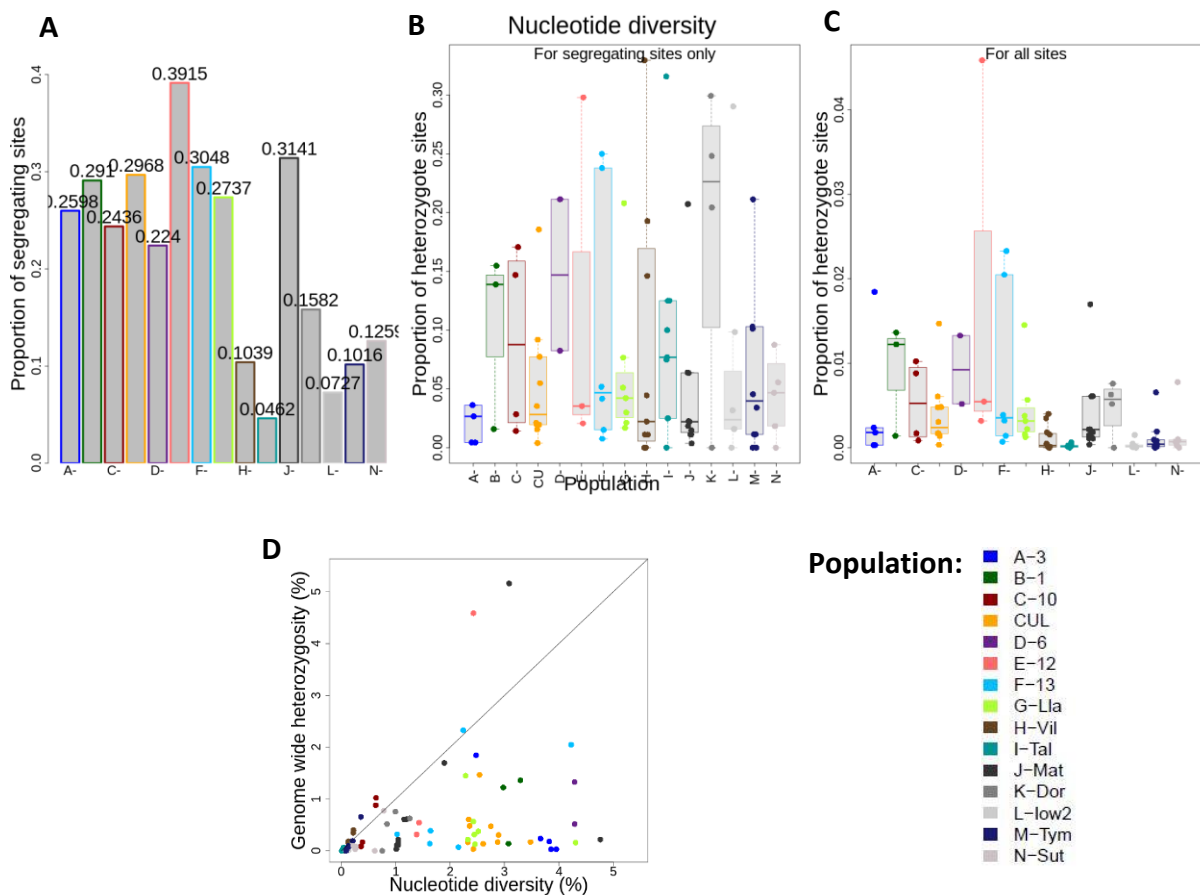
3485   Figure 8. Summaries of nucleotide diversity measures for our *Linum* samples. 8A. Bar chart to show the

3486   proportion of segregating sites within each population levels in our samples based on regions. 8B. Box plots to

3487   show the proportion of heterozygous alleles within the segregating sites, within each population in our

3488   samples based on regions. 8C. Box plots to show the proportion of heterozygous alleles for all sites within the

3489   genome. 8D. A scatterplot to show the nucleotide diversity or pi (%) against the genome wide heterozygosity

3490   (%) for all individuals colour coded by population. The colours on every part of this Figure corresponds to

3491   population list as inferred in the Figure.

3492

3493   ### 4.4.1ii Multi-Locus Heterozygosity

3494   A measure of multi-locus heterozygosity can also be observed with a boxplot (Figure 9). In

3495   the box plot, an observation of multi-locus heterozygosity was illustrated. In the x-axis, the

3496   populations were lined alphabetically according to their latitude, with the most Southern

3497   populations to the left. Southern populations such as population B-1, C-10, E-12, F-13, and D-6 were

3498   seen to have a more diverse multi-locus heterozygosity, agreeing with the sequence heterozygosity

3499   as observed in previous section. When looking at the Northern population, the box plots are forming

3500   closer to 0. This infers less multi-locus heterozygosity in the more Northern individuals. Interestingly,

3501   population 3, which is a Southern population region-wise, is trending more with the Northern

3502   population and the Northern population "K-Dor" seems to have a more diverse multi-locus

3503   heterozygosity. Less multi-locus heterozygosity in a population is thought to illustrate more

3504   inbreeding within that population and less outcrossing. It could be that the breeding strategy of

3505   population 3 is genetically influenced by the more Northern population and are therefore more

3506   inbred than the rest of the Southern population and the opposite is true for some Northern

3507   populations such as "K-Dor".

3508
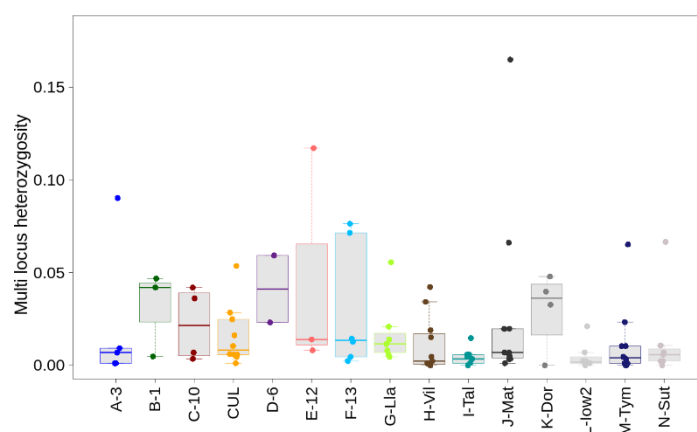
3509   Figure 9. A boxplot to reveal multi-locus heterozygosity at a population level.

3510

3511   **4.4.1iii Numberof private alleles**

3512   A measure of the number of private alleles can suggest population divergence as higher

3513   mutation rate may be implied (Szpiech & Rosenberg, 2011). We reveal a higher number of private

3514   alleles for our cultivars than most of the wild populations observed within this study except

3515   population E-12 (Figure 10).

3516   Our analysis suggests that most of our wild population have around 228 – 246 out of 866

3517   SNPs which are private alleles. Our cultivars have more, at 269 SNPs showing as private alleles. The

3518   Mediterannean population E-12 suggests an even higher number of SNPs showing private alleles at

3519   374 SNPs. This suggests more discinction in this population, which suggests population divergence

3520   for the Mediterranean population.

SNPs: 0.866 k

Population	Value
N-Sut	0.23
M-Tym	0.228
L-Iow2	0.229
K-Dor	0.234
J-Mat	0.228
I-Tal	0.231
H-Vil	0.231
G-Lla	0.246

Population	Value
F-13	0.239
E-12	0.374
D-6	0.241
CUL	0.269
C-10	0.233
B-1	0.233
A-3	0.228
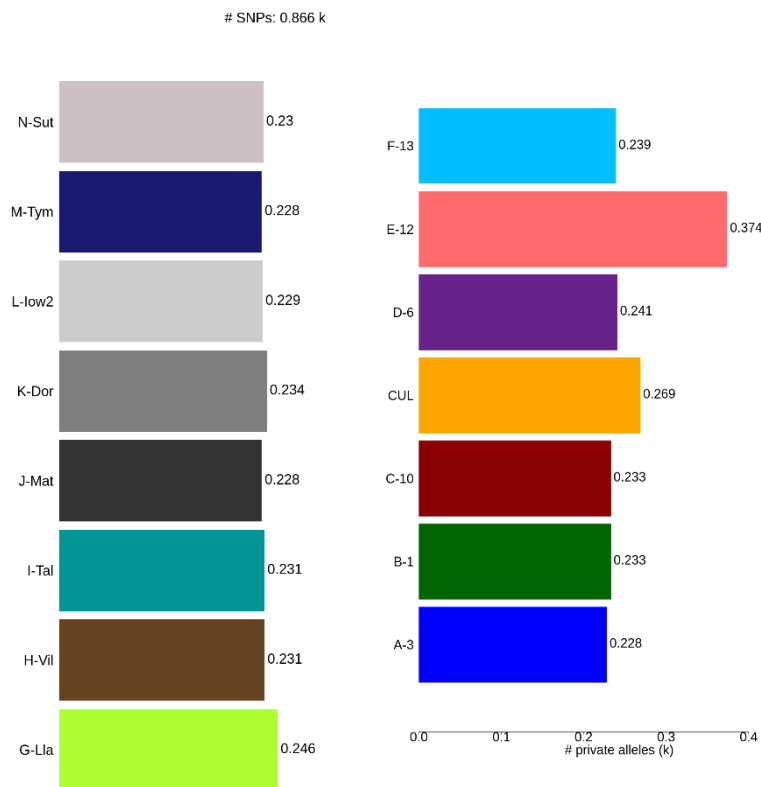
private alleles (k)

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

3521

3522   **4.4.1iv Variation in Segregating Sites (Tajima's D)**

3523   One way to look at population divergence is to examine variation within segregating sites for

3524   the population. This will infer whether selection have occurred which have removed variation or

3525   there is potential selection in which variation is maintained within population. To look at this, we can

3526   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))*.
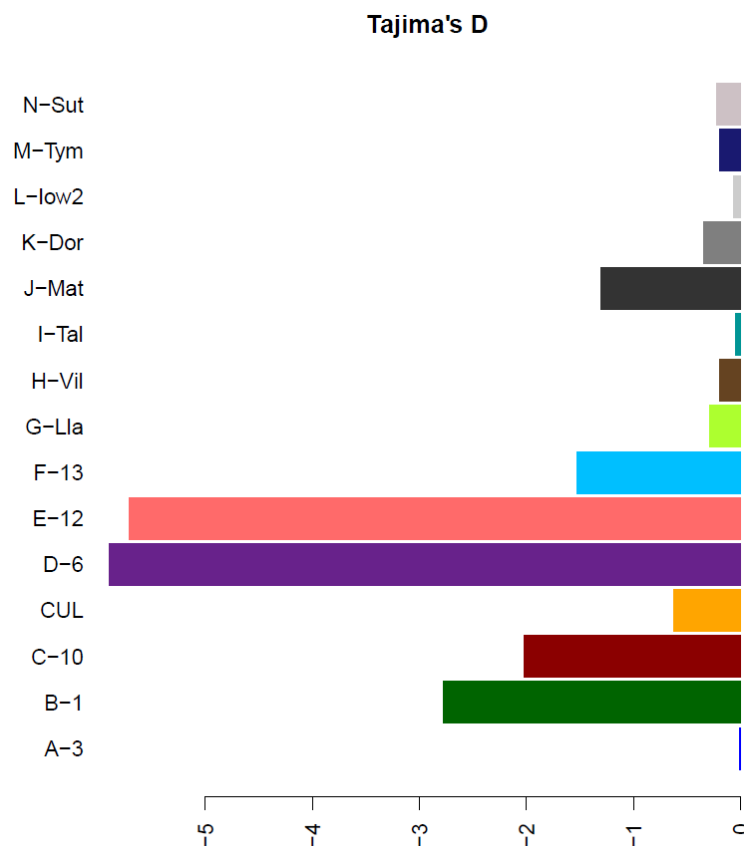
Figure 11. A bar chart to show Tajima's D statistics for the different populations.

3544

4.4.2 Results: Genetic Structure

3545

3546   When the data was fully prepared with potential paralogs removed (top 1% most

3547   heterozygous loci removed), it is with further confidence that we can carry out structure analysis

3548   without paralogous alleles. To reveal population structure, we constructed a Ward D neighbour

3549   joining tree. When genotyped heterozygous filters were applied, the tree suggests genetic

3550   structuring between the different populations (Figure 12).

3551   The Ward.D neighbour joining tree reveals that there are three main population clusters.

3552   The tree reveals that most of the cultivars are grouped together (Figure 12). The Southern

3553   population were grouped further away from the Northern and Cultivar individuals in our sample set.

3554   In terms of the wild, there is an observation of two groups whereby Southern population were

3555   grouping as genetically closer to each other and the Northern population grouping on together. This

3556   reveals that, genetically, in terms of their population the two Southern and Northern populations

3557   are potentially genetically distanced to each other. The Southern group are also more genetically

3558   distanced to the cultivars and may have a higher allelic richness. The cultivar "Suz" grouped with the

3559   Northern cluster. In this cluster, there also occurred some Southern individuals such as the Southern

3560   population "3". This could be due to unexpected history of migration in population A-3. We suspect

3561   that this is due to human errors such as seed labelling and labelling in the laboratory that may have

3562   led to these individuals not grouping to their clusters as expected. Future studies could examine

3563   structuring between more individuals from these population that may infer migration or

3564   hybridisation in population A-3.

3565   In other observations, the population highlighted in red circle are the only "Mediterranean"

3566   population observed within our samples (population E-12). On the heterozygous-filtered, unrooted

3567   tree, this group seems to be structurally different to the rest of the samples. They do not group with

3568   either Northern or Southern wild populations. This Mediterranean population seems to have a

3569   unique genetic structuring to the rest of the wild groups as well as the cultivars observed in this

3570   study. This reveals that in the samples contained within our RADSeq analysis, population E-12 is

3571   potentially unique to the rest of the wild populations found in the West, which are closely grouped

3572   to either the North or South cline of the population.

3573   In summary, the genetic network reveals that within the Western Europe wild *Linum*

3574   poulations, there were observations of genetic structuring, generally between populations

3575   originating in the North and South of Western Europe. There are some individuals that are out of

3576   place from their expected cluster, which could indicate potential gene flow or human errors within

3577   the data. The discovery of population E-12 being structurally unique to either North or South

3578   populations is an indication that the more Eastern, Mediterranean populations are structurally

3579   different to the rest of the wild populations in the West.
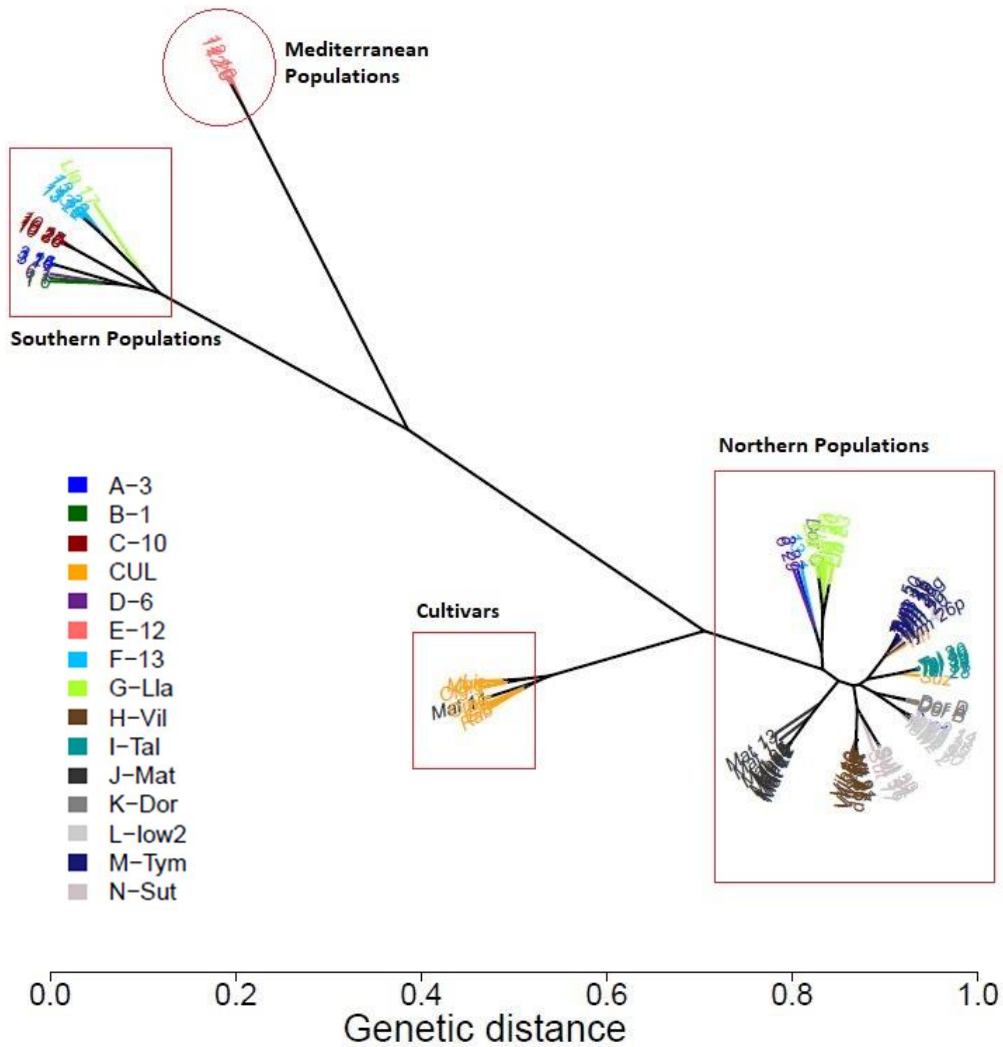
3580   Figure 12. A Ward.D (Ward's minimum variance method) Neighbour-joining tree for the illustration of genetic

3581   distance on an individual level. The genetic tree reveals structuring in our samples with Northern and Southern

3582   main clines for the wild. Cultivars are revealed to be more closely related to the Northern cline.

3583

3584

4.4.2i Principle Coordinate Analysis (PcoA)

3586   When looking at genetic structure, dissimilarity between the individuals based on Nei's

3587 genetic distance can also be observed. A principal coordinate analysis (PcoA) was conducted, with

3588 PC1 (representing 46.7% variance) and PC2 (representing 29% variance) plotted to examine genetic

3589 similarity (Figure 12). Another multivariate statistical technique that will go hand to hand with the

3590 PcoA analysis is the more descriptive Correspondence analysis (CA), whereby another plot can be

3591 illustrated in appendix 26. There are observations here of clusters for both PcoA and CA analysis,

3592 highlighted in red boxes. Southern and Northern populations are mostly clustering with what seem

3593 like a cline between the two clusters highlighted (Figure 13 and Appendix 26). In addition,

3594 population E-12 seem to be clustering separately in what we have suggested as "Mediterranean

3595 population".

3596   Some Southern individuals from populations such as F-13 and A-3 can be observed clustering

3597 closer to the Northern cluster, suggesting potential gene flow between Northern and Southern

3598 populations. Some Northern populations such as the French population "Mat", can also be observed

3599 to have an individual outside of the cluster, closer to the Southern populations. The Spanish

3600 population of "Lla" are clustering with their relatives in the North. This suggests that this population

3601 may have implication of gene flow through past genetic events.

3602   Another point of interest is the separation of the population E-12 which we suggest as being

3603 the "Mediterranean population". In this analysis, this was the only "Mediterranean population" able

3604 to be analysed due to DNA material availability. They seem to be structuring differently to both

3605 Northern and Southern groups of the Western European *L. bienne* samples in this study (Figure 13

3606 and Appendix 26).
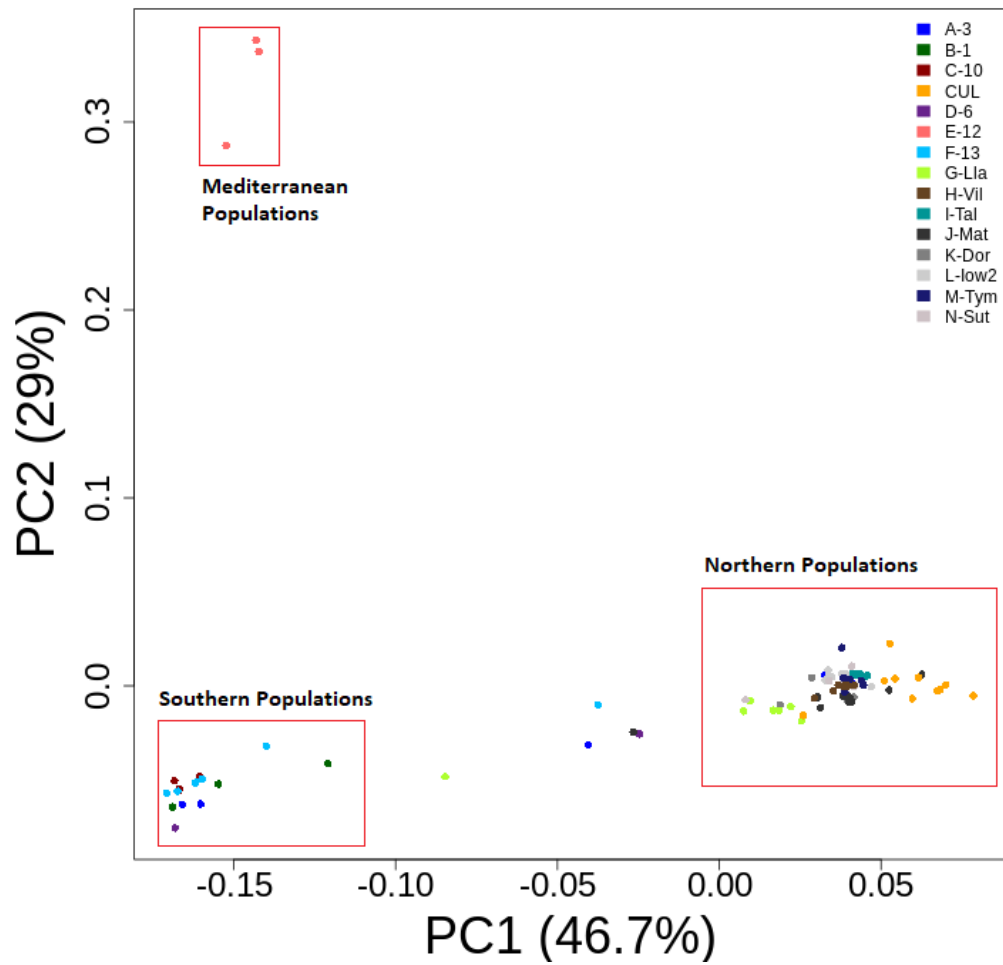
Figure 13. A PcoA scatter plot with PC1=47.3% of total variance and PC2=34.9% of total variance. The red boxes highlight potential clustering of populations. The plot agrees with the structuring suggested by the genetic distance tree, in addition, a few intermediate individuals between the Northern and Southern populations were also suggested, forming a potential cline between the Southern and Northern population.

3613   <u>4.4.2ii Sequence Dissimilarity</u>

3614   One way to illustrate genetic distance is to look at the proportion of dissimilarities between

3615   individual sequences or otherwise known as pi. This is the average pairwise difference between

3616   individuals. A sequence dissimilarity measure (pi) was illustrated as a matrix to illustrate

3617   dissimilarities between individuals and within their sequences, with Southern populations showing

3618   more dissimilarities (Figure 14).

3619   In terms of genetic structuring, there is an observation that most of the Southern

3620   populations are showing a degree of sequence dissimilarities when compared to the Northern

3621   populations (Figure 14). The sequence dissimilarity measures agree with the finding observed with

3622   the genetic tree, that the Southern and Northern populations are genetically distanced. Most of the

3623   Northern individuals are showing little signs of dissimilarity between themselves. For example,

3624   sequence dissimilarity between individuals in the Northern population "N-Sut" and "I-Tal" is close to

3625   0. Dissimilarities are observed to occur on the more Southern population with our *Linum* sample set

3626   (Figure 14).

3627   The sequence dissimilarity (pi) matrix (Figure 14) also reveals that population E-12 (pink strip

3628   on the matrix) has the highest level of sequence dissimilarity (pi) comparatively between all the

3629   other individuals analysed in this study. This is highlighted by the more intense shading of yellow

3630   when comparing individuals belonging to population E-12 to the rest of the individuals in the sample

3631   set (revealing pi values between 0.385 to 0.44). What's more interesting is that this dissimilarity is

3632   even more than the observed difference for cultivated and wild individuals between in our samples,

3633   when genotyping for highly heterozygous genes were implied to the data.
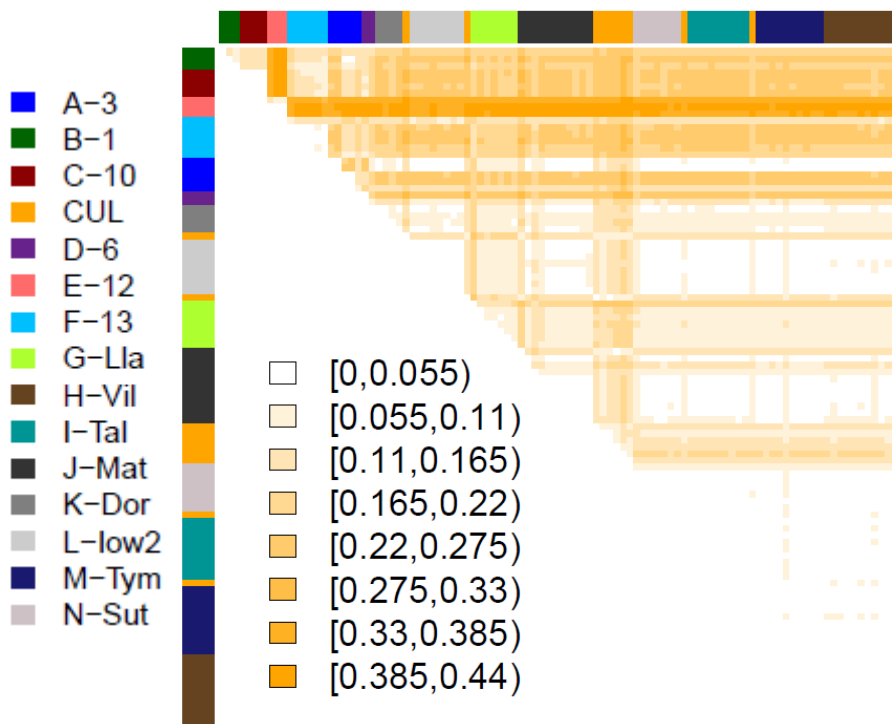
Figure 14. A sequence dissimilarity (pi) matrix for individuals summarised into populations. The colour chart on the left indicates the population each colour strip is representing in the matrix.

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

3652 Nei's genetic distance would be at 0.201 between the cultivated population and population 13. In

3653 the literature, a Nei's genetic distance of more than 0.250 is considered as significant genetic

3654 distance between populations observed (Nei M., 1972; Wright S, 1978). This finding suggests that

3655 the Mediterranean population E-12 is considered as genetically more distanced to the rest of the

3656 samples observed within this study.
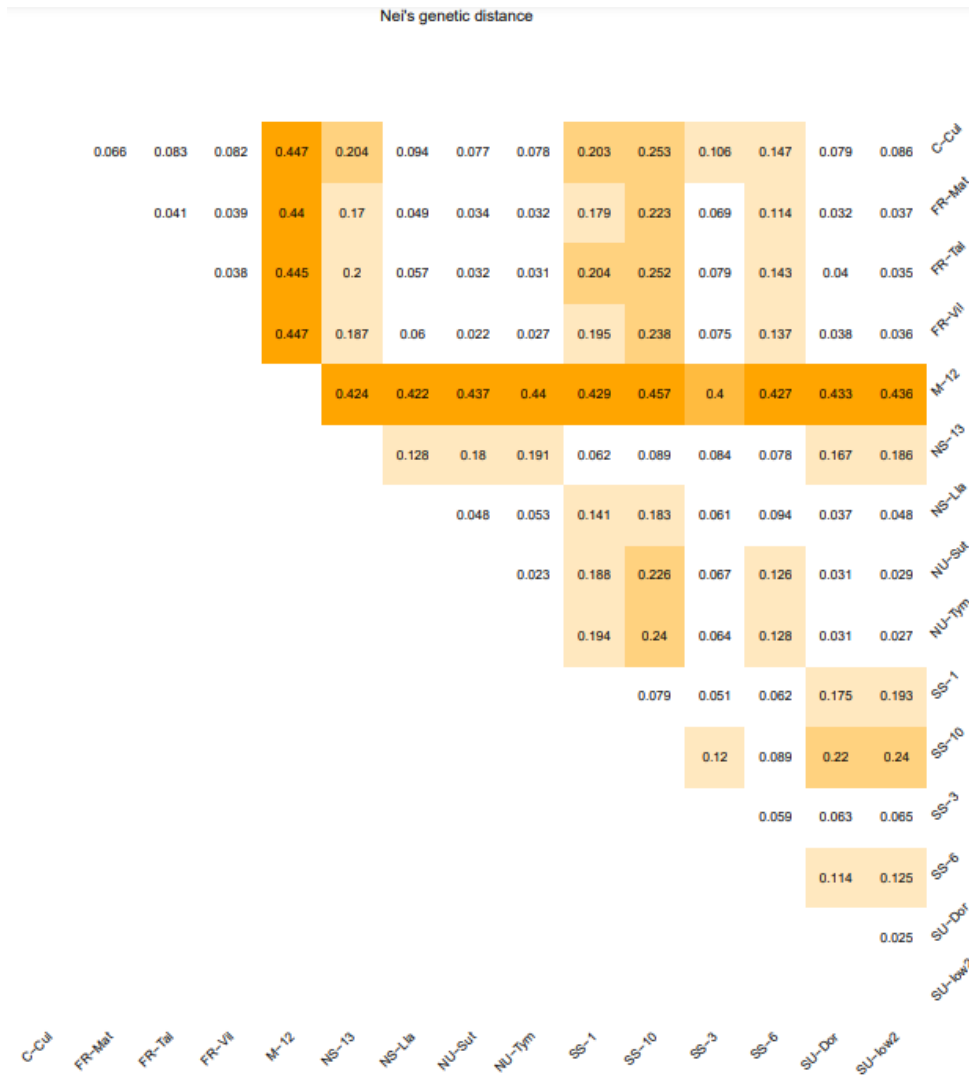
3657 Figure 15. A Nei's genetic matrix between the populations observed in this sample. The matrix reveals a higher

3658 genetic distance between the Northern and Southern populations. Darker orange shades reveals a more

3659 significant genetic differences.

3660

3661 <u>4.4.2iii DAPC Analysis</u>

3662   As an addition, we used a multivariate method to illustrate population structure. One

3663 method is the Discriminant Analysis of Principal Component (DAPC) (Miller et al., 2020). The DAPC is

3664 calculated in SambaR by scoring the number of retained Principle Components (PCs), interpolation of

3665 this scoring and scoring cumulative variance explained by the Principle Component Analysis (PCA)

3666   (De jong M. et al., 2021). The summary DAPC results for our genotype filtered data can be found

3667   under Appendix 27. A geographical map can be inferred using DAPC whereby population clustering

3668   can be illustrated under a population based on their latitude and longitude of origin (Figure 16).

3669   SambaR outputs pie charts on a map based on our population's locality to illustrate DAPC

3670   scoring (Figure 16). The Figure is separated by the number of clustering per population or otherwise

3671   known as 'K'. The cultivar individual "Mar" was excluded from this analysis as this was an exclusive

3672   Canadian population in the sample and we are interested only on Western European Latitudes for

3673   this study. From the Figure, we can observe groupings based on a DAPC analysis. When *K=2* there

3674   was already some grouping of the Southern population. However, the cultivars are not distinctive

3675   yet. When *K=3*, the population 'E-12', highlighted in red became distinct. When *K=4* populations

3676   showed further grouping whereby the four groups (Northern, Southern, Cultivars, Mediterranean)

3677   were observed as can be implied by the genetic distance tree. Here, the cultivars are seen as

3678   grouping in yellow, Northern populations in 'colourless', Southern population in brown and the

3679   population 'E-12' in red. Higher K's can also be inferred whereby some of the individuals from

3680   Southern populations are grouping closer to the Northern populations as inferred in K=5. When K=6

3681   the Southern populations are further split, separating away from each other. When K=4, it can be

3682   inferred that some cultivars populations such as those in France and Netherlands are grouping with

3683   the Northern populations. The Netherlands population is the cultivar population "Suz", which if we

3684   refer to the genetic distance tree, is also grouping with the Northern clines, suggesting genetic

3685   relationship with Northern population and perhaps cultivation from these populations.

3686

3687
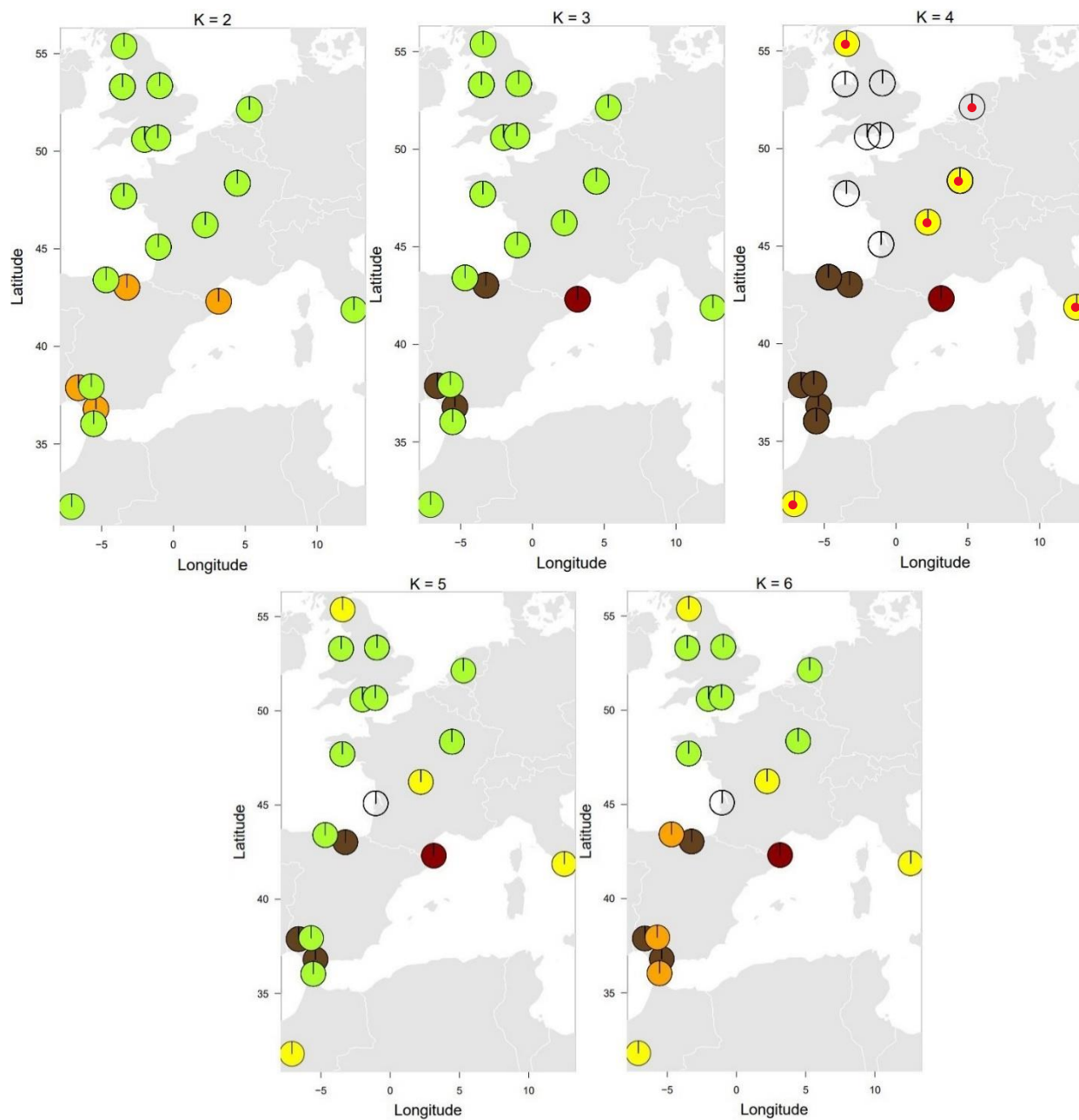
3688

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.

3694   4.4.2iv Landscape and Ecological Association

3695   As part of the population structure analysis, a Landscape and Ecological Association analysis

3696   (LEA) were utilised whereby ancestry coefficients can be implied between populations. Number of

3697   sub-populations (K) was described at 1-6. A Cross-entropy criterion can be used to determine the

3698   best run for a fixed value of K. The plot below illustrates the cross-entropy criterion for our dataset.

3699   The lower the cross-entropy value the better prediction capability a K value has.

3700   The Cross-entropy criterion plot reveals that for our dataset, at higher K-values the minimal

3701   cross-entropy was the lowest (Figure 17). This starts to level off when K=4. This levelling off minimal

3702   cross-entropy suggests that the optimal number of populations of K would be equal to 4 for further

3703   LEA analysis. An LEA bar plot with ancestry coefficients can be constructed for K=1-6 to imply any

3704   admixture within our dataset.
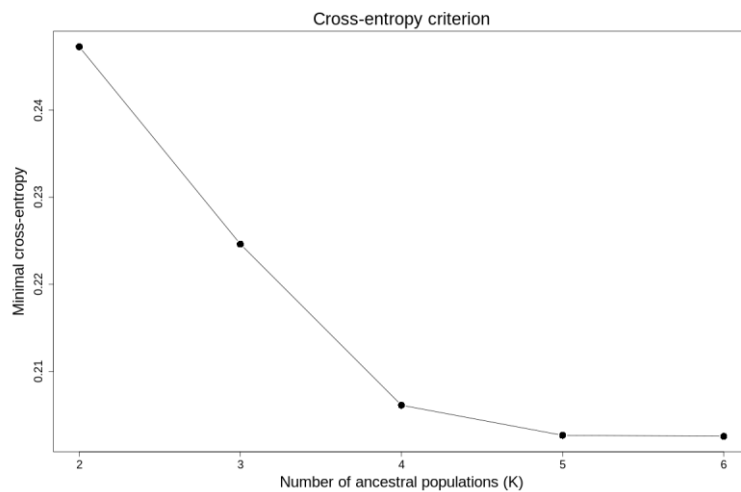
3705   Figure 17. Cross-entropy criterion graph illustrating the best number of population (K) is K=4. When the trend

3706   levels off.

3707   We further illustrated genetic relationships between our sample using an LEA barplot

3708   revealing some admixture within the Southern and Northern populations when *K≥4* (Figure 14).

3709   Admixture is suggested by the presence of sequences/SNPs from multiple genetic clusters for an

3710   individual. The Northern population were less admixed. The cultivar populations were more similar

3711   to the Northern accessions (Figure 18) when *K≤3*. They begin to differentiate when *K=≤4*. Admixture

3712   was also observed with the Northern accessions showing most of the coefficient changes when

3713   *K=≤4*.

3714   In the LEA plot there was also observation of divergence in population E-12. This is observed

3715   by the diverging bars on population E-12, observed when *K=≥3* (Figure 18). The LEA agrees with the

3716   distinction of population 12 from the other groups observed in the other genetic structure analyses

3717   above.
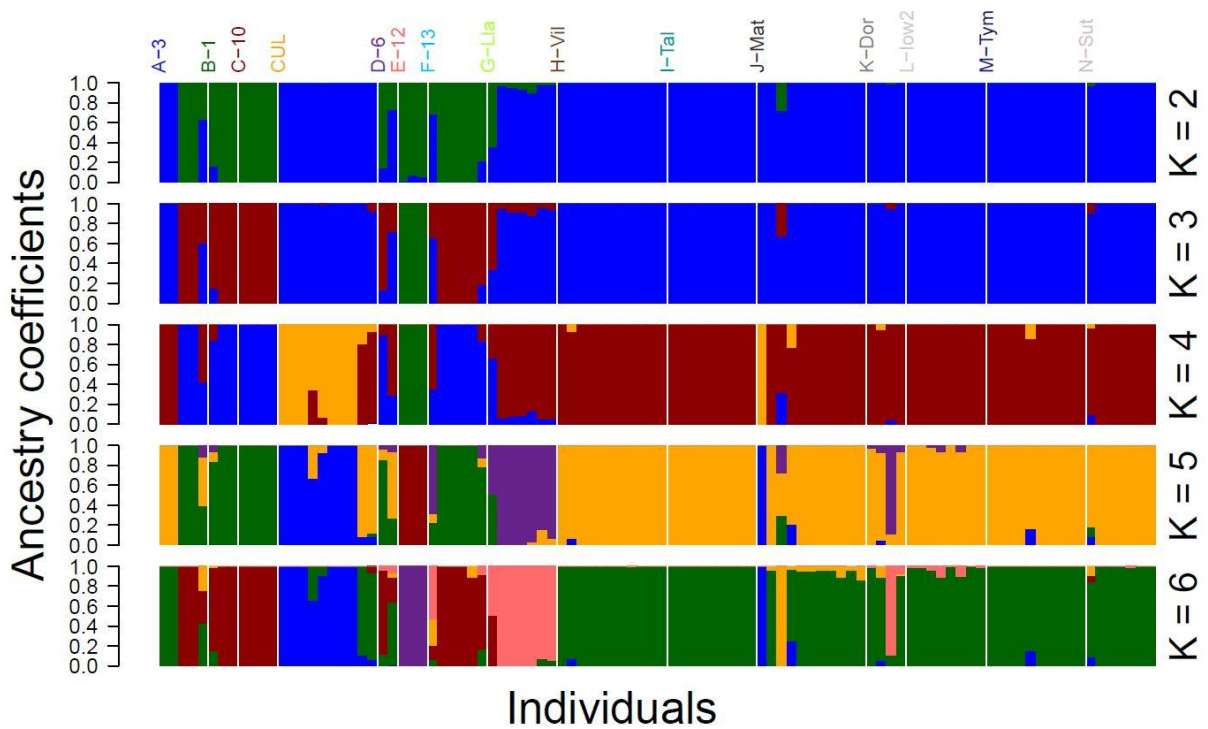
Figure 18. A LEA bar chart for every individual, separated into populations for K=2-6. The population names are

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

3732   populations "Suz" and "Tin", suggesting genetic relatedness of these cultivars it's wild ancestors

3733   (Figure 19).
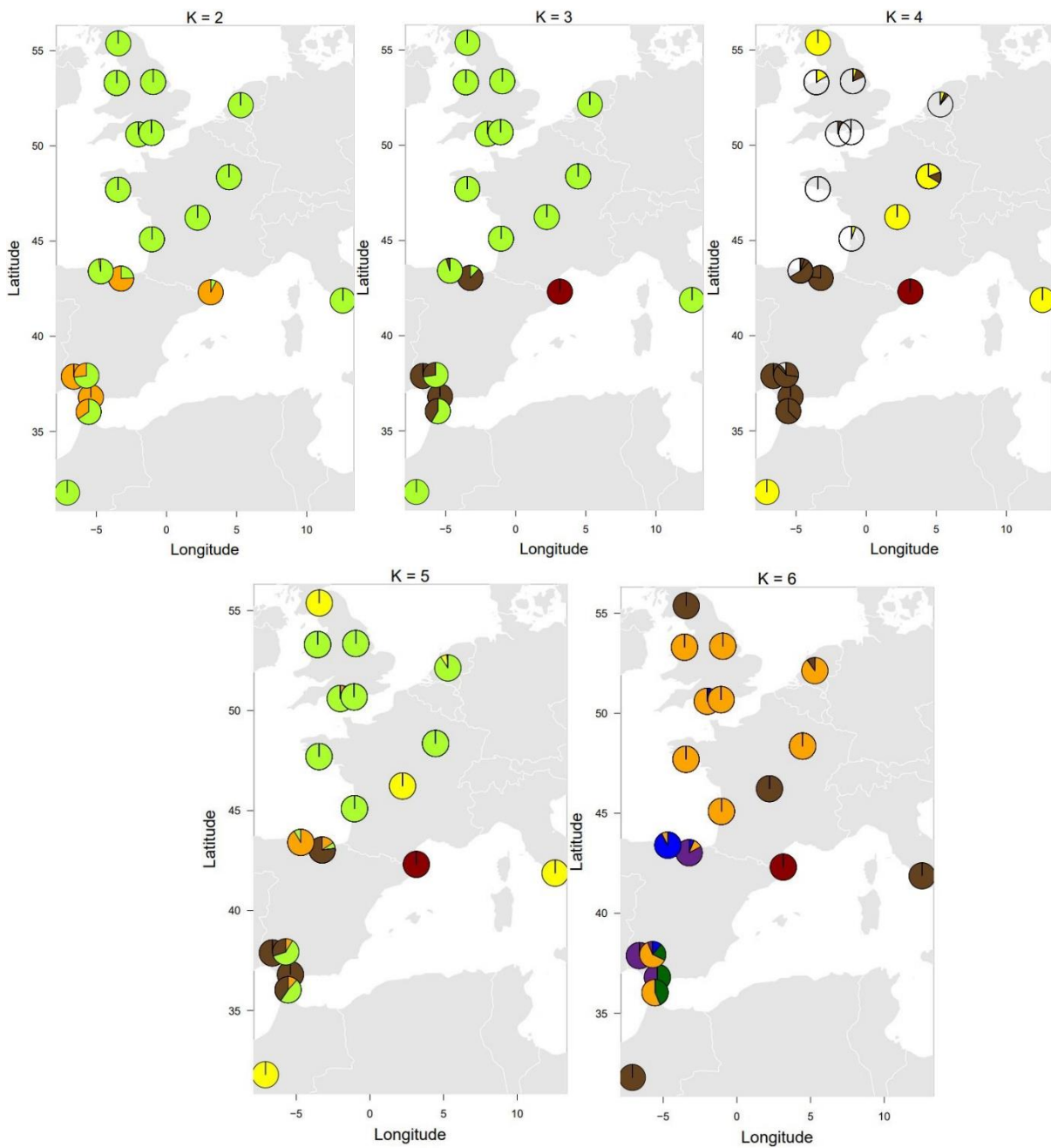
3734   Figure 19. A geographical map of all the populations conveying ancestry differences based on LEA for K=2-6.

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

3755 Figure 20. A matrix depicting Wright's Fst value. Higher value is highlighted in darker orange shades

3756

3757 4.4.3i Population Dissimilarity

3758   We observed dissimilarity of Sequence between individual population. It is with interest for
3759 this study to observe the most divergence populations by observing the population which has the
3760 most dissimilar SNPs with the other populations. As a starting point, variation and divergence can be
3761 looked at for the individuals within the same population.

3762　　　　The boxplot (Figure 21) reveals the overall sequence (Figure 21A) and SNP only (Figure 21B)

3763　　dissimilarity of individuals within a population. The dissimilarity is low overall for all populations

3764　　(>5% for overall sequence dissimilarity and >0.2 proportions for SNPs only). Although this was the

3765　　case, there was some differences which can be observed. In the earlier more Southern populations,

3766　　the SNP dissimilarity within populations were more diverse than that of most of the Northern

3767　　populations, except for the French population "J-Mat". Population 3 is showing the most

3768　　dissimilarity within the Southern population. There is potential here that within population 3, either

3769　　they are more diverging than other populations, or there was contamination within the sampling

3770　　processes whereby other population may have been unintentionally selected as population 3.
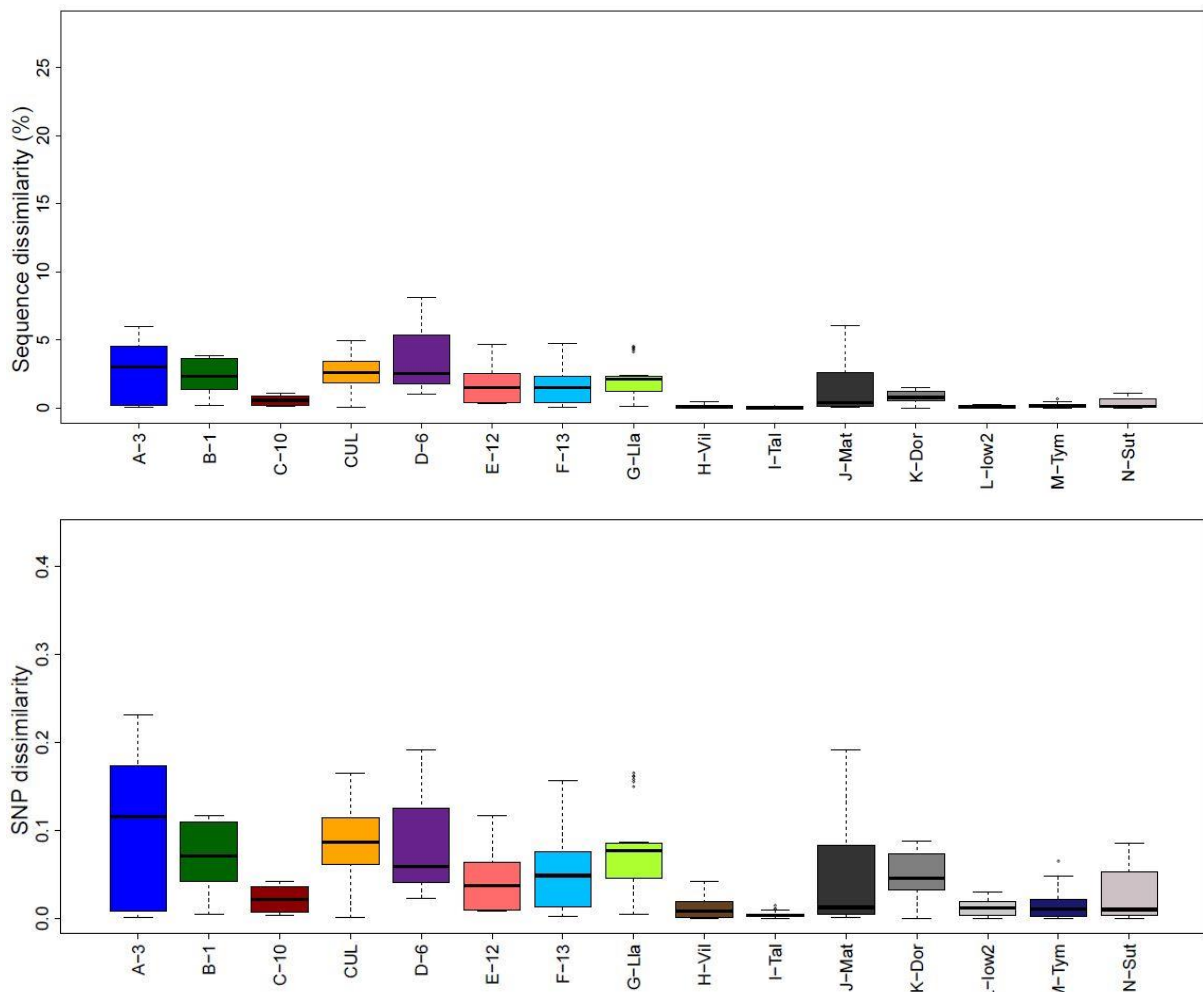
3771　Figure 21. A boxplot to show Sequence (A) and SNP (B) dissimilarity between the populations in this study.

3772　　　　For completion, a measure of nucleotide diversity (pi), as was mentioned under section

3773　　"*4.4.2 i Genome-wide Diversity*" of this chapter can be compared and mean pi for each population

3774　　can be extracted from SambaR (Figure 22). The mean pi was tested using a two sample t-test. The t-

3775　　test revealed that there was a significant mean pi difference between the wild population groups

3776   (Northern and Southern populations) *(P(T<=t) two-tail = 0.019 (<0.05))*. This illustrates divergence in

3777   the genetic diversity of Southern and Northern populations.

3778
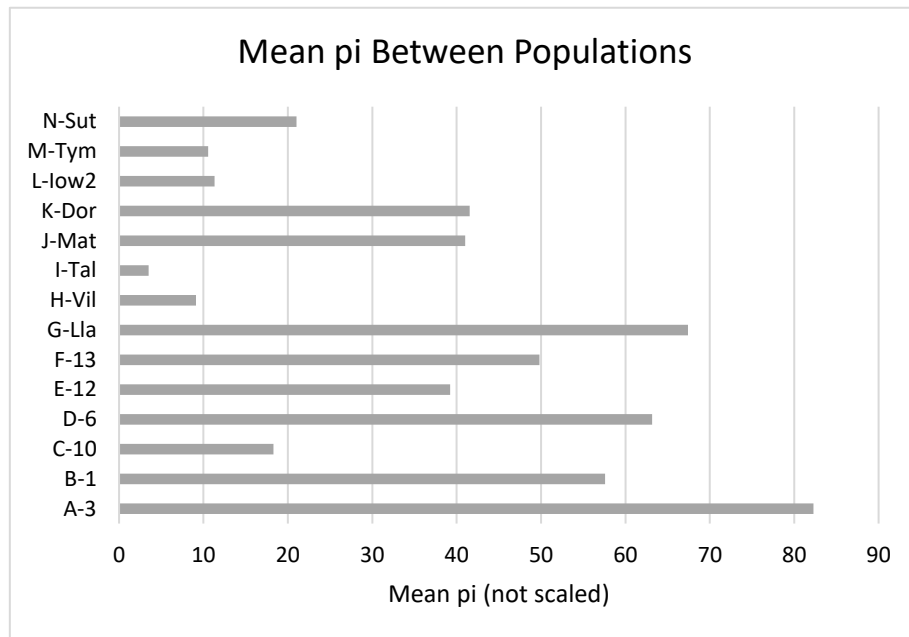
3779   Figure 22. A bar chart illustrating mean pi for each population in our sample as outputted by SambaR.

3780

3781   ## 4.4.3ii Minor Allele Frequency

3782   Another measure of population divergence can be done using the minor allele frequency

3783   (MAF). Minor Allele Frequency (MAF) is widely known as the frequency on which the second most

3784   common allele occurs in a sequence of a given population. They have been shown to play a role in

3785   population selection and divergence because MAF variants occurs once, and they drive a significant

3786   amount of selection (Hernandez et al., 2019). This would give an idea of how varied a genotype is for

3787   a given SNP. This can be used to differentiate between common and rare variants in the population

3788   (Linck & Battey., 2019). If the MAF is low, it may imply that the major allele for the SNP is conserved.

3789   A high presence of common alleles may also reflect signs of genetic bottlenecks (Marth et al., 2004).

3790   Alternatively, a high presence of rare alleles may suggest that a population is expanding (Marth et

3791   al., 2004).

3792   A matrix was drawn to illustrate MAF variance between the population samples within this

3793   study (Figure 23 on the next page). Higher variance in MAF variance is revealed in the more

3794   Southern population highlighted in shades of orange when MAF variance = >0.04. This reveals that

3795   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))*.

MAF variance

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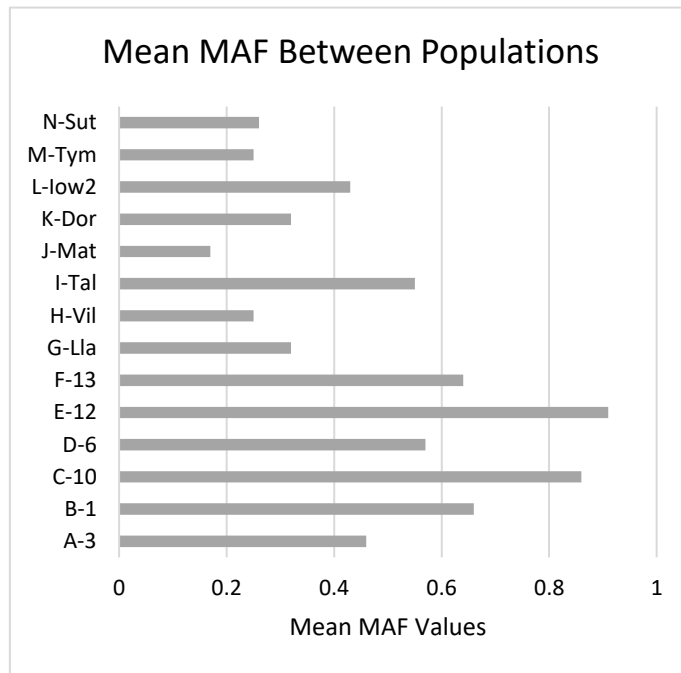
Figure 24. A Bar chart to represent the mean MAF values over each population.

150

4.5 DISCUSSION AND CONCLUSSION

4.5.1 Genetic structure

3813   We hypothesized that within our population samples, there are population groupings based
3814   on their geographic origin. The Ward.D neighbour-joining tree revealed this, showing genetic
3815   structuring between northern and southern populations, and between wild and cultivars within our
3816   sample set. The PCoA based on Nei's genetic distance also supports this. This genetic structuring
3817   suggests possible genetic variation within Western European *Linum bienne* populations. The PcoA
3818   and CA analysis revealed grouping based on northern and southern regions, with some intermediary
3819   individuals in between the clusters. In addition, the pie-charts based on the DAPC agreed that for up
3820   to four clusters (K=4), there was structuring related to the geography of the *Linum* populations of
3821   this study. The sequence dissimiliarity matrix revealed that the northern populations were
3822   differentiated from those of the southern populations. These results confirm structuring according
3823   to the geography of the wild *Linum*. Although this is the case, filtering for heterozygosity may impact
3824   measures such as the dissimilarity matrix and genetic distancing that maybe of true nature as
3825   opposed to paralogs. In this thesis, we were conservative towards avoidance of paralogs, based on
3826   the knowledge that our study models are highly-selfing and prone to gene duplication events such as
3827   polyploidy, which can complicate assembly of our sequences (Mastretta-Yanes et al., 2014). With a
3828   1% heterozygous filter we managed to remove the most heterozygous individuals whilst retaining
3829   heterozygousities that may suggest dissimilarities important for population diversity and divergence
3830   analyses. We also filtered at 5% heterozygousity where we begin to see populations dropping off for
3831   further analyses. Future studies may look at a range of heterozygous filter for comparison to this
3832   study.

3833   The geographic distribution of a plant population can lead to variation within species, so it is
3834   not surprising to find this in our wild *Linum* samples. In the model *Arabidopsis thaliana*, it has been
3835   illustrated that polymorphisms revealed differences within the population genetics of northern and
3836   southern populations (Fodorenko et al., 2001). In addition, studies of the model *Arabidopsis thaliana*
3837   have revealed patterns of genetic structuring that signals evolutionary processes such as migrations
3838   (Shirsekar et al., 2021). Additionally, Mediterranean populations of *Arabidopsis thaliana* revealed
3839   relationship that are closer to their relatives further South in Morocco and North Africa (Brennan et
3840   al., 2014). In that study, there was genetic structuring between northern and southern accessions.
3841   The more southern Mediterranean population were revealed to be related to populations further
3842   south of the range of the samples studied. There could be similar case made here when looking at
3843   the northern and southern clustering of wild *Linum* populations within our samples, with the
3844   Southern population suggesting genetic distinction from the Northern clusters. When looking at

3845   genetic structuring, these could be influenced by interaction of ecological and genetic processes

3846   such as local adaptation to seed dispersal. Ecological barriers to seed dispersal, due to the

3847   biogeography of Northern and Southern Europe may also play a part in the structuring seen in these

3848   Western European wild *Linum* populations. Seed dispersal maybe limited in Southern and Norhern

3849   regions of Europe thus limiting gene flow between the two wild groups, causing this genetic

3850   structuring. This barrier to disperal was observed in the Southeast Asian mangroves (Wee et al.,

3851   2020) and in peatmosses (*Sphagnum*) (Kyrkjeeide et al., 2016). Studies suggests that climate is the

3852   dominant determinant of plant range in Europe but in addition, species dispersal plays an important

3853   role in the genetic flow of a population (Normand et al., 2011).

3854   Another case made in the arctic-alpine plant species *Lloydia serotina*, was that reproductive

3855   biology can result in population structuring (Jones and Giddeon., 1999). Genetic differentiation that

3856   suggests variation in breeding system can also be observed in studies using three orchid species (Sun

3857   and Wong., 2001). In this study we observed signals of geneflow between Northern and Southern

3858   populations. We also saw that Southern population maybe more outcrossing in terms of breeding

3859   strategy. It maybe plausible that there is variation in breeding strategies between Northern and

3860   Southen clines of wild *L. bienne* observed in this study, this may be resulting from a seed dispersal

3861   barrier (ecological and geographical) ultimately resulting in genetic structuring. Studies on breeding

3862   strategies of wild *L. bienne* is scarce. We think that Southern population are more outcrossing based

3863   on higher heterozygousity. However, more samples needed to be studied in the future before this

3864   can be confirmed.

3865   Genetic structure analysis suggested that some individuals did not cluster with either

3866   northern or southern groups. This is potentially a signal for geneflow. If this signal is gene flow

3867   between Southern and Northern population, this would impede local adaptations within the

3868   Northern wild *Linum* populations. In *Arabidopsis*, it has been suggested that different populations

3869   under different environments showed this genetic structure (Hämälä et al., 2019). Another

3870   possibility is that historically, Northern populations diverged from these populations with some

3871   individuals not clustering under PCoA. However, the number of these individuals not clustering is

3872   limited to a few individuals. We saw no individuals forming a cline with the Mediterranean

3873   population E-12. This suggest that this population is divergent, and perhaps more possibility of local

3874   adaptation occured in the Mediterranean populations.

3875   We saw in our neighbour joining tree, PcoA and DAPC analysis that the Northern and

3876   Southern wild population are more divergent from each other than the cultivars are to either

3877   population groups. The cultivars are more closely related to the Northern wild individuals. If these

3878   cultivars are genetically closer to the Northern groups, we expect them to have some phenotypic

3879 similarities observed in the other chapters such as their pollen viabilities and the relative gene

3880 expression to flowering initiation. We didn't find sensitivity to vernalization requirement in our

3881 cultivars as we did in the Northern wild populations. There wasn't any phenotypic case where it may

3882 suggest that the cultivar population are closely related to the wild Northern populations. However, it

3883 may be that the cultivars are genetically more related to the wild Northern populations than they

3884 are to the Southern populations due to cultivation from the Northern populations. Artificial selection

3885 in the cultivars may have implications on the variation on phenotypes between the two species.

3886   The cultivars "Suz" and "Tin" were found to be grouping with Northern wild populations as

3887 opposed to the cultivar group (Figure 12). We think that there's potential mislabelling, either in

3888 previous seed collections or in the laboratory, with regards to the cultivars "Suz" and "Tin" as seen in

3889 our Ward.D tree. These cultivars were grouped with the Northern population, however, are very

3890 unlikely to be doing so since there was no phenotypic case suggesting this. We take caution with any

3891 results regarding these cultivars.

3892

3893 4.5.2 Population Diversity and Divergence

3894   Genome-wide diversity measures looking at proportion of segregating sites suggests a lower

3895 level in most Northern populations and a higher population in the Southern population. A higher

3896 amount of segregating site can suggest a higher mutation rate for the Northern populations and

3897 expansion further North (Fu Y., 1995). Segregating sites can suggest evidence of positive selection in

3898 a population (Gilad Y. et al., 2002; Przeworski M., 2002; Booker T. et al., 2017). Suggestion of more

3899 occurrence in positive selection could be interpreted by a higher number of segregating sites within

3900 our Northern groups, potentially driven by the local adaptation to the environment (Figure 11).

3901 Positive selection is often associated with selective sweep, suggesting less genetic diversity in our

3902 Northern populations (Booker T. et al., 2017). Less genetic diversity could imply more inbreeding in

3903 these Northern population.

3904   Furthermore, studies using *Camellia sinensis* (Tea) and their wild relatives suggests genetic

3905 divergence in the wild relatives by looking at proportion of segregating sites and in addition, the

3906 proportion of heterozygous sites (Yang H. et al., 2016). We looked at the proportion of

3907 heterozygous sites for all sites in our study model. Most of these sites were filtered due to filtering

3908 of paralogs. However, for all sites, surviving heterozygous alles were higher in the Southern

3909 populations, suggesting allelic richness (Figure 8C). This suggest that the Southern populations are

3910 genetically more diverse than the Northern population. Northern population may be more inclined

3911   to be conservative towards local adaptation to colder seasons. This was perhaps less desirable in the

3912   Southern populations due to warmer climates.

3913   The Mediterranean population E-12 was the most distinct among our samples. Based on

3914   Nei's genetic distance these Mediterranean individuals are clustering on their own as opposed to

3915   either Northern or Southern clines. It is thought that this Mediterranean population is genetically

3916   distinct from the rest of the Western European wild population. It is possible that the more

3917   genetically distinct population E-12 was closer to wild relatives in the eastern Europe region, but this

3918   was not tested in this study. We did not have materials for the wild *Linum* population originating in

3919   eastern European region such as Turkey and Italy, in which both wild and cultivated flax have been

3920   observed. There is even a possibility that population E-12 is more related to population as far as the

3921   eastern Mediterannean and the Middle East regions, because the cultivated flax (*L. usitatissimum*) is

3922   thought to be native to these regions before they were introduced to the more Northern climates

3923   (Sen, T. and Reddy, H.J., 2011). It would be of interest for future research to include more

3924   Mediterranean and Middle Eastern populations, to determine their genetic relationships with our

3925   samples. The distinction of population E-12 was also supported by the Tajima's D statistics, whereby

3926   population E-12 had a more negative Tajima's D than the rest of the populations sampled in this

3927   study. This may suggest that this population were diverged in the past, and this was a result of

3928   population expansion in wild *Linum*. In past studies, low Tajima's D in the fir species *Cunninghamia*

3929   *konishii* have been inferred as a post-glacial expansion with their populations (Hwang et al., 2003). A

3930   study in *L. flavum* suggested a post-glacial migration history in *Linum,* which may result in a

3931   population expansion during the post-glacial period (Plenk et al., 2017).

3932

3933   ### 4.5.3 Breeding Strategies in Western European *L. bienne*

3934   Population diversity measures such as the measures of heterozygosity along with the multi

3935   locus heterozygosity also differed between the southern and northern populations. When filtering

3936   for excessive heterozygous alleles, remaining heterozygousity were observed mostly in Southern

3937   populations, with most of the Northern population infering no heterozygousity (Figure 9). In addition

3938   to the structuring observed in the genetic distance tree and the PcoA, wild populations might also be

3939   differentiating in the way they breed. We observed a higher heterozygousity in Southern plants in all

3940   sites (Figure 8c and Figure 9). The more heterozygous populations suggest less inbreeding and more

3941   outcrossing than the less heterozygous northern populations. Sharry and Lord (1996) also explored

3942   heterozygosity, concluding that less heterozygous populations may have been more inbred than

3943   those of the more heterozygous populations (Sherry & Lord., 1996). We think that our Southern *L.*

3944 *bienne* populations are more outcrossing and that our Northern population are inbreeding, perhaps

3945 to conserve local adaptations or perhaps due to climatic barriers for seed dispersals. This

3946 differentiation would also be supported by the difference seen in Wright's Fst matrix. It is seen that

3947 Fst values between Northern and Southern populations are higher as highlighted in shades of

3948 orange. In the literature it is accepted that an Fst value greater than 0.15 between populations can

3949 be considered as significant in differentiating populations (Luo et al., 2019). This adds to the genetic

3950 structuring revealed by the genetic distance, PcoA and CA analysis that northern and southern

3951 populations are genetically different. In addition, there is potential that Southern population are

3952 more outcrossing in terms of their breeding strategy. This agrees with the findings on genetic

3953 structuring observed in section 4.41 "Results: Genetic Structure" and can so therefore observation in

3954 genetic diversity supports the finding that there is potential breeding strategy variation in Northern

3955 and Southern clines of *L. bienne*, causing genetic structuring between the two Western European

3956 wild *Linum* populations.

3957

3958 4.5.4 Genetic variation in Western European *L. bienne*

3959 Molecular diversity and association analyses in the past have revealed the potential of

3960 genetic variation in wild flax for the development of cultivated flax (Soto-Cerda et al., 2014). It is

3961 interesting that the cultivar populations within our sample sets seemed to be more genetically

3962 closely related to the Northern populations than to the Southern populations. The genetic distance

3963 tree and the ancestry coefficients on the LEA bar plot agreed that the cultivars in our samples were

3964 more closely related to the more northern wild accessions than so the southern accessions.

3965 Variation in genetic makeup within the *L. bienne* gene pool represents the potential of genetic

3966 variation for the improvement of cultivated flax in Northern Europe. Traits that are conserved in

3967 Northern populations in response to colder climate can be useful for flax cultivation in different

3968 growing seasons.

3969 To add to the potential genepool resources, we can infer some ecological interpretation for

3970 our wild populations. This includes selection and bottlenecks events that may have arisen in the

3971 past. Tajima's D suggest that there is less variation compared to what the population can sustain. In

3972 population genetics studies, negative Tajima's D have previously suggested population expansions

3973 and differentiation after a bottleneck (De Jong et al., 2011; Gunther et al., 2016). This negative

3974 Tajima's D was strongest in the more southern populations, particularly in population E-12. This may

3975 suggest that some southern populations, such as population E-12 have diverged through past

3976 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;

157

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

4071   all these in mind, it is evident that in the literature, variation in seed sizes occurs in many plant study

4072   species with local environmental variables, potentially playing a role in natural selection of seed sizes

4073   in plant populations.

4074   Seed size is a trait of interest in *Linum*. Variation in seed size due to vernalization treatment

4075   in the wild may indicate local adaptation to the environment in terms of seed sizes. In addition to this,

4076   seed sizes of *Linum* wild relatives can be useful in their cultivar relatives. This is because linseed is one

4077   of the biproduct of *Linum*. We will investigate if our wild *L. bienne* vary in seed sizes when under

4078   different vernalization treatments. Germination of plants are suggested to be related to their seed

4079   sizes, so seed size can also count for population fitness (Keddy & Constable., 1986). A study on onions

4080   (*Allium cepa I.*) also suggests that vernalization temperatures, duration and bulb size significantly

4081   influenced seed yield of all cultivars (Muthamia., 1994). This suggests that in many plant species

4082   requirements of vernalization will have a significant effect on seed health which suggests seed sizes

4083   and seed yield which is of interest in agriculture. We hypothesize that seed sizes for different *Linum*

4084   populations will have been affected by vernalization and this will have represented a correlation

4085   between seed size and environmental variables such as latitude of localities and climatic variables.

4086   With the above in mind, it is of both ecological and agricultural interest to investigate the

4087   consequences of vernalization in *Linum* traits. Findings here can be linked to local adaptation in the

4088   wild *L. bienne* and implications of potential loss of vernalization in cultivated *L. usitatissimum*. In this

4089   chapter we investigate whether vernalization treatments affects the fitness of wild *L. bienne*. For

4090   traits, we expect to find variation between vernalized and non-vernalized individuals. We expect seed

4091   sizes to also vary between Northern and Southern individuals, suggesting local adaptation in seed

4092   sizes. Implication of requirement of vernalization can then suggest that any adverse traits observed

4093   within the more Northern individuals due to vernalization stress are avoided by the potential loss of

4094   vernalization as was observed under chapter 2 of this thesis.

4095

4096

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

4130   used to observe *Linum* seeds using a colour camera mounted on a Leica DMI-3000 Microscope. The

4131   Microscope was hardwired to a Windows computer running Leica's LAS X software, where

4132   measurements of seed length and width can take place. Measurement of the seed area is done by

4133   using ImageJ to measure pixels and calculate area. To do this, images of 5 seeds per-individual were

4134   captured using a microscope with a 10×10 objective. We used Leica's LAS X software to capture and

4135   save images before processing in ImageJ. In ImageJ, the images were individually measured. We

4136   calibrated ImageJ to each image by drawing a 1mm line and setting a scale to that line as 1mm. The

4137   image is then converted into 8-bit format and made binary (Process → Binary → Make Binary).

4138   Although ImageJ usually outputs binary images without holes, there are occasions where holes in

4139   the seed can be observed. Potential holes were able to manually be filled or automatically via the

4140   "fill holes" function (Process → Binary → Fill Holes). Images were then adjusted by threshold (Image

4141   → Adjust → Threshold) and analyzed by using the "Analyze particles" function of ImageJ. This

4142   function can analyze particles (pixels) and calculate area of a of each binary colour (black and white).

4143   The seeds were measured as the area in black. Values were inputted into Microsoft office and

4144   formatted for further analysis in R.

4145

5.2.1 Data Analysis

4147   All phenotypic measures were statistically tested during this chapter and all additional data

4148   on this chapter were analysed using R v 4.1.2 (R Core Team, 2022) and Rstudio (Rstudio Team, 2020)

4149   as a graphical interface. For modelling and correlation purposes the R package "ggpubr"

4150   (Kassambara, 2022), "ggplot2" (Wickham and Sievert, n.d.), and "car" (Fox and Weisberg, 2011)

4151   were used to execute general-linear modelling (GLM). There are several assumptions made for this

4152   GLM; the data is normally distributed, and that values of trait measurements, and other variables

4153   are independent of each other. We also assumed that the variance in data residual is the same. For

4154   summary of statistics, model can be summarised in terms of adjusted r^2, F-statistics, and p-values.

4155   The full R commands for the analysis during this chapter is available in appendix 30. Barcharts can be

4156   illustrated using the "barplot" function of R. A two-tailed t-test were then used to compare data

4157   under vernalized and non-vernalized treatments.

4158   For correlation analysis to environmental variables, data for latitude were collected for each

4159   population sampled from the wild at the place of collection by their respective collectors, as per

4160   chapter two of this thesis. Principle component values variables were able to be gathered for

4161   different climatic variables as was described under chapter 2 of this thesis. Scatterplot for

4162   correlation between traits and environments were illustrated using "ggscatter" function of R with a

4163   GLM modelling with a Pearson's correlation method. The same is applied for analyses of seed sizes

4164   against environmental variables.

4165   _____

4166

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.
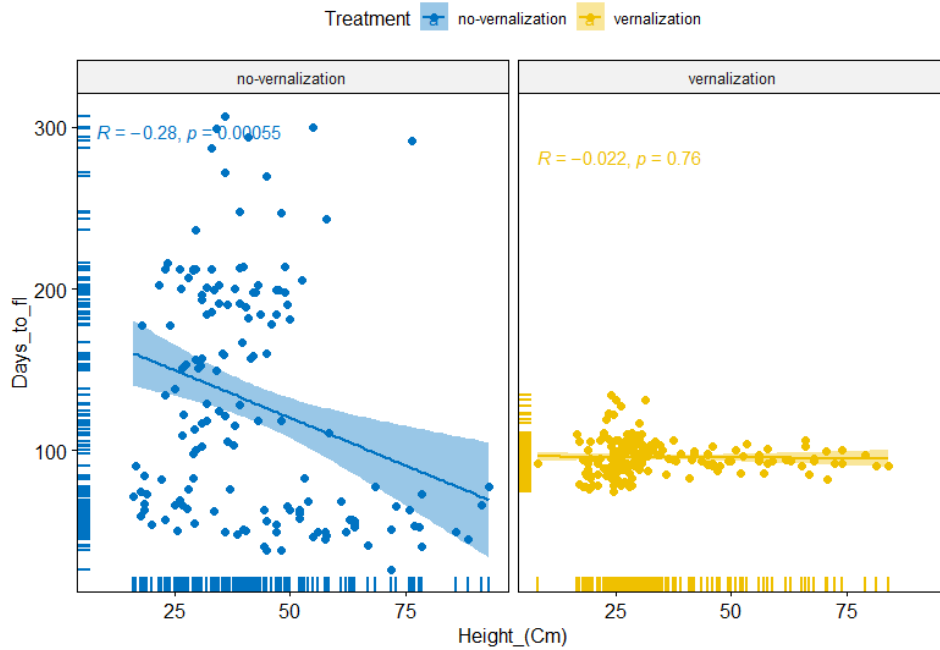
4191   Figure 1. Two scatterplots to illustrate the relationship between the number of days a plant takes to flower

4192   (Days_to_fl) and he plant height (Height_(Cm)). The plot revealed a contrast between no-vernalization and

4193   vernalization treatments, with a significant negative correlation under the no-vernalization treatment (*R=-*

4194   *0.28, p=<0.001*).

4195

4196   Furthermore, a comparison between the plant height and the different treatments was

4197   illustrated using a two-sample t-test on Figure 2 (assuming unequal variance). The t-test suggests

4198   that there was a significant difference on the plant height between the two treatments (*P(T<=t) two-*

4199   *tail=<0.001*).
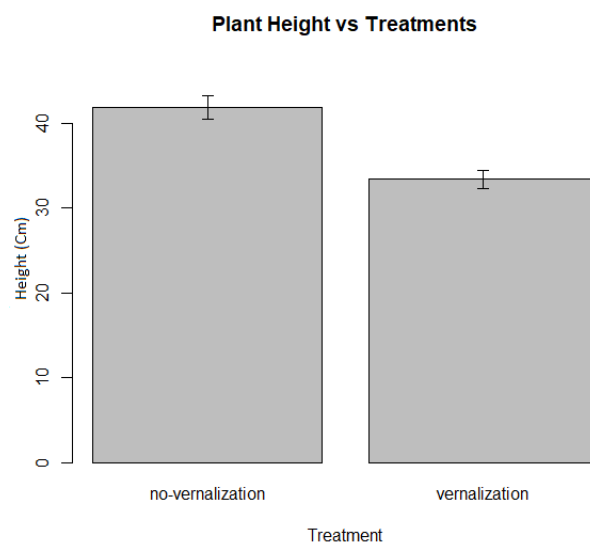
4200   Figure 2. A bar chart to illustrate the significant difference between no-vernalization and vernalization

4201   treatments on mean plant height (Height (Cm) (*P(T<=t) two-tail=<0.001*).

4202   Another trait we measured during the study is the number of stems at first flowering. The

4203   result suggests that when no-vernalization and vernalization treatments were applied to the *Linum*

4204   samples, a positive correlation can be observed between the number of days a plant takes to flower

4205   and the number of observed stems on first flowering (Figure 3). However, the strength of the

4206   relationship seems to have been reduced when plants were vernalized. R values for the vernalized

4207   treatments were reduced from *R=0.69* to *R=0.18.*

4208
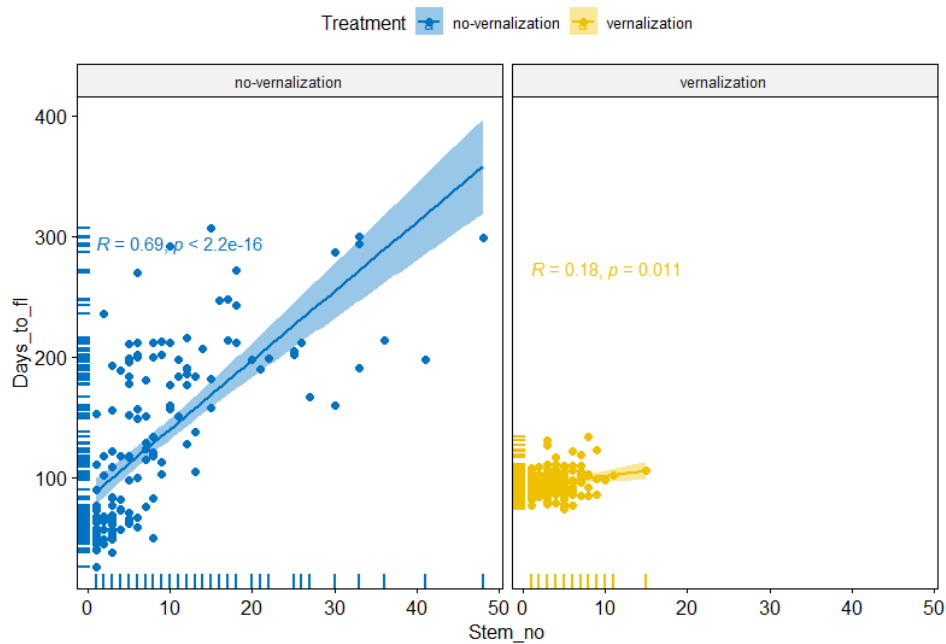
Figure 3. Scatterplots to illustrate the relationship between the number of days a plant takes to flower (Days_to_fl) and the number of observed stems at first flowering (Stem_no) for no-vernalization and vernalization treatments. For both treatments, the correlation was significantly positive (no-vernalization *R=0.69, p=<0.001;* vernalization *R=0.18, p=0.011*).

4209

4210   To observe whether this change of stem number was significant, a two-sample t-test

4211   (assuming unequal variance) can be implied for the two sets of data (Figure 4). The t-test revealed

4212   that the number of stems observed after first flowering was significantly lower when individuals

4213   were vernalized (*P(T<=t) two-tail=<0.001*). These findings suggest that the number of stems

4214   observed when vernalized is significantly reduced, and the positive correlation between the number

4215   of days to flowering and the number of stems observed was also reduced. However, the correlation

4216   was still significant between the two traits, even when vernalized (*p=<0.05* for both treatments).

4217

Stem Number vs Treatments
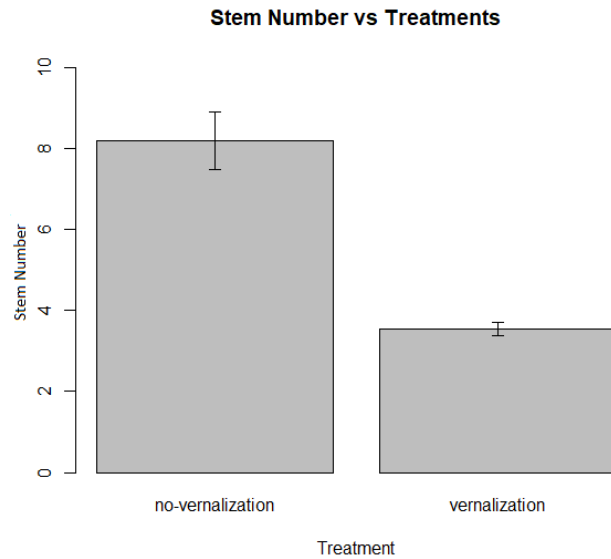
4218 Figure 4. A bar chart to illustrate the mean stem number differences between the two treatments, showing a
4219 significant reduction in stem number when vernalization occurred (*P(T<=t) two-tail=<0.001*).

4220

4221   Another plant traits which were observed during the vernalization study is the number of
4222 buds. The relationship between the number of days to flowering and the number of observed buds
4223 on the first flowering suggests that there was a relationship when plants were introduced to
4224 vernalization treatments (Figure 5). Interestingly, this reduction corresponds with the of number of
4225 days to flower due to vernalization where it was also reduced. Perhaps vernalization is also affecting
4226 the ability of buds to be formed. The correlation between the number of days a plant takes to flower
4227 and the observed number of buds on first flowering suggests that when vernalized they're negatively
4228 correlating with each other (*R=-0.25, p=<0.001*). The result in Figure 5 demonstrates that under
4229 vernalization, plants that flowers longer tend to have less buds. When non-vernalized this trend was
4230 not observed.
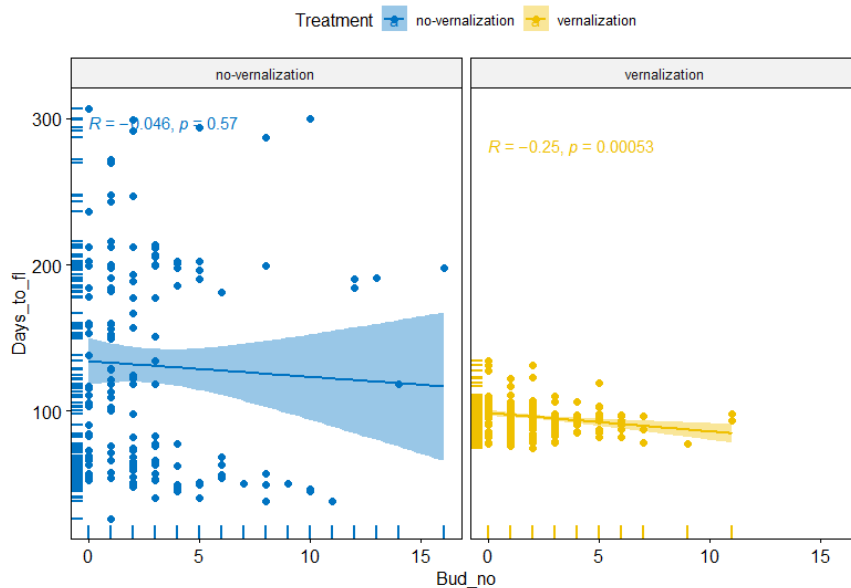
Figure 5. Scatterplots with a linear model to illustrate the relationship between the number of days a plant takes to flower (Days_to_fl) to the number of observed buds on first flowering (Bud_no). The correlation suggests that there was no significant relationship under the no-vernalization treatments ($R= 0.046, p=0.57$). 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).
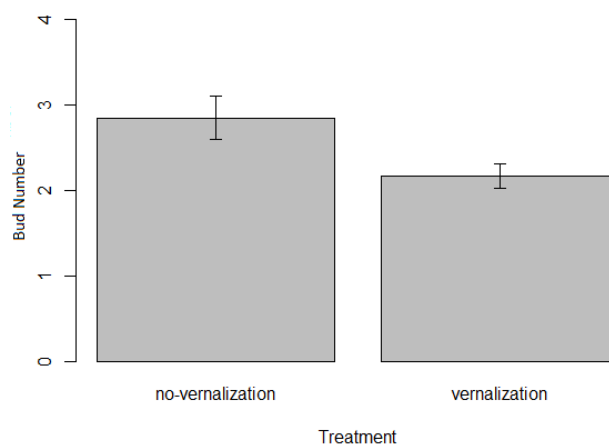
Figure 6. A barchart to illustrate the difference in mean bud number under the two vernalization treatments, revealing a significant difference between the two treatments ($P(T<=t)$ two-tail=0.021).

167

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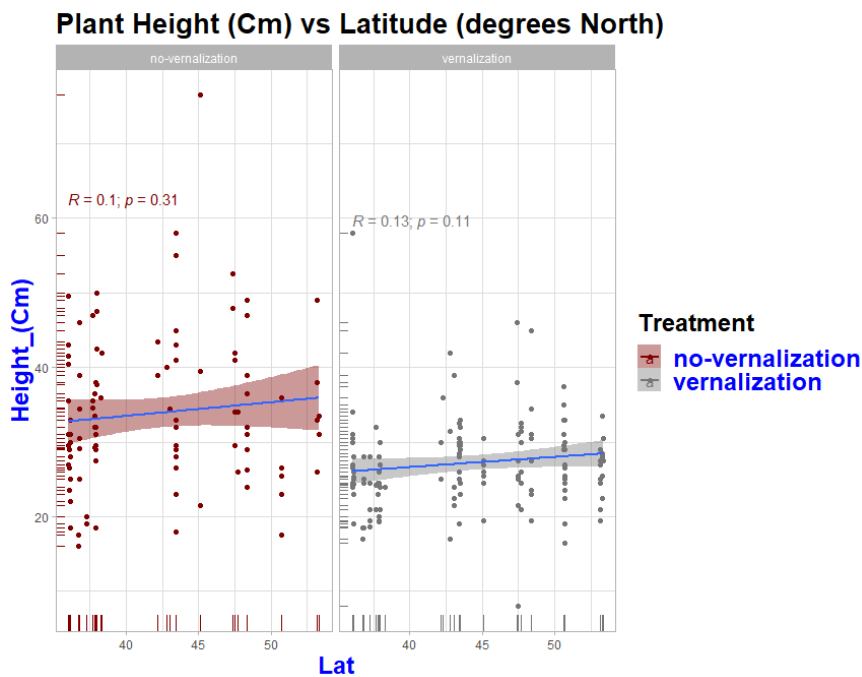
Figure 7. Scatterplots with a linear model to illustrate the relationship between the height of individual plants and their latitude of locality, showing no significant correlation between plant height and latitude (for no-vernalization R=0.1, p=0.31, for vernalization R=0.13, p=0.11).

4259   The number of stems were also observed in relation with the latitude of the individuals in

4260   the two separate experiments. The stem numbers were observed to be less strongly correlated with

4261   the latitude of individuals when *Linum* samples were vernalised.

4262   The result suggests relationship between the number of stems observed and the latitude

4263   locality for individuals under both no-vernalization and vernalization treatments (Figure 8). Under

4264   both treatments the relationship between stem number and the latitude is positive, with

4265   significance under no-vernalization treatment ($R=0.4, p=<0.001$). However, the significance of this

4266   relationship was not seen under the vernalization experiment ($R=0.15, p=0.06$). This suggests that

4267   when *Linum* samples are treated with vernalization, the stem number observed in some individuals

4268   is reduced to the point that significant correlation with larger stem number observed in more

4269   Northern individuals were marginally observable under vernalization.
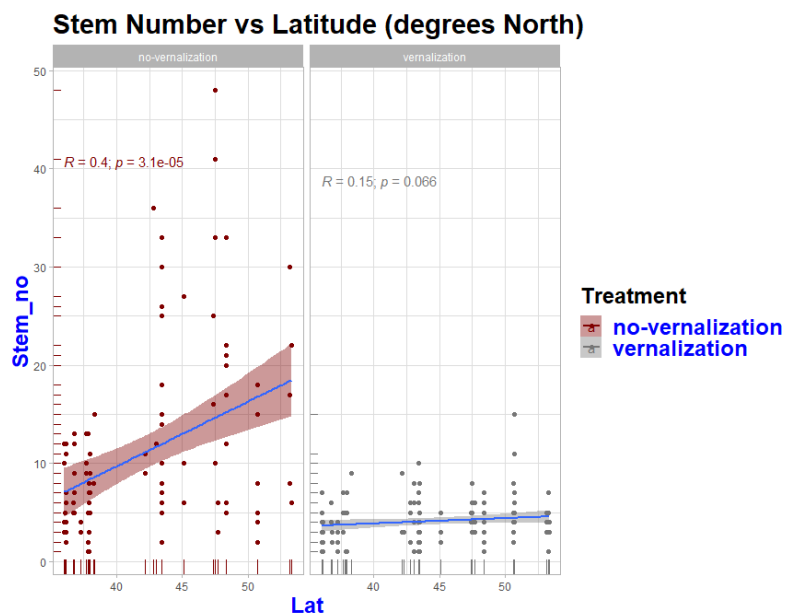
4270   Figure 8. Scatterplots with linear model to illustrate the number of stems observed over the two treatments.

4271   The model reveals that when under no-vernalization treatments, the number of stems observed, is positively

4272   correlated with latitude ($R=0.4, p=<0.001$). This significance was not seen in the vernalization treatments

4273   ($R=0.15, p=0.06$).

4274

4275   Another trait which was observed during this study was the number of buds for every

4276   individual. This was also observed based on the treatments of the individuals, to see whether there

4277   was a trend between the number of buds observed and the latitude of different individuals.

4278   Linear modelling on a scatterplot illustrates that the number of buds observed was

4279   significantly correlated with the latitude of locality under the no-vernalization treatments ($R=0.23$,

4280 $p=0.017$) (Figure 9). However, this relationship was contradicted when *Linum* samples were

4281 vernalized. There was a negative correlation between the number of observed buds and the latitude

4282 of individuals. Although this was the case, the negative correlation observed was not significant (*R=-*

4283 *0.086, p=0.29*). This result suggests that correlation between the bud number and latitude differs in

4284 the two treatments. There is a positive trend, that the more northern an individual is localized, the

4285 more buds were observed when treated with no vernalization. However, when vernalized, this trend

4286 was not observed,
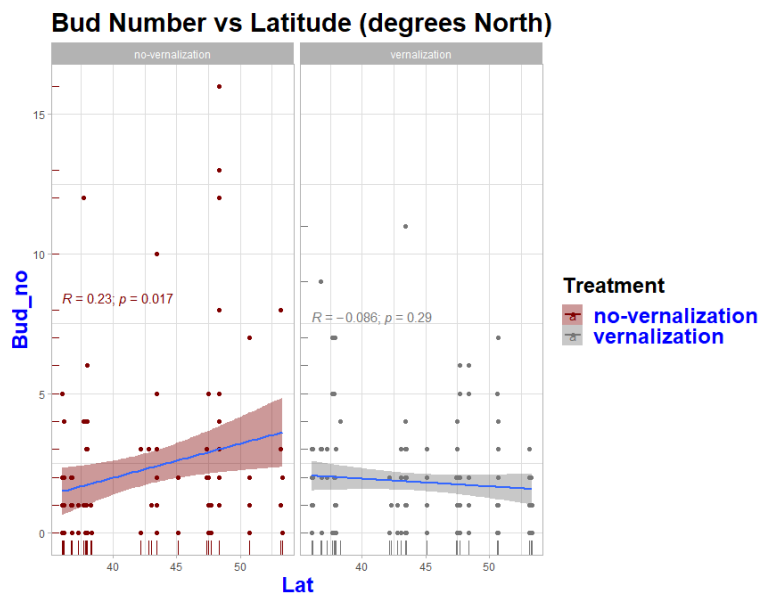
4287 Figure 9. Scatterplots with a linear model to reveal the relationship between the number of buds observed and

4288 latitude between the individuals under no-vernalization and vernalization treatments. The Figure revealed that

4289 there is a positive correlation between the number of bud and latitude on the no-vernalization treatments

4290 (*R=0.23, p=0.017*). The trend became negative when vernalized, although not significantly correlated (*R=-*

4291 *0.086, p=0.29*).

4292

4293

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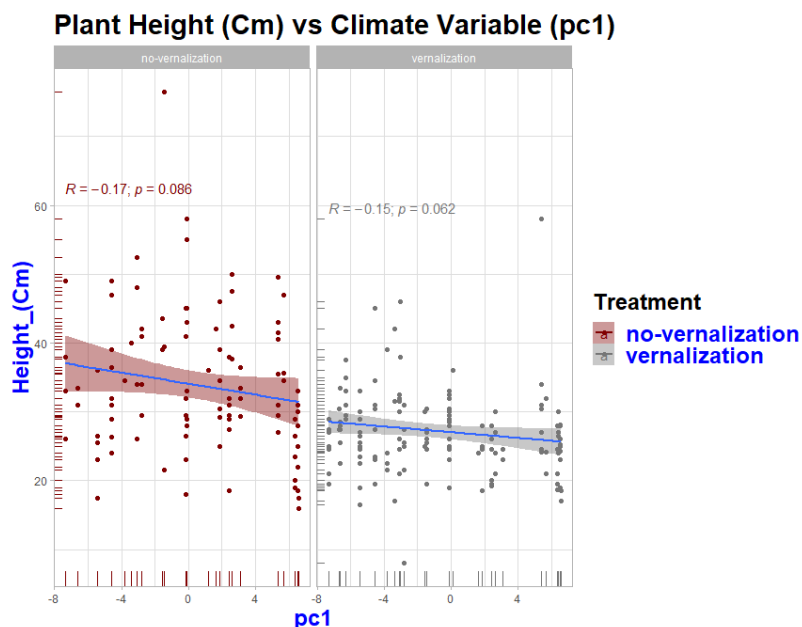
Figure 10. Scatterplots with a linear model fitted to illustrate relationships between plant height (Cm) against climate variable (pc1) for different vernalization treatments. There were negative correlations between the no-vernalization and vernalization treatments with the correlation on the no-vernalization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

4316   reduced for individuals experiencing more negative pc1 climatic variable values. The correlation

4317   between the stem number and the climatic variable are still found to be statistically significant under

4318   α=0.05 (*R=-0.18, p=0.028*).
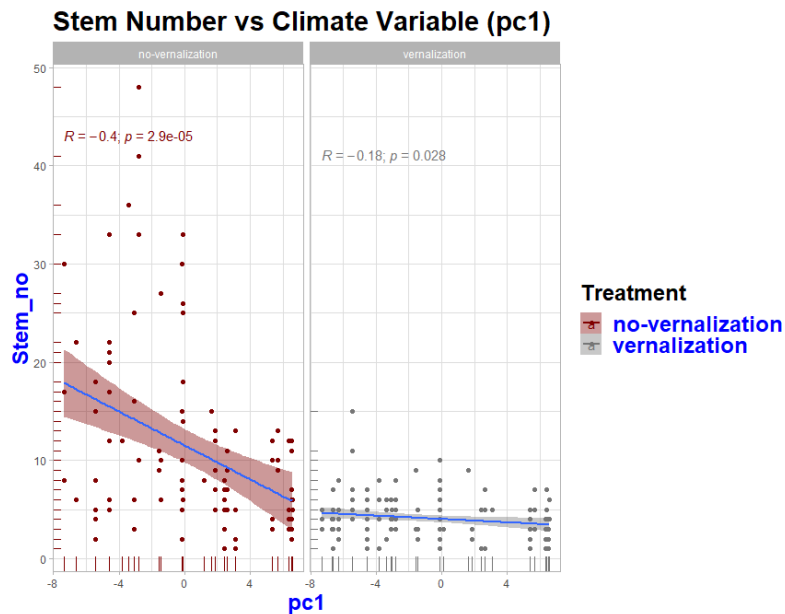
4319

4320   Figure 11. Scatterplot with a fitted linear model to suggest the correlation between stem number and the

4321   climatic variables of their local environments (in terms of pc1). The correlation reveals negative correlation

4322   between the no-vernalization and vernalization treatments, with a bigger significance under the no-

4323   vernalization treatments (*R=-0.4, p=<0.001*).

4324

4325   The number of buds observed was also seen to be negatively correlated with climatic

4326   variables (pc1) under the no-vernalization treatment (Figure 12). This corelation was found to be

4327   statistically significant when modelled (*R=-0.22, p=0.028*). In contrast with what was observed with

4328   the stem number though, the correlation between bud numbers and pc1 seems to have changed

4329   directions under the vernalization treatments. Albeit, not statistically significant (*R=0.12, p=0.15*).

4330   This difference suggests that vernalization affects the number of buds observed during this study to

4331   a point that it no longer correlates with local environments as observed under the no-vernalization

4332   treatments. This reveals when all plants were vernalized, the individuals with the most numbers of

4333   buds observed under the no-vernalization treatments, reduced their number of buds drastically

4334   through the observed effects of vernalization.

Bud Number vs Climate Variable (pc1)
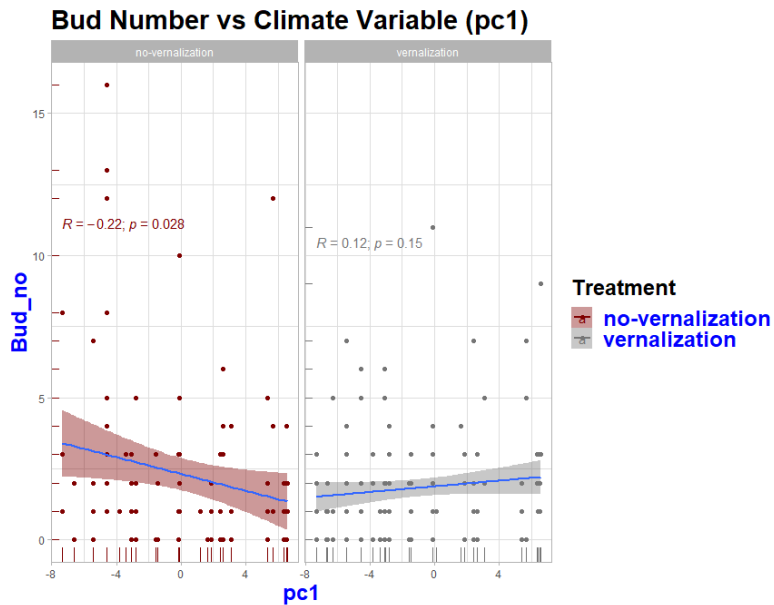
4335 Figure 12. Scatterplots with fitted linear models which suggests negative correlation between observed bud

4336 numbers and climatic variables (pc1) under the no-vernalization treatments ($R=-0.22$, $p=0.028$). The

4337 correlation changed directions when vernalization was introduced. However, this correlation was not

4338 statistically significant ($R=0.12$, $p=0.15$).

4339

5.3.4 Seed Area vs Latitude

4341   As a part of this chapter, we examined seed size variation in wild *Linum* based on different

4342 latitudes. It is of interest to see whether seed sizes of the different wild *Linum* populations suggests

4343 significant difference to each other. We investigated whether seed size (seed area) forms a

4344 correlation to environmental variables. We investigate the question of; "to what extent can

4345 localization to a local environment in wild *Linum* be of an affect to their seed size?"

4346 To explore this, a preliminary seed size measurements were able to be carried out, using data from

4347 non-vernalization, vernalization and as an addition, wild individuals. Microscopic images of 5 seeds

4348 were taken. The images were taken with the same light levels, making sure all 5 seeds were visible

4349 on the image. After the image were captured, it was saved into a drive and were further processed

4350 using ImageJ (Schneider C. et al., 2012). Seed areas were able to be highlighted and calculated in

4351 ImageJ. A macro was written to semi-automate the process. The macro can be found under

4352 appendix 32. In this case, vernalized individuals were labelled as "Vern18", and non-vernalized

4353 individuals were labelled as "nonvern18". Wild S0 individuals were also able to be counted observed,

4354 labelled as "Wild".

4355

4356   When looking at the correlation between seed area and latitude, it is found that in all

4357 treatments, seed areas are significantly correlated for all treatments (nonvern18 *R=0.51, p=<0.001*;

4358 vern18 *R=0.44, p=<0.001*; Wild *R=0.61, p=<0.001*) (Figure 13). The wild seeds were also correlating

4359 stronger than the vernalized seeds. This suggests in all cases seed area is positively correlated to

4360 latitude and that initial vernalization treatments will result in no significant change to this

4361 correlation.
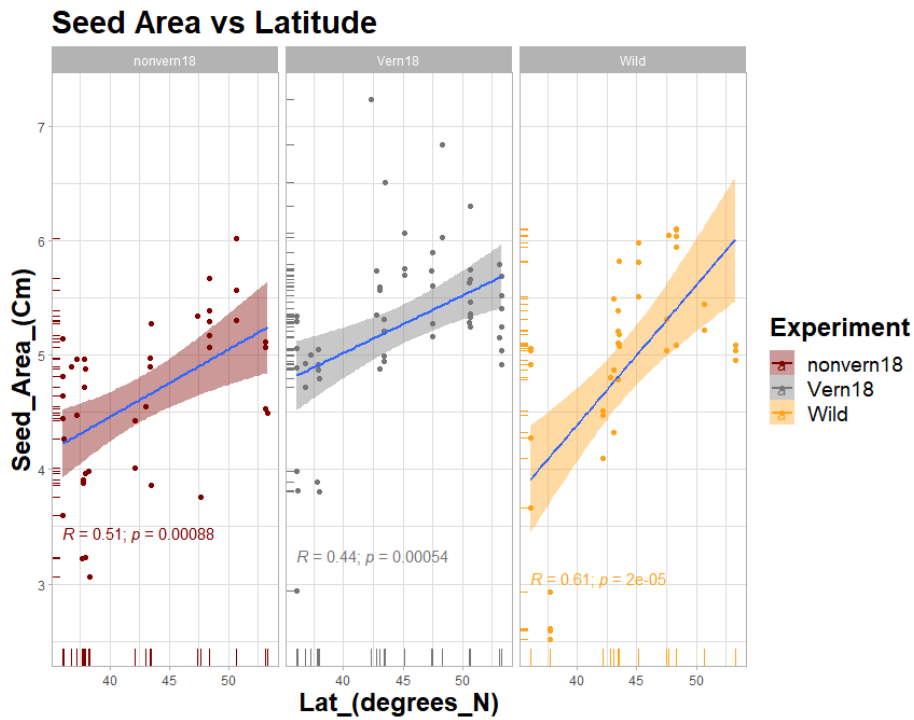
4362 Figure 13. Scatterplots with a linear model to show the seed area (Cm) in relation to latitude for our individual

4363 *Linum* samples. The correlation illustrates that seed area is significantly correlated with latitude of locality for

4364 all our treatments, the correlation was considerable higher in the wild (nonvern18 *R=0.51, p=<0.001*; vern18

4365 *R=0.44, p=<0.001*; Wild *R=0.61, p=<0.001*).

4366

5.3.5 Seed Area vs Climate Variable (pc1)

4368   For completion, it is with interest to further see whether seed size in terms of total seed

4369   area observed for different individual sampled during this study. The scatterplots with linear

4370   modelling suggest negative correlations between seed area and climatic variables (Figure 14).

4371   Negative pc1 values represent loading values for colder climates (see chapter 2, under section 2.2.1

4372   Samples and Experiment"). This data suggests that in all treatments, the seed sizes seem to be larger

4373   in area in individuals found in a colder climate and that vernalization have little effects to change this

4374   significance. In the wild individuals, the correlation seems to be stronger.
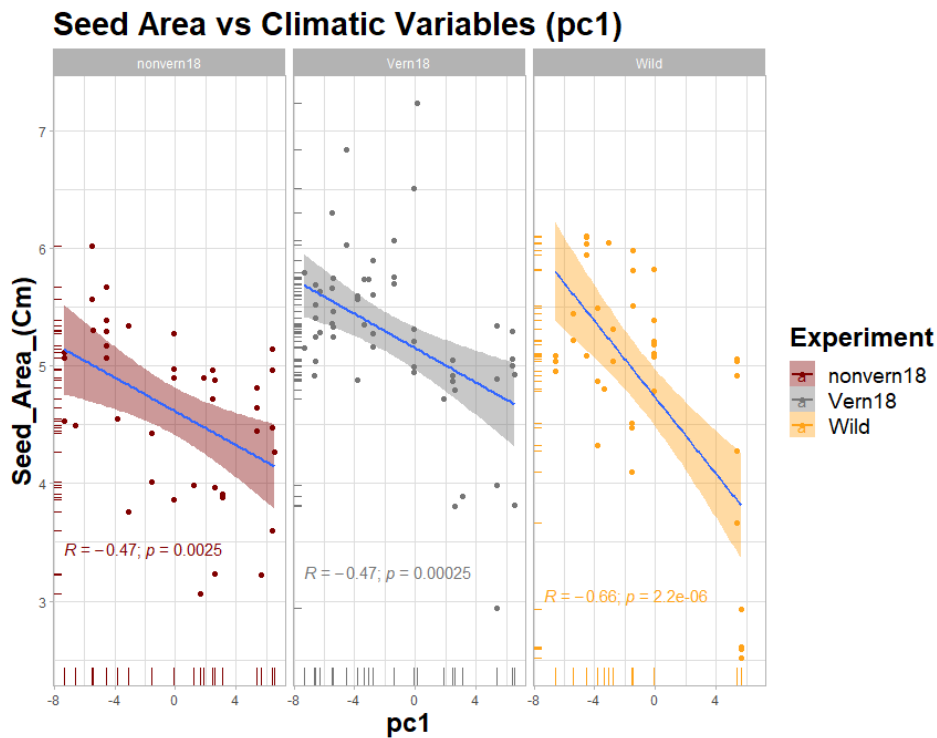
4375   Figure 14. Scatterplot with a linear model to illustrate the correlation between the seed area (Cm) and the

4376   climatic variables (pc1). The correlation suggests that there is a negative correlation between seed area and

4377   climatic variables in all treatments (nonvern18 *R=-0.47, p=0.0025*; vern18).

4378

4379

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

4412

4413   There was also a positive correlation between the number of days individuals takes to flower

4414 and stem number under the no-vernalization treatments. It seems that the longer an individual

4415 takes to flower the more stem it can develop before first flowering. This correlation was reduced

4416 when vernalization was implied to the wild individuals, with individuals under vernalization

4417 treatments significantly forming a smaller number of stems. However, the correlation between the

4418 number of days to flower and the number of stems was still significant. This shows that stem

4419 number is highly correlated to the number of days individuals can flower initially even when wild

4420 individuals were exposed to vernalization. When testing against different treatments, individuals

4421 exposed to the vernalization experiments have significantly a smaller number of stems. This suggests

4422 an effect of vernalization in the number of stems wild *Linum* individuals can form. There is

4423 suggestion in the literature that different plant populations may have different number of stems

4424 according to their local environment (Sultan, 2000; Szakiel et al., 2010; ATES., 2011). Vernalization

4425 suggests stark differences in stem number when compared to non-vernalization treatments, which

4426 suggests that in *L. bienne*, this is also true. In other cases, vernalization affects the morphology of

4427 stem formation in *Phleum pratense* . (Seppänen et al., 2010) This illustrates the potential of

4428 vernalization as an abiotic stress to affect developments of stems in wild *L. bienne* found across

4429 Western Europe.

4430

5.4.3 Number of Days to Flower and Bud Numbers

4431

4432   In terms of the number of buds forming, our results suggest that under the no-vernalization

4433 treatment this measure was not significantly correlated to the number of days an individual is able

4434 to form their first flower. When exposed to vernalization however, the number of buds observed

4435 seems to negatively correlate with the number of first flowering as they were significantly reduced

4436 due to vernalization as an abiotic stress. This suggests that under vernalization requirements, there

4437 may be groupings of plants which can form a higher number of buds with a shortened number of

4438 days to first flowering. In all three cases of the trait measures, including number of buds, variation in

4439 correlation between the number of days to flower and traits measured in this chapter occurred . In

4440 all cases though, vernalization reduced the number of observed bud numbers significantly. This

4441 suggests that vernalization is an abiotic stress that wild population are potentially locally adapting

4442 into.

4443

5.4.4 Effects of Vernalization in Plant Traits

4444
4445   With the above additional results for the plant traits (plant height, stem number, bud
4446 number) considered, it is suggested that the effect of vernalization leads to significantly lower values
4447 in all three trait measures. This suggests that under vernalization, wild *Linum* samples in this study
4448 experienced significant reductions in desired plants traits such as height and plant stem number. In
4449 the model *Arabidopsis thaliana*, it has been observed that traits such as leaf sizes were reduced
4450 under vernalization (Hopkins., 2008). The study also found a latitudinal cline to this response in
4451 vernalization. This suggests that morphology of plant traits may be affected by vernalization and that
4452 environmental variables may also play a part. In the study for this chapter, it is found in wild *Linum*
4453 *bienne* samples across Western Europe that plant height, stem number, and bud number is negative
4454 affected by vernalization. Findings in these traits under vernalization, adds to the suggestion that
4455 vernalization is an abiotic stress and that variation in sensitivity observed can be seen as potential
4456 local adaptation mechanisms between wild *L. bienne* populations found across Western Europe. To
4457 thrive in a colder niche there could be a case that some wild *L. bienne* populations are more locally
4458 adapted to colder climates through requirement of vernalization. In addition, manipulation to
4459 flowering under different seasons/environments can also be studied further as suggested under
4460 other studies observing vernalization in *Calandrinia* (Cave & Johnston, 2010). Reduction in desirable
4461 traits also suggests that the loss of vernalization under *L. usitatissimum*, observed in chapter two of
4462 this thesis could be a benefit in terms of agriculture. Loss of vernalization could mean that negative
4463 effects on traits such as plant height, stem numbers and bud numbers observed under vernalization
4464 in the wild *L. bienne* is selectively avoided in Northern European Winter and Spring *L. usitatissimum*
4465 types. Although this was the case, we assumed that the data is normally distributed, and that values
4466 of trait measurements, and other variables are independent of each other. Traits observed may be
4467 dependent of each other and that the same genes may influence multiple traits observed in this
4468 chapter. These analyses do not consider these possibilities.

4469   The response to plant height in both vernalization treatments, suggests there was no
4470 correlation with the latitude of locality for these plants. However, when looking at local climatic
4471 variables, there was negative correlation between the local climate variable against plant height
4472 when no-vernalization occurred, with individuals from warmer climate (bigger pc1 values) being
4473 taller. This correlation became less significant under vernalization treatment. This suggests that
4474 perhaps climatic variables are having stronger effects to plant height, in our case, individuals under
4475 no-vernalization seems to be taller in colder climate (smaller pc1 values). Vernalization also plays a
4476 part in this, when plants were exposed to vernalization, it seems that this correlation became less
4477 significant. It would be viable to suggest that the process of vernalization induces stress to wild *L.*

4478 *bienne* individuals. To undergo vernalization and flower at the correct time, individual requiring

4479 vernalization may have less favourable traits. Loss of vernalization requirement in the cultivated *L.*

4480 *usitatissimum* in Northern Europe (as discussed in chapter two), could mean that this process is

4481 bypassed in some cultivar types and stress induced by cold for plants to require vernalization is not a

4482 factor. This may form as one of the factors enabling the cultivars to have form better in traits of

4483 agricultural interests for their biproducts. We didn't compare traits between cultivars and wild

4484 species when vernalized as cultivated samples were not included in this chapter for analysis. This will

4485 be of interest for future study.

4486

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

4487

4488 When looking at other traits against latitude, such as the stem number, there was also

4489 suggestions that the number of stems observed is correlated with the latitude of locality for the

4490 individuals for the no-vernalization treatment. In this case the correlation seems to be positive, with

4491 more Northern plants revealing a greater number of stems. This was also supported when looking at

4492 the climatic variable. A negative correlation was observed which suggests that individual from

4493 warmer climates (higher pc1 values) having a smaller number of stems. When looking at individuals

4494 experiencing vernalization, the stem number able to be formed was not significantly correlated to

4495 either latitude of locality or local climatic variables. This suggests varied morphology in response to

4496 vernalization for our *Linum* collection. The bud number observations suggest there was a positive

4497 correlation between the observed bud number and latitude of localities under no-vernalization

4498 treatments, this suggests more Northern individuals were able to form more buds under no-

4499 vernalization. When vernalized however, the correlation became negative, with more Northern

4500 individuals suggesting for be forming less buds. However, this change of direction in correlation, was

4501 not observed to be significant. When looking at bud number and pc1, it is supporting the traits seen

4502 with latitude, with the two treatments showing contrasting correlation. In this case negative

4503 correlation between bud number and climatic variable (pc1) suggests that individuals found in

4504 warmer climate seems to form less buds. The correlation changed direction under vernalization

4505 treatments, however, as seen with the latitudinal correlation, this change in correlation direction

4506 was not statistically significant.

4507 In all three cases of the measured traits, there is suggestion of differentiation in strategy for

4508 different individuals and when no-vernalization and vernalization treatments were implied. This

4509 suggests there may be genetic interest for looking at these differences in morphology, as was found

4510 under chapter two of this thesis when looking at relative gene expression of flowering time genes

4511 and the number of days an individual takes to flower. In a study with the majorly cultivated durum
4512 wheat, a study has found different phenology and trait strategies in different landraces. This was
4513 also linked to differences in genetic variation responses for the vernalization gene (*VRN1*) and
4514 photoperiod sensitivity gene (*PPD1*) (Royo et al., 2020). There is suggestion in the literatures which
4515 highlight different morphology and phenotype of species of plants due to vernalization and suggests
4516 that this is also linked to genetic variation. It is likely that differences seen in plant height, stem
4517 number, and bud number in wild *Linum* is a result of genetic variation due to vernalization
4518 responses. It will be with great interest in the future to look at specific genes which may play a part
4519 in plant growth, specifically for stems and buds' formation to see whether variation between
4520 individuals still correlate and to see whether variation seen between no-vernalization and
4521 vernalization treatments can be linked genetic factors.

4522

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

4524   We also observed seed sizes in terms of seed areas, against latitude of locality for every
4525 individual as well as local climatic variable for them. It is suggested that there was significant
4526 variation between seed area, and these were all correlated with latitude and local climatic variables
4527 (pc1). The correlation was still retained under no-vernalization and vernalization treatments. Positive
4528 correlation with latitude suggests that in all treatment cases, the more Northern individual's seeds
4529 were bigger than those of their Southern counterparts. This also translates to the correlation seen in
4530 climatic variables, with negative correlations, suggesting that individuals from colder regions having
4531 the bigger seeds. Bigger seeds may imply better fitness in plant reproduction, smaller seeds are
4532 however, easier to produce, suggesting this trade-off between seed size and the number of seeds a
4533 plant can make for a given amount of energy (MOLES et al., 2004). In a study with wild barley
4534 (*Hordeum vulgare ssp.*), there is a suggestion that variation in seed sizes contributes to plant growth
4535 and reproduction, with larger seeds often found to have better fitness (Giles., 1990). A study on
4536 different *Glycine* species found that seed mass tends to be larger at Northern latitudes. They also
4537 found that the major contributing climatic variable to this are temperature and daylight availability
4538 (Murray et al., 2004). These agree with our finding that *Linum* seeds also found to be larger in
4539 Northern populations. When looking at the wild seeds in our observations, the correlation between
4540 seed size and latitude and climate variables were found to be stronger. In sand rice (*Agriophyllum*
4541 *squarrosum*), there was also geographic variation of seed size due in wild populations. In *Linum*, the
4542 seed sizes are strongly correlated to local variation in climatic variables. It would be of interest to
4543 genetically evaluate this to see whether the strong correlation found in this thesis had genetic
4544 implication for genetic basis which may reveal evidence of genetic variance. In the past, studies

4545   looking at genetic variance for variation in seed sizes in different plant species have found no

4546   correlation between genetic variation and variation in seed sizes, suggesting that seed size could not

4547   have evolved due to natural selection (Wolfe., 1995) (Schwaegerle & Levin., 1990). It would be of

4548   good interest to genetically evaluate the seed size variation under different environmental variables

4549   as illustrated in this thesis.

4550

4551

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

4585   population are locally adapting to the environmental variables such as precipitation, solar radiation,

4586   and temperatures. Wang J. et al., 2018 illustrates that for perennial plants, the locus *FT* forms part of

4587   the genomic region responsible for local adaptation (Wang J. et al., 2018). We make a case that

4588   evidence of local adaptation can be observed in wild *L. bienne* across Western Europe, based on the

4589   relationship seen between *FT* expression and latitude and climatic variables. Although this is the

4590   case, there is a possibility that variation in requirement to vernalize can arise from mechanisms such

4591   as genetic drift. We didn't have precise location data for our cultivars. It would be of further interest

4592   to compare the relationship of *FT* expression with latitude/climatic variables for the different *Linum*

4593   species. We expect that, due to domestication and the loss of requirement to vernalize observed in

4594   this study, our *L. usitatissimum* variants would not form a relationship with latitude/climatic

4595   variables. This would suggest a contrast between the wild *L. bienne* and the cultivated *L.*

4596   *usitatissimum*. Whilst the requirement to vernalize may be artificially selected in the cultivars, the

4597   wild suggests that flowering initation and requirement to vernalize may be driven by local

4598   adaptation under various climate. Although this was our finding, we suggest taking the results with

4599   precautions. Along with a relatively small sample size, relative gene expressions were comparative

4600   only against one HKG. We propose a future study with bigger sample sizes from a wide range of

4601   latitude, specifically more Southern populations and their floral gene relative expression compared

4602   against multiple HKG's. This is to robustly reflect the correlation with *FT* expression and latitude. In

4603   addition to this, it would be ideal to observe expression of genes directly in the vernalization

4604   pathway such as *FRI*, *VRN1* and *FLC*. We expect that the relative expression of these genes is related

4605   to the latitude and local climate in which the wild population are found since they affect *FT*

4606   expression.

4607   We saw in chapter two that the number of days wild *L. bienne* takes to flower is related to

4608   the latitude. This implies that temperature, as a variable, may control for the number of days to

4609   flower. To investigate whether temperature had further effects on flowering mechanisms, we

4610   continued to observe responses in floral organs. We explored pollen viability in response to

4611   reduction (-5°C from initial temperature) and addition (+5°C from initial temperature). We see a

4612   response in reduction of pollen count for both reduction and addition of temperature. When looking

4613   at pollen tube count, this varied within all our temperature treated samples. However, when scaling

4614   this with the pollen (proportion of pollen tube forming to pollen count) only the addition in

4615   temperature shows significant reduction in the proportion of pollen tube forming. This suggests that

4616   *Linum* pollen tubes are potentially sensitive to increases in temperature. To reinforce this, we found

4617   the proportion of pollen tube able to reach the ovary to be significantly reduced under the +5°C

4618   temperature treatment. This was observed in both species. This may suggest that pollen ability to

4619 germinate is limited to temperature or the female floral organs are not able to sustain pollen tubes

4620 under heat stress. This suggests the interest in looking at ability of female counterparts to sustain

4621 pollen tube growth under different temperature for future studies in this area.

4622   We concluded this chapter by observing whether latitude/climate variables are correlated

4623 with the amount of pollen and pollen tubes observed in *L. bienne* when treated to the different

4624 temperature treatments. We found no correlation between the number of pollen and

4625 latitude/climate. However, the number of pollen tube and its proportion to the number of pollen

4626 observed a correlation with latitude and climate in the 20°C (-5°C) treatment. Furthermore,

4627 correlation with latitude is found to be positive, illustrating that more Northern individual can form

4628 more pollen tubes. This may suggest potential adaptation to the colder temperature for the more

4629 Northern individual in relation to its ability to form pollen tubes and thus increasing its chances of

4630 germinating. We suggest that this also formed because of local adaptation found in the Northern

4631 population of Western European *L. bienne*. From this chapter we reinforced the suggestion that local

4632 adaptation between the North and South wild *L. bienne* in Western Europe occurs, as was illustrated

4633 in chapter two. For future research, it would be of interest to look at this from the female

4634 counterpart's perspective. We suspect that ovary would be sensitive to heat-stress and stigma will

4635 have varying ability to sustain pollen tube growth. In addition to this, it would also be beneficial to

4636 conduct a GWAS study to identify the genes which may play a role in cold resistance in the wild

4637 which maybe linked to the higher amount of pollen tube able to be observed under wild individuals.

4638   In chapter four we analysed 100 Western European *Linum* individuals and their population

4639 genetics. One of the main results from this chapter was the observation of Northern and Southern

4640 groups in the wild *L. bienne*. There seems to be structuring between the Western European wild *L.*

4641 *bienne* population as was inferred in section 4.4.1 "Results: Genetic Structure". We observed initial

4642 signal of potential variation in breeding strategies of *L. bienne* found across Western Europe in our

4643 population diversity analysis. There is potential that the Southern group of wild *L. bienne* is more

4644 outcrossing in terms of their breeding strategy and thus causing this genetic structuring we see

4645 when illustrating genetic distance tree. In the literatures, it is found that various mode of seed

4646 dispersal often affects a plant's breeding strategy. They found that animal-dispersed plant species

4647 exhibited higher levels of genetic diversity and lack of inbreeding (Nazareno et al., 2021). This further

4648 suggests that there is potentially more variation within Southern *L. bienne* inidividuals in Western

4649 Europe. All measures of population diversity suggests that heterozygousity is higher in the Southern

4650 cline, with relatively little amount of heterozygousity observed in the Northern group. Measures of

4651 population divergence such as Tajima's D also suggest that Southern population is more recently

4652 expanded and are expanding more than the Northern group. These findings suggests that there is

4653   potential variation in breeding strategies and dispersal of North and South groups of *L. bienne* found

4654   in Western Europe. The implication of variation in breeding strategies could result in the variation

4655   we saw in chapter two whereby Northern individuals tends to be more sensitive to vernalization as

4656   was expressed by their variation in relative ge expression under different latitudes. Furthermore, we

4657   saw that ability of pollen tube forming and germination processes is correlated with individual's

4658   latitude when treated to "cold" treatments (see section 3.4.8 "Results: Pollen and Latitude of

4659   Origin"). Genetic structuring and diversity analysis in chapter four supports the findings that

4660   Northern and Southern clines of *L. bienne* found in Western Europe is potentially locally adapted to

4661   their environment.

4662   In addition to the observation between Northern and Southern clusters, we observed a

4663   Mediterranean population which is significantly diverged from other populations. This could be a

4664   result of increased rate of outcrossing in nature we saw in the Southern clines when looking at

4665   population diversity measures. It would be of interest to include population from the Eastern

4666   European region and see if this cline between North and South population could also be observed

4667   between Western and Eastern European wild *L. bienne*, with Eastern population potentially more

4668   outcrossing in breeding strategy. In future studies investigating Western and Eastern *L. bienne*

4669   population, wild *L. bienne* seeds needs to be collected from Eastern European regions such as Turkey

4670   and the Eastern Mediterranean. There is no availability of S0 seeds from these wild population to

4671   date.

4672   We also acknowledge that we may have lost some valuable data by mapping to the *L.*

4673   *usitatissimum* whole genome. This was inferred from the higher number of SNPs gained from *de*

4674   *novo* mapping (see section 4.5.4 "*De_novo* mapping"). In addition, it is worth exploring other NGS

4675   techniques to imply further structuring based on a larger coverage of the genome. This is because of

4676   disadvantages to using ddRADSeq to sequence our short reads. Our ddRADSeq only studies 250-

4677   500bp of sequences per individual cut by the digestion process of ddRADSeq and require high quality

4678   DNA which may impact the number of individuals able to be analysed and outputted from the

4679   ddRADSeq procedures. We also looked at two species with potential divergence in our wild

4680   populations. Other recent short-reads NGS techniques that can be looked at for observation of

4681   mulptiple species is Sequence capture. Sequence capture holds more promise for obtaining data sets

4682   that are comparable across species and for calibrating parameter estimates for demographic or

4683   phylogenetic studies (Harvey et al., 2016). Sequence capture also doesn't require a high quality of

4684   DNA as opposed to RAD Sequencing processes. There is potential that individuals with a relatively

4685   lower quality of DNA could be analysed further with Sequence capture (Harvey et al., 2016). In

4686   addition, studies using Sequence capture can capture larger sizes of sequences (300-1200bp) which

4687   covers 972 genes (Sanderson et al., 2020). Future studies can look at the potential of using Sequence

4688   capture for population genetic analysis in *L. bienne* across Western Europe.

4689   When looking at traits in chapter five, we found evidence that plant height, stem number,

4690   and bud number is significantly reduced by the introduction of vernalization. This suggest that

4691   vernalization is an abiotic stress variable in wild *L. bienne* found across Western Europe. When under

4692   no-vernalization, we saw positive correlation between stem and buds formed at first flowering and

4693   latitude. This illustrates that under no vernalization stress, more Northern individuals can form more

4694   stems and buds. However, under vernalization there was no observed correlation between traits and

4695   latitude/climate. This suggest that, in the wild, vernalization acts as abiotic stress and as was found

4696   in chapter two, vernalization responses varies across the wild *L. bienne*. In terms of local adaptation

4697   in wild *L. bienne*, the ability Northern individuals to form more stems and buds, under no-

4698   vernalization can suggest that seasonal queues, especially regarding vernalization, may occur in the

4699   Northern cline. This ability to form more stems and buds under no-vernalization was seen less in the

4700   Southern cline, suggesting the potential for local adaptation regarding vernalization requirement in

4701   the Northern cline, which are expected to be more sensitive to vernalization due to colder climates.

4702   This was also supported by higher expression of the floral integrator *FT* as was suggested in chapter

4703   two. In future studies, it is of interest to look at expression of genes directly affecting number of

4704   stems and buds to reinforce the variation seen within this thesis with genetic expression studies.

4705   We further support the potential local adaptation between Northern and Southern clines of

4706   *L. bienne* population using seed sizes. Under three growth conditions, seeds areas are positively

4707   correlated with latitude and negatively correlated with climatic variables, mostly attributed to colder

4708   temperatures. This suggest that Northern individuals seems to have a bigger seed area to the

4709   Southern individuals, which can infer that trade-off between seed size and potential to produce

4710   more seeds between the two clines. In the wild S0 seeds, this correlation between seed area and

4711   latitude is more strongly observed, which may suggest that seed area is a variable for measuring

4712   seed size variation in wild *L. bienne* for future studies regarding local adaptation and breeding

4713   strategies of wild *L. bienne*.

4714   With the above five chapters concluded, we illustrated potential signal of local adaptation in

4715   wild *L. bienne* to Northern climates, requiring vernalization. We observed no sensitivity of *L.*

4716   *usitatissimum* in our sample set to vernalization. We expect there are variation between various *L.*

4717   *usitatissimum* variety regarding vernalization sensitivity, especially for non-winter types. As was seen

4718   in *L. bienne*, this vernalization requirement can suggest reduction in traits that are beneficial to

4719   argriculture, and so loss of vernalization in winter types maybe beneficial for these traits of interest.

4720   We imply that the loss of vernalization is agriculturally beneficial, however, this may limit growth

4721   time and cycle of plants under different seasons, thus resulting in seasonal types of *L. usitatissimum*

4722   varieties. Wild Northern population could be studied to understand mechanisms of vernalization in

4723   *Linum*, this may add to breeding purposes, providing *Linum* cultivars that are perhaps more resistant

4724   to colder climates an can be grown in different seasons.

4725   In terms of wild *L. bienne* population found across Western Europe, population structure

4726   suggests that gene flow between the Northern and Southern group in our samples are limited due to

4727   the grouping suggested by the Ward's D tree based on Nei's genetic. Genetically, we saw little

4728   evidence of geneflow between the wild population which is a signal of locally adapting populations

4729   (Boshier et al., 2015). Phenotypically, we saw variation for flowering traits for Nortern and Southern

4730   group of population within our sample set as a signal of this local adaptation.

4731   Although this was the case, in our PCoA we saw a few outliers to these groupings suggesting

4732   a cline between the Southern and Northern populations. This suggests a limited geneflow between

4733   the two groups, albeit we think that this geneflow is limited to one or two individuals in our sample

4734   set. We suggest future studies to look at more wild population originating in Northern Spain, as this

4735   is where we saw potential outliers which may suggest gene flow between the Southern and

4736   Northern groups of wild *Linum* observed in this study. Further divergence can be seen in our

4737   Mediterranean population. Population divergence analyses suggests that the population E-12 in our

4738   sample set is distinct to both Northern and Southern group with a higher number of private alleles

4739   and heterozygousities. This suggest that perhaps this population is part of a diverging group of wild

4740   *L. bienne*. Higher allelic richness can also suggest a different approach to breeding strategies than

4741   seen in the rest of the Western European wild population which may allude to differentiation in

4742   mechanisms such as seed dispersals.

4743   Less heterozygous alleles in the Northern group of population also infer less allelic richness

4744   and potentially more inbreeding, suggesting that these Northern group have a more conserved

4745   genetic makeup, potentially counting for local adaptation to a colder climate as we initially observed

4746   under flowering mechanisms such as vernalization. However, it is interesting to find that there are

4747   intermediate individuals forming a cluster between the Southern and Northen population in our

4748   PCoA analyses. This suggests that perhaps several populations based in Northern Spain acts as a

4749   geneflow meditator, connecting the two distinct Western European wild flax population. This could

4750   suggest that, although the Northern group of wild *L. bienne* population are more conserved and

4751   locally adapted to Northern climates, they have diverged from a Southern population. This suggests

4752   implications for *Linum* to diverge into colder climates and are locally adapting to these climates. This

4753   implies the importance of colder climates for natural selection of wild plants and their

4754   diversification. This could impact availability of a wider genepool for study on mechanisms to

4755 counter cold climates in plants, vernalization to name one. These wider genepool are useful for

4756 future breeding purposes agriculturally. In a warming climate, these wider genepool may be

4757 threatened as they are found to be more conservative and inbred than their Southern relatives.

4758   Other studies suggests that the Mediterranean is a "hotspot" for recent plant diversification

4759 (Buira et al., 2020). We think that wild *L. bienne* population originating further in the Mediterranean

4760 as opposed to Western Europe may be more diverse and distinct than previously thought for *L.*

4761 *bienne* across Western European region. Diverging Mediterranean populations may also adapt at a

4762 different rate to their environment and are more outcrossing (as was observed in population E-12).

4763 We think that Mediterranean wild *L. bienne* population may provide a wider genepool for genetic

4764 study of *Linum* and have the potential to provide diversification to breeding strategies in *Linum*. To

4765 develop this further, we suggest collection of further wild *L. bienne* from across the Mediterranean

4766 region to confirm this genetic diversity from the Western European relatives, seen in this study.

4767   We take these implications within this thesis with precautions as the number of sample sets

4768 are relatively small and minimum representative for population genetics purposes was only seen to

4769 three individuals. For future relative gene expression studies, a confirmation using multiple HKG's

4770 needs to be conducted as future research towards the results found within this thesis. However,

4771 signals of local adaptations can still be observed even with these limitations in the wild *L. bienne*

4772 found across Western Europe. This would imply potential traits of interest in relation to local

4773 adaptation which maybe playing a role towards resistance to abiotic factors such as vernalization.

4774 There are also implications for further studies observing cold-resistance genes and pollen viability in

4775 wild *L. bienne* to observe mechanisms of cold resistance in future *Linum* GWAS studies that maybe

4776 beneficial for implication to *Linum* agriculture.

4777

4778

APPENDICES

4780

4781 **Appendix 1: A list of *L. bienne* and *L. usitatissimum* samples collected for this thesis**

4782 *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

4783

4784

4785

4786 *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

4787

4788

Appendix 2: Sample lists for vernalization study
4789

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
low1_10	low1	Bienne
IOW1_11	low1	Bienne
low1_17	low1	Bienne
low2_2	low2	Bienne
low2_25	low2	Bienne
low2_26	low2	Bienne
low2_30	low2	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

4790

4791

4792

4793

4794

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

4796

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
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

Loadings (vertical label spanning the climatic-variable rows)

4797

4798

4799 **Appendix 4: List of Primers and its sequences for flowering time experiments in Linum**

4800

4801

4802

4803

4804

4805

4806

4807

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

4808

4809

4810

4811

4812

4813

4814

Appendix 5:

Primer efficiency

comparisons:

(We used #3)

#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

4815

4816 **Appendix 6: Relative Gene Expression R commands:**

4817 library(ggpubr)

4818 library("gridExtra")

4819 library("ggplot2")

4820 library("cowplot")

4821 library("dplyr")

```
4822    library("ggpubr")

4823    library("viridis")

4824    library(readr)

4825    combined_vern_same_ind_clean1 <- read_csv("combined_vern_same_ind_clean1.csv",

4826                        col_types = cols(Days_to_flower = col_number(),

4827                            Experiment = col_character()))

4828    View(combined_vern_same_ind_clean1)

4829

4830    ggplot(combined_vern_same_ind_clean1, aes(x=Experiment, y=Days_to_flower, fill=Experiment)) +

4831    geom_boxplot() + labs(title = "Difference in Days to Flower Between Vernalization Years", x = "Vernalization

4832    Year", y = "Number of Days to Flower")

4833

4834    #multi-gene:

4835    library(ggplot2)

4836    ggplot(rge_input_luco1, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene

4837    Expression - LuCO1", x = "Sample", y = "Relative Gene Expression")+  stat_compare_means(method = "anova")

4838    ggplot(rge_input_luco2, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene

4839    Expression - LuCO2", x = "Sample", y = "Relative Gene Expression")

4840    ggplot(rge_input_luft1, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene

4841    Expression - LuFT1", x = "Sample", y = "Relative Gene Expression")

4842    ggplot(rge_input_luft2, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene

4843    Expression - LuFT2", x = "Sample", y = "Relative Gene Expression")

4844    ggplot(rge_input_lugi11, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative Gene

4845    Expression - LuGI1.1", x = "Sample", y = "Relative Gene Expression")

4846

4847

4848    #delta-delta:

4849    setwd("G:/Linum Project/rdir")

4850    #luco1:

4851    library(readr)
```

```
4852    rge_input_luco1_deltadelta <- read_csv("rge_input_luco1_deltadelta_withoutoutlyer.csv",

4853                        col_types = cols(RGE = col_number()))

4854    View(rge_input_luco1_deltadelta)

4855    library(ggplot2)

4856    ggplot(rge_input_luco1_deltadelta, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative
4857    Gene Expression - LuCO1", x = "Sample", y = "Relative Gene Expression")+ stat_compare_means(label =
4858    "p.signif", method = "t.test",ref.group = ".all.")

4859    #luco2:

4860    library(readr)

4861    rge_input_luco2_deltadelta <- read_csv("rge_input_luco2_deltadelta_withoutoutlyer.csv",

4862                        col_types = cols(RGE = col_number()))

4863    View(rge_input_luco2_deltadelta)

4864    library(ggplot2)

4865    ggplot(rge_input_luco2_deltadelta, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative
4866    Gene Expression - LuCO2", x = "Sample", y = "Relative Gene Expression")+stat_compare_means(label =
4867    "p.signif", method = "t.test",ref.group = ".all.")

4868    #lugi1.1:

4869    library(readr)

4870    rge_input_lugi1_1_deltadelta <- read_csv("rge_input_lugi1_1_deltadelta_withoutoutlyer.csv",

4871                        col_types = cols(RGE = col_number()))

4872    View(rge_input_lugi1_1_deltadelta)

4873    library(ggplot2)

4874    ggplot(rge_input_lugi1_1_deltadelta, aes(x=group, y=RGE, fill=group)) + geom_boxplot() + labs(title = "Relative
4875    Gene Expression - LuGI1.1", x = "Sample", y = "Relative Gene Expression")+stat_compare_means(label =
4876    "p.signif", method = "t.test",ref.group = ".all.")

4877    #Luft1:

4878    #without Saf_10:

4879    library(readr)

4880    rge_input_luft1_deltadelta_withoutoutlyer <- read_csv("rge_input_luft1_deltadelta_withoutoutlyer.csv",

4881                        col_types = cols(RGE = col_number()))
```

```
4882    View(rge_input_luft1_deltadelta_withoutoutlyer)

4883

4884    library(ggplot2)

4885    ggplot(rge_input_luft1_deltadelta_withoutoutlyer, aes(x=group, y=RGE, fill=group)) + geom_boxplot() +

4886    labs(title = "Relative Gene Expression - Luft1", x = "Sample", y = "Relative Gene

4887    Expression")+stat_compare_means(label = "p.signif", method = "t.test",ref.group = ".all.")

4888

4889    #luft2:

4890    library(readr)

4891    rge_input_luft2_deltadelta <- read_csv("rge_input_luft2_deltadelta_withoutoutlier.csv",

4892                    col_types = cols(RGE = col_number()))

4893    View(rge_input_luft2_deltadelta)

4894    #without saf_10:

4895    library(readr)

4896    rge_input_luft2_deltadelta_withoutoutlier <- read_csv("rge_input_luft2_deltadelta_withoutoutlier.csv",

4897                    col_types = cols(RGE = col_number()))

4898    View(rge_input_luft2_deltadelta_withoutoutlier)

4899    library(ggplot2)

4900    ggplot(rge_input_luft2_deltadelta_withoutoutlier, aes(x=group, y=RGE, fill=group)) + geom_boxplot() +

4901    labs(title = "Relative Gene Expression - Luft2", x = "Sample", y = "Relative Gene

4902    Expression")+stat_compare_means(label = "p.signif", method = "t.test",ref.group = ".all.")

4903

4904

4905    #luco1:

4906    rge_input_luco1$group <- factor(rge_input_luco1$group, levels = c("control_cul", "vern_cul", "control_wil",

4907    "vern_wil"))

4908    meanRGEluco1 <- tapply(rge_input_luco1$RGE, rge_input_luco1$group, mean)

4909    sdevRGEluco1 <- tapply(rge_input_luco1$RGE, rge_input_luco1$group, sd)

4910    numberRGEluco1 <- tapply(rge_input_luco1$RGE, rge_input_luco1$group, length)

4911    data.frame(mean=meanRGEluco1, std.dev=sdevRGEluco1, n=numberRGEluco1)
```

```
4912

4913   #anova:

4914   rge_luco1_anova <- lm(RGE ~ group, data = rge_input_luco1)

4915   anova(rge_luco1_anova)

4916

4917   #luco1 cultivar comparison:

4918   library(readr)

4919   rge_input_luco1_cul <- read_csv("rge_input_luco1_cul.csv",

4920                    +    col_types = cols(RGE = col_number()))

4921

4922   #t-test:

4923   t.test(RGE~group, data = rge_input_luco1_cul)

4924

4925   #luco1 wild comparison:

4926   library(readr)

4927   rge_input_luco1_wil <- read_csv("rge_input_luco1_wil.csv",

4928                    col_types = cols(RGE = col_number()))

4929   View(rge_input_luco1_wil)

4930

4931   #t-test:

4932   t.test(RGE~group, data = rge_input_luco1_wil)

4933

4934

4935   #luco2:

4936   rge_input_luco2$group <- factor(rge_input_luco2$group, levels = c("control_cul", "vern_cul", "control_wil",
4937   "vern_wil"))

4938   meanRGEluco2 <- tapply(rge_input_luco2$RGE, rge_input_luco2$group, mean)

4939   sdevRGEluco2 <- tapply(rge_input_luco2$RGE, rge_input_luco2$group, sd)

4940   numberRGEluco2 <- tapply(rge_input_luco2$RGE, rge_input_luco2$group, length)
```

```
4941    data.frame(mean=meanRGEluco2, std.dev=sdevRGEluco2, n=numberRGEluco2)

4942

4943    #anova:

4944    rge_luco2_anova <- lm(RGE ~ group, data = rge_input_luco2)

4945    anova(rge_luco2_anova)

4946

4947    #luco2 cultivar comparison:

4948    library(readr)

4949    rge_input_luco2_cul <- read_csv("rge_input_luco2_cul.csv",

4950                        col_types = cols(RGE = col_number()))

4951    View(rge_input_luco2_cul)

4952

4953    #t-test:

4954    rge_input_luco2_cul$group <- factor(rge_input_luco2_cul$group, levels = c("control_cul", "vern_cul"))

4955    rge_input_luco2_cul_ttest <- lm(RGE ~ group, data = rge_input_luco2_cul)

4956    t.test(RGE~group, data = rge_input_luco2_cul)

4957

4958    #luco2 wild comparison:

4959    library(readr)

4960    rge_input_luco2_wil <- read_csv("rge_input_luco2_wil.csv",

4961                        col_types = cols(`1.053672824` = col_number()))

4962    View(rge_input_luco2_wil)

4963

4964    t.test(RGE~group, data = rge_input_luco2_wil)

4965

4966    #luft2:

4967    rge_input_luft2$group <- factor(rge_input_luft2$group, levels = c("control_cul", "vern_cul", "control_wil",

4968    "vern_wil"))

4969    meanRGEluft2 <- tapply(rge_input_luft2$RGE, rge_input_luft2$group, mean)
```

```
4970    sdevRGEluft2 <- tapply(rge_input_luft2$RGE, rge_input_luft2$group, sd)

4971    numberRGEluft2 <- tapply(rge_input_luft2$RGE, rge_input_luft2$group, length)

4972    data.frame(mean=meanRGEluft2, std.dev=sdevRGEluft2, n=numberRGEluft2)

4973

4974    #anova

4975    rge_luft2_anova <- lm(RGE ~ group, data = rge_input_luft2)

4976    anova(rge_luft2_anova)

4977

4978    #luft2 cultivar comparison:

4979    library(readr)

4980    rge_input_luft2_cul <- read_csv("rge_input_luft2_cul.csv",

4981                        col_types = cols(RGE = col_number()))

4982    View(rge_input_luft2_cul)

4983

4984    #t-test:

4985    rge_input_luft2_cul$group <- factor(rge_input_luft2_cul$group, levels = c("control_cul", "vern_cul"))

4986    rge_input_luft2_cul_ttest <- lm(RGE ~ group, data = rge_input_luft2_cul)

4987    t.test(RGE~group, data = rge_input_luft2_cul)

4988

4989    #luft2 wild comparison:

4990    library(readr)

4991    rge_input_luft2_wil <- read_csv("rge_input_luft2_wil.csv",

4992                        +    col_types = cols(Ind = col_character(),

4993                                +    RGE = col_number(), group = col_character()))

4994    View(rge_input_luft2_wil)

4995

4996

4997    #luft1:
```

```
4998    rge_input_luft1$group <- factor(rge_input_luft1$group, levels = c("control_cul", "vern_cul", "control_wil",

4999    "vern_wil"))

5000    meanRGEluft1 <- tapply(rge_input_luft1$RGE, rge_input_luft1$group, mean)

5001    sdevRGEluft1 <- tapply(rge_input_luft1$RGE, rge_input_luft1$group, sd)

5002    numberRGEluft1 <- tapply(rge_input_luft1$RGE, rge_input_luft1$group, length)

5003    data.frame(mean=meanRGEluft1, std.dev=sdevRGEluft1, n=numberRGEluft1)

5004

5005    #anova

5006    rge_luft1_anova <- lm(RGE ~ group, data = rge_input_luft1)

5007    anova(rge_luft1_anova)

5008

5009

5010    #t-test:

5011    rge_input_luft2_wil$group <- factor(rge_input_luft2_wil$group, levels = c("control_cul", "vern_cul"))

5012    rge_input_luft2_wil_ttest <- lm(RGE ~ group, data = rge_input_luft2_wil)

5013    t.test(RGE~group, data = rge_input_luft2_wil)

5014

5015    #luft1 cultivar comparison:

5016    library(readr)

5017    rge_input_luft1_cul <- read_csv("rge_input_luft1_cul.csv",

5018                    col_types = cols(...1 = col_character(),

5019                            RGE = col_number(), group = col_character()))

5020    View(rge_input_luft1_cul)

5021

5022    #t-test:

5023    t.test(RGE~group, data = rge_input_luft1_cul)

5024

5025    #luft1 wild comparison:

5026    library(readr)
```

```
5027    rge_input_luft1_wil <- read_csv("rge_input_luft1_wil.csv",
5028                      col_types = cols(RGE = col_number()))
5029    View(rge_input_luft1_wil)
5030
5031    #t-test:
5032    t.test(RGE~group, data = rge_input_luft1_wil)
5033
5034    #lugi11:
5035    rge_input_lugi11$group <- factor(rge_input_lugi11$group, levels = c("control_cul", "vern_cul", "control_wil",
5036    "vern_wil"))
5037    meanRGElugi11 <- tapply(rge_input_lugi11$RGE, rge_input_lugi11$group, mean)
5038    sdevRGElugi11 <- tapply(rge_input_lugi11$RGE, rge_input_lugi11$group, sd)
5039    numberRGElugi11 <- tapply(rge_input_lugi11$RGE, rge_input_lugi11$group, length)
5040    data.frame(mean=meanRGElugi11, std.dev=sdevRGElugi11, n=numberRGElugi11)
5041
5042    #anova
5043    rge_lugi11_anova <- lm(RGE ~ group, data = rge_input_lugi11)
5044    anova(rge_lugi11_anova)
5045
5046    #lugi11 cultivar comparison:
5047    library(readr)
5048    rge_input_lugi11_cul <- read_csv("rge_input_lugi11_cul.csv",
5049                      col_types = cols(RGE = col_number()))
5050    View(rge_input_lugi11_cul)
5051
5052    #t-test:
5053    t.test(RGE~group, data = rge_input_lugi11_cul)
5054
5055    #lugi11 wild comparison:
```

```
5056    library(readr)

5057    rge_input_lugi11_wil <- read_csv("rge_input_lugi11_wil.csv",

5058            col_types = cols(RGE = col_number()))

5059    View(rge_input_lugi11_wil)

5060

5061    #t-test:

5062    t.test(RGE~group, data = rge_input_lugi11_wil)

5063
```

5064   **Appendix 7: Relative Gene Expression GLM R Commands:**

```
5065    library(readr)

5066    combined_vern_same_ind <- read_csv("combined_vern_same_ind.csv",

5067            col_types = cols(Lat = col_number(),

5068                    Alt = col_number(), Height = col_number(),

5069                    Stem_no = col_number(), Bud_no = col_number(),

5070                    Days_to_fl = col_number(), pc_1 = col_number(),

5071                    pc_2 = col_number(), pc_3 = col_number()))

5072    View(combined_vern_same_ind)

5073

5074    #plot:

5075    x <- combined_vern_same_ind$Height

5076    y <- combined_vern_same_ind$Days_to_fl

5077    plot(x, y, main = "Height against Days to Flowering",

5078        xlab = "Plant Height (Cm)", ylab = "Days to Flowering",

5079        pch = 19, frame = FALSE)

5080

5081    #scatterplot

5082    install.packages("car")

5083    library("car")

5084    #height:
```

```
5085    View(combined_vern_same_ind)

5086    scatterplot(Days_to_fl ~ Height | Experiment, data = combined_vern_same_ind,

5087         smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Plant Height (Cm)", ylab = "Days to

5088    Flowering")

5089    abline(lm(Height ~ Days_to_fl data = combined_vern_same_ind),col="red")

5090

5091    ggscatter(combined_vern_same_ind, x = "Height", y = "Days_to_fl", size = 2.0,

5092         rug = TRUE,                    # Add marginal rug

5093         color = "Experiment", palette = "jco", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +

5094      stat_cor(aes(color = Experiment), method = "pearson")

5095

5096    #stem no:

5097    scatterplot(Days_to_fl ~ Stem_no | Experiment, data = combined_vern_same_ind,

5098         smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Stem number", ylab = "Days to

5099    Flowering")

5100    abline(lm(Stem_no ~ Days_to_fl data = combined_vern_same_ind),col="red")

5101

5102

5103    ggscatter(combined_vern_same_ind, x = "Stem_no", y = "Days_to_fl", size = 2.0,

5104         rug = TRUE,                    # Add marginal rug

5105         color = "Experiment", palette = "jco", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +

5106      stat_cor(aes(color = Experiment), method = "pearson")

5107

5108    #bud no:

5109    scatterplot(Days_to_fl ~ Bud_no | Experiment, data = combined_vern_same_ind,

5110         smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Bud number", ylab = "Days to

5111    Flowering")

5112    abline(lm(Stem_no ~ Days_to_fl data = combined_vern_same_ind),col="red")

5113
```

```r
5114

5115    ggscatter(combined_vern_same_ind, x = "Bud_no", y = "Days_to_fl", size = 2.0,

5116         rug = TRUE,                    # Add marginal rug

5117         color = "Experiment", palette = "jco", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +

5118      stat_cor(aes(color = Experiment), method = "pearson")

5119

5120    #pc1:

5121    scatterplot(pc_1 ~ Days_to_fl | Experiment, data = combined_vern_same_ind,

5122         smooth = FALSE, grid = FALSE, frame = FALSE, xlab = "Days to First Flower", ylab = "Climate (PC1)")

5123    abline(lm(pc_1 ~ Days_to_fl data = combined_vern_same_ind),col="red")

5124

5125    plot(pc_1 ~ Days_to_fl, data = combined_vern_same_ind)

5126    abline(lm(pc_1 ~ Days_to_fl, data = combined_vern_same_ind),col="red")

5127    summary(abline)

5128

5129    #Alt:

5130    scatterplot(Days_to_fl ~ Alt | Experiment, data = combined_vern_same_ind,

5131         smooth = FALSE, grid = FALSE, frame = FALSE, xlab = "Days to First Flower", ylab = "Altitude (M)")

5132    abline(lm(Alt ~ Days_to_fl data = combined_vern_same_ind),col="red")

5133

5134    plot(Alt ~ Days_to_fl, data = combined_vern_same_ind)

5135    abline(lm(Alt ~ Days_to_fl, data = combined_vern_same_ind),col="red")

5136    summary(abline)

5137

5138

5139    #Model summary

5140    summary1 <- summary(glm(Stem_no ~ Days_to_fl, data = combined_vern_same_ind))

5141    adjRsq <- summary1$adj.r.squared

5142    fStat <- summary1$statistic
```

```
5143    pValue <- pf(fStat[summary1])

5144    summary(summary1)

5145    #PC1 - Days to fl LM:

5146    pc1daysmod<-cbind(combined_vern_same_ind$Days_to_fl, combined_vern_same_ind$pc_1)

5147    pc1daysmod2<-pc1daysmodlm<-lm(pc1daysmod~Experiment,data=combined_vern_same_ind)

5148    summary(pc1daysmodlm)

5149

5150

5151    #Alt - Days to fl GLM:

5152    altdaysmod<-cbind(combined_vern_same_ind$Days_to_fl, combined_vern_same_ind$Alt)

5153    altdaysmod2<-altdaysmodlm<-lm(altdaysmod~Experiment,data=combined_vern_same_ind)

5154    summary(altdaysmodlm)

5155

5156    #Lat - Days to fl GLM:

5157    latdaysmod<-cbind(combined_vern_same_ind$Days_to_fl, combined_vern_same_ind$Lat)

5158    latdaysmod2<-latdaysmodlm<-lm(altdaysmod~Experiment,data=combined_vern_same_ind)

5159    summary(latdaysmodlm)

5160

5161    #RGE - Days to fl GLM:

5162    library(readr)

5163    combined_loci_wild_deltadelta_ft2clear <- read_csv("combined_loci_wild_deltadelta_ft2clear.csv",

5164                            col_types = cols(`RGE_c` = col_number(),

5165                                    `RGE_v` = col_number(), `RGE_Difference` = col_number(),

5166                                    Days_to_fl = col_number(), Lat = col_number(),

5167                                    Alt = col_number(), pc = col_number(),

5168                                    pc2 = col_number(), pc3 = col_number()))

5169    View(combined_loci_wild_deltadelta_ft2clear)

5170

5171    #Cultivars:
```

```r
5172    library(readr)

5173    combined_loci_cult_deltadelta_fl <- read_csv("combined_loci_cult_deltadelta_fl.csv",

5174                            col_types = cols(`RGE_c` = col_number(),

5175                                `RGE_v` = col_number(), `RGE_Difference` = col_number(),

5176                                Lat = col_number(), Days_to_fl = col_number(),

5177                                `Height_(Cm)` = col_number()))

5178    View(combined_loci_cult_deltadelta_fl)

5179

5180

5181    #Gene expression influence in flowering time:

5182    #Wild

5183    #subset data to loci first:

5184    Luco1_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuCO1")

5185    Luco2_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuCO2")

5186    Luft1_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuFT1")

5187    Luft2_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuFT2")

5188    Lugi11_data <- subset(combined_loci_wild_deltadelta_ft2clear, Loci == "LuGI1.1")

5189

5190    #GLM:

5191    modluco1wild<-glm(Luco1_data$Days_to_fl ~ Luco1_data$"RGE_Difference")

5192    modluco2wild<-glm(Luco2_data$Days_to_fl ~ Luco2_data$"RGE_Difference")

5193    modluft1wild<-glm(Luft1_data$Days_to_fl ~ Luft1_data$"RGE_Difference")

5194    modluft2wild<-glm(Luft2_data$Days_to_fl ~ Luft2_data$"RGE_Difference")

5195    modlugi11wild<-glm(Lugi11_data$Days_to_fl ~ Lugi11_data$"RGE_Difference")

5196

5197    summary(modluco1wild)

5198    summary(modluco2wild)

5199    summary(modluft1wild)

5200    summary(modluft2wild)
```

```
5201   summary(modlugi11wild)

5202

5203   #Cultivars

5204   Luco1_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuCO1")

5205   Luco2_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuCO2")

5206   Luft1_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuFT1")

5207   Luft2_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuFT2")

5208   Lugi11_cult_data <- subset(combined_loci_cult_deltadelta_fl, Loci == "LuGI1.1")

5209

5210   #GLM:

5211   modluco1cul<-glm(Luco1_cult_data$Days_to_fl ~ Luco1_cult_data$"RGE_Difference")

5212   modluco2cul<-glm(Luco2_cult_data$Days_to_fl ~ Luco2_cult_data$"RGE_Difference")

5213   modluft1cul<-glm(Luft1_cult_data$Days_to_fl ~ Luft1_cult_data$"RGE_Difference")

5214   modluft2cul<-glm(Luft2_cult_data$Days_to_fl ~ Luft2_cult_data$"RGE_Difference")

5215   modlugi11cul<-glm(Lugi11_cult_data$Days_to_fl ~ Lugi11_cult_data$"RGE_Difference")

5216

5217   summary(modluco1cul)

5218   summary(modluco2cul)

5219   summary(modluft1cul)

5220   summary(modluft2cul)

5221   summary(modlugi11cul)

5222

5223   #charts:

5224   View(Luco1_data)

5225

5226   write.csv(Luco1_data,"G:/Linum Project/rdir/Luco1_data_only.csv", row.names = FALSE)

5227

5228   #

5229   #Delta-delta data:
```

```
5230    #Load packages

5231    library(ggplot2)

5232    library(ggpubr)

5233    library(viridis)

5234    library(readr)

5235    #Cultivars

5236    setwd("G:/Linum Project/rdir")

5237

5238

5239    ggscatter(combined_loci_cult_deltadelta_fl, x = "Days_to_fl", y = "RGE_Difference", size = 1.0,

5240        rug = TRUE,                    # Add marginal rug

5241        color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +

5242     stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.95, method =

5243    "pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+

5244    theme_light()+ggtitle("Cultivar RGE-Flowering")

5245

5246    #Lat:

5247    ggscatter(combined_loci_cult_deltadelta_fl, x = "Lat", y = "RGE_Difference", size = 1.0,

5248        rug = TRUE,                    # Add marginal rug

5249        color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +

5250     stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.95, method =

5251    "pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+

5252    theme_light()+ggtitle("Cultivar RGE-Latitude")

5253

5254    #GLM:

5255    modluco1cullat<-glm(Luco1_cult_data$Lat ~ Luco1_cult_data$"RGE_Difference")

5256    modluco2cullat<-glm(Luco2_cult_data$Lat ~ Luco2_cult_data$"RGE_Difference")

5257    modluft1cullat<-glm(Luft1_cult_data$Lat ~ Luft1_cult_data$"RGE_Difference")

5258    modluft2cullat<-glm(Luft2_cult_data$Lat ~ Luft2_cult_data$"RGE_Difference")

5259    modlugi11cullat<-glm(Lugi11_cult_data$Lat ~ Lugi11_cult_data$"RGE_Difference")
```

```
5260
5261    summary(modluco1cullat)
5262    summary(modluco2cullat)
5263    summary(modluft1cullat)
5264    summary(modluft2cullat)
5265    summary(modlugi11cullat)
5266
5267    #Wild:
5268    ggscatter(combined_loci_wild_deltadelta_ft2clear, x = "Days_to_fl", y = "RGE_Difference", size = 1.0,
5269         rug = TRUE,                    # Add marginal rug
5270         color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
5271      stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.32, method =
5272    "pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
5273    theme_light()+ggtitle("Wild RGE-Flowering")
5274
5275    #Lat
5276    ggscatter(combined_loci_wild_deltadelta_ft2clear, x = "Lat", y = "RGE_Difference", size = 1.0,
5277         rug = TRUE,                    # Add marginal rug
5278         color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +
5279      stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc="centre", method =
5280    "pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
5281    theme_light()+ggtitle("Wild RGE-Latitude")
5282
5283    #GLM:
5284    modluco1willat<-glm(Luco1_data$Lat ~ Luco1_data$"RGE_Difference")
5285    modluco2willat<-glm(Luco2_data$Lat ~ Luco2_data$"RGE_Difference")
5286    modluft1willat<-glm(Luft1_data$Lat ~ Luft1_data$"RGE_Difference")
5287    modluft2willat<-glm(Luft2_data$Lat ~ Luft2_data$"RGE_Difference")
5288    modlugi11willat<-glm(Lugi11_data$Lat ~ Lugi11_data$"RGE_Difference")
5289
```

```
5290    summary(modluco1willat)

5291    summary(modluco2willat)

5292    summary(modluft1willat)

5293    summary(modluft2willat)

5294    summary(modlugi11willat)

5295

5296

5297    #PC1:

5298    ggscatter(combined_loci_wild_deltadelta_ft2clear, x = "pc", y = "RGE_Difference", size = 1.0,

5299        rug = TRUE,                    # Add marginal rug

5300        color = "Loci", palette = "uchicago", facet.by = "Loci",add = "reg.line", conf.int = TRUE) +

5301     stat_cor(aes(color = Loci), label.sep = ";", label.x.npc = "left", label.y.npc=0.38, method =
5302    "pearson")+geom_point(aes(color = Loci))+stat_smooth(method="glm",se=FALSE)+
5303    theme_light()+ggtitle("Wild RGE-PC1")

5304

5305    #GLM:

5306    modluco1wilpc<-glm(Luco1_data$pc ~ Luco1_data$"RGE_Difference")

5307    modluco2wilpc<-glm(Luco2_data$pc ~ Luco2_data$"RGE_Difference")

5308    modluft1wilpc<-glm(Luft1_data$pc ~ Luft1_data$"RGE_Difference")

5309    modluft2wilpc<-glm(Luft2_data$pc ~ Luft2_data$"RGE_Difference")

5310    modlugi11wilpc<-glm(Lugi11_data$pc ~ Lugi11_data$"RGE_Difference")

5311

5312    summary(modluco1wilpc)

5313    summary(modluco2wilpc)

5314    summary(modluft1wilpc)

5315    summary(modluft2wilpc)

5316    summary(modlugi11wilpc)

5317

5318    #scatterplot
```

```
5319    install.packages("tidyverse")

5320    library(tidyverse)

5321    library(ggplot2)

5322

5323    ggplot(combined_loci_cult_deltadelta_fl, aes(x=RGE_Difference, y=Days_to_fl, shape=Loci, color=Loci)) +

5324     geom_point(size=2.5, shape=18) + ggtitle("Cultivar Flowering initiation")+labs(y= "Days to Flowering", x =

5325    "Relative Gene Expression Difference")

5326

5327    #wild

5328    ggplot(combined_loci_wild_deltadelta_ft2clear, aes(x=RGE_Difference, y=Days_to_fl, shape=Loci, color=Loci))

5329    +

5330     geom_point(size=2.5, shape=18) + ggtitle("Wild Flowering initiation")+labs(y= "Days to Flowering", x =

5331    "Relative Gene Expression Difference")

5332

5333    #

5334    s3d <- scatterplot3d(combined_vern_same_ind, pch = 16, color=colors)

5335    legend(s3d$xyz.convert(7.5, 3, 4.5), legend = levels(combined_vern_same_ind$Experiment),

5336        col =  c("#999999", "#E69F00", "#56B4E9"), pch = 16)

5337
```

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

5339   Relative gene expression comparison for the different genes tested in relation to each

5340   species and treatment:

5341

5342

5343

5344

5345

5346

5347

5348

5349

5350

5351

5352

5353

5354

5355

5356

5357

5358
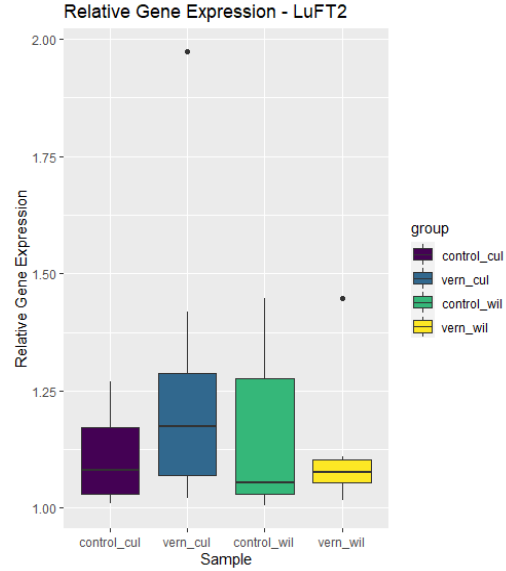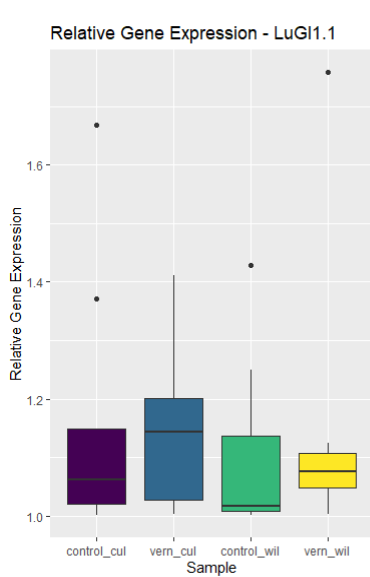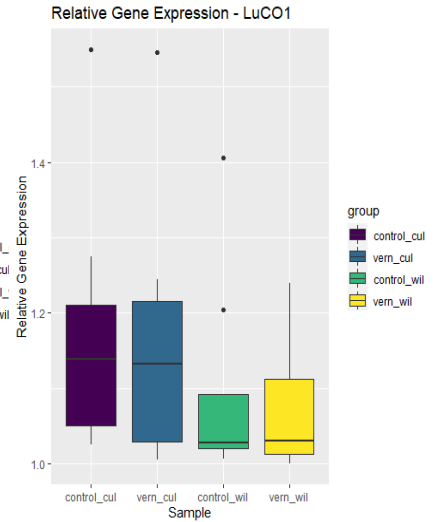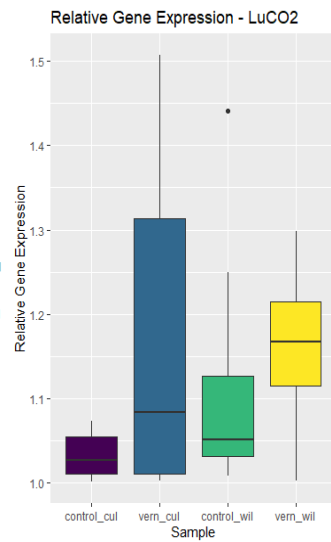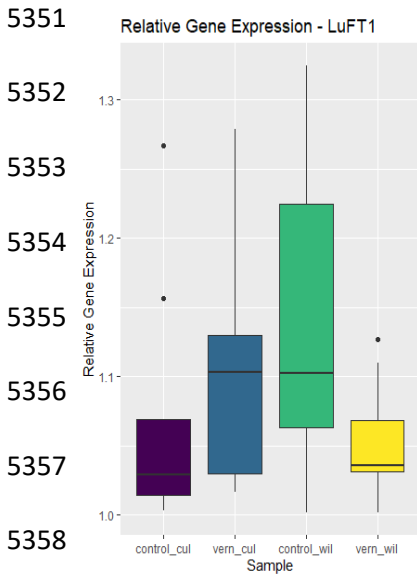
5359

5360

5361

5362   **Appendix 9: Scatterplot for gene expression and latitude correlation under two**

5363   **Housekeeping genes**

5364   Scatterplot showing relationships between latitude and relative gene expression (RGE) differences

5365   for five flowering time genes. A = wild types, B = Cultivars. RGE differences were calculated using 2

5366   HKGs:

5367
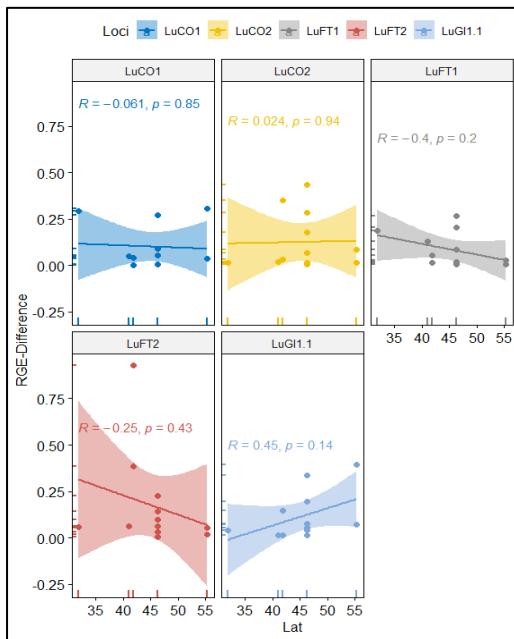
5368   A   B
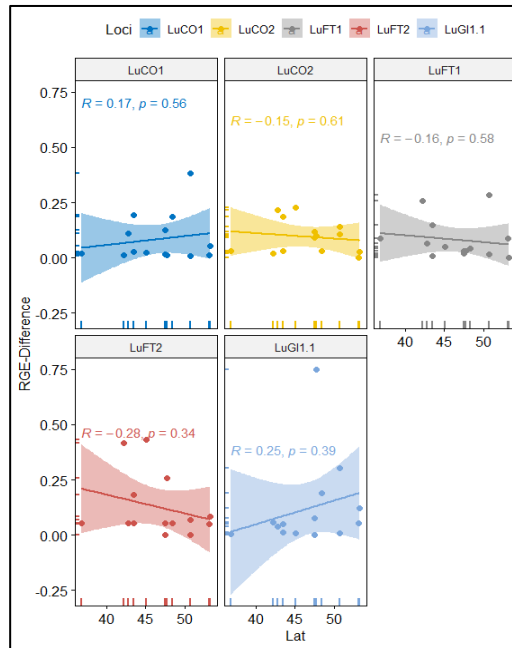
5378

5379

5380

5381

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

5386

5387

5388

5389

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

5393

5394

5395 **Appendix 11: Fluorescent microscopy observation of a *Linum* flower conserved at 4ºC in 70%EtOH**
5396 **solution, showing pollen and tube observations after 72 hours preservation.**

5397
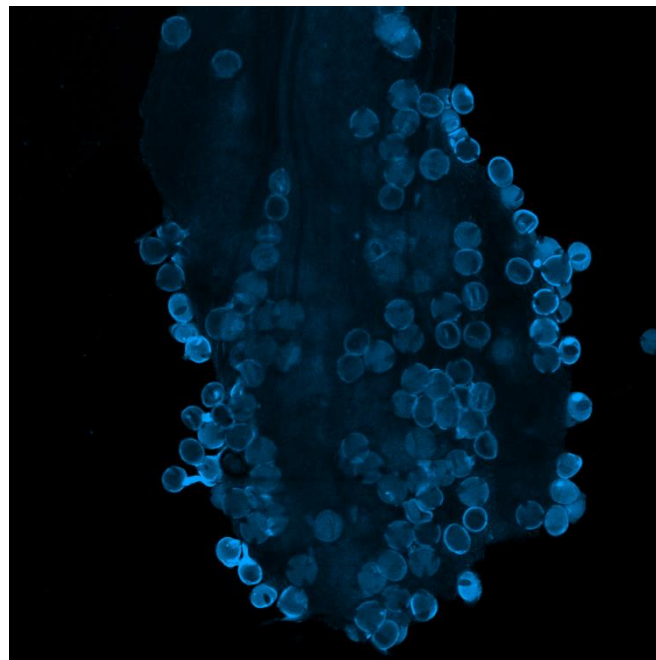
5398

5399

5400

5401

5402

5403

5404

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.

A

B
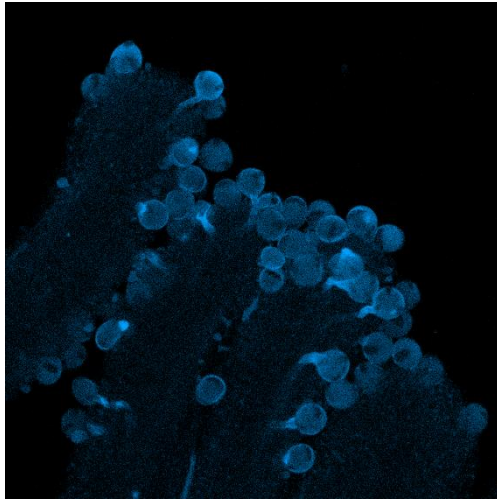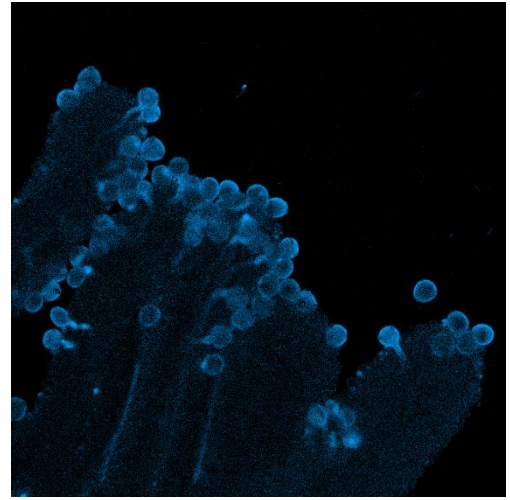
Appendix 13: Pollen Tube Observation Using 0.5% (w/v) Aniline blue in Potassium Phosphate (KH2PO4)
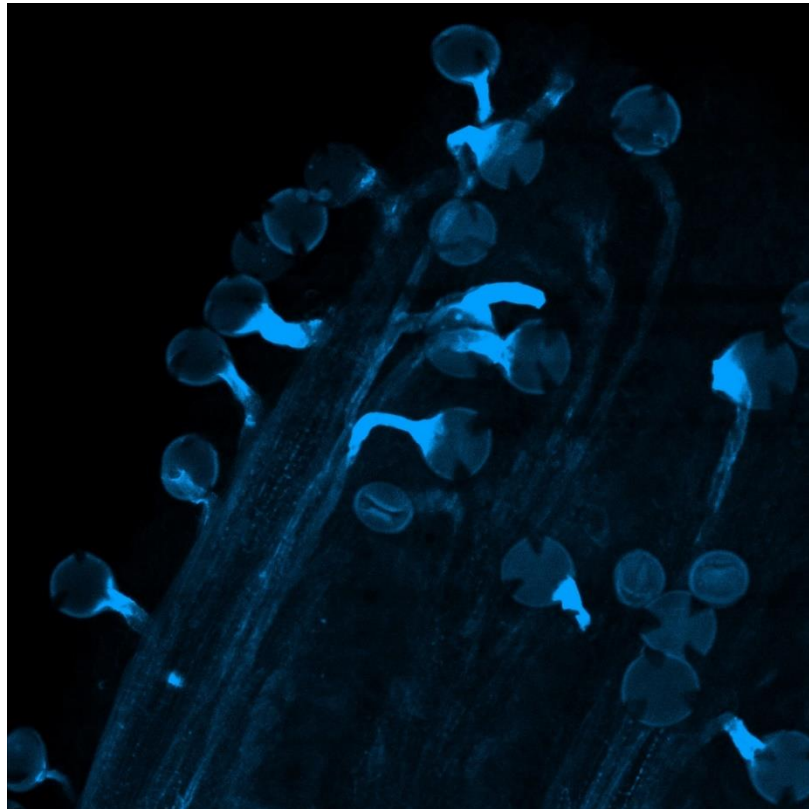
Appendix 14: Pollen Modelling Commands

5422

5423   #test for normality (pollen tube)

5424   plot(density(Pollen$Percent))

5425   hist(Pollen$Percent, main="Histogram for Percentage of Pollen Tube %", xlab="% of Pollen Tube Counts")

5426   densityPollen<-density(Pollen$Percent)

5427   lines(densityPollen$x,densityPollen$y*5000)

5428   shapiro.test(Pollen$Percent)

5429   #data is not normally distributed

5430   #adjust for non-normally distributed data (tukey's)

5431

5432   #Data for 20C - Cold (pollen tube)

5433   setwd("G:/Linum Project/Pollen germination")

5434   library(readxl)

5435   Pollencold <- as.data.frame(read_excel("Pollen_and_tube_count2.xlsx", sheet = "20c", na = "NA"))

5436   str(Pollencold)

5437   View(Pollencold)

5438   Pollencold$Treatment<-as.factor(Pollencold$Treatment)

5439

5440   #test for normality - Cold

5441   plot(density(Pollencold$Percent))

5442   hist(Pollencold$Percent, main="Histogram for Percentage of Pollen Tube %", xlab="% of Pollen Tube Counts")

5443   densityPollencold<-density(Pollencold$Percent)

5444   lines(densityPollencold$x,densityPollencold$y*5000)

5445   shapiro.test(Pollencold$Percent)

5446   #data is not normally distributed

5447   #adjust for non-normally distributed data (tukey's)

5448

5449   #Pollen number:

5450   plot(density(Pollen$Count))

```
5451    hist(Pollen$Count, main = "Pollen Count Distribution", xlab = "Pollen Count")

5452    densityCount<-density(Pollen$Count)

5453    lines(densityCount$x,densityCount$y*5000)

5454    shapiro.test(Pollen$Count)

5455

5456    #test pollen count versus treatment, quasibinomial model - TOTAL

5457    count<-Pollen$Count

5458    modcount<-lm(Pollen$Count~Treatment, data=Pollen)

5459    summary(modcount)

5460    tukcount<-glht(modcount,mcp(Treatment="Tukey"))

5461    summary(tukcount)

5462    Props<-Pollen$Tube/Pollen$Count

5463    Pollen<-data.frame(Pollen,count)

5464    str(Pollen)

5465    barcentres<-
5466    barplot(tapply(Pollen$Count,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Count,Pollen$Treatment,m
5467    ean)+5)), ylab = "Pollen Count", xlab = "Treatment", main = "Number of Pollen vs Treatment")

5468    means<-tapply(Pollen$Count,Pollen$Treatment,mean)

5469    ses<-tapply(Pollen$Count,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Count,Pollen$Treatment,length))

5470    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5471

5472    #test pollen count versus treatment, Poisson correction - TOTAL

5473    count<-Pollen$Count

5474    modcountpoisson<-glm(Pollen$Count~Treatment, family=poisson(), data=Pollen,)

5475    summary(modcountpoisson)

5476    tukcount<-glht(modcountpoisson,mcp(Treatment="Tukey"))

5477    summary(tukcount)

5478    Props<-Pollen$Tube/Pollen$Count

5479    Pollen<-data.frame(Pollen,count)
```

```r
5480    str(Pollen)

5481    barcentres<-
5482    barplot(tapply(Pollen$Count,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Count,Pollen$Treatment,m
5483    ean)+5)), ylab = "Pollen Count", xlab = "Treatment", main = "Number of Pollen vs Treatment")

5484    means<-tapply(Pollen$Count,Pollen$Treatment,mean)

5485    ses<-tapply(Pollen$Count,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Count,Pollen$Treatment,length))

5486    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5487

5488    #test pollen tube count versus treatment, Poisson correction - TOTAL

5489    tubecount<-Pollen$Tube

5490    modtubepoisson<-glm(Pollen$Tube~Treatment, family=poisson(), data=Pollen,)

5491    summary(modtubepoisson)

5492    tuktube<-glht(modtubepoisson,mcp(Treatment="Tukey"))

5493    summary(tuktube)

5494    Props<-Pollen$Tube/Pollen$Count

5495    Pollentube<-data.frame(Pollen,tubecount)

5496    str(Pollentube)

5497    barcentres<-
5498    barplot(tapply(Pollen$Tube,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Tube,Pollen$Treatment,mea
5499    n)+2)), ylab = "Pollen Tube Count", xlab = "Treatment", main = "Number of Pollen Tubes vs Treatment")

5500    means<-tapply(Pollen$Tube,Pollen$Treatment,mean)

5501    ses<-tapply(Pollen$Tube,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Tube,Pollen$Treatment,length))

5502    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5503

5504    #test reached versus treatment, binomial model

5505    mod1<-glm(Reached~Treatment,family=binomial(link='logit'),data=Pollen)

5506    summary(mod1)

5507    library(multcomp)

5508    tuk1<-glht(mod1,mcp(Treatment="Tukey"))

5509    summary(tuk1)
```

```
5510    barplot(tapply(Pollen$Reached,Pollen$Treatment,mean))

5511    barcentres<-

5512    barplot(tapply(Pollen$Reached,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Reached,Pollen$Treatm

5513    ent,mean)+0.1)), ylab = "Proportion of Pollen Reaching Ovary", xlab = "Treatment", main = "Proportion of

5514    Pollen Reaching Ovary vs Treatment")

5515    means<-tapply(Pollen$Reached,Pollen$Treatment,mean)

5516    ses<-tapply(Pollen$Reached,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Reached,Pollen$Treatment,length))

5517    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5518

5519    #test prop tube versus treatment, quasibinomial model

5520    Prop<-cbind(Pollen$Tube, Pollen$Count)

5521    mod2<-glm(Prop~Treatment, family=quasibinomial(link = 'logit'), data=Pollen)

5522    summary(mod2)

5523    tuk2<-glht(mod2,mcp(Treatment="Tukey"))

5524    summary(tuk2)

5525    Props<-Pollen$Tube/Pollen$Count

5526    Pollen<-data.frame(Pollen,Props)

5527    str(Pollen)

5528    barcentres<-

5529    barplot(tapply(Pollen$Props,Pollen$Treatment,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$Treatment,m

5530    ean)+0.1)), ylab = "Proportion of Pollen Tube Number", xlab = "Treatment", main = "Proportion of Pollen Tube

5531    vs Treatment")

5532    means<-tapply(Pollen$Props,Pollen$Treatment,mean)

5533    ses<-tapply(Pollen$Props,Pollen$Treatment,sd)/sqrt(tapply(Pollen$Props,Pollen$Treatment,length))

5534    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5535

5536    #test proptubes versus treatment x latitude

5537    modlat<-glm(Prop~Treatment*Lat, family=quasibinomial(link = 'logit'), data=Pollen)

5538    summary(modlat)

5539    anova(modlat,test = "F")
```

```
5540    interactlat<-interaction(Pollen$Treatment,Pollen$Lat)

5541    Pollen<-data.frame(Pollen,interactlat)

5542    View(Pollen)

5543    modlatx<-glm(Prop~-1+interactlat, family=quasibinomial(link = 'logit'), data=Pollen)

5544    summary(modlatx)

5545    tuklatx<-glht(modlatx,mcp(interactlat="Tukey"))

5546    summary(tuklatx)

5547    Props<-Pollen$Tube/Pollen$Count

5548    Pollen<-data.frame(Pollen,Props)

5549    barcentres<-
5550    barplot(tapply(Pollen$Props,Pollen$interactlat,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$interact,mean
5551    )+1.0)), ylab = "Proportion of Pollen Tube Number/Pollen Count", xlab = "Treatment and Latitude", main =
5552    "Proportion of Pollen Tube per Latitude and Treatment")

5553

5554    #test prop tubes/count versus treatment x species

5555    mod3<-glm(Props~Treatment*Species, family=quasibinomial(link = 'logit'), data=Pollen)

5556    summary(mod3)

5557    anova(mod3,test="F")

5558    interact<-interaction(Pollen$Treatment,Pollen$Species)

5559    Pollen<-data.frame(Pollen,interact)

5560    mod3x<-glm(Props~-1+interact, family=quasibinomial(link = 'logit'), data=Pollen)

5561    tuk3x<-glht(mod3x,mcp(interact="Tukey"))

5562    summary(tuk3x)

5563    Props<-Pollen$Tube/Pollen$Count

5564    Pollen<-data.frame(Pollen,Props)

5565    str(Pollen)

5566    barcentres<-
5567    barplot(tapply(Pollen$Props,Pollen$interact,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$interact,mean)+
5568    0.1)), ylab = "Proportion of Pollen Tube Number/Pollen Count", xlab = "Treatment and Species", main =
5569    "Proportion of Pollen Tube per species and treatment")
```

```
5570    means<-tapply(Pollen$Props,Pollen$interact,mean)

5571    ses<-tapply(Pollen$Props,Pollen$interact,sd)/sqrt(tapply(Pollen$Props,Pollen$interact,length))

5572    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5573

5574

5575

5576    #Analyse species separately for comparison:

5577    #bienne:

5578    library(readr)

5579    bienne_all <- read_csv("bienne_all.csv",

5580               col_types = cols(Count = col_number(),

5581                      Tube = col_number(), Percent = col_number(),

5582                      Reached = col_number(), Lat = col_number(),

5583                      pc1 = col_number()))

5584    View(bienne_all)

5585    bienne_all$Treatment<-as.factor(bienne_all$Treatment)

5586

5587    Propsbienne<-bienne_all$Tube/bienne_all$Count

5588    Propbienne<-cbind(bienne_all$Tube, bienne_all$Count)

5589    modpropbienne<-glm(Propbienne~Treatment, family=quasibinomial(link = 'logit'), data=bienne_all)

5590    summary(modpropbienne)

5591    tukpropbienne<-glht(modpropbienne,mcp(Treatment="Tukey"))

5592    summary(tukpropbienne)

5593    Propsb<-bienne_all$Tube/bienne_all$Count

5594    bienne_all<-data.frame(bienne_all,Propsb)

5595    str(bienne_all)

5596    barcentres<-

5597    barplot(tapply(bienne_all$Propsb,bienne_all$Treatment,mean),ylim=c(0,max(tapply(bienne_all$Propsb,bienn
```

```
5598   e_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Tube Number", xlab = "Treatment", main =

5599   "Proportion of Pollen Tube vs Treatment for L.bienne")

5600   means<-tapply(bienne_all$Propsb,bienne_all$Treatment,mean)

5601   ses<-

5602   tapply(bienne_all$Propsb,bienne_all$Treatment,sd)/sqrt(tapply(bienne_all$Propsb,bienne_all$Treatment,len

5603   gth))

5604   arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5605

5606   #usitatissimum:

5607   library(readr)

5608   usitatissimum_all <- read_csv("usitatissimum_all.csv",

5609                   col_types = cols(Count = col_number(),

5610                           Tube = col_number(), Percent = col_number(),

5611                           Reached = col_number(), Lat = col_number(),

5612                           pc1 = col_number()))

5613   View(usitatissimum_all)

5614   usitatissimum_all$Treatment<-as.factor(usitatissimum_all$Treatment)

5615

5616   Propscul<-usitatissimum_all$Tube/usitatissimum_all$Count

5617   Propcul<-cbind(usitatissimum_all$Tube, usitatissimum_all$Count)

5618   modpropcul<-glm(Propcul~Treatment, family=quasibinomial(link = 'logit'), data=usitatissimum_all)

5619   summary(modpropcul)

5620   tukpropcul<-glht(modpropcul,mcp(Treatment="Tukey"))

5621   summary(tukpropcul)

5622   Propsc<-usitatissimum_all$Tube/usitatissimum_all$Count

5623   usitatissimum_all<-data.frame(usitatissimum_all,Propsc)

5624   str(usitatissimum_all)

5625   barcentres<-

5626   barplot(tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,mean),ylim=c(0,max(tapply(usitatissim
```

```
5627   um_all$Propsc,usitatissimum_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Tube Number", xlab =

5628   "Treatment", main = "Proportion of Pollen Tube vs Treatment for L.usitatissimum")

5629   means<-tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,mean)

5630   ses<-

5631   tapply(usitatissimum_all$Propsc,usitatissimum_all$Treatment,sd)/sqrt(tapply(usitatissimum_all$Propsc,usitati

5632   ssimum_all$Treatment,length))

5633   arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5634

5635

5636   ###test reached versus treatment x species with mixed model, ind as random effect

5637   library(lmerTest)

5638   library("MuMIn")

5639   mod4<-glmer(Reached~Treatment*Species+(1|Ind), family=binomial(link='logit'), data=Pollen)

5640   summary(mod4)

5641   anova(mod4,test="F")

5642   r.squaredGLMM(mod4)

5643   mod4x<-glm(Reached~Treatment*Species, family=binomial(link='logit'), data=Pollen)

5644   anova(mod4, mod4x, test="LRT") #likelihood ratio test

5645   library(multcomp)

5646   tuk4<-glht(mod4,mcp(Treatment="Tukey"))

5647   summary(tuk4)

5648   interact<-interaction(Pollen$Treatment,Pollen$Species)

5649   Pollen<-data.frame(Pollen,interact)

5650   mod4y<-glmer(Reached~-1+interact+(1|Ind), family=binomial(link='logit'), data=Pollen)

5651   tuk4y<-glht(mod4y,mcp(interact="Tukey"))

5652   summary(tuk4y)

5653   barcentres<-

5654   barplot(tapply(Pollen$Reached,Pollen$interact,mean),ylim=c(0,max(tapply(Pollen$Reached,Pollen$interact,m

5655   ean)+0.1)),ylab = "Proportion of Pollen Tube Reached", xlab = "Treatment and Species", main = "Proportion of

5656   Pollen Tube Reaching Ovaries For Each Treatment and Species")
```

```
5657    means<-tapply(Pollen$Reached,Pollen$interact,mean)

5658    ses<-tapply(Pollen$Reached,Pollen$interact,sd)/sqrt(tapply(Pollen$Reached,Pollen$interact,length))

5659    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

5660    write.table(tuk4y)

5661

5662    #Pollen Reached for bienne:

5663    library(readr)

5664    bienne_all <- read_csv("bienne_all.csv",

5665              col_types = cols(Count = col_number(),

5666                      Tube = col_number(), Percent = col_number(),

5667                      Reached = col_number(), Lat = col_number(),

5668                      pc1 = col_number()))

5669    View(bienne_all)

5670    bienne_all$Treatment<-as.factor(bienne_all$Treatment)

5671

5672    modbreach<-glm(Reached~Treatment,family=binomial(link='logit'),data=bienne_all)

5673    summary(modbreach)

5674    library(multcomp)

5675    tukbreach<-glht(modbreach,mcp(Treatment="Tukey"))

5676    summary(tukbreach)

5677    barplot(tapply(bienne_all$Reached,bienne_all$Treatment,mean))

5678    barcentres<-
5679    barplot(tapply(bienne_all$Reached,bienne_all$Treatment,mean),ylim=c(0,max(tapply(bienne_all$Reached,bie
5680    nne_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Reaching Ovary", xlab = "Treatment", main =
5681    "Proportion of Pollen Reaching Ovary vs Treatment for L.bienne")

5682    means<-tapply(bienne_all$Reached,bienne_all$Treatment,mean)

5683    ses<-
5684    tapply(bienne_all$Reached,bienne_all$Treatment,sd)/sqrt(tapply(bienne_all$Reached,bienne_all$Treatment,l
5685    ength))

5686    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)
```

```
5687
5688   #Pollen Reached for usitatissimum:
5689   library(readr)
5690   usitatissimum_all <- read_csv("usitatissimum_all.csv",
5691                   col_types = cols(Count = col_number(),
5692                           Tube = col_number(), Percent = col_number(),
5693                           Reached = col_number(), Lat = col_number(),
5694                           pc1 = col_number()))
5695   View(usitatissimum_all)
5696   usitatissimum_all$Treatment<-as.factor(usitatissimum_all$Treatment)
5697
5698   modcreach<-glm(Reached~Treatment,family=binomial(link='logit'),data=bienne_all)
5699   summary(modcreach)
5700   library(multcomp)
5701   tukcreach<-glht(modcreach,mcp(Treatment="Tukey"))
5702   summary(tukcreach)
5703   barplot(tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean))
5704   barcentres<-
5705   barplot(tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean),ylim=c(0,max(tapply(usitatissi
5706   mum_all$Reached,usitatissimum_all$Treatment,mean)+0.1)), ylab = "Proportion of Pollen Reaching Ovary",
5707   xlab = "Treatment", main = "Proportion of Pollen Reaching Ovary vs Treatment for L.usitatissimum")
5708   means<-tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,mean)
5709   ses<-
5710   tapply(usitatissimum_all$Reached,usitatissimum_all$Treatment,sd)/sqrt(tapply(usitatissimum_all$Reached,us
5711   itatissimum_all$Treatment,length))
5712   arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)
5713
5714
5715   #test proportion tubes versus treatment x species with mixed model, ind as random effect
5716   Prop<-cbind(Pollen$Tube, Pollen$Count)
```

```
5717    mod5<-glmer(Prop~Treatment*Species+(1|Ind), family=quasibinomial(link='logit'), data=Pollen)

5718    summary(mod5)

5719    tuk5<-glht(mod5,mcp(Treatment="Tukey"))

5720    summary(tuk5)

5721    interact<-interaction(Pollen$Treatment,Pollen$Species)

5722    mod5y<-glmer(Prop~-1+interact+(1|Ind), family=binomial(link='logit'), data=Pollen)

5723    summary(mod5y)

5724    tuk5y<-glht(mod5y,mcp(interact="Tukey"))

5725    summary(tuk5y)

5726    Props<-Pollen$Tube/Pollen$Count

5727    Pollen<-data.frame(Pollen,Props)

5728    str(Pollen)

5729    barcentres<-
5730    barplot(tapply(Pollen$Props,Pollen$interact,mean),ylim=c(0,max(tapply(Pollen$Props,Pollen$interact,mean)+
5731    0.1)))

5732    means<-tapply(Pollen$Props,Pollen$interact,mean)

5733    ses<-tapply(Pollen$Props,Pollen$interact,sd)/sqrt(tapply(Pollen$Props,Pollen$interact,length))

5734    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3

5735
```

5736   **Appendix 15 : Adapter sequences used for the ddRADSeq process**

5737   Adapter_1 = AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT --
5738   Adapter_2 = CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

5739

5740   **Appendix 16 : P1 and P2 Barcode sequences for each samples**

5741   P1   P2   Sample

5742   gcatga   tcatgc   Sample_18

5743   gcatga   cggatc   Sample_19

5744   gcatga   taagac   Sample_20

5745   gcatga   actccg   Sample_49

5746   gcatga   aacgtg   Sample_50

5747   gcatga   ctgtat   Sample_69

5748   gcatga   gtaaca   Sample_70

5749   gcatga   gagcgt   Sample_71

5750   gcatga   tgccaa   Sample_106

5751   gcatga   gttact   Sample_83

5752   gcatga   cgcttg   Sample_21

5753   gcatga   tagcat   Sample_22

5754   gatccg   tcatgc   Sample_23

5755   gatccg   cggatc   Sample_74

5756   gatccg   taagac   Sample_84

5757   gatccg   actccg   Sample_85

5758   gatccg   aacgtg   Sample_92

5759   gatccg   ctgtat   Sample_93

5760   gatccg   gtaaca   Sample_107

5761   gatccg   gagcgt   Sample_110

5762   gatccg   tgccaa   Sample_111

5763   gatccg   tagcat   Sample_87

5764   gatccg   cgcttg   Sample_86

5765   gtctta   tcatgc   Sample_88

5766   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

5848   agtaac gtaaca Sample_2

5849   agtaac gagcgt Sample_3

5850   agtaac tgccaa Sample_4

5851   agtaac gttact   Sample_5

5852   agtaac cgcttg   Sample_32

5853   agtaac tagcat   Sample_76

5854   caagcg tcatgc Sample_121

5855   caagcg cggatc Sample_122

5856   caagcg taagac Sample_123

5857   caagcg actccg Sample_124

5858   caagcg aacgtg Sample_125

5859   caagcg ctgtat   Sample_126

5860   caagcg gtaaca Sample_150

5861   caagcg gagcgt Sample_151

5862   caagcg tgccaa Sample_152

5863

5864   **Appendix 17: Process_radtags command**

5865   process_radtags -P -1 /nobackup/mnkz72/raw/merged_rad_raw_R1.fastq.gz -2

5866   /nobackup/mnkz72/raw/merged_rad_raw_R2.fastq.gz -b /nobackup/mnkz72/barcodes/index_p1p2new.txt -o

5867   /nobackup/mnkz72/raw/demultiplexed -r -c -q --barcode_dist_2 2 --disable_rad_check --inline_inline --renz_1

5868   pstI --renz_2 mseI --adapter_1 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT -

5869   -adapter_2 CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

5870

5871   **Appendix 18: BWA-MEM Command**

5872   bwa mem -t16 /nobackup/mnkz72/1st_genome_asm/cultivar/GCA_000224295.2_ASM22429v2_genomic.fna

5873   /nobackup/mnkz72/raw/demultiplexed_run2/clean/Sample_1.1.fq.gz

5874   /nobackup/mnkz72/raw/demultiplexed_run2/clean/Sample_1.1.fq.gz | samtools view -b -h | samtools sort -

5875   @12 -o /nobackup/mnkz72/raw/aligned_run2/Sample_1.bam

5876

Appendix 19: Population list for popmap file

5877

5878	Sample_1	Vil	FRA
5879	Sample_10	Tym	UKN
5880	Sample_100	Mat	FRA
5881	Sample_101	Mat	FRA
5882	Sample_103	Lla	SNO
5883	Sample_104	15	SNO
5884	Sample_105	15	SNO
5885	Sample_106	15	SNO
5886	Sample_107	12	MED
5887	Sample_108	15	SNO
5888	Sample_109	15	SNO
5889	Sample_11	Tym	UKN
5890	Sample_110	12	MED
5891	Sample_111	12	MED
5892	Sample_112	Lla	SNO
5893	Sample_113	Lla	SNO
5894	Sample_114	Mat	FRA
5895	Sample_115	Mat	FRA
5896	Sample_116	Tym	UKN
5897	Sample_117	Tym	UKN
5898	Sample_118	Tym	UKN
5899	Sample_119	Tym	UKN
5900	Sample_12	Sut	UKN
5901	Sample_120	Tym	UKN
5902	Sample_121	Vil	FRA
5903	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

5985	Sample_90	13	SNO
5986	Sample_91	13	SNO
5987	Sample_92	10	SSB
5988	Sample_93	3	SSB
5989	Sample_94	Lla	SNO
5990	Sample_95	Lla	SNO
5991	Sample_96	Lla	SNO
5992	Sample_99	Tal	FRA
5993	Sample_97	Tal	FRA
5994	Sample_98	Tal	FRA

5995

Appendix 20: ref_map.pl pipeline command

5997   ref_map.pl -T 16 -r 0.8 -X populations:--fstats --structure --plink --vcf --samples

5998   /nobackup/mnkz72/raw/aligned_samples --popmap

5999   /nobackup/mnkz72/popmap/popmap/popmap_new_0322.txt -o /nobackup/mnkz72/new_refmap/run_2

6000

Appendix 21: PLINK command

6002   module load bioinformatics

6003   module load plink

6004   #Change path to workind directory:

6005   cd /nobackup/mnkz72/new_refmap/rdir/run3

6006   #Allow for extra choromosomes

6007   plink --file populations.plink --out run3_pops --recodeA --allow-extra-chr --noweb

6008   #Make BED:

6009   plink --file populations.plink --out run3_pops --make-bed --allow-extra-chr --noweb

6010

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
low2_10	-1.07492	50.68183
low2_13	-1.07492	50.68183
low2_14	-1.07492	50.68183
low2_2	-1.07492	50.68183
low2_25	-1.07492	50.68183
low2_26	-1.07492	50.68183
low2_7	-1.07492	50.68183
low2_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
```

6012

6013

6014   **Appendix 23: Command line to filter for heterozygousity**

6015   #Prepare and format file in bcftools

6016   #First, convert multiallelic sites to biallelic using bcftools:

6017   bcftools norm -m - "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" >
6018   allsamples_biallellic_convert.vcf

6019   #Filter the alternative alleles under certain value

6020   bcftools view -e "FORMAT/AD[:1]<2 && INFO/AD[1]<5"
6021   "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/allsamples_biallellic_convert.vcf" > biallelic-filtered.vcf

6022   #Compress vcf output with bgzip (Samtools):

6023   bgzip -@8 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/allsamples_biallellic_convert.vcf"

6024   bgzip -@8 "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf"

6025   #Produce vcf file statistics for files:

6026   bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf.gz" > run3_stats.txt

6027   bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/allsamples_biallellic_convert.vcf.gz" >
6028   biallelic_convert_stats.txt

6029   #het-filtered vcf:

6030   bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/allpopulations.hetfiltered.vcf.gz" >
6031   hetfiltered_stats.txt

6032   bcftools stats "/nobackup/mnkz72/new_refmap/bcf_out/allsamples.hetfiltered.vcf.gz" > het_bigfile_stats.txt

6033   #Extract Genotypes per sample:

```
6034    bcftools query -f '%CHROM %POS[\t%GT]\n' "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/biallelic-
6035    filtered.vcf" > genotypesbysample_run3.txt

6036    head -n 1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" | cut -c 1-
6037    1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" >
6038    genotypesbysample_cut_1000.txt

6039    bcftools query -f '%CHROM %POS[\t%GT]\n'
6040    "/nobackup/mnkz72/new_refmap/vcftools/allele_filter/biallelic_filtered_populations_snps.vcf" >
6041    genotypesbysample_biallelic.txt

6042    head -n 1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" | cut -c 1-
6043    1000 "/nobackup/mnkz72/new_refmap/bcf_out/allele_filter/genotypesbysample_run3.txt" >
6044    genotypesbysample_run3_cut_1000.txt

6045    #score hets as 1, homs as 0

6046    sumx<-rep(NA,nrow(genosx))

6047    sumy<-NULL

6048    for (i in c(1:ncol(genosx))){

6049      for (j in c(1:nrow(genosx))){

6050        if(genosx[j,i]=="1/0" | genosx[j,i]=="0/1"){

6051          sumx[j]<-1

6052        } else if (genosx[j,i]=="0/0" | genosx[j,i]=="1/1"){

6053          sumx[j]<-0

6054        }

6055      }

6056      sumy<-cbind(sumy,sumx)

6057      sumx<-rep(NA,nrow(genosx))

6058    }

6059    sumy

6060

6061    plot(c(1:nrow(sumy)),sumy[,3])

6062

6063    #proportion of hets per ind moving average
```

```
6064    sumz<-NULL

6065    mod<-50 # number of SNPs for moving average

6066    means<-NULL

6067    sumzz<-NULL

6068    inds<-NULL

6069    for (k in c(1:ncol(sumy))){

6070      for (l in c(1:nrow(sumy))){

6071        if (l %% mod == 0){

6072          meanx<-sum(sumy[c((l-mod):l),k],na.rm=T)/mod #mean over window size including missing data

6073          means<-c(means,meanx)

6074        }

6075      }

6076      sumz<-cbind(sumz,means)

6077      ind<-rep(k,l %/% mod)

6078      inds<-c(inds,ind)

6079      sumzz<-c(sumzz,means)

6080      means<-NULL

6081    }

6082    sumz

6083    sumzzz<-data.frame(inds,sumzz)

6084    head(sumzzz)

6085

6086    boxplot(sumzzz$sumzz~sumzzz$inds,xlab="sample",ylab="prop het")

6087    #view individuals of interest from boxplot to find regions of high het (replace number in sumz[,])

6088    plot(c(1:nrow(sumz)),sumz[,15],type="l",xlab="SNP_position",ylab="prop het")

6089    #view genos in region of interest (SNP_position*mod)

6090    sumy[(13*mod):(14*mod),15]

6091

6092    #plot all inds
```

```
6093    plot(c(1:nrow(sumz)),rep(max(sumz,na.rm=T),nrow(sumz)),ylim=c(-

6094    0.05,max(sumz,na.rm=T)),type="l",col="white",xlab="SNP_position",ylab="prop het")

6095    for (m in c(1:ncol(sumz))){

6096      lines(c(1:nrow(sumz)),sumz[,m],type="l")

6097    }

6098

6099    #proportion of hets per location moving average

6100    pop<-c(1:19)# list of inds to include (same population), numbers are order in the sample list

6101    lenp<-length(pop)

6102    sump<-sumy[,pop]

6103    meanp<-NULL

6104    for (m in c(1:nrow(sump))){

6105      meany<-sum(sump[m,],na.rm=T)/lenp #mean over population including missing data

6106      meanp<-c(meanp,meany)

6107    }

6108    meanp

6109

6110    plot(meanp,type="l",xlab="SNP_position",ylab="prop het")

6111

6112    #find het loci

6113    hetlocs<-which(meanp>quantile(meanp,0.99))

6114    hetvals<-meanp[meanp>quantile(meanp,0.99)]

6115    hetgenos<-data.frame(hetlocs,hetvals,sump[hetlocs,])

6116    hetgenos

6117

6118    #test with two pops

6119    #proportion of hets per location moving average

6120    pop<-list(c(1:8),c(9:19))# list of inds to include (same population), numbers are order in the sample list

6121    meanp<-NULL
```

245

```
6122    meanps<-NULL

6123    for (n in 1:length(pop)){

6124      lenp<-length(pop[[n]])

6125      sump<-sumy[,pop[[n]]]

6126      for (m in c(1:nrow(sump))){

6127        meany<-sum(sump[m,],na.rm=T)/lenp #mean over population including missing data

6128        meanp<-c(meanp,meany)

6129      }

6130      meanps<-cbind(meanps,meanp)

6131      meanp<-NULL

6132    }

6133    meanps

6134

6135    plot(meanps[,1],type="p",pch=1,xlab="SNP_position",ylab="prop het")

6136    for (p in 2:ncol(meanps)){

6137      points(meanps[,p],pch=p)

6138    }

6139

6140    #find het loci (top 99% hets)

6141    hetlocs<-NULL

6142    for (q in 1:length(pop)){

6143      hetloc<-which(meanps[,q]>quantile(meanps,0.99))

6144      hetlocs<-sort(union(hetlocs, hetloc))

6145    }

6146    hetvals<-meanps[hetlocs,]

6147    hetgenos<-data.frame(hetlocs,hetvals)

6148    hetgenos

6149
```

6150   **Appendix 24: Command line to filter by loci location in vcftools**

```
6151    #vcftools:

6152    #frequency analysis

6153    vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --freq --out
6154    run3_populations_freqanalysis

6155    #depth:

6156    vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --depth --out
6157    run3_populations_depthanalysis

6158    #allele counts:

6159    vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --counts --out
6160    run3_populations_countsanalysis

6161    #heterozygousity:

6162    vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --het --out
6163    run3_populations_hetanalysis

6164    #site-quality:

6165    vcftools --gzvcf /nobackup/mnkz72/new_refmap/vcftools/run3/populations.snps.vcf.gz --site-quality --out
6166    run3_populations_sitequalityanalysis

6167    #Allele-het-depth counts for raw vcg from ref_map:

6168    vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --freq --out
6169    allpopulations_freqanalysis_freq

6170    vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --depth --out
6171    allpopulations_freqanalysis_depth

6172    vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --het --out
6173    allpopulations_freqanalysis_het

6174    vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --counts --out
6175    allpopulations_freqanalysis_counts

6176    vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --site-quality --out
6177    allpopulations_freqanalysis_site-quality

6178    vcftools --vcf "/nobackup/mnkz72/new_refmap/bcf_out/populations.snps.vcf" --missing-indv --out
6179    allpopulations_freqanalysis_missingindv

6180    #Filter for bi-allelic alleles:

6181    vcftools --vcf /nobackup/mnkz72/new_refmap/vcftools/allele_filter/allsamples_cp.vcf --min-alleles 2 --max-
6182    alleles 2 --out biallelic_filtered_populations_snps –recode
```

6183

6184

6185 **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

6186

6187

6188

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

6190
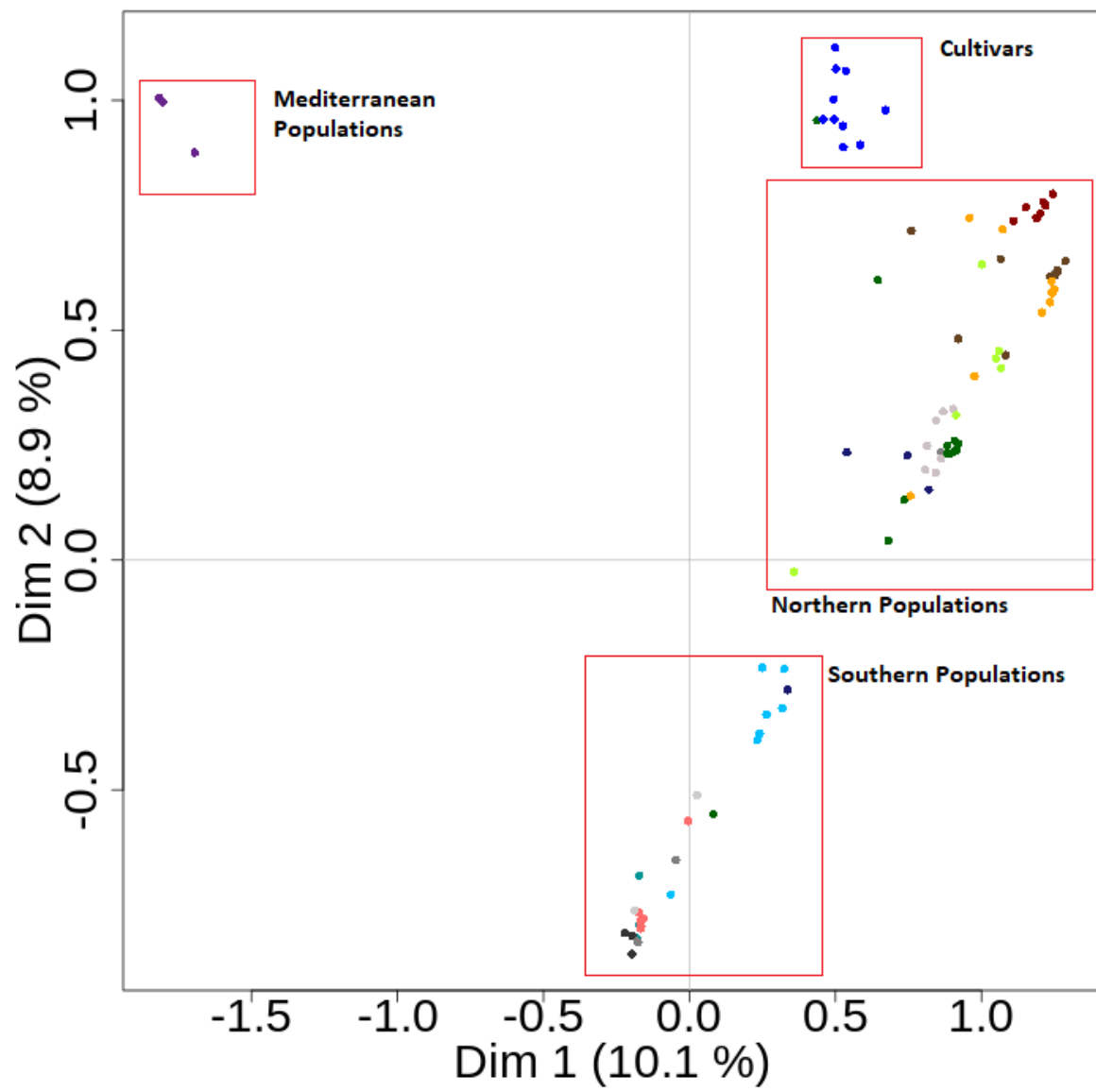
6191   **Appendix 27: DAPC Summary**
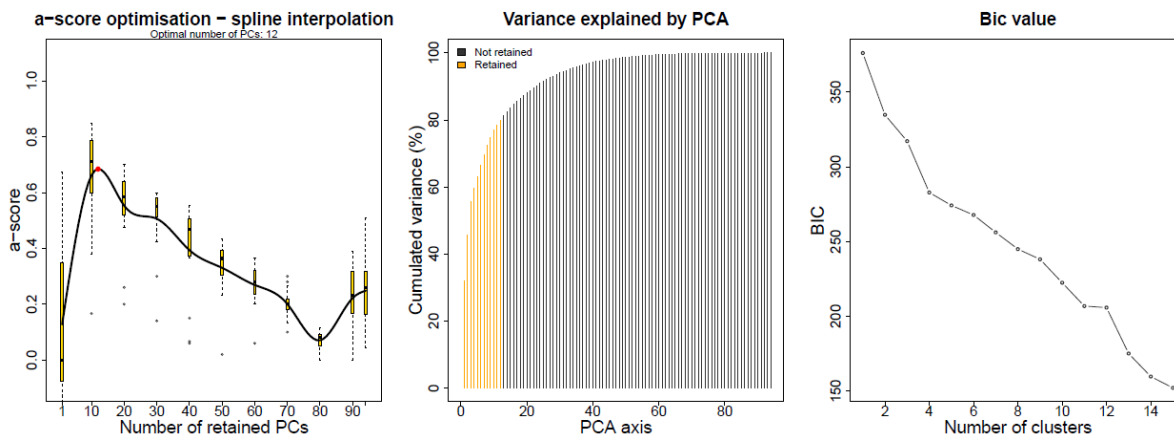
6192

6193

6194   **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			

6195

6196   **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

6197

Appendix 30: Chapter 5 R Commands

6198

6199 #Seed Measurements:

6200 setwd("G:/Linum Project/rdir")

6201 library(ggplot2)

6202 library(ggpubr)

6203 library(readr)

6204 seed_measurements_wild_data_pop <- read_csv("seed_measurements_wild_data_pop.csv",

6205   col_types = cols(`Lat_(degrees_N)` = col_number(),

6206   Alt = col_number(), `Seed_length_(Cm)` = col_number(),

6207   `Seed_width_(Cm)` = col_number(),

6208   `Seed_Area_(Cm)` = col_number(),

6209   pc1 = col_number()))

6210 View(seed_measurements_wild_data_pop)

6211

6212 #Modelling for seed area vs latitude:

6213 g <- ggscatter(seed_measurements_wild_data_pop, x = "Lat_(degrees_N)", y = "Seed_Area_(Cm)", size = 1.0,

6214   rug = TRUE,   # Add marginal rug

6215   color = "Experiment", palette = "uchicago", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +

6216   stat_cor(aes(color = Experiment), label.sep = ";", label.x.npc = "left", label.y.npc=0.20, method =

6217 "pearson")+geom_point(aes(color = Experiment))+stat_smooth(method="glm",se=FALSE)+

6218 theme_light()+ggtitle("Seed Area vs Latitude")

6219

```
6220    g +

6221      font("title", size = 22, color = "black", face = "bold")+

6222      font("xlab", size = 18, color = "black", face = "bold")+

6223      font("ylab", size = 18, color = "black", face = "bold")+

6224      font("legend.title", size = 18, color = "black", face = "bold")+

6225      font("legend.text", size = 14, color = "black",)

6226

6227    #Modelling for seed area vs pc1:

6228    h <- ggscatter(seed_measurements_wild_data_pop, x = "pc1", y = "Seed_Area_(Cm)", size = 1.0,

6229          rug = TRUE,                    # Add marginal rug

6230          color = "Experiment", palette = "uchicago", facet.by = "Experiment",add = "reg.line", conf.int = TRUE) +

6231     stat_cor(aes(color = Experiment), label.sep = ";", label.x.npc = "left", label.y.npc=0.20, method =

6232    "pearson")+geom_point(aes(color = Experiment))+stat_smooth(method="glm",se=FALSE)+

6233    theme_light()+ggtitle("Seed Area vs Climatic Variables (pc1)")

6234

6235    h +

6236      font("title", size = 22, color = "black", face = "bold")+

6237      font("xlab", size = 18, color = "black", face = "bold")+

6238      font("ylab", size = 18, color = "black", face = "bold")+

6239      font("legend.title", size = 18, color = "black", face = "bold")+

6240      font("legend.text", size = 14, color = "black",)

6241

6242    #Vern/Non-vern on other traits:

6243    library("car")

6244    setwd("G:/Linum Project/rdir")

6245    library(readr)

6246    combined_vern_control_2018_only <- read_csv("combined_vern_control_2018_only.csv",

6247                    col_types = cols(Lat = col_number(),

6248                        Lon = col_number(), pc1 = col_number(),
```

```
6249                                `Height_(Cm)` = col_number(), Stem_no = col_number(),

6250                                Bud_no = col_number(), Days_to_fl = col_number()))

6251    View(combined_vern_control_2018_only)

6252    scatterplot(Days_to_fl ~ Height_(Cm) | Treatment, data = combined_vern_control_2018_only,

6253         smooth = TRUE, regLine = TRUE, grid = FALSE, frame = FALSE, xlab = "Plant Height(Cm)", ylab = "Days to
6254    Flowering")

6255    abline(lm(Height ~ Days_to_fl data = combined_vern_control_2018_only),col="red")

6256

6257    ggscatter(combined_vern_control_2018_only, x = "Bud_no", y = "Days_to_fl", size = 2.0,

6258         rug = TRUE,                    # Add marginal rug

6259         color = "Treatment", palette = "jco", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6260      stat_cor(aes(color = Treatment), method = "pearson")

6261

6262    #Model summary

6263    summary1 <- summary(glm(Stem_no ~ Days_to_fl, data = combined_vern_control_2018_only))

6264    adjRsq <- summary1$adj.r.squared

6265    fStat <- summary1$statistic

6266    pValue <- pf(fStat[summary1])

6267    summary(summary1)

6268

6269    #Height - Days to fl test:

6270    Heightdaysmod<-cbind(combined_vern_control_2018_only$Days_to_fl,
6271    combined_vern_control_2018_only$'Height_(Cm)')

6272    Heightaysmod2<-glm(Heightdaysmod~Treatment, family=quasibinomial,
6273    data=combined_vern_control_2018_only)

6274    summary(Heightaysmod2)

6275    barcentres<-
6276    barplot(tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatm
6277    ent,mean),ylim=c(0,max(tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2
```

```
6278   018_only$Treatment,mean)+5)), ylab = "Height (Cm)", xlab = "Treatment", main = "Plant Height vs

6279   Treatments")

6280   means<-

6281   tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatment,mea

6282   n)

6283   ses<-

6284   tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatment,sd)/

6285   sqrt(tapply(combined_vern_control_2018_only$'Height_(Cm)',combined_vern_control_2018_only$Treatment

6286   ,length))

6287   arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

6288

6289   #Stem_no - Days to fl test:

6290   Stemdaysmod<-cbind(combined_vern_control_2018_only$Stem_no,

6291   combined_vern_control_2018_only$Days_to_fl)

6292   Stemdaysmod2<-glm(Stemdaysmod~Treatment, family=quasibinomial,

6293   data=combined_vern_control_2018_only)

6294   summary(Stemdaysmod2)

6295   barcentres<-

6296   barplot(tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,

6297   mean),ylim=c(0,max(tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only

6298   $Treatment,mean)+2)), ylab = "Stem Number", xlab = "Treatment", main = "Stem Number vs Treatments")

6299   means<-

6300   tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,mean)

6301   ses<-

6302   tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,sd)/sqrt(

6303   tapply(combined_vern_control_2018_only$Stem_no,combined_vern_control_2018_only$Treatment,length))

6304   arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

6305

6306   #Bud_no - Days to fl test:

6307   buddaysmod<-cbind(combined_vern_control_2018_only$Days_to_fl,

6308   combined_vern_control_2018_only$Bud_no)

6309   buddaysmod2<-buddaysmodlm<-lm(buddaysmod~Treatment,data=combined_vern_control_2018_only)
```

```
6310    summary(pc1daysmodlm)

6311    barcentres<-

6312    barplot(tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,m

6313    ean),ylim=c(0,max(tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Tr

6314    eatment,mean)+2)), ylab = "Bud Number", xlab = "Treatment", main = "Bud Number vs Treatments")

6315    means<-

6316    tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,mean)

6317    ses<-

6318    tapply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,sd)/sqrt(t

6319    apply(combined_vern_control_2018_only$Bud_no,combined_vern_control_2018_only$Treatment,length))

6320    arrows(barcentres, means-ses, barcentres, means+ses, length=0.05, angle=90, code=3)

6321

6322    #PC1 - Height Correlation:

6323    library(readr)

6324    wildonly_vern18 <- read_csv("wildonly_vern18.csv",

6325                    col_types = cols(Lat = col_number(),

6326                            Lon = col_number(), pc1 = col_number(),

6327                            `Height_(Cm)` = col_number(), Stem_no = col_number(),

6328                            Bud_no = col_number(), Days_to_fl = col_number()))

6329    View(wildonly_vern18)

6330

6331    #pc1 - height correlation:

6332    a <- ggscatter(wildonly_vern18, x = "pc1", y = "Height_(Cm)", size = 1.0,

6333         rug = TRUE,                     # Add marginal rug

6334         color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6335      stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.8, method =

6336    "pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+

6337    theme_light()+ggtitle("Plant Height (Cm) vs Climate Variable (pc1)")

6338

6339    a +

6340      font("title", size = 22, color = "black", face = "bold")+
```

```
6341    font("xlab", size = 18, color = "blue", face = "bold")+

6342    font("ylab", size = 18, color = "blue", face = "bold")+

6343    font("legend.title", size = 18, color = "black", face = "bold")+

6344    font("legend.text", size = 18, color = "blue", face = "bold")

6345

6346

6347    #Lat - height correlation:

6348    b <- ggscatter(wildonly_vern18, x = "Lat", y = "Height_(Cm)", size = 1.0,

6349         rug = TRUE,                    # Add marginal rug

6350         color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6351    stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.8, method =

6352    "pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+

6353    theme_light()+ggtitle("Plant Height (Cm) vs Latitude (degrees North)")

6354

6355    b +

6356    font("title", size = 20, color = "black", face = "bold")+

6357    font("xlab", size = 18, color = "blue", face = "bold")+

6358    font("ylab", size = 18, color = "blue", face = "bold")+

6359    font("legend.title", size = 18, color = "black", face = "bold")+

6360    font("legend.text", size = 18, color = "blue", face = "bold")

6361

6362    #pc1 - Stem number correlation:

6363    c <- ggscatter(wildonly_vern18, x = "pc1", y = "Stem_no", size = 1.0,

6364         rug = TRUE,                    # Add marginal rug

6365         color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6366    stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.9, method =

6367    "pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+

6368    theme_light()+ggtitle("Stem Number vs Climate Variable (pc1)")

6369

6370    c +
```

```
6371    font("title", size = 22, color = "black", face = "bold")+

6372    font("xlab", size = 18, color = "blue", face = "bold")+

6373    font("ylab", size = 18, color = "blue", face = "bold")+

6374    font("legend.title", size = 18, color = "black", face = "bold")+

6375    font("legend.text", size = 18, color = "blue", face = "bold")

6376

6377    #Lat - Stem number correlation:

6378    d <- ggscatter(wildonly_vern18, x = "Lat", y = "Stem_no", size = 1.0,

6379         rug = TRUE,                    # Add marginal rug

6380         color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6381    stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.85, method =

6382    "pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+

6383    theme_light()+ggtitle("Stem Number vs Latitude (degrees North)")

6384

6385    d +

6386     font("title", size = 22, color = "black", face = "bold")+

6387     font("xlab", size = 18, color = "blue", face = "bold")+

6388     font("ylab", size = 18, color = "blue", face = "bold")+

6389     font("legend.title", size = 18, color = "black", face = "bold")+

6390     font("legend.text", size = 18, color = "blue", face = "bold")

6391

6392    #pc1 - Bud number correlation:

6393    e <- ggscatter(wildonly_vern18, x = "pc1", y = "Bud_no", size = 1.0,

6394         rug = TRUE,                    # Add marginal rug

6395         color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6396    stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.7, method =

6397    "pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+

6398    theme_light()+ggtitle("Bud Number vs Climate Variable (pc1)")

6399

6400    e +
```

```
6401    font("title", size = 22, color = "black", face = "bold")+

6402    font("xlab", size = 18, color = "blue", face = "bold")+

6403    font("ylab", size = 18, color = "blue", face = "bold")+

6404    font("legend.title", size = 18, color = "black", face = "bold")+

6405    font("legend.text", size = 18, color = "blue", face = "bold")

6406

6407    #Lat - Bud number correlation:

6408    f <- ggscatter(wildonly_vern18, x = "Lat", y = "Bud_no", size = 1.0,

6409         rug = TRUE,                    # Add marginal rug

6410         color = "Treatment", palette = "uchicago", facet.by = "Treatment",add = "reg.line", conf.int = TRUE) +

6411     stat_cor(aes(color = Treatment), label.sep = ";", label.x.npc = "left", label.y.npc=0.53, method =

6412    "pearson")+geom_point(aes(color = Treatment))+stat_smooth(method="glm",se=FALSE)+

6413    theme_light()+ggtitle("Bud Number vs Latitude (degrees North)")

6414

6415    f +

6416     font("title", size = 22, color = "black", face = "bold")+

6417     font("xlab", size = 18, color = "blue", face = "bold")+

6418     font("ylab", size = 18, color = "blue", face = "bold")+

6419     font("legend.title", size = 18, color = "black", face = "bold")+

6420     font("legend.text", size = 18, color = "blue", face = "bold")
```

6421 **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

6422

Appendix 32: Macro for ImageJ processing (Chapter 5)

6424   #Opening file and setting known distance unit:

6425   open("D:\\Linum Project\\Cap and seed photos\\Seed photos Aug 2018\\1_1 Seeds.tif");

6426   makeLine(996, 210, 1014, 324);

6427   run("Set Scale...", "distance=115.41 known=1 pixel=1 unit=mm");

6428   run("Split Channels");

6429   selectWindow("1_1 Seeds.tif (blue)");

6430   run("Color Balance...");

6431   run("Apply LUT");

```
6432    run("Analyze Particles...", "size=25-Infinity pixel circularity=0.00-1.00 show=Outlines display");

6433    setAutoThreshold("Default");

6434    //run("Threshold...");

6435    setThreshold(0, 254);

6436    run("Convert to Mask");

6437    setAutoThreshold("Default");

6438    setThreshold(1, 255);

6439    run("Convert to Mask");

6440    run("Analyze Particles...", "size=25-Infinity pixel circularity=0.00-1.00 show=Outlines display");

6441    saveAs("Results", "D:\\Linum Project\\Cap and seed photos\\Seed photos Aug 2018\\1_1 Seeds.xls");

6442

6443    #_____

6444    _____

6445

6446    #Aplying LUT and making binary:

6447

6448    run("Apply LUT");

6449    run("Make Binary");

6450    run("Analyze Particles...", "size=25-Infinity pixel circularity=0.00-1.00 show=Outlines display");

6451

6452
```

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.

6518 Beltrán, R. *et al.* (2019) 'Effect of temperature on pollen germination for several Rosaceae species: Influence of

6519 freezing conservation time on germination patterns', *PeerJ*, 2019(12). Available at:

6520 https://doi.org/10.7717/PEERJ.8195/SUPP-1.

6521 Bigot, S. *et al.* (2018) 'Pivotal roles of environmental sensing and signaling mechanisms in plant responses to

6522 climate change', *Global Change Biology*, 24(12), pp. 5573–5589. Available at:

6523 https://doi.org/10.1111/GCB.14433.

6524 Bilton, T.P. *et al.* (2018) 'Linkage Disequilibrium Estimation in Low Coverage High-Throughput Sequencing

6525 Data', *Genetics*, 209(2), p. 389. Available at: https://doi.org/10.1534/GENETICS.118.300831.

6526 Bird, K.A. *et al.* (2017) 'Population structure and phylogenetic relationships in a diverse panel of Brassica rapa

6527 L.', *Frontiers in Plant Science*, 8. Available at: https://doi.org/10.3389/fpls.2017.00321.

6528 Blackman, B.K., Michaels, S.D. and Rieseberg, L.H. (2011) 'Connecting the sun to flowering in sunflower

6529 adaptation', *Molecular ecology*, 20(17), pp. 3503–3512. Available at: https://doi.org/10.1111/J.1365-

6530 294X.2011.05166.X.

6531 Blanquart, F. *et al.* (2013) 'A practical guide to measuring local adaptation', *Ecology Letters*, 16(9), pp. 1195–

6532 1205. Available at: https://doi.org/10.1111/ELE.12150.

6533 Blázquez, M., Koornneef, M. and Putterill, J. (2001) 'Flowering on time: genes that regulate the floral

6534 transition: Workshop on the molecular basis of flowering time control', *EMBO Reports*, 2(12), p. 1078.

6535 Available at: https://doi.org/10.1093/EMBO-REPORTS/KVE254.

6536 Blümel, M., Dally, N. and Jung, C. (2015) 'Flowering time regulation in crops — what did we learn from

6537 Arabidopsis?', *Current Opinion in Biotechnology*, 32, pp. 121–129. Available at:

6538 https://doi.org/10.1016/J.COPBIO.2014.11.023.

6539 Boavida, L.C. and McCormick, S. (2007) 'TECHNICAL ADVANCE: Temperature as a determinant factor for

6540 increased and reproducible in vitro pollen germination in Arabidopsis thaliana', *The Plant Journal*, 52(3), pp.

6541 570–582. Available at: https://doi.org/10.1111/J.1365-313X.2007.03248.X.

6542 Bolsheva, N.L. *et al.* (2017) 'Evolution of blue-flowered species of genus Linum based on high-throughput

6543 sequencing of ribosomal RNA genes', *BMC Evolutionary Biology*, 17(2), pp. 23–36. Available at:

6544 https://doi.org/10.1186/S12862-017-1105-X/FIGURES/6.

6545 Booker, T.R., Jackson, B.C. and Keightley, P.D. (2017) 'Detecting positive selection in the genome', *BMC*

6546 *Biology*, 15(1), pp. 1–10. Available at: https://doi.org/10.1186/S12915-017-0434-Y/FIGURES/2.

6547 Boshier, D. et al. (2015) 'Is local best? examining the evidence for local adaptation in trees and its scale',

6548 Environmental Evidence, 4(1). doi:10.1186/s13750-015-0046-3.

6549 Bougeard, S. and Dray, S. (2018) 'Supervised Multiblock Analysis in R with the ade4 Package', *Journal of*

6550 *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.

6584   Chambers, J. and Hastie, T., 1992. Statistical models in S. London: Chapman & Hall

6585   Chandler, J., Wilson, A. and Dean, C. (1996) 'Arabidopsis mutants showing an altered response to
6586   vernalization', *The Plant Journal*, 10(4), pp. 637–644. Available at: https://doi.org/10.1046/J.1365-
6587   313X.1996.10040637.X.

6588   Chapman, J.R. *et al.* (2009) 'A quantitative review of heterozygosity–fitness correlations in animal populations',
6589   *Molecular Ecology*, 18(13), pp. 2746–2765. Available at: https://doi.org/10.1111/J.1365-294X.2009.04247.X.

6590   Chen, C.-C. and Gibson, P.B. (1973) 'Effect of Temperature on Pollen-Tube Growth in Trifolium repense after
6591   Cross- and Self-Pollinations1', *Crop Science*, 13(5), pp. 563–566. Available at:
6592   https://doi.org/10.2135/CROPSCI1973.0011183X001300050021X.

6593   Chen, D. *et al.* (2018) 'Architecture of gene regulatory networks controlling flower development in Arabidopsis
6594   thaliana', *Nature Communications 2018 9:1*, 9(1), pp. 1–13. Available at: https://doi.org/10.1038/s41467-018-
6595   06772-3.

6596   Chiluwal, A. *et al.* (2020) 'Deterioration of ovary plays a key role in heat stress-induced spikelet sterility in
6597   sorghum', *Plant Cell and Environment*, 43(2), pp. 448–462. Available at: https://doi.org/10.1111/pce.13673.

6598   Cho, L.H., Yoon, J. and An, G. (2017) 'The control of flowering time by environmental factors', *The Plant
6599   Journal*, 90(4), pp. 708–719. Available at: https://doi.org/10.1111/TPJ.13461.

6600   Chouard, P. (1960) 'Vernalization and its Relations to Dormancy', *Annual Review of Plant Physiology*, 11(1), pp.
6601   191–238. Available at: https://doi.org/10.1146/ANNUREV.PP.11.060160.001203.

6602   Clarke, J.D. (2009) 'Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation', *Cold
6603   Spring Harbor Protocols*, 2009(3), p. pdb.prot5177. Available at: https://doi.org/10.1101/PDB.PROT5177.

6604   Cloutier, S. *et al.* (2009) 'Development and analysis of EST-SSRs for flax (Linum usitatissimum L.)', *Theoretical
6605   and Applied Genetics*, 119(1), pp. 53–63. Available at: https://doi.org/10.1007/S00122-009-1016-3/FIGURES/4.

6606   Cloutier, S. *et al.* (2012) 'Integrated consensus genetic and physical maps of flax (Linum usitatissimum L.)',
6607   *Theoretical and Applied Genetics 2012 125:8*, 125(8), pp. 1783–1795. Available at:
6608   https://doi.org/10.1007/S00122-012-1953-0.

6609   Clua, A.A. and Gimenez, D.O. (2003) 'Environmental factors during seed development of narrow-leaved bird's-
6610   foot-trefoil (Lotus tenuis) influences subsequent dormancy and germination', *Grass and Forage Science*, 58(4),
6611   pp. 333–338. Available at: https://doi.org/10.1046/j.1365-2494.2003.00385.x.

6612   Conner, J.K. (2002) 'Genetic mechanisms of floral trait correlations in a natural population', *Nature*, 420(6914),
6613   pp. 407–410. Available at: https://doi.org/10.1038/NATURE01105.

6614   Corbesier, L. and Coupland, G. (2006) 'The quest for florigen: a review of recent progress', *Journal of
6615   Experimental Botany*, 57(13), pp. 3395–3403. Available at: https://doi.org/10.1093/JXB/ERL095.

6616   Cox, M. and Doudna, J., 2015. Molecular Biology:2nd Revised edition. New York: W.H. FREEMAN & CO LTD,
6617   pp.221-238, 369-376, 592-593.

6618   Çopur, O., Gur, M.A., Karakus, M. and Demirel, U., 2006. Determination of correlation and path analysis among
6619   yield components and seed yield in oil flax varieties (Linum usitatissimum L.). *Journal of Biological Sciences*,
6620   6(4), pp.738-743.

6621   Craufurd, P.Q. and Wheeler, T.R. (2009) 'Climate change and the flowering time of annual crops', *Journal of*
6622   *experimental botany*, 60(9), pp. 2529–2539. Available at: https://doi.org/10.1093/JXB/ERP196.

6623   Crean, A.J. and Immler, S. (2021) 'Evolutionary consequences of environmental effects on gamete
6624   performance', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1826). Available at:
6625   https://doi.org/10.1098/RSTB.2020.0122.

6626   Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*, 27(15), pp. 2156–2158.
6627   Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR330.

6628   Darapuneni, M.K. *et al.* (2014) 'Effect of vernalization and photoperiod on flax flowering time', *Euphytica*,
6629   195(2), pp. 279–285. Available at: https://doi.org/10.1007/s10681-013-0996-x.

6630   Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional*
6631   *Genomics*, 9(5-6), 416–423. https://doi.org/10.1093/bfgp/elq031

6632   Davey, J.W. *et al.* (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation
6633   sequencing', *Nature Reviews Genetics 2011 12:7*, 12(7), pp. 499–510. Available at:
6634   https://doi.org/10.1038/nrg3012.

6635   Debieu, M. *et al.* (2013) 'Co-Variation between Seed Dormancy, Growth Rate and Flowering Time Changes with
6636   Latitude in Arabidopsis thaliana', *PLoS ONE*, 8(5). Available at: https://doi.org/10.1371/journal.pone.0061075.

6637   Delph, L. and Havens, K., 1998. Sperm Competition and Sexual Selection. Elsevier Inc. Pp.149-173.

6638   Dempewolf, H. *et al.* (2017) 'Past and Future Use of Wild Relatives in Crop Breeding', *Crop Science*, 57(3), pp.
6639   1070–1082. Available at: https://doi.org/10.2135/CROPSCI2016.10.0885.

6640   Descamps, C., Quinet, M. and Jacquemart, A.L. (2021) 'Climate Change–Induced Stress Reduce Quantity and
6641   Alter Composition of Nectar and Pollen From a Bee-Pollinated Species (Borago officinalis, Boraginaceae)',
6642   *Frontiers in Plant Science*, 12, p. 2264. Available at: https://doi.org/10.3389/FPLS.2021.755843/BIBTEX.

6643   Devasirvatham, V. *et al.* (2012) 'Effect of high temperature on the reproductive development of chickpea
6644   genotypes under controlled environments', *Functional Plant Biology*, 39(12), pp. 1009–1018. Available at:
6645   https://doi.org/10.1071/FP12033.

6646   Diamond, J. (2002) 'Evolution, consequences and future of plant and animal domestication', *Nature*,
6647   418(6898), pp. 700–707. Available at: https://doi.org/10.1038/NATURE01019.

6648   Dickinson, H., Rodriguez-Enriquez, J. and Grant-Downton, R. (2018) 'Pollen Germination and Pollen Tube
6649   Growth of Arabidopsis thaliana: in vitro and Semi in vivo Methods', *Bio-protocol*, 8(16). Available at:
6650   https://doi.org/10.21769/BIOPROTOC.2977.

6651   Diederichsen, A. (2001) 'Comparison of genetic diversity of flax (Linum usitatissimum L.) between Canadian
6652   cultivars and a world collection', *Plant Breeding*, 120(4), pp. 360–362. Available at:
6653   https://doi.org/10.1046/J.1439-0523.2001.00616.X.

6654   Diederichsen, A. and Raney, J.P. (2006) 'Seed colour, seed weight and seed oil content in Linum usitatissimum
6655   accessions held by Plant Gene Resources of Canada', *Plant Breeding*, 125(4), pp. 372–377. Available at:
6656   https://doi.org/10.1111/J.1439-0523.2006.01231.X.

6657   Diederichsen, A. and Ulrich, A. (2009) 'Variability in stem fibre content and its association with other
6658   characteristics in 1177 flax (Linum usitatissimum L.) genebank accessions', *Industrial Crops and Products*, 30(1),
6659   pp. 33–39. Available at: https://doi.org/10.1016/j.indcrop.2009.01.002.

6660   Dmitriev, A.A. *et al.* (2021) 'Genome Sequencing of Fiber Flax Cultivar Atlant Using Oxford Nanopore and
6661   Illumina Platforms', *Frontiers in Genetics*, 11, p. 1487. Available at:
6662   https://doi.org/10.3389/FGENE.2020.590282/BIBTEX.

6663   Dole, J.M. and Wilkins, H.F. (1994) "Interaction of bulb vernalization and shoot photoperiod on `Nellie White'
6664   Easter Lily," HortScience, 29(3), pp. 143–145. Available at: https://doi.org/10.21273/hortsci.29.3.143.

6665   Donohue, K. *et al.* (2005) 'Environmental and genetic influences on the germination of Arabidopsis thaliana in
6666   the field', *Evolution*, 59(4), pp. 740–757. Available at: https://doi.org/10.1111/J.0014-3820.2005.TB01750.X.

6667   Donohue, K. *et al.* (2010) 'Germination, postgermination adaptation, and species ecological ranges', *Annual
6668   Review of Ecology, Evolution, and Systematics*, 41, pp. 293–319. Available at: https://doi.org/10.1146/annurev-
6669   ecolsys-102209-144715.

6670   Dray, S. and Dufour, A.B. (2007) 'The ade4 Package: Implementing the Duality Diagram for Ecologists', *Journal
6671   of Statistical Software*, 22(4), pp. 1–20. Available at: https://doi.org/10.18637/JSS.V022.I04.

6672   Dudley, S.A. (1996) 'THE RESPONSE TO DIFFERING SELECTION ON PLANT PHYSIOLOGICAL TRAITS: EVIDENCE
6673   FOR LOCAL ADAPTATION', *Evolution*, 50(1), pp. 103–110. Available at: https://doi.org/10.1111/J.1558-
6674   5646.1996.TB04476.X.

6675   Ejsmond, M.J. *et al.* (2011) 'Does climate affect pollen morphology? Optimal size and shape of pollen grains
6676   under various desiccation intensity', *Ecosphere*, 2(10), pp. 1–15. Available at: https://doi.org/10.1890/ES11-
6677   00147.1.

6678   El-Beltagi, H. and Salama, Z. (2007) *Evaluation of fatty acids profile and the content of some secondary
6679   metabolites in seeds of different flax cultivars (Linum usitatissimum L.) Cytotoxic and antioxidant properties of
6680   active principals isolated from water hyacinth against four cancer cells lines View project Ascorbic Acid*

6681 *Enhances Growth and Yield of Sweet Peppers (Capsicum annum) by Mitigating Salinity Stress View project*.

6682 Available at: https://www.researchgate.net/publication/255606154.

6683 Elfving, N. *et al.* (2011) 'The Arabidopsis thaliana Med25 mediator subunit integrates environmental cues to

6684 control plant development', *Proceedings of the National Academy of Sciences of the United States of America*,

6685 108(20), pp. 8245–8250. Available at: https://doi.org/10.1073/PNAS.1002981108.

6686 Ellstrand, N.C., Prentice, H.C. and Hancock, J.F. (2003) 'Gene Flow and Introgression from Domesticated Plants

6687 into their Wild Relatives', *https://doi.org/10.1146/annurev.ecolsys.30.1.539*, 30, pp. 539–563. Available at:

6688 https://doi.org/10.1146/ANNUREV.ECOLSYS.30.1.539.

6689 Essemine, J., Ammar, S. and Bouzid, S. (2010) 'Impact of heat stress on germination and growth in higher

6690 plants: physiological, biochemical and molecular repercussions and mechanisms of defence.', *Journal of

6691 Biological Sciences*, 10(6), pp. 565–572.

6692 Fábián, A. *et al.* (2019) 'Stigma functionality and fertility are reduced by heat and drought co-stress in wheat',

6693 *Frontiers in Plant Science*, 10. Available at: https://doi.org/10.3389/FPLS.2019.00244/FULL.

6694 Falconer, D.S., 1996. Introduction to quantitative genetics. Pearson Education India.

6695 FAOSTAT data (2006) Food and Agriculture Organiza-tion of the United Nations, FAO Statistical Data-

6696 bases, URL: http://faostat.fao.org/

6697 Fedorenko, O.M., Savushkin, A.I. and Olimpienko, G.S. (2001) 'Genetic Diversity in Natural Populations of

6698 Arabidopsis thaliana (L.) Heynh. from Karelia', *Russian Journal of Genetics 2001 37:2*, 37(2), pp. 162–167.

6699 Available at: https://doi.org/10.1023/A:1009041922186.

6700 Firon, N., Nepi, M. and Pacini, E. (2012) 'Water status and associated processes mark critical stages in pollen

6701 development and functioning', *Annals of Botany*, 109(7), pp. 1201–1214. Available at:

6702 https://doi.org/10.1093/AOB/MCS070.

6703 Flowers, J.M. *et al.* (2009) 'Population Genomics of the Arabidopsis thaliana Flowering Time Gene Network',

6704 *Molecular Biology and Evolution*, 26(11), pp. 2475–2486. Available at:

6705 https://doi.org/10.1093/MOLBEV/MSP161.

6706 Fragallah, S.A.D.A. *et al.* (2019) 'Effects of Sucrose, Boric Acid, pH, and Incubation Time on in Vitro Germination

6707 of Pollen and Tube Growth of Chinese fir (Cunnighamial lanceolata L.)', *Forests 2019, Vol. 10, Page 102*, 10(2),

6708 p. 102. Available at: https://doi.org/10.3390/F10020102.

6709 Frichot, E. *et al.* (2014) 'Fast and efficient estimation of individual ancestry coefficients', *Genetics*, 196(4), pp.

6710 973–983. Available at: https://doi.org/10.1534/GENETICS.113.160572/-/DC1.

6711 Frichot, E. and François, O. (2015) 'LEA: An R package for landscape and ecological association studies',

6712 *Methods in Ecology and Evolution*, 6(8), pp. 925–929. Available at: https://doi.org/10.1111/2041-210X.12382.

6713 Fu, Y.-B., Diederichsen, A. and Allaby, R.G. (2011) "Locus-specific view of flax domestication history," Ecology

6714 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.

6748   Ginzinger, D.G. (2002) 'Gene quantification using real-time quantitative PCR: An emerging technology hits the

6749   mainstream', *Experimental Hematology*, 30(6), pp. 503–512. Available at: https://doi.org/10.1016/S0301-

6750   472X(02)00806-8.

6751   Giorno, F. *et al.* (2013) 'Ensuring Reproduction at High Temperatures: The Heat Stress Response during Anther

6752   and Pollen Development', *Plants 2013, Vol. 2, Pages 489-506*, 2(3), pp. 489–506. Available at:

6753   https://doi.org/10.3390/PLANTS2030489.

6754   Girden, E.R., 1992. ANOVA: Repeated measures (No. 84). sage.

6755   Glantz, S., Slinker, B. and Neilands, T., n.d. Primer of applied regression et analysis of variance.

6756   Godwin, I.D., Aitken, E.A.B. and Smith, L.W. (1997) 'Application of inter simple sequence repeat (ISSR) markers

6757   to plant genetics', *ELECTROPHORESIS*, 18(9), pp. 1524–1528. Available at:

6758   https://doi.org/10.1002/ELPS.1150180906.

6759   Goudenhooft, C., Bourmaud, A. and Baley, C. (2019) 'Flax (Linum usitatissimum L.) fibers for composite

6760   reinforcement: Exploring the link between plant growth, cell walls development, and fiber properties',

6761   *Frontiers in Plant Science*. Frontiers Media S.A. Available at: https://doi.org/10.3389/fpls.2019.00411.

6762   Grapin, A. *et al.* (1998) 'Diploid Musa acuminata genetic diversity assayed with sequence-tagged microsatellite

6763   sites', *ELECTROPHORESIS*, 19(8–9), pp. 1374–1380. Available at: https://doi.org/10.1002/ELPS.1150190829.

6764   Grassein, F., Lavorel, S. and Till-Bottraud, I. (2014) 'The importance of biotic interactions and local adaptation

6765   for plant response to environmental changes: field evidence along an elevational gradient', *Global Change

6766   Biology*, 20(5), pp. 1452–1460. Available at: https://doi.org/10.1111/GCB.12445.

6767   Greer, S. *et al.* (2010) 'Housekeeping genes; expression levels may change with density of cultured cells',

6768   *Journal of Immunological Methods*, 355(1–2), pp. 76–79. Available at:

6769   https://doi.org/10.1016/J.JIM.2010.02.006.

6770   Gregis, V. *et al.* (2009) 'The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress

6771   class B and C floral homeotic genes', *The Plant Journal*, 60(4), pp. 626–637. Available at:

6772   https://doi.org/10.1111/J.1365-313X.2009.03985.X.

6773   Guan, S. *et al.* (2021) 'Genetic analysis and population structure of wild and cultivated wishbone flower

6774   (Torenia fournieri Lind.) lines related to specific floral color', *PeerJ*, 7, p. e11702. Available at:

6775   https://doi.org/10.7717/PEERJ.11702/SUPP-6.

6776   Günther, T. *et al.* (2016) 'Genomic and phenotypic differentiation of Arabidopsis thaliana along altitudinal

6777   gradients in the North Italian Alps', *Molecular Ecology*, 25(15), pp. 3574–3592. Available at:

6778   https://doi.org/10.1111/MEC.13705.

6779   Guo, D. *et al.* (2020) 'Resequencing 200 Flax Cultivated Accessions Identifies Candidate Genes Related to Seed

6780   Size and Weight and Reveals Signatures of Artificial Selection', *Frontiers in Plant Science*, 10. Available at:

6781   https://doi.org/10.3389/fpls.2019.01682.

6782 Guo, Q. (2012) 'Incorporating latitudinal and central–marginal trends in assessing genetic variation across

6783 species ranges', *Molecular Ecology*, 21(22), pp. 5396–5403. Available at: https://doi.org/10.1111/MEC.12012.

6784 Gupta, S. *et al.* (2021) 'Identification and characterization of a novel long juvenile resource AGS 25', *Genetic*

6785 *Resources and Crop Evolution*, 68(3), pp. 1149–1163. Available at: https://doi.org/10.1007/S10722-020-01055-

6786 7/TABLES/2.

6787 Gutaker, R.M. *et al.* (2019) 'Flax latitudinal adaptation at LuTFL1 altered architecture and promoted fiber

6788 production', *Scientific Reports 2019 9:1*, 9(1), pp. 1–12. Available at: https://doi.org/10.1038/s41598-018-

6789 37086-5.

6790 Gutiérrez-Valencia, J. *et al.* (2022) 'Genomic analyses of the Linum distyly supergene reveal convergent

6791 evolution at the molecular level', *Current Biology*, 32(20), pp. 4360-4371.e6. Available at:

6792 https://doi.org/10.1016/J.CUB.2022.08.042.

6793 El Haddad, N. *et al.* (2021) 'Crop wild relatives in durum wheat breeding: Drift or thrift?', *Crop Science*, 61(1),

6794 pp. 37–54. Available at: https://doi.org/10.1002/CSC2.20223.

6795 Hall, M.C. and Willis, J.H. (2006) 'DIVERGENT SELECTION ON FLOWERING TIME CONTRIBUTES TO LOCAL

6796 ADAPTATION IN MIMULUS GUTTATUS POPULATIONS', *Evolution*, 60(12), pp. 2466–2477. Available at:

6797 https://doi.org/10.1111/J.0014-3820.2006.TB01882.X.

6798 Hämälä, T. and Savolainen, O. (2019) 'Genomic Patterns of Local Adaptation under Gene Flow in Arabidopsis

6799 lyrata', *Molecular Biology and Evolution*, 36(11), pp. 2557–2571. Available at:

6800 https://doi.org/10.1093/MOLBEV/MSZ149.

6801 Hampe, A. and Petit, R.J. (2005) 'Conserving biodiversity under climate change: the rear edge matters', *Ecology*

6802 *Letters*, 8(5), pp. 461–467. Available at: https://doi.org/10.1111/J.1461-0248.2005.00739.X.

6803 Han, Q. *et al.* (2021) 'Association between RsFT, RsFLC and RsCOL5 (A&B) expression and flowering regulation

6804 in Japanese wild radish', *AoB PLANTS*, 13(4). Available at: https://doi.org/10.1093/aobpla/plab039.

6805 Hanley, M.E. *et al.* (2008) 'Breeding system, pollinator choice and variation in pollen quality in British

6806 herbaceous plants', *Functional Ecology*, 22(4), pp. 592–598. Available at: https://doi.org/10.1111/J.1365-

6807 2435.2008.01415.X.

6808 Harvey, M.G. et al. (2016) "Sequence capture versus restriction site associated DNA sequencing for Shallow

6809 Systematics," Systematic Biology, 65(5), pp. 910–924. Available at: https://doi.org/10.1093/sysbio/syw036.

6810 Hayat, M.J. and Higgins, M. (2014) 'Understanding poisson regression', *Journal of Nursing Education*, 53(4), pp.

6811 207–215. Available at: https://doi.org/10.3928/01484834-20140325-04.

6812 He, Y., Chen, T. and Zeng, X. (2020) 'Genetic and Epigenetic Understanding of the Seasonal Timing of

6813 Flowering', *Plant Communications*, 1(1). Available at:

6814 https://doi.org/10.1016/J.XPLC.2019.100008/GENETIC_AND_EPIGENETIC_UNDERSTANDING_OF_THE_SEASON

6815 AL_TIMING_OF_FLOWERING.PDF.

6816 Hedhly, A., Hormaza, J.I. and Herrero, M. (2004) 'Effect of temperature on pollen tube kinetics and dynamics in

6817 sweet cherry, Prunus avium (Rosaceae)', *American Journal of Botany*, 91(4), pp. 558–564. Available at:

6818 https://doi.org/10.3732/ajb.91.4.558.

6819 Hedhly, A., Hormaza, J.I. and Herrero, M. (2005) 'The effect of temperature on pollen germination, pollen tube

6820 growth, and stigmatic receptivity in peach', *Plant Biology*, 7(5), pp. 476–483. Available at:

6821 https://doi.org/10.1055/S-2005-865850.

6822 Heisler, M.G. and Jönsson, H. (2007) 'Modelling meristem development in plants', *Current Opinion in Plant*

6823 *Biology*, 10(1), pp. 92–97. Available at: https://doi.org/10.1016/J.PBI.2006.11.005.

6824 Herburger, K. and Holzinger, A. (2016) 'Aniline blue and Calcofluor white staining of callose and cellulose in the

6825 streptophyte green algae Zygnema and Klebsormidium', *Bio-protocol*, 6(20). Available at:

6826 https://doi.org/10.21769/BIOPROTOC.1969.

6827 Hernández, F. *et al.* (2019) 'Genetic Diversity and Population Structure of Wild Sunflower (Helianthus annuus

6828 L.) in Argentina: Reconstructing Its Invasion History', *Journal of Heredity*, 110(6), pp. 746–759. Available at:

6829 https://doi.org/10.1093/JHERED/ESZ047.

6830 Hernandez, R.D. *et al.* (2019) 'Ultrarare variants drive substantial cis heritability of human gene expression',

6831 *Nature Genetics 2019 51:9*, 51(9), pp. 1349–1355. Available at: https://doi.org/10.1038/s41588-019-0487-7.

6832 Higgins, J.A., Bailey, P.C. and Laurie, D.A. (2010) 'Comparative genomics of flowering time pathways using

6833 brachypodium distachyon as a model for the temperate Grasses', *PLoS ONE*, 5(4). Available at:

6834 https://doi.org/10.1371/journal.pone.0010065.

6835 Higuchi, R. *et al.* (1992) 'Simultaneous amplification and detection of specific DNA sequences', *Bio/technology*

6836 *(Nature Publishing Company)*, 10(4), pp. 413–417. Available at: https://doi.org/10.1038/NBT0492-413.

6837 Hjelmqvist, H. (1950) 'The flax weeds and the origin of cultivated flax.', *Botaniska Notiser*, 2, pp. 257–98.

6838 Hoban, S. *et al.* (2016) 'Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future

6839 directions', *American Naturalist*, 188(4), pp. 379–397. Available at:

6840 https://doi.org/10.1086/688018/ASSET/IMAGES/LARGE/FG4.JPEG.

6841 Hopkins, R., Schmitt, J. and Stinchcombe, J.R. (2008) 'A latitudinal cline and response to vernalization in leaf

6842 angle and morphology in Arabidopsis thaliana (Brassicaceae)', *New Phytologist*, 179(1), pp. 155–164. Available

6843 at: https://doi.org/10.1111/J.1469-8137.2008.02447.X.

6844 Hothorn, T., Bretz, F. and Westfall, P. (2008a) 'Simultaneous Inference in General Parametric Models',

6845 *Biometrical Journal*, 50(3), pp. 346–363. Available at: https://doi.org/10.1002/BIMJ.200810425.

6846 Hothorn, T., Bretz, F. and Westfall, P. (2008b) 'Simultaneous Inference in General Parametric Models',

6847 *Biometrical Journal*, 50(3), pp. 346–363. Available at: https://doi.org/10.1002/BIMJ.200810425.

6848 Hu, T. *et al.* (2021) 'Next-generation sequencing technologies: An overview', *Human Immunology*, 82(11), pp.

6849 801–811. Available at: https://doi.org/10.1016/J.HUMIMM.2021.02.012.

6850   Huggett, J. *et al.* (2005) 'Real-time RT-PCR normalisation; strategies and considerations', *Genes & Immunity*
6851   *2005 6:4*, 6(4), pp. 279–284. Available at: https://doi.org/10.1038/sj.gene.6364190.

6852   Huis, R., Hawkins, S. and Neutelings, G. (2010) 'Selection of reference genes for quantitative gene expression
6853   normalization in flax (Linum usitatissimum L.)', *BMC Plant Biology*, 10(1), pp. 1–14. Available at:
6854   https://doi.org/10.1186/1471-2229-10-71/FIGURES/6.

6855   Ibrahim, A. *et al.* (2018) 'A regulator of early flowering in barley (Hordeum vulgare L.)', *PLOS ONE*, 13(7), p.
6856   e0200722. Available at: https://doi.org/10.1371/JOURNAL.PONE.0200722.

6857   IDT. 2022. Tips for resuspending and diluting your oligonucleotides. [online] Available at:
6858   <https://eu.idtdna.com/pages/education/decoded/article/tips-for-resuspending-and-diluting-your-
6859   oligonucleotides> [Accessed 2 June 2022].

6860   Iguchi, A. *et al.* (2020) 'Deep-sea amphipods around cobalt-rich ferromanganese crusts: Taxonomic diversity
6861   and selection of candidate species for connectivity analysis', *PLOS ONE*, 15(2), p. e0228483. Available at:
6862   https://doi.org/10.1371/JOURNAL.PONE.0228483.

6863   Innan, H. *et al.* (1999) 'A Method for Estimating Nucleotide Diversity From AFLP Data', *Genetics*, 151(3), pp.
6864   1157–1164. Available at: https://doi.org/10.1093/GENETICS/151.3.1157.

6865   Innes, P. *et al.* (2022) 'Assessment of biogeographic variation in traits of Lewis flax (Linum lewisii) for use in
6866   restoration and agriculture', *AoB PLANTS*, 14(2). Available at: https://doi.org/10.1093/aobpla/plac005.

6867   Ionkova, I. *et al.* (2013) 'Linum narbonense: A new valuable tool for biotechnological production of a potent
6868   anticancer lignan Justicidine B', *Pharmacognosy Magazine*, 9(33), p. 39. Available at:
6869   https://doi.org/10.4103/0973-1296.108138.

6870   Iovane, M. *et al.* (2022) 'High temperature and humidity affect pollen viability and longevity in olea europaea
6871   l.', *Agronomy*, 12(1). Available at: https://doi.org/10.3390/agronomy12010001.

6872   Iwaizumi, M.G. and Takahashi, M. (2012) 'Effects of pollen supply and quality on seed formation and
6873   maturation in Pinus densiflora', *Journal of plant research*, 125(4), pp. 517–525. Available at:
6874   https://doi.org/10.1007/S10265-011-0463-5.

6875   Jaakola, L. and Hohtola, A. (2010) 'Effect of latitude on flavonoid biosynthesis in plants', *Plant, Cell &*
6876   *Environment*, 33(8), pp. 1239–1247. Available at: https://doi.org/10.1111/J.1365-3040.2010.02154.X.

6877   Jahnke, M.R. and Etterson, J.R. (2019) 'Autonomous self-fertilization in Linum sulcatum, a native annual with a
6878   previously unknown mating system', *Journal of plant research*, 132(1), pp. 57–67. Available at:
6879   https://doi.org/10.1007/S10265-018-1076-Z.

6880   Jarman, S.N., Ward, R.D. and Elliott, N.G. (2002) 'Oligonucleotide primers for PCR amplification of coelomate
6881   introns', *Marine Biotechnology*, 4(4), pp. 347–355. Available at: https://doi.org/10.1007/s10126-002-0029-6.

6882 Jensen, H. *et al.* (2007) 'Multilocus heterozygosity and inbreeding depression in an insular house sparrow
6883 metapopulation', *Molecular ecology*, 16(19), pp. 4066–4078. Available at: https://doi.org/10.1111/J.1365-
6884 294X.2007.03452.X.

6885 Jhala, A. and Hall, L.M. (2014) *Flax (Linum usitatissimum L.): Current uses and future applications*. Available at:
6886 https://www.researchgate.net/publication/228952435.

6887 Jhala, A.J. and Hall, L.M. (2010) 'Flax (Linum usitatissimum L.): Current Uses and Future Applications',
6888 *Australian Journal of basic and Applied Sciences*, 4(9), pp. 4304–4312.

6889 Jiang, F., Cadotte, M.W. and Jin, G. (2021) 'Individual-level leaf trait variation and correlation across biological
6890 and spatial scales', *Ecology and Evolution*, 11(10), p. 5344. Available at: https://doi.org/10.1002/ECE3.7425.

6891 Johnson, M.A., Harper, J.F. and Palanivelu, R. (1AD) 'A Fruitful Journey: Pollen Tube Navigation from
6892 Germination to Fertilization'. Available at: https://doi.org/10.1146/annurev-arplant-050718.

6893 Johnstone, A., 2001. Biology. Oxford: Oxford University Press, pp.94-95.

6894 Jombart, T. and Ahmed, I. (2011) 'adegenet 1.3-1: new tools for the analysis of genome-wide SNP data',
6895 *Bioinformatics*, 27(21), pp. 3070–3071. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR521.

6896 Jombart, T. and Bateman, A. (2008) 'adegenet: a R package for the multivariate analysis of genetic markers',
6897 *Bioinformatics*, 24(11), pp. 1403–1405. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTN129.

6898 Jones, B. and Gliddon, C. (1999) *Reproductive biology and genetic structure in Lloydia serotina*, *Plant Ecology*.

6899 de Jong, M.A. *et al.* (2011) 'Mitochondrial DNA Signature for Range-Wide Populations of Bicyclus anynana
6900 Suggests a Rapid Expansion from Recent Refugia', *PLOS ONE*, 6(6), p. e21385. Available at:
6901 https://doi.org/10.1371/JOURNAL.PONE.0021385.

6902 de Jong, M.J. *et al.* (2021) 'SambaR: An R package for fast, easy and reproducible population-genetic analyses
6903 of biallelic SNP data sets', *Molecular Ecology Resources*, 21(4), pp. 1369–1379. Available at:
6904 https://doi.org/10.1111/1755-0998.13339.

6905 Jordon-Thaden, I.E. *et al.* (2020) 'A basic ddRADseq two-enzyme protocol performs well with herbarium and
6906 silica-dried tissues across four genera', *Applications in Plant Sciences*, 8(4), p. e11344. Available at:
6907 https://doi.org/10.1002/APS3.11344.

6908 Jung, J.H. *et al.* (2007) 'The GIGANTEA-Regulated MicroRNA172 Mediates Photoperiodic Flowering
6909 Independent of CONSTANS in Arabidopsis', *The Plant Cell*, 19(9), pp. 2736–2748. Available at:
6910 https://doi.org/10.1105/TPC.107.054528.

6911 Kakani, V.G. *et al.* (2002) 'Response of in vitro pollen germination and pollen tube growth of groundnut
6912 (Arachis hypogaea L.) genotypes to temperature', *Plant, Cell & Environment*, 25(12), pp. 1651–1661. Available
6913 at: https://doi.org/10.1046/J.1365-3040.2002.00943.X.

6914   Kakani, V.G. *et al.* (2005) 'Differences in in vitro Pollen Germination and Pollen Tube Growth of Cotton

6915   Cultivars in Response to High Temperature', *Annals of Botany*, 96(1), pp. 59–67. Available at:

6916   https://doi.org/10.1093/AOB/MCI149.

6917   Kamvar, Z.N., Tabima, J.F. and Gr̈unwald, N.J. (2014a) 'Poppr: An R package for genetic analysis of populations

6918   with clonal, partially clonal, and/or sexual reproduction', *PeerJ*, 2014(1), pp. 1–14. Available at:

6919   https://doi.org/10.7717/PEERJ.281/TABLE-6.

6920   Kamvar, Z.N., Tabima, J.F. and Gr̈unwald, N.J. (2014b) 'Poppr: An R package for genetic analysis of populations

6921   with clonal, partially clonal, and/or sexual reproduction', *PeerJ*, 2014(1), pp. 1–14. Available at:

6922   https://doi.org/10.7717/PEERJ.281/TABLE-6.

6923   Kassambara, A., 2022. 'ggplot2' Based Publication Ready Plots [R package ggpubr version 0.4.0]. [online]

6924   Cran.r-project.org. Available at: <https://cran.r-project.org/web/packages/ggpubr/index.html> [Accessed 9

6925   October 2022].

6926   Kayes, J.M. and Clark, S.E. (1998) 'CLAVATA2, a regulator of meristem and organ development in Arabidopsis',

6927   *Development*, 125(19), pp. 3843–3851. Available at: https://doi.org/10.1242/DEV.125.19.3843.

6928   Keddy, P.A. and Constabel, P. (1986) 'Germination of Ten Shoreline Plants in Relation to Seed Size, Soil Particle

6929   Size and Water Level: An Experimental Study', *The Journal of Ecology*, 74(1), p. 133. Available at:

6930   https://doi.org/10.2307/2260354.

6931   Khan, M.R.G., Ai, X.Y. and Zhang, J.Z. (2014) 'Genetic regulation of flowering time in annual and perennial

6932   plants', *Wiley Interdisciplinary Reviews: RNA*. Blackwell Publishing Ltd, pp. 347–359. Available at:

6933   https://doi.org/10.1002/wrna.1215.

6934   Khanday, I. and Sundaresan, V. (2021) 'Plant zygote development: recent insights and applications to clonal

6935   seeds', *Current opinion in plant biology*, 59. Available at: https://doi.org/10.1016/J.PBI.2020.101993.

6936   Kim, D.H. *et al.* (2009) 'Vernalization: Winter and the Timing of Flowering in Plants',

6937   *http://dx.doi.org/10.1146/annurev.cellbio.042308.113411*, 25, pp. 277–299. Available at:

6938   https://doi.org/10.1146/ANNUREV.CELLBIO.042308.113411.

6939   Kinmonth-Schultz, H. *et al.* (2021a) 'Flowering Times of Wild Arabidopsis Accessions From Across Norway

6940   Correlate With Expression Levels of FT, CO, and FLC Genes', *Frontiers in Plant Science*, 12, p. 2376. Available at:

6941   https://doi.org/10.3389/FPLS.2021.747740/BIBTEX.

6942   Kinmonth-Schultz, H. *et al.* (2021b) 'Flowering Times of Wild Arabidopsis Accessions From Across Norway

6943   Correlate With Expression Levels of FT, CO, and FLC Genes', *Frontiers in Plant Science*, 12, p. 2376. Available at:

6944   https://doi.org/10.3389/FPLS.2021.747740/BIBTEX.

6945   Koonin, E. V. (2005) 'Orthologs, paralogs, and evolutionary genomics', *Annual review of genetics*, 39, pp. 309–

6946   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.

6979 Leijten, W. *et al.* (2018) 'Translating Flowering Time from Arabidopsis thaliana to Brassicaceae and Asteraceae
6980 Crop Species', *Plants 2018, Vol. 7, Page 111*, 7(4), p. 111. Available at:
6981 https://doi.org/10.3390/PLANTS7040111.

6982 Leimu, R. and Fischer, M. (2008a) 'A meta-analysis of local adaptation in plants', *PLoS ONE*, 3(12). Available at:
6983 https://doi.org/10.1371/journal.pone.0004010.

6984 Leimu, R. and Fischer, M. (2008b) 'A Meta-Analysis of Local Adaptation in Plants', *PLOS ONE*, 3(12), p. e4010.
6985 Available at: https://doi.org/10.1371/JOURNAL.PONE.0004010.

6986 Lemuth, K. and Rupp, S. (2015) 'Microarrays as Research Tools and Diagnostic Devices', *RNA Technologies*, pp.
6987 259–280. Available at: https://doi.org/10.1007/978-3-319-17305-4_13/FIGURES/2.

6988 Li, H. (2011) 'A statistical framework for SNP calling, mutation discovery, association mapping and population
6989 genetical parameter estimation from sequencing data', *Bioinformatics*, 27(21), p. 2987. Available at:
6990 https://doi.org/10.1093/BIOINFORMATICS/BTR509.

6991 Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform',
6992 *Bioinformatics (Oxford, England)*, 25(14), pp. 1754–1760. Available at:
6993 https://doi.org/10.1093/BIOINFORMATICS/BTP324.

6994 Li, L. *et al.* (2015a) 'Flowering responses to light and temperature', *Science China Life Sciences 2016 59:4*, 59(4),
6995 pp. 403–408. Available at: https://doi.org/10.1007/S11427-015-4910-8.

6996 Li, L. *et al.* (2015b) 'Flowering responses to light and temperature', *Science China Life Sciences 2016 59:4*, 59(4),
6997 pp. 403–408. Available at: https://doi.org/10.1007/S11427-015-4910-8.

6998 Li, Y.X. *et al.* (2016) 'Identification of genetic variants associated with maize flowering time using an extremely
6999 large multi-genetic background population', *The Plant Journal*, 86(5), pp. 391–402. Available at:
7000 https://doi.org/10.1111/TPJ.13174.

7001 Lian, H., Roy, E. and Zheng, H. (2016) 'Microglial Phagocytosis Assay', *BIO-PROTOCOL*, 6(21). Available at:
7002 https://doi.org/10.21769/BIOPROTOC.1988.

7003 Liang, W. *et al.* (2018a) 'Selection and evaluation of reference genes for qRT-PCR analysis in Euscaphis konishii
7004 Hayata based on transcriptome data', *Plant Methods*, 14(1), pp. 1–9. Available at:
7005 https://doi.org/10.1186/S13007-018-0311-X/FIGURES/4.

7006 Liang, W. *et al.* (2018b) 'Selection and evaluation of reference genes for qRT-PCR analysis in Euscaphis konishii
7007 Hayata based on transcriptome data', *Plant Methods*, 14(1), pp. 1–9. Available at:
7008 https://doi.org/10.1186/S13007-018-0311-X/FIGURES/4.

7009 Linck, E. and Battey, C.J. (2019) 'Minor allele frequency thresholds strongly affect population structure
7010 inference with genomic data sets', *Molecular Ecology Resources*, 19(3), pp. 639–647. Available at:
7011 https://doi.org/10.1111/1755-0998.12995.

7012 Lischer, H.E.L. and Excoffier, L. (2012) 'PGDSpider: an automated data conversion tool for connecting

7013 population genetics and genomics programs', *Bioinformatics (Oxford, England)*, 28(2), pp. 298–299. Available

7014 at: https://doi.org/10.1093/BIOINFORMATICS/BTR642.

7015 Liu, J. *et al.* (2015) 'Genetic and epigenetic control of plant heat responses', *Frontiers in Plant Science*, 6(APR),

7016 p. 267. Available at: https://doi.org/10.3389/FPLS.2015.00267/BIBTEX.

7017 Liu, Y. *et al.* (2019) 'Genome-Wide Single Nucleotide Polymorphisms are Robust in Resolving Fine-Scale

7018 Population Genetic Structure of the Small Brown Planthopper, Laodelphax striatellus (Fallén) (Hemiptera:

7019 Delphacidae)', *Journal of Economic Entomology*, 112(5), pp. 2362–2368. Available at:

7020 https://doi.org/10.1093/JEE/TOZ145.

7021 Livak, K.J. and Schmittgen, T.D. (2001a) 'Analysis of Relative Gene Expression Data Using Real-Time

7022 Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402–408. Available at:

7023 https://doi.org/10.1006/METH.2001.1262.

7024 Livak, K.J. and Schmittgen, T.D. (2001b) 'Analysis of Relative Gene Expression Data Using Real-Time

7025 Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402–408. Available at:

7026 https://doi.org/10.1006/METH.2001.1262.

7027 Loha, A., Tigabu, M. and Fries, A. (2008) 'Genetic variation among and within populations of Cordia africana in

7028 seed size and germination responses to constant temperatures', *Euphytica 2008 165:1*, 165(1), pp. 189–196.

7029 Available at: https://doi.org/10.1007/S10681-008-9806-2.

7030 Lotz, L.A.P. (1990) 'The Relation Between Age and Size at First Flowering of Plantago Major in Various

7031 Habitats', *The Journal of Ecology*, 78(3), p. 757. Available at: https://doi.org/10.2307/2260897.

7032 Loveless, M.D. and Hamrick, J.L. (1984a) *Ecological Determinants of Genetic Structure in Plant Populations*,

7033 *Source: Annual Review of Ecology and Systematics*. Available at: https://about.jstor.org/terms.

7034 Loveless, M.D. and Hamrick, J.L. (1984b) *Ecological Determinants of Genetic Structure in Plant Populations*,

7035 *Source: Annual Review of Ecology and Systematics*. Available at: https://about.jstor.org/terms.

7036 Lovell, J.T. *et al.* (2013) 'Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation', *Proceedings*

7037 *of the Royal Society B: Biological Sciences*, 280(1763). Available at: https://doi.org/10.1098/rspb.2013.1043.

7038 Lu, P. *et al.* (2014) 'The Arabidopsis CALLOSE DEFECTIVE MICROSPORE1 Gene Is Required for Male Fertility

7039 through Regulating Callose Metabolism during Microsporogenesis', *Plant Physiology*, 164(4), pp. 1893–1904.

7040 Available at: https://doi.org/10.1104/PP.113.233387.

7041 Luo, Z. *et al.* (2019) 'Genetic diversity and population structure of a Camelina sativa spring panel', *Frontiers in*

7042 *Plant Science*, 10, p. 184. Available at: https://doi.org/10.3389/FPLS.2019.00184/BIBTEX.

7043 Lütfi, P. (2002) *The Effects of Temperature on Pollen Germination and Pollen Tube Growth of Apricot and Sweet*

7044 *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.

7078   Miller, J.M., Cullingham, C.I. and Peery, R.M. (2020) 'The influence of a priori grouping on inference of genetic
7079   clusters: simulation study and literature review of the DAPC method', *Heredity 2020 125:5*, 125(5), pp. 269–
7080   280. Available at: https://doi.org/10.1038/s41437-020-0348-2.

7081   Miranda, J.A. and Steward, G.F. (2017) 'Variables influencing the efficiency and interpretation of reverse
7082   transcription quantitative PCR (RT-qPCR): An empirical study using Bacteriophage MS2', *Journal of Virological*
7083   *Methods*, 241, pp. 1–10. Available at: https://doi.org/10.1016/J.JVIROMET.2016.12.002.

7084   Mittell, E.A. *et al.* (2020) 'Feral populations of Brassica oleracea along Atlantic coasts in western Europe',
7085   *Ecology and Evolution*, 10(20), p. 11810. Available at: https://doi.org/10.1002/ECE3.6821.

7086   Mobaxterm., 2022. MobaXterm free Xserver and tabbed SSH client for Windows. [online] Available at:
7087   <https://mobaxterm.mobatek.net/> [Accessed 21 September 2022].

7088   MOLES, ANGELA T. et al. (2004) "Small-seeded species produce more seeds per square metre of canopy per
7089   year, but not per individual per lifetime," Journal of Ecology, 92(3), pp. 384–396. Available at:
7090   https://doi.org/10.1111/j.0022-0477.2004.00880.x.

7091   Montalt, R. *et al.* (2019) 'Influence of temperature on the progamic phase in Citrus', *Environmental and*
7092   *Experimental Botany*, 166, p. 103806. Available at: https://doi.org/10.1016/J.ENVEXPBOT.2019.103806.

7093   Montesinos-Navarro, A., Picó, F. and Tonsor, S. (2012) 'Clinal variation in seed traits influencing life cycle
7094   timing in Arabidopsis thaliana', *Evolution*, 66.

7095   Moran, E. V. (2020) 'Simulating the effects of local adaptation and life history on the ability of plants to track
7096   climate shifts', *AoB PLANTS*, 12(1). Available at: https://doi.org/10.1093/aobpla/plaa008.

7097   Mori, A.S. *et al.* (2022) 'Feedbacks Between Biodiversity and Climate Change', *The Ecological and Societal*
7098   *Consequences of Biodiversity Loss*, pp. 281–304. Available at: https://doi.org/10.1002/9781119902911.CH13.

7099   Muravenko, O. V. *et al.* (2010a) 'Karyogenomics of species of the genus Linum L.', *Russian Journal of Genetics*
7100   *2010 46:10*, 46(10), pp. 1182–1185. Available at: https://doi.org/10.1134/S1022795410100091.

7101   Muravenko, O. V. *et al.* (2010b) 'Karyogenomics of species of the genus Linum L.', *Russian Journal of Genetics*
7102   *2010 46:10*, 46(10), pp. 1182–1185. Available at: https://doi.org/10.1134/S1022795410100091.

7103   Murray, B.R. *et al.* (2004) 'Geographical gradients in seed mass in relation to climate', *Journal of Biogeography*,
7104   31(3), pp. 379–388. Available at: https://doi.org/10.1046/J.0305-0270.2003.00993.X.

7105   Muthamia, K. (1994) 'Effect of bulb size and duration of vernalization on onion (allium cepa l.) seed yield'.
7106   Available at: http://erepository.uonbi.ac.ke/handle/11295/23061 (Accessed: 9 October 2022).

7107   Mylne, J. *et al.* (2004) 'Epigenetic Regulation in the Control of Flowering', *Cold Spring Harbor Symposia on*
7108   *Quantitative Biology*, 69, pp. 457–464. Available at: https://doi.org/10.1101/SQB.2004.69.457.

7109   Nag, S., Mitra, J. and Karmakar, P.G. (2015) ' An Overview on Flax (Linum usitatissimum L.) and its Genetic

7110   Diversity ', *International Journal of Agriculture, Environment and Biotechnology*, 8(4), p. 805. Available at:

7111   https://doi.org/10.5958/2230-732x.2015.00089.3.

7112   Nazareno, A.G. et al. (2021) 'By animal, water, or wind: Can dispersal mode predict genetic connectivity in

7113   riverine plant species?', *Frontiers in Plant Science*, 12. doi:10.3389/fpls.2021.626405.

7114   Nei, M. (1972a) *Genetic Distance between Populations*, *Source: The American Naturalist*.

7115   Nei, M. (1972b) *Genetic Distance between Populations*, *Source: The American Naturalist*. Available at:

7116   https://about.jstor.org/terms.

7117   Newbigin, E., Anderson, M.A. and Clarke', A.E. (1993) *Gametophytic Self-lncompatibility Systems*, *The Plant

7118   Cell*.

7119   Nievola, C.C. *et al.* (2017) 'Rapid responses of plants to temperature changes', *Temperature: Multidisciplinary

7120   Biomedical Journal*, 4(4), p. 371. Available at: https://doi.org/10.1080/23328940.2017.1377812.

7121   Normand, S. et al. (2011) 'Postglacial migration supplements climate in determining plant species ranges in

7122   Europe', *Proceedings of the Royal Society B: Biological Sciences*, 278(1725), pp. 3644–3653.

7123   doi:10.1098/rspb.2010.2769.

7124   Nôžková, J., Remeselníková, K. and Bjelková, M. (2014) 'Characterization and evaluation of flax seeds (Linum

7125   usitatissimum L.) on selected genotypes', *Journal of Central European Agriculture*, 15(1), pp. 193–207.

7126   Available at: https://doi.org/10.5513/JCEA01/15.1.1434.

7127   Odong, T.L. *et al.* (2011) 'Determination of genetic structure of germplasm collections: Are traditional

7128   hierarchical clustering methods appropriate for molecular marker data?', *Theoretical and Applied Genetics*,

7129   123(2), pp. 195–205. Available at: https://doi.org/10.1007/S00122-011-1576-X/FIGURES/6.

7130   Öhlund, E., Zurek, K. and Hammer, M. (2015) 'Towards Sustainable Agriculture? The EU framework and local

7131   adaptation in Sweden and Poland', *Environmental Policy and Governance*, 25(4), pp. 270–287. Available at:

7132   https://doi.org/10.1002/EET.1687.

7133   Olejniczak, P. *et al.* (2018) 'Seed size in mountain herbaceous plants changes with elevation in a species-

7134   specific manner', *PLoS ONE*, 13(6). Available at: https://doi.org/10.1371/journal.pone.0199224.

7135   Osnato, M. *et al.* (2022) 'Photoperiod Control of Plant Growth: Flowering Time Genes Beyond Flowering',

7136   *Frontiers in Plant Science*, 12, p. 3359. Available at: https://doi.org/10.3389/FPLS.2021.805635/BIBTEX.

7137   Ozhimkova, E., Uschapovsky, I. and Manaenkov, O. (2023) "Study of varietal differences in the composition of

7138   heteropolysaccharides of oil flax and fiber flax," Polysaccharides, 4(1), pp. 78–87. Available at:

7139   https://doi.org/10.3390/polysaccharides4010006.

7140   Pacini, E. (2008) 'Pollination', *Encyclopedia of Ecology*, pp. 562–565. Available at:

7141   https://doi.org/10.1016/B978-0-444-63768-0.00859-3.

7142 Pacini, E. and Dolferus, R. (2016) 'The Trials and Tribulations of the Plant Male Gametophyte — Understanding

7143 Reproductive Stage Stress Tolerance', *Abiotic and Biotic Stress in Plants - Recent Advances and Future*

7144 *Perspectives* [Preprint]. Available at: https://doi.org/10.5772/61671.

7145 Paina, C. *et al.* (2014) 'Vernalization Mediated Changes in the Lolium perenne Transcriptome', *PLOS ONE*, 9(9),

7146 p. e107365. Available at: https://doi.org/10.1371/JOURNAL.PONE.0107365.

7147 Pali, V. *et al.* (2014) 'Identification of microsatellite markers for fingerprinting popular Indian flax (Linum

7148 usitatissimum L.) cultivars and their utilization in seed genetic purity assessments', *Australian Journal of Crop*

7149 *Sciences*, 8(1).

7150 Pan, C. et al. (2018) 'Tomato stigma exsertion induced by high temperature is associated with the jasmonate

7151 signalling pathway', *Plant, Cell & Environment*, 42(4), pp. 1205–1221. Available at: doi:10.1111/pce.13444.

7152 Paradis, E. and Schliep, K. (2019a) 'ape 5.0: an environment for modern phylogenetics and evolutionary

7153 analyses in R', *Bioinformatics*, 35(3), pp. 526–528. Available at:

7154 https://doi.org/10.1093/BIOINFORMATICS/BTY633.

7155 Paradis, E. and Schliep, K. (2019b) 'ape 5.0: an environment for modern phylogenetics and evolutionary

7156 analyses in R', *Bioinformatics*, 35(3), pp. 526–528. Available at:

7157 https://doi.org/10.1093/BIOINFORMATICS/BTY633.

7158 Park, M.J. *et al.* (2016) 'LATE ELONGATED HYPOCOTYL regulates photoperiodic flowering via the circadian clock

7159 in Arabidopsis', *BMC Plant Biology*, 16(1), pp. 1–12. Available at: https://doi.org/10.1186/S12870-016-0810-

7160 8/FIGURES/5.

7161 Peet, M.M. and Bartholemew, M. (1996) *Effect of Night Temperature on Pollen Characteristics, Growth, and*

7162 *Fruit Set in Tomato*, *J. AMER. SOC. HORT. SCI*.

7163 Pembleton, L.W., Cogan, N.O.I. and Forster, J.W. (2013a) 'StAMPP: an R package for calculation of genetic

7164 differentiation and structure of mixed-ploidy level populations', *Molecular Ecology Resources*, 13(5), pp. 946–

7165 952. Available at: https://doi.org/10.1111/1755-0998.12129.

7166 Pembleton, L.W., Cogan, N.O.I. and Forster, J.W. (2013b) 'StAMPP: an R package for calculation of genetic

7167 differentiation and structure of mixed-ploidy level populations', *Molecular Ecology Resources*, 13(5), pp. 946–

7168 952. Available at: https://doi.org/10.1111/1755-0998.12129.

7169 Peterson, B.K. *et al.* (2012a) 'Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and

7170 Genotyping in Model and Non-Model Species', *PLOS ONE*, 7(5), p. e37135. Available at:

7171 https://doi.org/10.1371/JOURNAL.PONE.0037135.

7172 Peterson, B.K. *et al.* (2012b) 'Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and

7173 Genotyping in Model and Non-Model Species', *PLOS ONE*, 7(5), p. e37135. Available at:

7174 https://doi.org/10.1371/JOURNAL.PONE.0037135.

7175   Peterson, B.K. *et al.* (2012c) 'Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and

7176   Genotyping in Model and Non-Model Species', *PLOS ONE*, 7(5), p. e37135. Available at:

7177   https://doi.org/10.1371/JOURNAL.PONE.0037135.

7178   Pfaffl, M.W. (2001) 'A new mathematical model for relative quantification in real-time RT–PCR', *Nucleic Acids*

7179   *Research*, 29(9), p. e45. Available at: https://doi.org/10.1093/NAR/29.9.E45.

7180   Porebski, S., Bailey, L.G. and Baum, B.R. (1997) 'Modification of a CTAB DNA extraction protocol for plants

7181   containing high polysaccharide and polyphenol components', *Plant Molecular Biology Reporter 1997 15:1*,

7182   15(1), pp. 8–15. Available at: https://doi.org/10.1007/BF02772108.

7183   Posé, D. *et al.* (2013) 'Temperature-dependent regulation of flowering by antagonistic FLM variants', *Nature*

7184   *2013 503:7476*, 503(7476), pp. 414–417. Available at: https://doi.org/10.1038/nature12633.

7185   Promega.co.uk. 2022. GoTaq® qPCR and RT-qPCR Systems | qPCR Master Mix | RT-qPCR Master Mix. [online]

7186   Available at: <https://www.promega.co.uk/products/pcr/qpcr-and-rt-qpcr/qpcr-

7187   kits/?catNum=A6020#resources> [Accessed 30 May 2022].

7188   Przeworski, M. (2002) 'The Signature of Positive Selection at Randomly Chosen Loci', *Genetics*, 160(3), pp.

7189   1179–1189. Available at: https://doi.org/10.1093/GENETICS/160.3.1179.

7190   Purcell, S. *et al.* (2007a) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage

7191   Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at:

7192   https://doi.org/10.1086/519795.

7193   Purcell, S. *et al.* (2007b) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage

7194   Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at:

7195   https://doi.org/10.1086/519795.

7196   Purcell, S. *et al.* (2007c) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage

7197   Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. Available at:

7198   https://doi.org/10.1086/519795.

7199   Purugganan, M.D. and Fuller, D.Q. (2009) 'The nature of selection during plant domestication', *Nature*,

7200   457(7231), pp. 843–848. Available at: https://doi.org/10.1038/NATURE07895.

7201   Putterill, J., Laurie, R. and Macknight, R. (2004) 'It's time to flower: the genetic control of flowering time',

7202   *BioEssays*, 26(4), pp. 363–373. Available at: https://doi.org/10.1002/BIES.20021.

7203   Qaderi, M.M., Martel, A.B. and Dixon, S.L. (2019) 'Environmental Factors Influence Plant Vascular System and

7204   Water Regulation', *Plants*, 8(3), p. 65. Available at: https://doi.org/10.3390/PLANTS8030065.

7205   Qin, Y. *et al.* (2009) 'Penetration of the Stigma and Style Elicits a Novel Transcriptome in Pollen Tubes, Pointing

7206   to Genes Critical for Growth in a Pistil', *PLOS Genetics*, 5(8), p. e1000621. Available at:

7207   https://doi.org/10.1371/JOURNAL.PGEN.1000621.

7208   *QuantiFluor dsDNA System for High Concentration DNA* (no date). Available at:

7209   https://www.promega.co.uk/resources/pubhub/quantifluor-dsdna-system-for-high-concentration-dna/

7210   (Accessed: 3 May 2022).

7211   Rahimi, M.M., Zarei, M.A. and Arminian, A. (2011) 'Selection criteria of flax (Linum usitatissimum L.) for seed

7212   yield, yield components and biochemical compositions under various planting dates and nitrogen', *African*

7213   *Journal of Agricultural Research*, 6(13), pp. 3167–3175. Available at: https://doi.org/10.5897/AJAR11.382.

7214   Rao, X. *et al.* (2013a) 'An improvement of the 2ˆ(−delta delta CT) method for quantitative real-time

7215   polymerase chain reaction data analysis', *Biostatistics, bioinformatics and biomathematics*, 3(3), p. 71.

7216   Available at: /pmc/articles/PMC4280562/ (Accessed: 13 June 2022).

7217   Rao, X. *et al.* (2013b) 'An improvement of the 2ˆ(−delta delta CT) method for quantitative real-time

7218   polymerase chain reaction data analysis', *Biostatistics, bioinformatics and biomathematics*, 3(3), p. 71.

7219   Available at: /pmc/articles/PMC4280562/ (Accessed: 15 August 2022).

7220   Rastogi, A. *et al.* (2013) *Effect of auxin and gibberellic acid on growth and yield components of linseed (Linum*

7221   *usitatissimum L.)*, *Crop Breeding and Applied Biotechnology*.

7222   Rauf, S. *et al.* (2010) *Consequences of Plant Breeding on genetic diversity Bt cotton in Pakistan: Performance*

7223   *and adoption with imperfect information View project Drought tolerant Sunflower breeding material View*

7224   *project International Journal of Plant Breeding Consequences of Plant Breeding on Genetic Diversity*. Available

7225   at: https://www.researchgate.net/publication/235910237.

7226   Raven, P., 1982. Biology of plants. New York: Worth Publishers, pp.132-137.

7227   Ray, C. (1944) 'CYTOLOGICAL STUDIES ON THE FLAX GENUS, LINUM', *American Journal of Botany*, 31(4), pp.

7228   241–248. Available at: https://doi.org/10.1002/J.1537-2197.1944.TB08027.X.

7229   R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical

7230   Computing, Vienna, Austria. Available online at https://www.R-project.org/.

7231   Ream, T.S. *et al.* (2014) 'Interaction of Photoperiod and Vernalization Determines Flowering Time of

7232   Brachypodium distachyon', *Plant Physiology*, 164(2), p. 694. Available at:

7233   https://doi.org/10.1104/PP.113.232678.

7234   Reeves, P.H. and Coupland, G. (2000) 'Response of plant development to environment: control of flowering by

7235   daylength and temperature', *Current Opinion in Plant Biology*, 3(1), pp. 37–42. Available at:

7236   https://doi.org/10.1016/S1369-5266(99)00041-2.

7237   Rehman, H.U. *et al.* (2014) 'Seed Priming Influence on Early Crop Growth, Phenological Development and Yield

7238   Performance of Linola (Linum usitatissimum L.)', *Journal of Integrative Agriculture*, 13(5), pp. 990–996.

7239   Available at: https://doi.org/10.1016/S2095-3119(13)60521-3.

7240 Remans, T. *et al.* (2008) 'Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis

7241 thaliana exposed to increased metal concentrations', *Planta*, 227(6), pp. 1343–1349. Available at:

7242 https://doi.org/10.1007/S00425-008-0706-4/FIGURES/1.

7243 Reynolds, W.W. and Casterlin, M.E. (1980) 'The Role of Temperature in the Environmental Physiology of

7244 Fishes', *Environmental Physiology of Fishes*, pp. 497–518. Available at: https://doi.org/10.1007/978-1-4899-

7245 3659-2_19.

7246 Riedel, G. *et al.* (2014a) 'An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes

7247 Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells', *PLOS ONE*, 9(3), p. e93031. Available

7248 at: https://doi.org/10.1371/JOURNAL.PONE.0093031.

7249 Riedel, G. *et al.* (2014b) 'An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes

7250 Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells', *PLOS ONE*, 9(3), p. e93031. Available

7251 at: https://doi.org/10.1371/JOURNAL.PONE.0093031.

7252 Riedel, G. *et al.* (2014c) 'An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes

7253 Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells', *PLOS ONE*, 9(3), p. e93031. Available

7254 at: https://doi.org/10.1371/JOURNAL.PONE.0093031.

7255 Del Río, J.C. *et al.* (2011) 'Structural characterization of guaiacyl-rich lignins in flax (Linum usitatissimum) fibers

7256 and shives', *Journal of Agricultural and Food Chemistry*, 59(20), pp. 11088–11099. Available at:

7257 https://doi.org/10.1021/JF201222R/ASSET/IMAGES/LARGE/JF-2011-01222R_0001.JPEG.

7258 Roach, D.A. (1987) *Variation in Seed and Seedling Size in Anthoxanthum odoratum*, *Source: The American

7259 Midland Naturalist*.

7260 Rodriguez-Enriquez, M.J. *et al.* (2013) 'A novel method for efficient in vitro germination and tube growth of

7261 Arabidopsis thaliana pollen', *New Phytologist*, 197(2), pp. 668–679. Available at:

7262 https://doi.org/10.1111/NPH.12037.

7263 Royo, C. *et al.* (2020) 'Agronomic performance of durum wheat landraces and modern cultivars and its

7264 association with genotypic variation in vernalization response (Vrn-1) and photoperiod sensitivity (Ppd-1)

7265 genes', *European Journal of Agronomy*, 120, p. 126129. Available at:

7266 https://doi.org/10.1016/J.EJA.2020.126129.

7267 Rozhmina, T. *et al.* (2021) 'A comprehensive dataset of flax (Linum uitatissimum L.) phenotypes', *Data in Brief*,

7268 37. Available at: https://doi.org/10.1016/j.dib.2021.107224.

7269 RStudio Team, 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL

7270 http://www.rstudio.com/.

7271 Rúa, M.A. *et al.* (2016) 'Home-field advantage? evidence of local adaptation among plants, soil, and arbuscular

7272 mycorrhizal fungi through meta-analysis', *BMC Evolutionary Biology*, 16(1). Available at:

7273 https://doi.org/10.1186/s12862-016-0698-9.

7274 Ruane, L.G. and Donohue, K. (2007) 'Environmental effects on pollen–pistil compatibility between Phlox

7275 cuspidata and P. drummondii (Polemoniaceae): implications for hybridization dynamics', *American Journal of*

7276 *Botany*, 94(2), pp. 219–227. Available at: https://doi.org/10.3732/AJB.94.2.219.

7277 Rychel, S. *et al.* (2019) 'FLOWERING LOCUS T, GIGANTEA, SEPALLATA, and FRIGIDA homologs are candidate

7278 genes involved in white lupin (Lupinus albus L.) early flowering', *Molecular Breeding*, 39(3), pp. 1–17. Available

7279 at: https://doi.org/10.1007/S11032-019-0952-0/FIGURES/2.

7280 Sa, R. *et al.* (2021) 'Chromosome-Level Genome Assembly and Annotation of the Fiber Flax (Linum

7281 usitatissimum) Genome', *Frontiers in Genetics*, 12, p. 735690. Available at:

7282 https://doi.org/10.3389/FGENE.2021.735690/FULL.

7283 Saini, H.S., Sedgley, M. and Aspinall, D. (1983) 'Effect of Heat Stress During Floral Development on Pollen Tube

7284 Growth and Ovary Anatomy in Wheat (Triticum aestivum L.)', *Functional Plant Biology*, 10(2), pp. 137–144.

7285 Available at: https://doi.org/10.1071/PP9830137.

7286 Sallam, A.H. et al. (2021) "Cold conditioned: Discovery of novel alleles for low-temperature tolerance in the

7287 Vavilov Barley Collection," Frontiers in Plant Science, 12. Available at:

7288 https://doi.org/10.3389/fpls.2021.800284.

7289 Sand Korneliussen, T. *et al.* (2013) *Calculation of Tajima's D and other neutrality test statistics from low depth*

7290 *next-generation sequencing data*. Available at: http://www.biomedcentral.com/1471-2105/14/289.

7291 Sanderson, B.J. et al. (2020) "A targeted sequence capture array for phylogenetics and population genomics in

7292 the salicaceae," Applications in Plant Sciences, 8(10). Available at: https://doi.org/10.1002/aps3.11394.

7293 Santamaría, L. *et al.* (2003) 'PLANT PERFORMANCE ACROSS LATITUDE: THE ROLE OF PLASTICITY AND LOCAL

7294 ADAPTATION IN AN AQUATIC PLANT', *Ecology*, 84(9), pp. 2454–2461. Available at: https://doi.org/10.1890/02-

7295 0431.

7296 Sasaki, E. *et al.* (2017) 'The genetic architecture of the network underlying flowering time variation in

7297 Arabidopsis thaliana', *bioRxiv*, p. 175430. Available at: https://doi.org/10.1101/175430.

7298 Schena, M. *et al.* (1995) 'Quantitative monitoring of gene expression patterns with a complementary DNA

7299 microarray', *Science (New York, N.Y.)*, 270(5235), pp. 467–470. Available at:

7300 https://doi.org/10.1126/SCIENCE.270.5235.467.

7301 Schliep, K.P. (2011) 'phangorn: phylogenetic analysis in R', *Bioinformatics*, 27(4), pp. 592–593. Available at:

7302 https://doi.org/10.1093/BIOINFORMATICS/BTQ706.

7303 Schmittgen, T.D. and Livak, K.J. (2008) 'Analyzing real-time PCR data by the comparative CT method', *Nature*

7304 *Protocols 2008 3:6*, 3(6), pp. 1101–1108. Available at: https://doi.org/10.1038/nprot.2008.73.

7305 Schmitz, R.J. and Amasino, R.M. (2007) 'Vernalization: A model for investigating epigenetics and eukaryotic

7306 gene regulation in plants', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1769(5–6), pp.

7307 269–275. Available at: https://doi.org/10.1016/J.BBAEXP.2007.02.003.

7308   Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012a) 'NIH Image to ImageJ: 25 years of image analysis',

7309   *Nature Methods*, pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.

7310   Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012b) 'NIH Image to ImageJ: 25 years of image analysis',

7311   *Nature Methods 2012 9:7*, 9(7), pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.

7312   Schopf, J.W. *et al.* (1996) 'Activation of floral meristem identity genes in Arabidopsis', *Nature 1996 384:6604*,

7313   384(6604), pp. 59–62. Available at: https://doi.org/10.1038/384059a0.

7314   Schueler, S., Schlünzen, K.H. and Scholz, F. (2005) 'Viability and sunlight sensitivity of oak pollen and its

7315   implications for pollen-mediated gene flow', *Trees - Structure and Function*, 19(2), pp. 154–161. Available at:

7316   https://doi.org/10.1007/S00468-004-0376-1/FIGURES/4.

7317   Schuerger, A.C., Brown, C.S. and Stryjewski, E.C. (1997) *Anatomical Features of Pepper Plants (Capsicum*

7318   *annuum L.) Grown under Red Light-emitting Diodes Supplemented with Blue or Far-red Light, Sedgwick Rd.,*

7319   *Bldg*.

7320   Schwaegerle, K.E. and Levin, D.A. (1990) 'Quantitative genetics of seed size variation inPhlox', *Evolutionary*

7321   *Ecology 1990 4:2*, 4(2), pp. 143–148. Available at: https://doi.org/10.1007/BF02270911.

7322   Seppänen, M.M. et al. (2010) "Vernalization response of Phleum pratense and its relationships to stem

7323   lignification and floral transition," Annals of Botany, 106(5), pp. 697–707. Available at:

7324   https://doi.org/10.1093/aob/mcq174.

7325   Servedio, M.R. *et al.* (2014) 'Not Just a Theory—The Utility of Mathematical Models in Evolutionary Biology',

7326   *PLoS Biology*, 12(12), p. 1002017. Available at: https://doi.org/10.1371/JOURNAL.PBIO.1002017.

7327   Shachar-Hill, B. *et al.* (2013) 'Mercury-sensitive water channels as possible sensors of water potentials in

7328   pollen', *Journal of Experimental Botany*, 64(16), pp. 5195–5205. Available at:

7329   https://doi.org/10.1093/JXB/ERT311.

7330   SHAPIRO, S.S. and WILK, M.B. (1965) 'An analysis of variance test for normality (complete samples)',

7331   *Biometrika*, 52(3–4), pp. 591–611. Available at: https://doi.org/10.1093/BIOMET/52.3-4.591.

7332   Sharma, N. *et al.* (2020a) 'The molecular mechanism of vernalization in Arabidopsis and cereals: role of

7333   Flowering Locus C and its homologs', *Physiologia Plantarum*, 170(3), pp. 373–383. Available at:

7334   https://doi.org/10.1111/PPL.13163.

7335   Sharma, N. *et al.* (2020b) 'The molecular mechanism of vernalization in Arabidopsis and cereals: role of

7336   Flowering Locus C and its homologs', *Physiologia Plantarum*, 170(3), pp. 373–383. Available at:

7337   https://doi.org/10.1111/PPL.13163.

7338   Shavrukov, Y. *et al.* (2017) 'Early flowering as a drought escape mechanism in plants: How can it aid wheat

7339   production?', *Frontiers in Plant Science*, 8, p. 1950. Available at:

7340   https://doi.org/10.3389/FPLS.2017.01950/BIBTEX.

7341   Sheldon, C.C. *et al.* (2009) 'Mechanisms of gene repression by vernalization in Arabidopsis', *Plant Journal*,

7342   59(3), pp. 488–498. Available at: https://doi.org/10.1111/J.1365-313X.2009.03883.X.

7343   Sherry, R.A. and Lord, E.M. (1996) 'DEVELOPMENTAL STABILITY IN LEAVES OF CLARKIA TEMBLORIENSIS

7344   (ONAGRACEAE) AS RELATED TO POPULATION OUTCROSSING RATES AND HETEROZYGOSITY', *Evolution*, 50(1),

7345   pp. 80–91. Available at: https://doi.org/10.1111/J.1558-5646.1996.TB04474.X.

7346   Shindo, C. *et al.* (2005) 'Role of FRIGIDA and FLOWERING LOCUS C in Determining Variation in Flowering Time

7347   of Arabidopsis', *Plant Physiology*, 138(2), pp. 1163–1173. Available at: https://doi.org/10.1104/PP.105.061309.

7348   Shirsekar, G. *et al.* (2021) 'Fine-scale Population Structure of North American Arabidopsis thaliana Reveals

7349   Multiple Sources of Introduction from Across Eurasia', *bioRxiv*, p. 2021.01.22.427575. Available at:

7350   https://doi.org/10.1101/2021.01.22.427575.

7351   Singh, N. *et al.* (2013) 'Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population

7352   Structure of Indian Rice Varieties', *PLOS ONE*, 8(12), p. e84136. Available at:

7353   https://doi.org/10.1371/JOURNAL.PONE.0084136.

7354   Singh, N.B. *et al.* (1995) *RELATIONSHIP OF PLANT HEIGHT, DAYS TO FLOWERING AND MATURITY TO GRAIN*

7355   *YIELD IN SHORT-DURATION DETERMINATE PIGEONPEA* •. Available at: www.IndianJournals.com.

7356   Smýkal, P. *et al.* (2018) 'The Impact of Genetic Changes during Crop Domestication', *Agronomy 2018, Vol. 8,*

7357   *Page 119*, 8(7), p. 119. Available at: https://doi.org/10.3390/AGRONOMY8070119.

7358   Song, Y.H. *et al.* (2014) 'Distinct roles of FKF1, GIGANTEA, and ZEITLUPE proteins in the regulation of constans

7359   stability in Arabidopsis photoperiodic flowering', *Proceedings of the National Academy of Sciences of the*

7360   *United States of America*, 111(49), pp. 17672–17677. Available at:

7361   https://doi.org/10.1073/PNAS.1415375111/SUPPL_FILE/PNAS.201415375SI.PDF.

7362   Song, Y.H., Ito, S. and Imaizumi, T. (2013) 'Flowering time regulation: photoperiod- and temperature-sensing in

7363   leaves', *Trends in plant science*, 18(10), p. 575. Available at: https://doi.org/10.1016/J.TPLANTS.2013.05.003.

7364   Sork, V.L. (2018) 'Genomic Studies of Local Adaptation in Natural Plant Populations', *Journal of Heredity*,

7365   109(1), pp. 3–15. Available at: https://doi.org/10.1093/jhered/esx091.

7366   Soto-Cerda, B.J. *et al.* (2014) 'The potential of pale flax as a source of useful genetic variation for cultivated flax

7367   revealed through molecular diversity and association analyses', *Molecular Breeding*, 34(4), pp. 2091–2107.

7368   Available at: https://doi.org/10.1007/S11032-014-0165-5/FIGURES/3.

7369   South, A. (2011) *rworldmap: A New R package for Mapping Global Data*. Available at:

7370   http://www.un.org/millenniumgoals/bkgd.

7371   Speck, T. and Burgert, I. (2011) 'Plant Stems: Functional Design and Mechanics',

7372   *http://dx.doi.org/10.1146/annurev-matsci-062910-100425*, 41, pp. 169–193. Available at:

7373   https://doi.org/10.1146/ANNUREV-MATSCI-062910-100425.

7374   Sporemex, 2022. Spores and Pollens. [online] Sporomex.co.uk. Available at:

7375   <https://www.sporomex.co.uk/technology/spores-and-pollens> [Accessed 8 August 2022].

7376   Sreedharan, S.P., Kumar, A. and Giridhar, P. (2018) 'Primer design and amplification efficiencies are crucial for

7377   reliability of quantitative PCR studies of caffeine biosynthetic N-methyltransferases in coffee', *3 Biotech*, 8(11).

7378   Available at: https://doi.org/10.1007/s13205-018-1487-5.

7379   Starlabgroup.com. 2022. 96-Well PCR Plate, Skirted, Low Profile, White - STARLAB. [online] Available at:

7380   <https://www.starlabgroup.com/GB-en/product/96-well-pcr-plate-skirted-low-profile-white-e1403-

7381   5209.html> [Accessed 30 May 2022].

7382   Steel, R.G.D. and Torrie, J.H., 1960. Principles and procedures of statistics. Principles and procedures of

7383   statistics.

7384   Stinchcombe, J.R. *et al.* (2004) 'A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the

7385   flowering time gene FRIGIDA', *Proceedings of the National Academy of Sciences*, 101(13), pp. 4712–4717.

7386   Available at: https://doi.org/10.1073/PNAS.0306401101.

7387   Stöcklin, J., Kuss, P. and Pluess, A.R. (2009) 'Genetische Vielfalt, phänotypische Variabilität und lokale

7388   Adaptation in der Alpenlandschaft: Fallstudien mit vier alpinen Pflanzenarten', *Botanica Helvetica*, 119(2), pp.

7389   125–133. Available at: https://doi.org/10.1007/s00035-009-0065-1.

7390   Sultan, S.E. (2000) "Phenotypic plasticity for plant development, function and life history," Trends in Plant

7391   Science, 5(12), pp. 537–542. Available at: https://doi.org/10.1016/s1360-1385(00)01797-0.

7392   Sun, J. *et al.* (2019) 'Photoperiod sensitivity of Canadian flax cultivars and 5-azacytidine treated early flowering

7393   derivative lines', *BMC Plant Biology*, 19(1). Available at: https://doi.org/10.1186/s12870-019-1763-5.

7394   Sun, M. and Wong, K.C. (2001) 'Genetic structure of three orchid species with contrasting breeding systems

7395   using RAPD and allozyme markers', *American Journal of Botany*, 88(12), pp. 2180–2188. Available at:

7396   https://doi.org/10.2307/3558379.

7397   Sun, S. and Frelich, L.E. (2011) 'Flowering phenology and height growth pattern are associated with maximum

7398   plant height, relative growth rate and stem tissue mass density in herbaceous grassland species', *Journal of

7399   Ecology*, 99(4), pp. 991–1000. Available at: https://doi.org/10.1111/J.1365-2745.2011.01830.X.

7400   Sundaresan, V. (2005) 'Control of seed size in plants', *Proceedings of the National Academy of Sciences of the

7401   United States of America*, 102(50), p. 17887. Available at: https://doi.org/10.1073/PNAS.0509021102.

7402   Suter, L. *et al.* (2014) 'Gene Regulatory Variation Mediates Flowering Responses to Vernalization along an

7403   Altitudinal Gradient in Arabidopsis', *Plant Physiology*, 166(4), pp. 1928–1942. Available at:

7404   https://doi.org/10.1104/PP.114.247346.

7405   Sveinsson, S. *et al.* (2014a) 'Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (Linum)

7406   using transcriptomics', *Annals of Botany*, 113(5), pp. 753–761. Available at:

7407   https://doi.org/10.1093/AOB/MCT306.

7408 Szakiel, A., Pączkowski, C. and Henry, M. (2010) "Influence of environmental abiotic factors on the content of

7409 saponins in plants," Phytochemistry Reviews, 10(4), pp. 471–491. Available at:

7410 https://doi.org/10.1007/s11101-010-9177-x.

7411 Szpiech, Z.A. and Rosenberg, N.A. (2011) 'On the size distribution of private microsatellite alleles', *Theoretical*

7412 *Population Biology*, 80(2), pp. 100–113. doi:10.1016/j.tpb.2011.03.006.

7413 Tajima, F. (1989) 'Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism',

7414 *Genetics*, 123(3), p. 585. Available at: https://doi.org/10.1093/GENETICS/123.3.585.

7415 Tatham, S., 2022. PuTTY User Manual. [online] PuTTY. Available at:

7416 <https://the.earth.li/~sgtatham/putty/0.77/htmldoc/> [Accessed 21 September 2022].

7417 Teper-Bamnolker, P. and Samach, A. (2005) 'The Flowering Integrator FT Regulates SEPALLATA3 and FRUITFULL

7418 Accumulation in Arabidopsis Leaves', *The Plant Cell*, 17(10), p. 2661. Available at:

7419 https://doi.org/10.1105/TPC.105.035766.

7420 *The GNU Bourne-Again Shell* (no date). Available at: https://tiswww.case.edu/php/chet/bash/bashtop.html

7421 (Accessed: 5 May 2022).

7422 Thomashow, M.F. (1999) "Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms,"

7423 Annual Review of Plant Physiology and Plant Molecular Biology, 50(1), pp. 571–599. Available at:

7424 https://doi.org/10.1146/annurev.arplant.50.1.571.

7425 Tigano, A. and Friesen, V.L. (2016) 'Genomics of local adaptation with gene flow', *Molecular Ecology*. Blackwell

7426 Publishing Ltd, pp. 2144–2164. Available at: https://doi.org/10.1111/mec.13606.

7427 Toräng, P. *et al.* (2015) 'Large-scale adaptive differentiation in the alpine perennial herb Arabis alpina', *The*

7428 *New phytologist*, 206(1), pp. 459–470. Available at: https://doi.org/10.1111/NPH.13176.

7429 Tork, D.G. *et al.* (2022) 'Ideotype Selection of Perennial Flax (Linum spp.) for Herbaceous Plant Habit Traits',

7430 *Agronomy*, 12(12). Available at: https://doi.org/10.3390/agronomy12123127.

7431 Trelease, W., 1931. Winter botany. New York: Dover Publications.

7432 Tukey, J., 1949. Comparing Individual Means in the Analysis of Variance. Biometrics, 5(2), p.99.

7433 Unacademy, 2022. Draw a Labelled Diagram of the Pollen Grain. [online] Unacademy.com. Available at:

7434 <https://unacademy.com/content/question-answer/biology/draw-a-labelled-diagram-of-the-pollen-grain/>

7435 [Accessed 8 August 2022].

7436 Uysal, H. *et al.* (2010) 'Genetic diversity of cultivated flax (Linum usitatissimum L.) and its wild progenitor pale

7437 flax (Linum bienne Mill.) as revealed by ISSR markers', *Genetic Resources and Crop Evolution*, 57(7), pp. 1109–

7438 1119. Available at: https://doi.org/10.1007/S10722-010-9551-Y/FIGURES/4.

7439   Uysal, H. *et al.* (2012) 'Variation in phenotypic characters of pale flax (Linum bienne Mill.) from Turkey',

7440   *Genetic Resources and Crop Evolution*, 59(1), pp. 19–30. Available at: https://doi.org/10.1007/S10722-011-

7441   9663-Z/FIGURES/10.

7442   Valentim, F.L. *et al.* (2015a) 'A quantitative and dynamic model of the Arabidopsis flowering time gene

7443   regulatory network', *PloS one*, 10(2). Available at: https://doi.org/10.1371/JOURNAL.PONE.0116973.

7444   Vandesompele, J. *et al.* (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric

7445   averaging of multiple internal control genes', *Genome biology*, 3(7). Available at: https://doi.org/10.1186/GB-

7446   2002-3-7-RESEARCH0034.

7447   Venables, W.N. and Ripley, B.D. (2002a) 'Random and Mixed Effects', pp. 271–300. Available at:

7448   https://doi.org/10.1007/978-0-387-21706-2_10.

7449   Verdu, C.F. *et al.* (2016) 'Dealing with paralogy in RADseq data: in silico detection and single nucleotide

7450   polymorphism validation in Robinia pseudoacacia L.', *Ecology and Evolution*, 6(20), pp. 7323–7333. Available

7451   at: https://doi.org/10.1002/ECE3.2466.

7452   Vieira, M.L.C. *et al.* (2016) 'Microsatellite markers: what they mean and why they are so useful', *Genetics and*

7453   *Molecular Biology*, 39(3), p. 312. Available at: https://doi.org/10.1590/1678-4685-GMB-2016-0027.

7454   Villanueva, R.A.M. and Chen, Z.J. (2019) 'ggplot2: Elegant Graphics for Data Analysis (2nd ed.)',

7455   *https://doi.org/10.1080/15366367.2019.1565254*, 17(3), pp. 160–167. Available at:

7456   https://doi.org/10.1080/15366367.2019.1565254.

7457   Vuletin Selak, G. *et al.* (2013) 'The effect of temperature and genotype on pollen performance in olive (Olea

7458   europaea L.)', *Scientia Horticulturae*, 156, pp. 38–46. Available at:

7459   https://doi.org/10.1016/J.SCIENTA.2013.03.029.

7460   Walker, J.M. (1969) 'One-Degree Increments in Soil Temperatures Affect Maize Seedling Behavior', *Soil Science*

7461   *Society of America Journal*, 33(5), pp. 729–736. Available at:

7462   https://doi.org/10.2136/SSSAJ1969.03615995003300050031X.

7463   Walters, D., Keil, D. and Walters, B., 1996. Vascular plant taxonomy. Dubuque, Iowa: Kendall/Hunt, p.598.

7464   Wang, J. *et al.* (2018) 'A major locus controls local adaptation and adaptive life history variation in a perennial

7465   plant', *Genome Biology*, 19(1). Available at: https://doi.org/10.1186/s13059-018-1444-y.

7466   Wang, Z. *et al.* (2012) 'The genome of flax (Linum usitatissimum) assembled de novo from short shotgun

7467   sequence reads', *The Plant journal : for cell and molecular biology*, 72(3), pp. 461–473. Available at:

7468   https://doi.org/10.1111/J.1365-313X.2012.05093.X.

7469   Watanabe, S. *et al.* (2011) 'A Map-Based Cloning Strategy Employing a Residual Heterozygous Line Reveals that

7470   the GIGANTEA Gene Is Involved in Soybean Maturity and Flowering', *Genetics*, 188(2), pp. 395–407. Available

7471   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.

7538   Zinn, K.E., Tunc-Ozdemir, M. and Harper, J.F. (2010) 'Temperature stress and plant sexual reproduction:

7539   uncovering the weakest links', *Journal of Experimental Botany*, 61(7), p. 1959. Available at:

7540   https://doi.org/10.1093/JXB/ERQ053.

7541   Zipper, H. *et al.* (2004) 'Investigations on DNA intercalation and surface binding by SYBR Green I, its structure

7542   determination and methodological implications', *Nucleic Acids Research*, 32(12), p. e103. Available at:

7543   https://doi.org/10.1093/NAR/GNH101.

7544   Zohary, D. and Hopf, M. (2000) 'Domestication of plants in the Old World: the origin and spread of cultivated

7545   plants in West Asia, Europe and the Nile Valley.', *Domestication of plants in the Old World: the origin and*

7546   *spread of cultivated plants in West Asia, Europe and the Nile Valley.* [Preprint], (Ed.3).

7547   Zonia, L. and Munnik, T. (2004) 'Osmotically Induced Cell Swelling versus Cell Shrinking Elicits Specific Changes

7548   in Phospholipid Signals in Tobacco Pollen Tubes', *Plant Physiology*, 134(2), p. 813. Available at:

7549   https://doi.org/10.1104/PP.103.029454.

7550

7551

7552