THE GENETICS OF GREEN LENTIL (*LENS CULINARIS* MEDIK.) SEED COAT COLOUR QUALITY AND RETENTION

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

By

Matt D. Remenda

PERMISSION TO USE

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

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

OR

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

ABSTRACT

Green lentil is typically consumed whole; consequently, the colour of the seed coat is an important factor in determining its market value. However, green seed coat colour is not stable and tends to deteriorate towards dark brown over time. This reduces the desirability of the lentil seed and results in a lower market value. The purpose of this project was to generate information to assist breeders in developing strategies for improving seed coat colour retention in green lentil. The population used in this study (LR-06) consists of 160 recombinant inbred lentil lines (RILs) that were derived from a cross between two breeding lines: 1294m-23 and 1048-8R. The resulting progeny were grown in three replications, in two locations, during the summers of 2019, 2021, and 2022. Approximately 200 individual seeds from each plot were imaged using a high-throughput imaging device (BELT). Commission Internationale de l'Eclairage (CIE) L*a*b* is a three-dimensional colour space where L* is a scale of dark to light; a* is a scale of green to red; and b* is a scale of blue to yellow. These scores were extracted from images using BELT's accompanying software: PhenoSEED. The seeds were imaged at ≈ 6 months post-harvest for seed grown in 2022; ≈ 6 months, ≈ 12 months, and ≈ 18 months for seed harvested in 2021; and 18 months post-harvest for seeds grown in 2019. Mixed linear models were used to evaluate the predictability of seed coat colour quality (quantified using CIE L*a*b* scores), following aging, based on the initial screenings. The results of this study provide evidence that lentil seed coat colour quality, following aging, is predictable based on these initial screenings. Genotyping was performed using a legume single nucleotide polymorphism (SNP) chip in Australia and the genotypic data, combined with CIE L*a*b* scores, were used to perform quantitative trait locus (QTL) analysis. The results of QTL analysis provide evidence that seed coat colour quality is genetic and specific regions of the genome can be targeted during selection.

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my deepest thanks to my advisor Dr. Kirstin Bett for her instruction, guidance, trust, support, patience, and compassion throughout the years. I would like to thank my committee members Dr. Scott Noble and Dr. Ana Vargas and my external examiner Dr. Peter Pauls for all their advice, guidance, and patience. In addition, special mention must be given to Keith Halcro, Derek Wright, Robert Stonehouse, Akiko Tomita, Laura Jardine, Sarahi Viscara, Larissa Ramsay, the rest of the EVOLVES team, Brent Barlow, and the Pulse Crop field lab team. I would like to thank Gillian Murza, Scott Halpin, and Ellen Misfeld for keeping me employed as a teaching assistant. Finally, I would like to extend my thanks to my loving parents (Jayne and Andy), my band (Nathan, Mike, and Cam), Dom, my friends, family, and fellow grad students for their love, support, and friendship over the years.

I would also like to thank EVOLVES, the Rene Vandeveld Postgraduate Scholarship, the Peter & Dora Kushner Bursary in Agriculture, the Alexander and Jean Auckland Agricultural Bursary, the C. Paul W. and Marianne M. Ziehlke Postgraduate Bursary in Agriculture, and Gopalan Selvaraj Innovation Scholarship for Graduate Studies in Plant Sciences for their financial support.

DEDICATION

To my favourite little “biologist to be,” Laoghaire. Your insatiable curiosity and enthusiasm for life inspires me to see the natural world with a renewed awe and wonderment everyday. I don’t know where you are going, or what you will do, but I know it will be amazing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
BELT	High-throughput imaging device
CDC	Crop Development Centre
CIE	Commission Internationale de l'Eclairage
CTAB	Cetyltrimethylammonium bromide
DNA	Deoxyribonucleic acid
DTE	Days to Emergence
DTF	Days to flowering
DTM	Days to maturity
ED	Euclidian distance
F	Favourable allele
ICIM	Inclusive composite interval mapping
IM	Interval mapping
LG	Linkage group
LOD	Limit of detection
MAS	Marker assisted selection
POD	Period of deterioration
QTL	Quantitative trait loci
RIL	Recombinant inbred line
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
SR	Significant region
U	Unfavourable allele

1 INTRODUCTION

Green lentil (*Lens culinaris* Medik.) is typically consumed whole; consequently, the colour of the seed coat is an important factor in determining its market value. However, green seed coat colour is not stable and tends to deteriorate towards dark brown over time (Nozzolillo and De Bezada 1984). This reduces the desirability of the lentil seed and results in a lower market value. The Canadian Grain Commission (2021) defines the highest colour quality of lentils as having "good natural colour – lentils that are sound, well matured and have good natural colour," while the lowest quality for lentils is defined as "poor colour – lentils that do not meet the definition of fair colour but are without severely adhered soil or are severely discoloured (dark brown)."

Some compounds that have been linked to seed coat darkening in lentil may be colourless (Mirali et al. 2016). This means that phenotyping seeds for colour, closely following harvest, may not predict how much the colour will change over time. If this were the case, phenotyping would require biochemical profiling of lines to detect compounds related to colour deterioration or allowing seeds to age and measuring their colour quality at multiple time points. The reliability of biochemical profiling, for the purpose of making selections, assumes that the pertinent compounds related to colour deterioration have been identified and can be profiled. However, colour deterioration of lentil seed coats can continue to deteriorate for 12 to 18 months, therefore phenotyping for age-related colour deterioration is time consuming. An alternative to biochemical profiling or waiting for seed colour to deteriorate is to perform marker assisted selection (MAS), where alleles are associated with the favourability of traits and selections are made based on the presence of these favourable alleles. This method allows for avoiding biochemical profiling and making selections earlier than 12 to 18 months following aging.

This work involved using high-throughput and precision phenotyping to determine the rate of colour deterioration in green lentil seed coats and performing quantitative trait loci (QTL) mapping for the purpose of better understanding the genetics related to seed coat colour quality following aging. Using single nucleotide polymorphism (SNP) markers, this work provides knowledge and tools for breeders to perform MAS for colour quality of aged green lentil seed coats. This study expands on previous work, using the same lentil recombinant inbred line (RIL)

population (LR-06), where seed coat colour desirability (as defined by market preference) was determined to be highly heritable (Davey 2007), and the results provided preliminary evidence that seed coat colour quality is heritable (Bett and Vandenberg 2018).

1.1 Hypotheses

1. Objective colour scoring using Commission Internationale de l'Eclairage (CIE) L*a*b* scores will reveal that L* scores decrease and a* scores increase as lentil seeds age.

2. The L*a*b* scores of "Fresh" (6-months post-harvest) seed, will be predictive of the colour quality of "Old" (18-months post-harvest) seed.

3. The change of L* scores and a* scores, and the Euclidian distances between the colour scores of fresh and aged seeds from different lentil lines will demonstrate that pre-disposition to green seed coat colour deterioration is multigenic and controlled by distinct regions of the genome.

1.2 Objectives

The primary objectives of this work were to: 1) characterize the intensity of green colour in freshly harvested and aged seed of LR-06 lentil RILs and their parents; 2) determine colour changes over time using high-throughput and precision phenotyping to be completed using BELT; and 3) perform QTL mapping, using single nucleotide polymorphism (SNP) markers, to identify regions on the genome related to seed coat colour quality and deterioration.

2 LITERATURE REVIEW

2.1 Lentils

Lentil is a cool season pulse crop that is thought to have originated from the Fertile Crescent and spread into South Asia, North Africa, and Europe. The earliest evidence for the domestication of lentil dates to 10,000 – 8,000 years ago, within the Fertile Crescent region (Liber et al. 2021); however, there is evidence of its consumption as early as 19,000 years ago in the modern-day region of Israel (Kislev et al. 1992). Therefore, lentil is a very old crop, placing its domestication contemporary with wheat and barley.

2.2 Lentil Production

Lentil production has increased from 2.9 million tonnes in 1999 to 5.7 million tonnes in 2019 (FAOSTAT 2021). Saskatchewan, having produced 2.5 million tonnes in 2020, was the world's largest producer that year (Government of Saskatchewan 2021; FAOSTAT 2021). Red lentils tend to dominate production in Saskatchewan; however, green lentils still accounted for 29.2% (729,500 tonnes) of the province's total production in 2020 (Government of Saskatchewan 2021).

Seed coat browning can result in the extreme devaluation lentil seed. The difference in valuation between high quality (No 1.) and lower quality (No 3.) green lentils varies between years but it has been as high as 50% (Saskatchewan Crop Insurance Corporation 2023). Seed coat browning has been studied in pinto bean (Junk-Knievel et al. 2008) and this characteristic has been found to be highly heritable. However, efforts to determine genotypes and develop markers for the selection of green seed coat colour retention in lentil have been preliminary and restricted to the development of amplified fragment length polymorphism (AFLP) markers (Bett and Vandenberg 2018).

2.3 Genetics of Seed Coat Colour

Lentil seed coats can have a variety of colours and patterns. Vandenberg and Slinkard (1990) determined that phenotypes for brown, grey, tan, and green seed coats are controlled by four alleles across two independent loci. A dominant allele (*Ggc*) at one locus resulted in a grey phenotype, whereas a dominant allele (*Tgc*) at the other locus resulted in a tan phenotype. Brown phenotypes were the result of the double dominant alleles and green phenotypes were the result of the double homozygous recessive condition. Furthermore, Vaillancourt and Slinkard (1992)

and Emami and Sharma (2000) found evidence of a third locus, given the designation of *Blt* by Emami and Sharma (2000), for a black seed coat which occurs because of a dominant allele.

On top of these three major seed coat colour loci, Davey (2007) observed that the degree of "desirability" of green seed coat colour is genetic and quantitatively controlled in green lentil. Seeds with the highest desirability were identified as having a high level of greenness, while seeds with the lowest level of desirability were identified as having high levels of oxidation. Using an Acurum® colour sorter, calibrated to 20 qualitative classifications (based on market preferences for seed coat colour), the breeding line 1294m-23 was determined to have the highest colour quality.

2.4 Phenolic Compounds

Phenolic compounds have a strong relationship to browning in pinto beans (Beninger et al. 2005; Duwadi et al. 2018) and lentil (De Bezada 1981; Nozzolillo and De Bezada 1984; Mirali et al. 2016). They are secondary metabolites that serve many purposes in plants, including defense and structural roles (Erskine et al. 2009). Furthermore, some phenolic compounds in food have both positive and negative implications towards human health. Some phenolic compounds have been widely reported to negatively impact the bioavailability of iron in food (Gillooly et al. 1982; Hurrell et al. 1999; Sandberg 2002) by acting as chelators (Moran et al. 1996; Johnson et al. 2013; Elessawy et al. 2021). However, some phenolic compounds tend to correlate strongly with antioxidant properties in pulses (Beninger and Hosfield 2003; Elessawy et al. 2021).

Singleton (1995) and De Bezada (1981) suggested that browning in seed is caused by the polymerization of monomeric and dimeric phenolics to insoluble phenolics which bind to the cell wall and oxidize. In a study examining green lentil lines containing phenolic compounds (cv. Laird) and lines with no phenolic compounds (PI 345635), De Bezada (1981) reported that green lentil seed coat colour deteriorates rapidly when exposed to higher temperatures, light, and humidity when phenolic monomers are present. Suspending seed in a) 5°C, 0% humidity, and dark conditions or b) anoxic conditions can preserve seed coat colour for three or eleven months (respectively) following harvest. Furthermore, De Bezada determined that phenolic monomers are more abundant in fresh, green seed than aged, brown seed and that using water to remove monomeric phenolics can slow the rate at which seed coat colour deterioration occurs. Finally, cv. Laird darkened at a much faster rate than PI 345635. The results of this study are consistent with the mechanisms proposed by Singleton (1995) and De Bezada (1981); furthermore, the

differences in darkening rate between the lines suggests the possibility that seed coat colour deterioration is genetic in green lentil.

Phenolic profiles vary considerably between seed coats and cotyledons in lentil (DellaValle et al. 2013). The primary phenolics contained in cotyledons are non-flavonoids (mostly hydroxycinnamic acids and hydroxybenzoic acids), whereas the vast majority of phenolics in seed coats are flavonoids (mostly procyanidin dimers and catechins) (Dueñas et al. 2002). The latter group's relationship to browning has been studied by Duwadi et al. (2018) in pinto beans. Transcriptomes related to polymeric phenolic pathways (specifically the anthocyanidin pathway) are upregulated in a rapid darkening pinto cultivar (CDC Pintium) in contrast to a slow darkening pinto line (1533-15), and it is possible that genes related to these proteins (along with the other steps related to the proanthocyanidin pathway) are similarly involved in green lentil seed coat deterioration.

2.5 Quantifying Colour

2.5.1 Colour Interpretation

Colour is not an intrinsic property of an object but is the result of perception by an individual (Wu and Sun 2013). When light in the visible spectrum (400-700nm) reflects off an object into the human eye, it stimulates photo receptors (rods and cones), which send information to the brain to interpret (Sherwood et al. 2013). Rods are photoreceptors responsible for collecting information on the intensity of the light (brightness), whereas cones are photoreceptors responsible for collecting information on the wavelengths of the light (Sherwood et al. 2013). There are three types of cones within the human eye, which are maximally stimulated by different wavelengths of light: short cones (420-440nm – blue), middle cones (530-540nm – green), and long cones (560-580nm – red) (Hunt 1995). The interpretation of colour relies on the proportion of cone cell types stimulated. For example: the stimulation of 0.5 long cones, 0.5 short cones, and 0.0 middle cones would correspond to the interpretation of an approximation to the colour purple (resulting from the mixture of red and blue cones) (Sherwood et al. 2013). Identifying colours can be reasonably accomplished when they are classified in broad terms (red versus blue); however, humans have difficulty in distinguishing small differences and intermediary colours from each other (Emery et al. 2017). The abundance of each type of cone differs between individuals (Emery et al. 2017) and with age (Kilbride et al. 1986). In addition,

colour perception has been found to be influenced heavily by factors independent of proportional conical stimulation including psychology, context, socio-culture, and central processing in the brain – all of which (especially the latter) are poorly understood (Emery et al. 2017). Consequently, this presents challenges in reliability and replicability between individuals when evaluating colours based on observation.

2.5.2 Colour Spaces and CIE L*a*b* Scores

Because of the challenges associated with standardizing colour, there are many different scales used for depicting and interpreting colour. These scales are typically three-dimensional and fall into three different categories: a) Hardware oriented spaces – device dependant colour spaces used for image acquisition, storage, and display; b) Human oriented spaces – highly subjective methods favoured for their functionality and informality which are informed by human perception; and c) Instrumental spaces – device independent colour spaces used for image acquisition, and storage (Wu and Sun 2013). The Commission Internationale de l'Eclairage's (CIE) L*a*b system is an instrumental, three-dimensional colour space with three scales: L*, a*, and b*. L* is a scale of luminance or "lightness" (0-100); a* is a scale of green to red (negative to positive); and b* is a scale of blue to yellow (negative to positive) (Ly et al. 2020). The CIE L*a*b* colour space was developed as a method to approximate human perception of colour using standards set by CIE which form a uniform gradient and are device independent. The uniformity of gradient means that the Euclidean distance (ED) between two points (colours) on the CIE L*a*b* scale is proportional to what the standard observer would perceive as a difference in colours (this is not the case with other popular scales, including RGB and CIE XYZ) (Wu and Sun 2020). Combined, these factors are the reason why the CIE L*a*b* colour space is the favoured approach for evaluating colour in food products (Markovic et al. 2013).

2.5.3 Quantifying Colour Deterioration

Efforts to quantify colour deterioration in pulse products over time (using CIE L*a*b* scores) have been limited. However, a study by Jackson et al. (2021), examining how seed coats protect lentil cotyledons when exposed to different types of light and quantified changes in colour by using ΔL^*, Δa^*, and Δb^* (where $\Delta X = X_{t1} - X_{t0}$). Furthermore, the Euclidian distance ($\Delta E = [\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}]^{1/2}$) between L*, a*, and b* scores across two time points was used to

quantify the colour changes. This approach has been used commonly in other food products including meat (Larraín et al. 2008) and dairy products (Alpaslan et al. 2020).

2.6 Image Acquisition and Processing

BELT is a high-throughput phenotyping device developed by Halcro et al. (2020) at the University of Saskatchewan's College of Engineering. The operation of the machine involves loading seeds into a vibratory feeder, which passes seeds onto a conveyor, and under one of two cameras. The cameras can capture an image of the seed from two angles: above and the side (using a prism). The images are uploaded to BELT's accompanying software (PhenoSEED) where CIE L*a*b* scores are extracted for every seed. Consequently, the use of BELT and PhenoSEED allow for the processing of many more seed samples and much more precision when evaluating colour than can be accomplished through visual scoring by a researcher.

2.7 Quantitative Trait Locus (QTL) Analysis

2.7.1 Quantitative Traits

"Genetic architecture" is a broad term that refers to the relationship between a genotype and how it is expressed in the phenotype (Hansen 2006) (Futuyma 2013). The description of the genetic architecture of a trait can include references to the dominance, pleiotropic, or epistatic effects of genes or their alleles. Furthermore, the genetic architecture of a trait can refer to it being monogenic, polygenic, qualitative, or quantitative (Futuyma 2013).

Qualitative traits are traits that are controlled by one or a few genes, where the phenotypes of progeny form distinct categories (Xu 2022), such as pea seed shape (wrinkled versus round) or colour (green versus yellow) (Mendel 1866). This is the case for seed coat colour classes (green, grey, tan, brown, and black) with lentil (Vandenberg and Slinkard 1990) (Vaillancourt and Slinkard 1992) (Emami and Sharma 2000). Quantitative traits, on the other hand, tend to show normal distributions of phenotypic values that are measured as either integer or continuous values (Fisher 1918) that are in different regions across the genome; These regions are called quantitative trait loci (QTL). While seed coat colour classes are qualitative in lentil, the colour quality, particularly under the classification of green seed coat colour, is quantitative (Davey 2007).

2.7.2 Interval Mapping (IM)

There are three related (but distinct) terms that pertain to QTL analysis: 1) Genes, 2) QTL, and 3) Markers. Genes are generally thought of as being the DNA sequences on chromosomes that transcribe RNA; however, the term is ambiguous, and the definitions can vary based on context. A locus is the region in which a gene is located. Finally, markers are known sequences of DNA on the genome that may or may not transcribe RNA. For bi-parental populations, the process of QTL mapping involves genotyping many markers across the genomes of the population of individuals, with the expectation that there will be differences of alleles at these locations based on which parent's allele was inherited during fertilization. These markers may not fall with a gene, however, exploiting the idea that regions on the genome that are close to each other are likely to co-segregate in progeny, some markers may to be linked to genes that are important to the trait of interest (QTL). Consequently, QTL mapping (alone) is not capable of determining the exact location of a QTL or gene. However, it can provide evidence for the approximate location of genes involved in expression of a trait within an interval.

Interval mapping (IM), first proposed by Lander and Botstein (1989), involves relating quantitative phenotypes to regions on the genome that are flanked by two markers, using permutation tests. The idea behind this is that regions that are strongly related to the phenotypes are regions where QTL for the relevant traits of interest are located. However, the usefulness of identifying these regions is highly dependent on how far apart the markers flanking the region are located. If the markers flanking the region are far apart, it decreases the chance of either of them being strongly linked to or falling within the gene of interest, thus rendering MAS less effective.

Expanding on the earlier methods proposed by Lander and Botstein (1989), Inclusive Composite Interval Mapping (ICIM) has become a preferred and widely used method for performing QTL analysis. Using this method, all markers can be considered simultaneously, as opposed to independently as with IM (Li et al. 2010). This results in a lower rate of false detection of QTL but was only made possible using modern computing. Furthermore, the availability of free and user-friendly software that uses ICIM (e.g., ICIMapping), has popularized the use of this method.

3 RESEARCH STUDY

3.1 Experimental Approach

The seeds used in this study were harvested from plants grown 2019, 2021, and 2022 in two locations per year. Individual seeds were imaged at different ages using a high-throughput imaging device (BELT) and the CIE L*a*b* scores were extracted from the images using PhenoSEED (Halcro et al. 2020). Linear modeling was used to characterize the changes in L*, a*, and b* scores over time and determine the relative contributions of each channel to (overall) colour deterioration. Mixed effect linear modeling was used to determine if post-harvest seed colour predicts seed colour following aging. QTL mapping was performed to identify regions on the genome that are associated with L*, a*, and b* scores and alleles associated with desirable phenotypes (higher L*, lower a*, and lower b*) were identified.

3.1.1 Plant Material and Experimental Design

The population used in this study (LR-06) consists of 160 recombinant inbred lentil lines (RILs) that were derived from a cross between two breeding lines: 1294m-23 (maternal and subjectively the desirable parent) and 1048-8R (paternal and subjectively the undesirable parent) (Davey 2007). The resulting progeny were taken to F_7 by single seed decent and then seed from individual plants were bulked. Seeds were sown in meter-long plots, consisting of three rows in 2019, 2021, and 2022. The plots were arranged in a randomized complete block design, consisting of three blocks at two locations in Saskatchewan (with field designations of “SPG” and “Sutherland” in 2019; and field designations of “Rosthern” and “Sutherland” in 2021 and 2022). Planting and harvest dates (where known) are included in (Table 3.1). Each site-year was harvested within a single day in 2021 and 2022, excluding Sutherland 2021, which was harvested over two days. All plots were harvested in bulk, threshed, cleaned, and transferred to paper envelopes (henceforth referred to as “samples”). These samples were stored in the dark, at room temperature, between screenings.

Table 3.1. The planting, harvest dates, and the number of days between these dates, expressed as days to harvest (DTH), of LR-06 grown in 2019, 2021, and 2022.

Year	Location	Plots	Planting Date	Harvest Date	DTH
2019	Sutherland	All	2019-04-26	Unknown	-
	SPG	All	2019-05-01	Unknown	-
2021	Sutherland	1002 - 1295	2021-05-04	2021-08-16	104
	Sutherland	1297 - 1554	2021-05-04	2021-08-17	105
	Rosthern	All	2021-05-11	2021-08-19	100
2022	Sutherland	All	2022-05-11	2022-08-16	97
	Rosthern	All	2022-05-16	2022-08-23	99

3.1.2 Phenotyping

3.1.2.1 Digital Imagery

Seeds from each sample were imaged using BELT, a high-throughput seed phenotyping device developed by Halcro et al. (2020). Poor-quality photos (photos containing overlapping seeds, cracked or broken seeds, and an abundance of debris) were discarded such that ≈ 200 images were saved for each sample. Imaging was conducted at three different time points where "Fresh" (≈ 6 months post-harvest), "Mid" (≈ 12 months-post harvest) and "Old" (≈ 18 months following harvest). For seeds grown in 2019, imaging only took place when seed was Old. For the samples grown in 2021, seeds were imaged at all three time-points ("Fresh," "Mid," and "Old"); and seed grown in 2022, was only imaged when "Fresh." The batch of samples that was imaged from a single year, location, and at a specific age will be referred to as an imaging cycle (e.g. 2019-Sutherland-Old). To mitigate deterioration while imaging site-years during an imaging cycle, the samples were temporarily stored within a freezer. Samples harvested in 2021 were moved back to room temperature between imaging the three different imaging cycles. PhenoSEED software (Halcro et al. 2020) was used to automatically extract CIE L*a*b* scores and generated CSV files for the output data for 958,297 single-seed images.

3.1.2.2 Assessment of Deterioration

Deterioration is defined, within this work, as the changes in L*, a*, and b* scores (ΔL*, Δa*, Δb*), and the Euclidian distances (ED) over the periods of deterioration (POD). The PODs are defined as the times between imaging cycles on samples, where POD1 = "Fresh" to "Mid;" POD2 = "Mid" to "Old;" and POD3 = "Fresh" to "Old." The colour changes in each channel (ΔL*, Δa*, Δb*) were calculated by subtracting the initial value from the subsequent value. ED is

the Euclidian distance between two points in a three-dimensional space calculated from the following formula:

$$ED = [\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}]^{1/2} \quad (3.1)$$

Imaging cycles did not occur instantaneously. This means that the first sample to be imaged could have deteriorated less than samples that were imaged later within an imaging cycle. For this reason, the lm function from base R v4.2.3 (R Core Team 2021) was used to determine if there was significant deterioration during the imaging cycles. Models were fit using the mean L*, a*, and b* scores for each sample as the dependent variables and days to imagery as the explanatory variable in the following models:

$lm(L^*\mu_{sample} \sim$ Days to imagery)

$lm(a^*\mu_{sample} \sim$ Days to imagery)

$lm(b^*\mu_{sample} \sim$ Days to imagery)

Days to imagery was calculated based on the days between the first day of imaging in an imaging cycle and the day in which the imaging of a sample occurred (equation 3.2).

$$\text{Days to Imagery} = Day_x - Day_0 \quad (3.2)$$

3.1.2.3 Phenology

The days to flowering (DTF: Days after planting when 10% of plants in a plot have at least one open flower) and days to maturity (DTM: Days after planting when 10% of plants have 50% of their pods matured) were recorded for both locations in 2021 and 2022. In addition, the days to emergence (DTE: Days after planting that 10% of plants have emerged through the soil surface) were recorded for the Sutherland 2021 and 2022 years. All the phenology data are available on KnowPulse (https://knowpulse.usask.ca/experiment/phenomics/Phenology-Green-Seed-Coat-Colour-Retention). The cor.test function from base R was used to determine if there were any relationships between DTF and DTM and L*a*b* scores.

3.1.3 Genotyping

Using seed from SPG 2019, single plants of the parents and all lines were planted in horticultural trays with Sunshine Mix and grown under controlled environment conditions for 18 days. Leaf tissue was collected and Cetyltrimethylammonium bromide (CTAB) DNA extractions (Stonehouse et al. 2022) were performed on each of the lines. Genotyping was conducted by using a legume single nucleotide polymorphism (SNP) chip in Australia (Gebremedhin et al.

2024); data are available in KnowPulse (https://knowpulse.usask.ca/genotypes/download). A linkage map, consisting of 1069 SNP markers, was assembled by Robert Stonehouse using IciMapping (Meng et al. 2015) and the marker order was compared to the CDC Redberry (Lcu.2RBY; Ramsay et al. 2021) genome assembly to name linkage groups according to their lentil chromosome numbers. The relative positions of markers from this map can be found at https://knowpulse.usask.ca/study/Green-Seed-Coat-Colour-Retention and the full map will be made available at https://knowpulse.usask.ca/search/genetic-maps.

3.1.4 Statistical Approaches

3.1.4.1 Z-Score Filtering

To eliminate images of outliers that may have resulted from underdeveloped seeds and possible contaminants, Z-score filtering was performed on all images. Of the ≈ 200 seeds recovered from every plot, Z-scores (for each of volume, L*, a*, and b*) were calculated using equation 3.3 (where x = L*, a*, or b* score assigned to an individual seed; μ = the plot mean of L*, a*, or b*; and σ = standard deviation within the plot) (equation 3.3). Images of seeds with Z-scores that did not fall within the threshold of +/-3 were removed from further analysis. Furthermore, if z-score filtering resulted in a sample with fewer than 100 seeds, that sample was removed from further analysis. Through z-score filtering, removal of fills (lines planted that do not belong to the LR-06 population) and removal of lines labeled as the parents (due to suspected contamination), 900,770 images were carried forward for further analysis. The number of saved images, and the dates where imagery occurred can be found in Appendix A and accessed at https://knowpulse.usask.ca/experiment/phenomics/BELT-Green-Seed-Coat-Colour-Retention.

$$Z = (x - \mu)/\sigma \tag{3.3}$$

3.1.4.2 ANOVA for Replication/Location Effects

Analysis of Variance (ANOVA) and Tukey's honestly significant test (Tukey HSD) were performed using the "aov" and "tukeyHSD" functions in base R v4.2.3 (R Core Team 2021) to determine if significant differences in L*, a*, and b* scores existed between replications and site-years within (and across) time points. The Tukey HSD tests were performed on the imaging cycles (i.e. 2021-Rosthern-Fresh), to identify which replications differed from each other.

The broad sense heritability (H^2) was estimated using the "gen.var" function from the "variability" package in R (Popat et al. 2020), for imaging cycles with the same age ("Fresh," "Mid," or "Old").

3.1.4.3 Linear Modeling

Using the R packages "lme4" (Bates et al. 2015) and "lmertest" (Kuznetsova et al. 2017), the following linear mixed model was used to evaluate the relationships of DTM and L*a*b* scores (for plots harvested in 2021 and 2022), and between "Fresh" and "Old" L*a*b* scores for plots harvested in 2021: (Y ~ X + (1|Location/Replication)), where Y is the dependent variable, X is the fixed effect, and Location and Replication are both fit as random effects.

The "lm" function from base R v4.2.3 (R Core Team 2021) was used to determine how much ΔL*, Δa*, and Δb* contribute to ED. For this purpose, the following models were fit for all three PODs:

$lm(ED_{PODX} \sim \Delta L^*_{PODX} + \Delta a^*_{PODX} + \Delta b^*_{PODX})$

$lm(ED_{PODX} \sim \Delta L^*_{PODX})$

$lm(ED_{PODX} \sim \Delta a^*_{PODX})$

$lm(ED_{PODX} \sim \Delta b^*_{PODX})$

$lm(ED_{PODX} \sim \Delta L^*_{PODX} + \Delta a^*_{PODX})$

3.1.4.4 QTL Mapping

QTL Mapping was performed using the program IciMapping (Meng et al. 2015). IciMapping accepts an .xlsx file with tabs titled GeneralInfo, Chromosome, LinkageMap, Genotype, and Phenotype. GeneralInfo includes information regarding the population where the settings used were: 1, mapping; 4, RIL; 1, Kosambi; 1, intervals; 1, centimorgan; and 8, linkage groups. The linkage groups (LG1-LG6) corresponded to 6 of the 7 chromosomes of lentil. However, chromosome 7 had a large gap between the markers Lcu.2RBY.Chr7.479207507 (24.5 cM) and Lcu.2RBY.Chr7.483928632 (42.67 cM). Consequently, there were concerns about associations of phenotypes with this region. Therefore, chromosome 7 was split, between these 2 markers, into 2 linkage groups (LG7 and LG8).

To identify regions on the chromosome that are related to L*, a*, and b* scores at single time points, separate .xlsx files were loaded into IciMapping for every imaging cycle; LOD thresholds of 3.0000; and permutation times of 1000 were set as parameters. The program identifies markers that flank regions with strong associations (LOD ≥ 3.0000) to the phenotypes and generates an output file listing the marker names, the LOD scores, and the percent variances explained. Additional .xlsx files containing ΔL, Δa, Δb, and ED were loaded into IciMapping for separate site-years and PODs for samples grown in 2021 and QTL analysis was performed using the parameters previously described.

3.1.4.5 Significant Markers

The markers flanking regions on the genome where there were significant associations to L*, a*, and b* scores occurred in five of the ten imaging cycles. These were extracted from IciMapping and loaded into R where ANOVA was performed using the aov function (R Core Team 2021) to determine if there were significant differences between the phenotypes of lines possessing different SNP alleles at these locations. The favourable (F) alleles that were associated with desirable phenotypes for a given trait (higher L* scores, lower a* scores, and lower b* scores) were identified. Similarly, unfavorable alleles (U) that were associated with the opposite phenotypes (low L*, high a*, and high b*) were characterized. Finally, lines that possessed different combinations of markers associated with the desirable phenotypes were identified.

3.2 Results

3.2.1 There Was No Deterioration During Imaging Cycles

The time to complete an imaging cycle varied between 11 and 91 days. L*a*b* scores were stable within each of these cycles (Appendix B). Cold storage was possibly an effective strategy in preventing deterioration during periods where imagery was spread over a long period of time. This is the case for all imaging cycles, where the largest R^2 from the linear models was only 0.19 (found in L* for 2021-Rosthern-Mid).

3.2.2 Analysis of Variance (ANOVA)

The genotypes, site-years, replications (within site-years), and interactions between genotypes and site-years were significant ($p < 0.01$) for L*a*b* scores at each age, except for replications within site-years and genotype by site-year interactions in “Mid” aged L* values

(Appendix C). While the replication effects within site-years were significant ($p < 0.05$), the sums of squares were low relative to the sums of squares of site-years and genotypes. This indicated that replications (within site-years) did not play as an important of role in overall variance compared to site-years and genotypes (Appendix C). Furthermore, Tukey HSD results indicated that significant ($p < 0.05$) differences between replications, within imaging cycles, were only present in 3 of the 10 imaging cycles (2022-Rosthern-Fresh; 2019-SPG-Old; and 2019-Sutherland-Old) (Appendix D; Appendix E). Of the 90 comparisons between the L*, a*, or b* scores of replications within an imaging cycle, only 17 were different (7 for L*; 5 for a*; and 5 for b*) (Appendix D; Appendix E). For these reasons, the means of each site-year were used for subsequent analyses.

3.2.3 Seed Coat Color Changes Over Time

There was a decline in mean L* scores over time (Figure 3.1 - top), indicating a deterioration in seed coat color quality between successive imaging cycles of the same samples. Similarly, there was a consistent increase in a* scores with seed aging (Figure 3.1 - middle). Finally, there was a slight increase in b* scores with age, albeit less pronounced compared to L* and a* scores (Figure 3.1 - bottom).

The mean L*, a* and b* scores from plots, across timepoints, were used to calculate the ED and Δ values that are depicted in Figure 3.2. These data reinforce the trends where L* (top) tends to decrease and a* (centre-top) and b* (center-bottom) tend to increase, overtime. The ED and Δ values in POD3 are equal to the ED and Δ values of POD1 and POD2 added together. ED and Δ values in POD1 and POD2 differed ($p < 0.01$) from each other in all instances, where there as less deterioration in POD1 than in POD2 (Appendix F). These differences were more pronounced in Δb* for both locations and in ED, ΔL*, and Δa* within the Sutherland location.

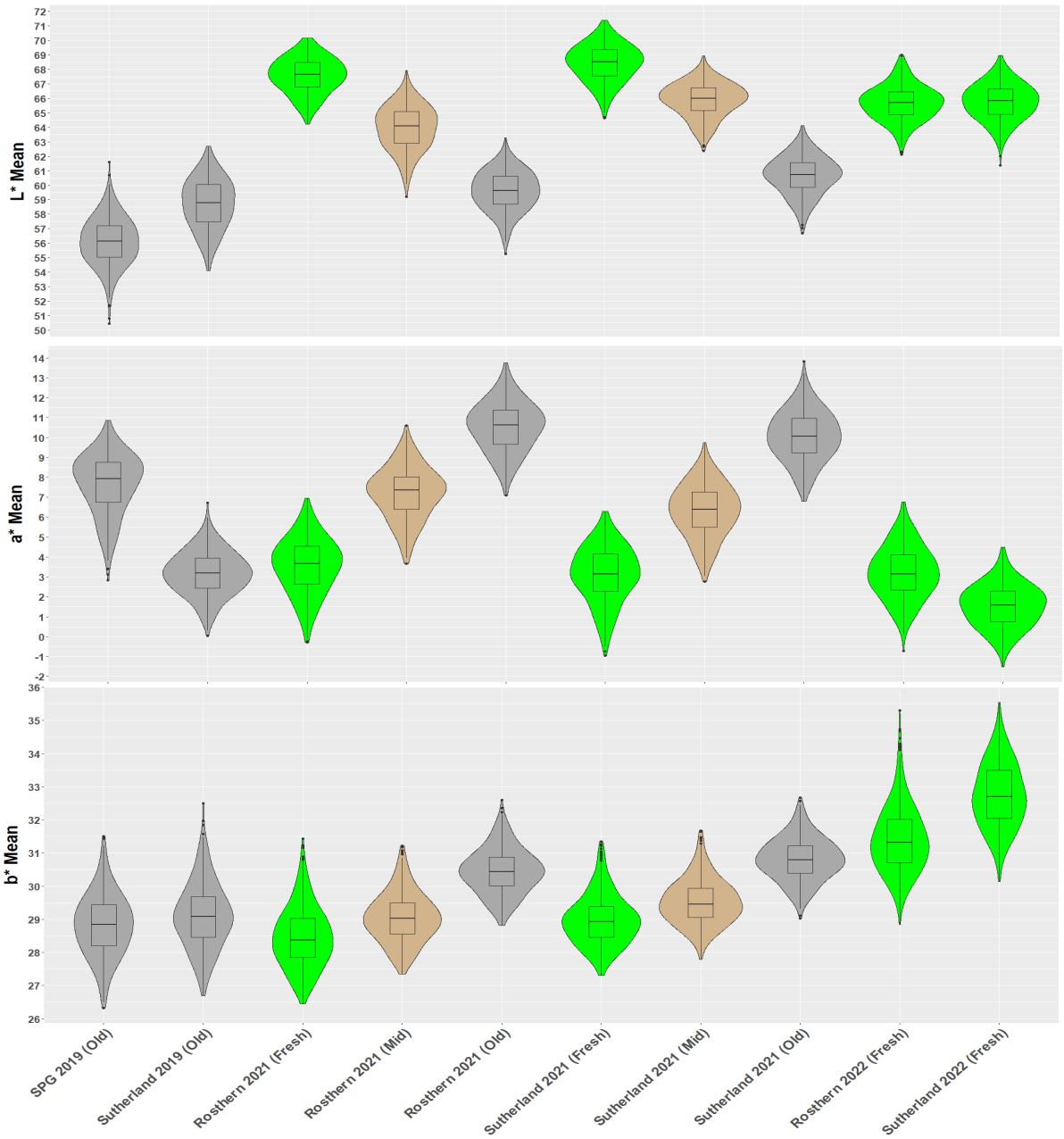

Figure 3.1: The mean CIE L* (top), a* (middle), and b* scores of 3 replications of ≈200 images of individual seeds, of 160 Lines of the lentil RIL (LR-06) population grown at six different site-years. Imagery occurred when seeds were "Fresh" (≈6 months post-harvest; green), "Mid" (≈12 months post-harvest; brown), and "Old" (≈18 months post-harvest; grey).

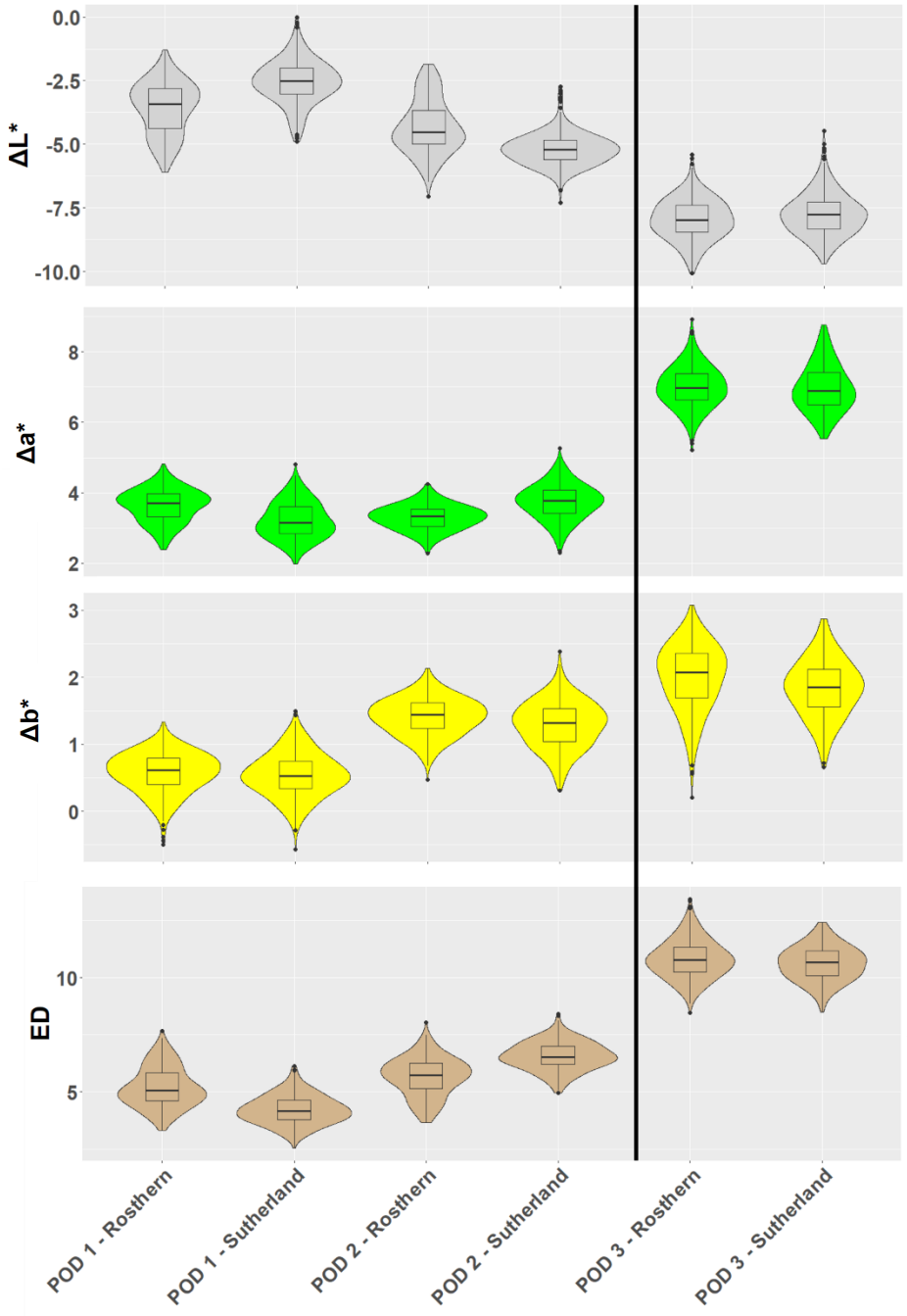

Figure 3.2: The mean Euclidean distance (ED), ΔL*, Δa* and Δb* values of seeds from plots grown in two locations (Rosthern and Sutherland) in 2021. The Δ values were calculated using "Mid" and "Fresh" values (POD1); "Old" and "Mid" values (POD2); and "Old" and "Fresh" values (POD3 i.e., POD1 + POD2). All ED and Δ means between PODs and locations differ significantly ($p < 0.05$), except for Δb* in POD1, Δa* in POD3, and Δa* between Sutherland POD1 and Rosthern POD2.

3.2.4 Linear Modeling of Seed Coat Colour Over Time

Linear mixed-effect models were effective at representing the relationship between seed coat colour and aging (Figure 3.3). L* scores exhibit an estimated fixed-effect of 0.87 ($p < 0.01$), indicating a strong predictive power of the mixed-effect model. The a* and b* scores had estimated fixed effects of 0.80 ($p < 0.01$) and 0.67 ($p < 0.01$), respectively, highlighting the robustness of the linear models for predicting a* and b* scores based on seed aging. Figure 3.4 provides a three-dimensional visualization of the L*a*b* scores for the 2021 data over the three ages with two representative genotypes highlighted, emphasizing the relationship between post-harvest seed coat color quality and color retention over time.

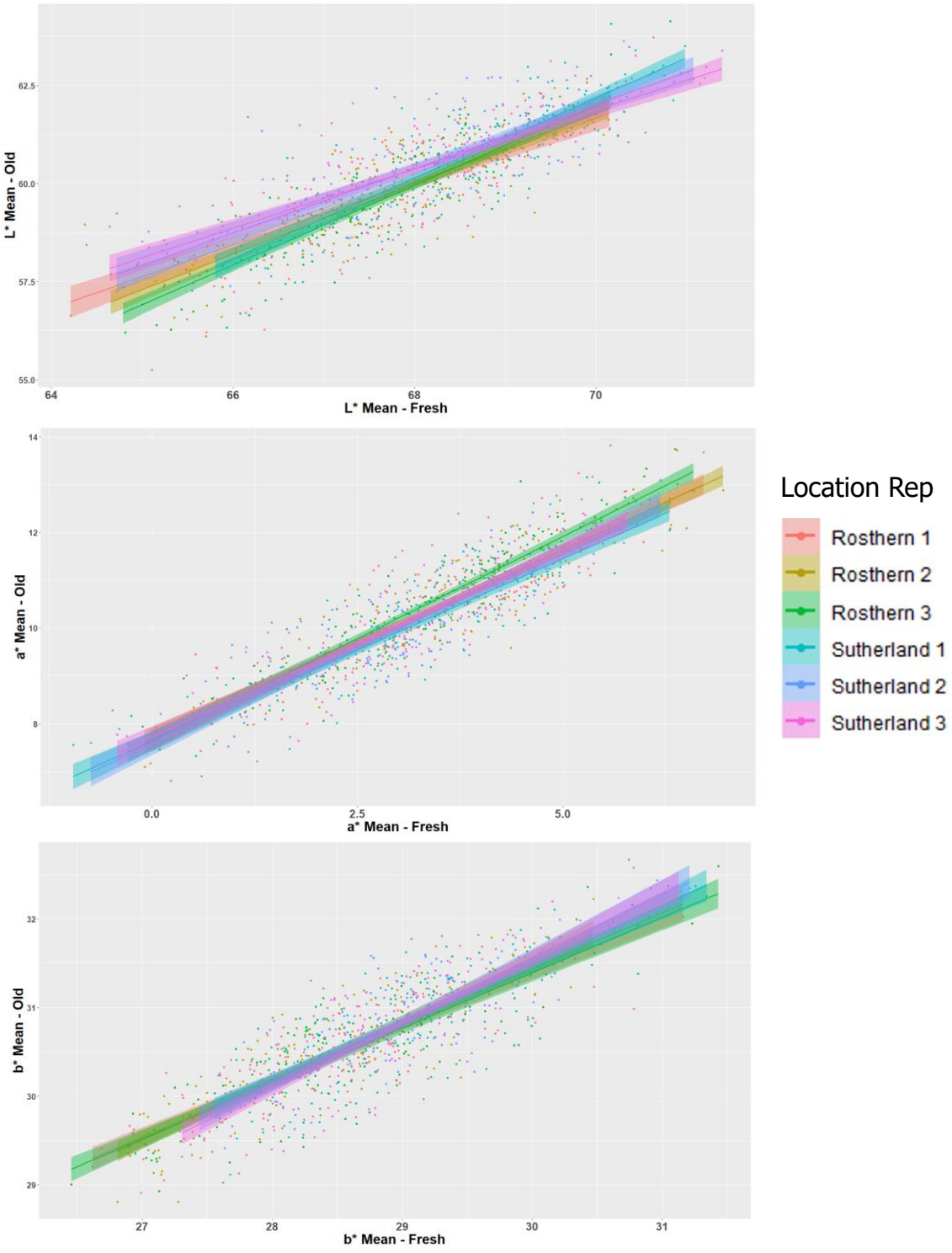

Figure 3.3: Mixed-effect linear models for "Old" L*a*b* scores as a function of "Fresh" L*a*b* scores for plot means of seed grown in 2021. "Fresh" L*, a*, b* scores were fit as fixed effects, whereas replications (1-3) and locations (Rosthern and Sutherland) were fit as random effects. The large bars surrounding the linear models depict standard error.

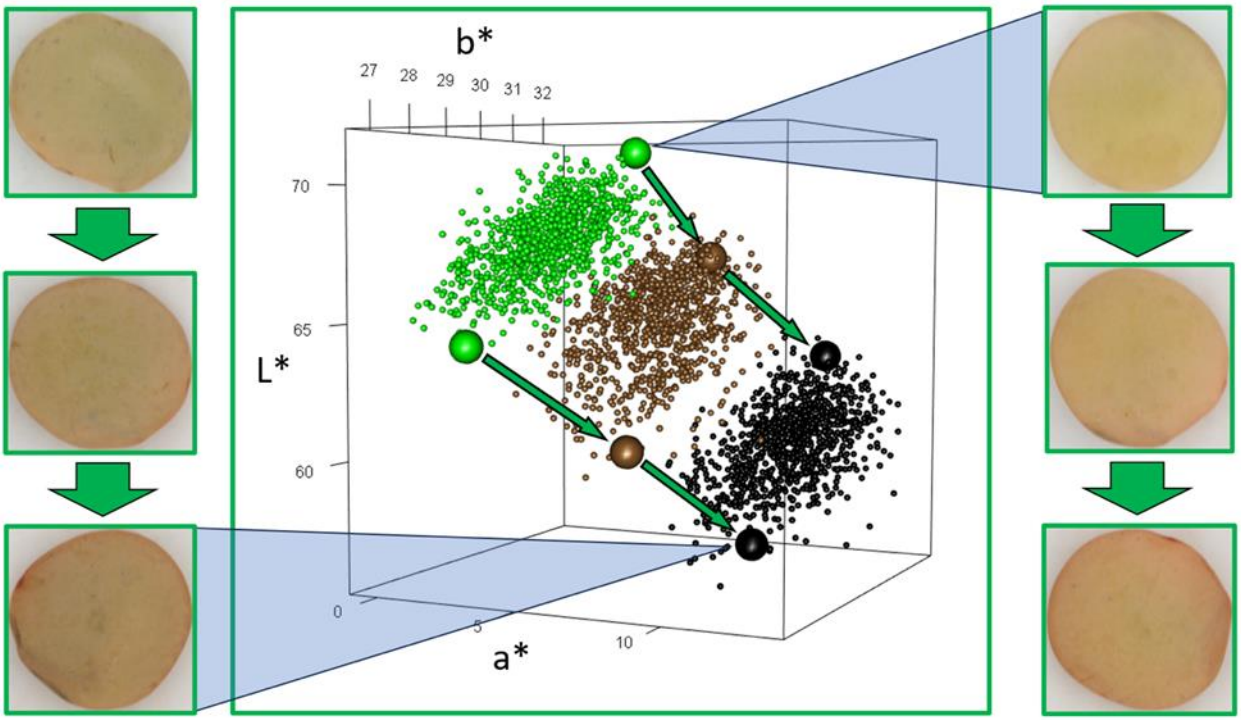

Figure 3.4: A three-dimensional representation of mean CIE L*a*b* scores of ≈200 seeds, from 160 RILs, across three time points, grown in 3 replications, across 2 locations, in 2021. The coloured clusters represent seed that was imaged when "Fresh" (green); "Mid" (brown); and "Old" (black). The six images of individual lentils to the periphery of the graph belong to one of two plots, imaged across the three time points. The samples from which these images were collected are represented on the graph by the three large orbs, connected by the arrow. The orb-arrow complexes represent the plots with the highest (top) and lowest (bottom) L* means in ≈6 months post-harvest seed.

3.2.5 Changes in b* Scores Do Not Appear to Contribute to Deterioration

Table 3.2 shows linear models where ED is fit as a function of ΔL*, Δa*, and Δb* for lines grown in 2021. ED is calculated from these numbers, therefore, fitting all three components together should have a very high R^2. However, to test the contribution of each parameter to this prediction, each were tested individually. Using ΔL* alone to model ED varies between R^2s of 0.71 and 0.9 for Rosthern during POD1 and POD2, respectively (Table 3.2). For Sutherland, the R^2s were lower at 0.43 and 0.65, for the same two PODs respectively (Table 3.2). Furthermore, using Δa* alone, to model ED varied between R^2s of 0.28 and 0.71 for Rosthern and R^2s of 0.46 and 0.62 for Sutherland (Table 3.2). The highest R^2 (between both locations) for Δb* is 0.04. Consequently, the removal of Δb* from the model only results in a marginal decrease in the R^2 value.

Table 3.2. Linear models for lines grown in Rosthern and Sutherland in 2021, where Euclidean Distance (ED) is fit as a function of L*a*b* scores across the three Periods of Deterioration (PODs).

Model	POD	R^2	p-values			
			Intercept	L*	a*	b*
Rosthern						
lm(ED ~ ΔL* + Δa* + Δb*)	1	0.99	0.279	< 0.01	< 0.01	< 0.01
	2	> 0.99	< 0.01	< 0.01	< 0.01	< 0.01
	3	> 0.99	< 0.05	< 0.01	< 0.01	< 0.01
lm(ED ~ ΔL*)	1	0.80	< 0.01	< 0.01		
	2	0.90	< 0.01	< 0.01		
	3	0.71	< 0.01	< 0.01		
lm(ED ~ Δa*)	1	0.44	< 0.01		< 0.01	
	2	0.28	< 0.01		< 0.01	
	3	0.71	< 0.01		< 0.01	
lm(ED ~ Δb*)	1	0.01	< 0.01			0.273
	2	0.01	< 0.01			0.178
	3	< 0.01	< 0.01			0.715
lm(ED ~ ΔL* + Δa*)	1	0.99	<0.01	< 0.01	< 0.01	
	2	0.98	<0.01	< 0.01	< 0.01	
	3	0.91	<0.01	< 0.01	< 0.01	
Sutherland						
lm(ED ~ ΔL* + Δa* + Δb*)	1	0.97	0.574	< 0.01	< 0.01	< 0.01
	2	> 0.99	<0.01	< 0.01	< 0.01	< 0.01
	3	> 0.99	0.269	< 0.01	< 0.01	< 0.01
lm(ED ~ ΔL*)	1	0.43	<0.01	< 0.01		
	2	0.76	<0.01	< 0.01		
	3	0.65	<0.01	< 0.01		
lm(ED ~ Δa*)	1	0.46	<0.01		< 0.01	
	2	0.62	<0.01		< 0.01	
	3	0.59	<0.01		< 0.01	
lm(ED ~ Δb*)	1	< 0.01	< 0.01			< 0.01
	2	0.04	< 0.01			< 0.01
	3	< 0.01	< 0.01			< 0.01
lm(ED ~ ΔL* + Δa*)	1	0.97	< 0.01	< 0.01	< 0.01	
	2	0.98	< 0.01	< 0.01	< 0.01	
	3	0.98	< 0.01	< 0.01	< 0.01	

3.2.6 Correlations Between L*a*b* Scores and Phenology Were Not Strong

The plots within site-years flowered within 4 to 8 days of each other (Appendix G). Correlation coefficients range between -0.29 and –0.22 for L*; -0.3 to -0.066 for a*; and -0.16 to 0.1 for b* and DTF (Appendix H). While most correlations were very low between L*a*b* scores and DTF, there was only one significant ($p < 0.01$) exception. This was for a* in the 2022-Rosthern-Fresh imaging cycle which possessed a correlation coefficient of 0.3.

The plots within site-years matured within 6 to 15 days of each other (Appendix G). While most correlation coefficients were negative and below absolute values of 0.3, there were two exceptions. There is an extremely low and positive coefficient of 0.0069 in 2022-Rosthern-Fresh, and a larger correlation coefficient (relative to the other imaging cycles) of -0.33 for 2022-Sutherland-Fresh, indicating there is a minor correlation with DTM for this imaging cycle. Correlations are more prevalent between a* scores and DTM. The correlations vary between -0.51 and -0.16, with half of the correlation coefficients being higher than absolute values of 0.3. Finally, b* scores and DTM correlation coefficients range between -0.05 and 0.41, with three imaging cycles (2021-Rosthern-Fresh, 2022-Rosthern-Fresh, and 2022-Sutherland-Fresh) having higher correlation coefficients than absolute values of 0.3. All b* correlations are positive with one exception (-0.05) being in 2021-Sutherland-Old.

3.2.7 QTL Analysis for Seed Coat Color Traits

Figure 3.5 is a summary of the results of QTL analyses for CIE L*a*b* scores across separate imaging cycles. Each of the eight boxes correspond to linkage groups (LGs) representing the seven chromosomes of lentil, with linkage groups seven and eight representing chromosome seven. Regions of the genome showing strong associations ($LOD \geq 3$) with seed coat color traits (L*, a*, and b*) are depicted with squares, with the color indicating the percent variance explained by these associations. A total of 105 QTL were identified across imaging cycles, where 29 were for L*; 36 were for a*; and 40 were for b*.

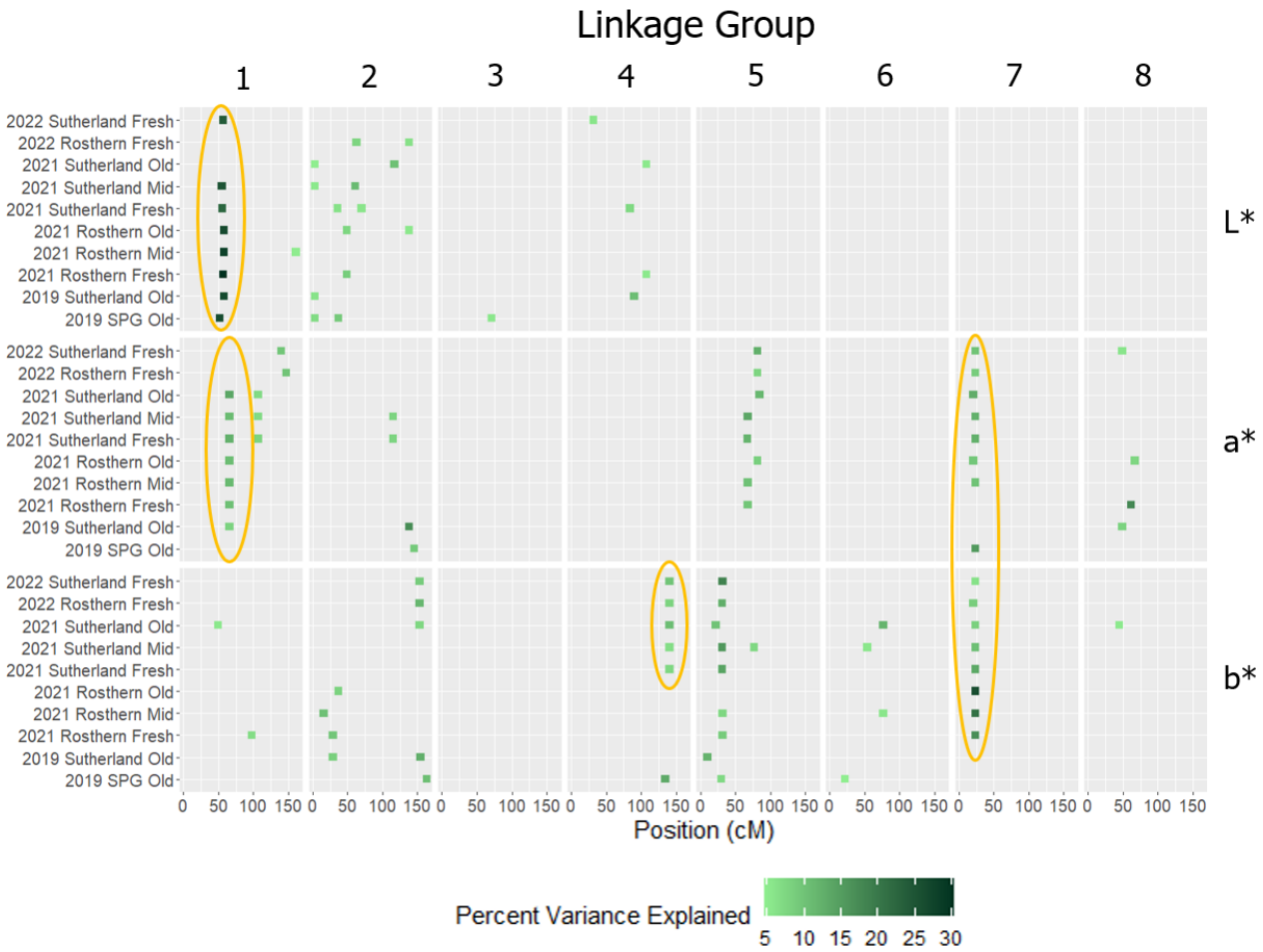

Figure 3.5: The results of QTL analysis for L*a*b* scores, across 10 imaging cycles of seed of LR-06. The linkage groups correspond to the 7 lentil chromosomes, with linkage groups 7 and 8 corresponding to chromosome 7. Data were extracted from IciMapping with significant (LOD ≥ 3) associations between regions on the genome with L* (top), a* (middle), and b* (bottom) scores. The percent variance explained by these regions is coded based on colour (with light to dark corresponding to percent variance explained). Regions with many (≥ 5) imaging cycles associated with the same markers are considered, within this work, important for determining colour quality (and are circled in orange).

3.2.8 Significant Regions (SRs) of the Genome

Four regions on the genome were identified as having common (≥ 5) occurrences across imaging cycles for L*, a*, or b* scores (Figure 3.5). These regions include one association with L* within LG1; one association with a* within LG1; one association with b* within LG4; and associations with a* and b* within the same region on LG7. These regions, termed onward as "Significant Regions" included SR1 for the L* association on LG1; SR2 for the a* association on LG1; SR3 for the b* association on LG4; and SR4 for the a* and b* associations on LG7.

The genotypes carrying the favourable (F) versus unfavourable (U) alleles were confirmed to differ significantly from each other at each of these loci (Figure 3.6 and Appendix I). Taking the means of the relevant L*a*b* scores for all lines (regardless of location, year, and age) reveals that higher L* scores are associated with an AA-AA (left marker-right marker) genotype at SR1. Lower a* scores are associated with a CC-TT genotype at SR2. Lower b* scores are associated with an AA-TT genotype at SR3 and lower a* and b* scores are associated with a CC-GG genotype at SR4 (Appendix I).

There were 138 lines in possession of genotypic information at all 4 SRs. The other 22 lines were either heterozygous or had missing data at one or more loci. Only one line (LR-06-145) possessed the F alleles for all 4 SRs, and only two additional lines (LR-06-057 and LR-06-136) possessed all F alleles at the SRs relevant to L* and a* scores (Figure 3.6).

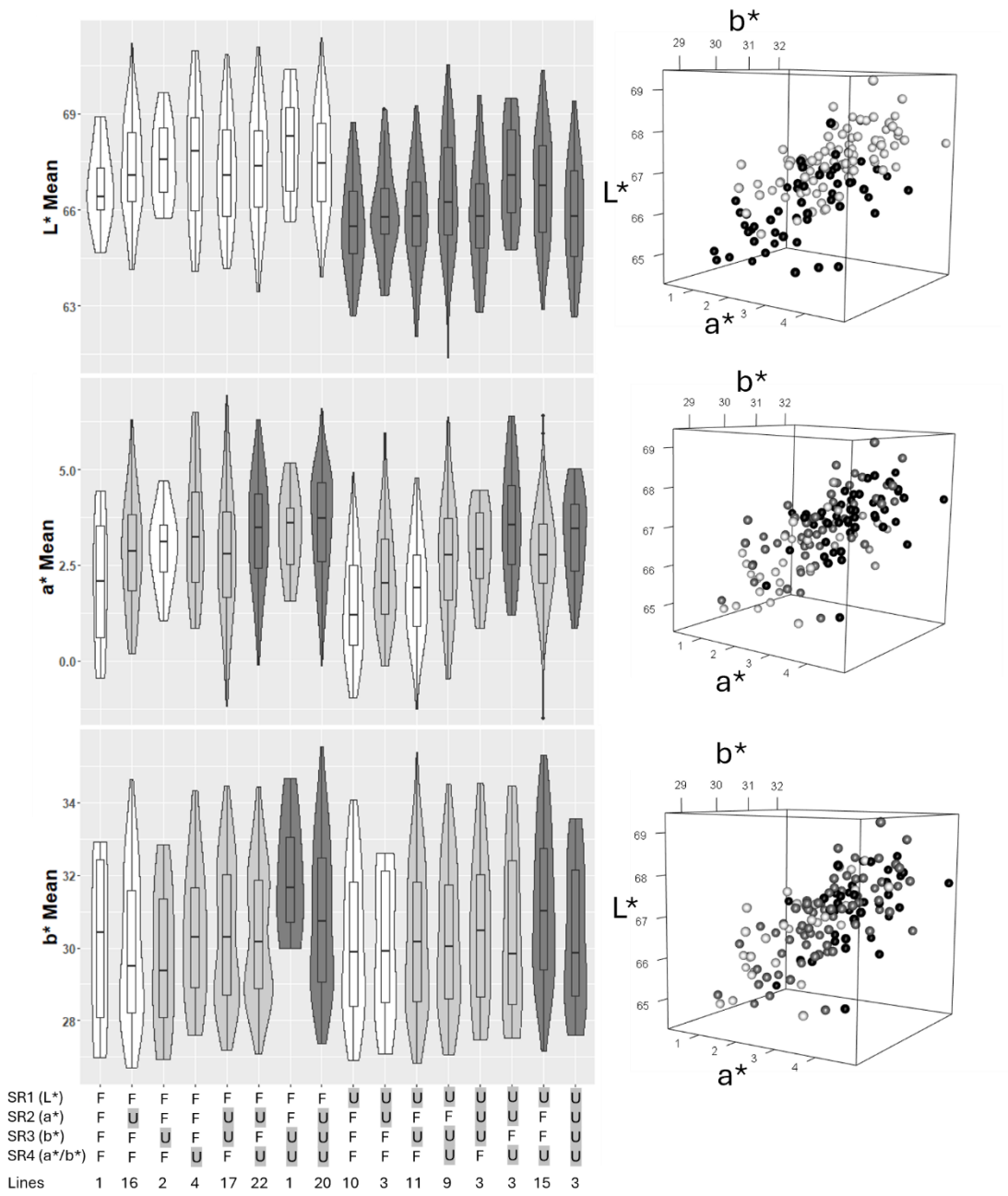

Figure 3.6: The plot means (≈ 200 seeds) CIE L* (top-left), a* (middle-left), and *b* (bottom-left) scores of Fresh-Sutherland-2021 LR-06 genotypes, in the 16 combinations of alleles, between the 4 significant regions (SRs). Favourable (F) alleles are associated with higher L* values and lower a* and b* values; unfavorable (U) alleles are associated with lower L* values and higher a* and b* values. The shades of the violin plots depict the number of favourable alleles possessed by each genotype, where white indicates that a genotype possesses all F alleles for a specific trait (L*, a* or b*); light grey indicates that a genotype possesses a combination of F and U alleles; and dark grey indicates that a genotype possesses all U alleles for that trait. The plots on the right depict the L*a*b* scores of the different allelic combinations in a 3-dimensional space, where the colours correspond to the genotypes as follows: white = all F alleles; grey = a combination of F and U alleles; and black = all U alleles for a given trait.

3.3 Discussion

3.3.1 Colour Changes Over Time

Green lentil seed coat colour quality is not stable and tends to deteriorate towards dark brown over time (Nozzolio and De Bezada 1984). This study demonstrated that these changes in seed coat colour can be quantified using CIE L*a*b* scores, in which seed coats tend to decrease in L*, increase in a*, and increase in b* values, coinciding with decreases in "lightness," decreases in "greenness," and increases in "yellowness," respectively (Figure 3.1).

While all three dimensions in the L*a*b* colour system change as seed ages, b* does not change as much, or as consistently as L* and a*. While b* scores tended to increase with age, some samples had negative Δb* scores (Figure 3.2). This indicates that the b* scores of some samples decreased over time. The lack of a consistent change in b* scores suggests that they are not as reliable of a metric to quantify colour deterioration in green lentil as L* and a* scores, which more reliably move in one direction as seed ages.

The minimal importance of b* (compared to L* and a*) is further demonstrated by linear modelling (Table 3.2). Given that ED is calculated from ΔL*, Δa*, and Δb*, it was expected that fitting ED as the dependent variable and the ΔL*, Δa*, and Δb* scores (together) as the explanatory variables would have R^2 that are (or are close to) 1. However, excluding b* from the models resulted in R^2 values that were still close to 1 (ranging from 0.97 to > 0.99). This means that changes in L* and a*, without b*, can be used to quantify deterioration (within this population).

In addition to varying between ages ("Fresh," "Mid," and "Old"), the colour quality of seed coats varied between lines within ages. The initial L*a*b* scores of "Fresh" lentil seed coats were highly predictive of the seed coat colour quality following aging (Figures 3.3 and 3.4). This means that, despite the decline in L* scores and increase in a* scores, the relative position of each sample (within an imaging cycle) was similar in successive imaging cycles.

A distinction between colour quality (CIE L*a*b* scores at a single time point) and deterioration (ΔL*a*b* scores and ED) was emphasized in the methodology of this work because, prior to this study, it was not known if initial colour quality would be linked to colour deterioration, in green lentil. Both chlorophyll and polyphenols play a role legume seed colour

quality and deterioration. Chlorophyll deterioration occurs through the replacement of the magnesium with hydrogen ions in its porphyrin ring, over time (Joslyn and Mackinney 1907). This results in the green colour, associated with chlorophyll turning to a dark brown. Martins and Silva (2002) found that while chlorophyll content is related to the initial colour of pods, it does not predict the quality of colour following aging in green beans (*Phaseolus vulgaris*). Furthermore, while Davey (2007) found that chlorophyll content was highly correlated to seed coat colour desirability in the LR-06 green lentil population, they did not investigate how chlorophyll content related to desirability across multiple time-points. In addition, darkening of seed has been linked to phenolic compounds in pinto beans (Beninger et al. 2005; Duwadi et al. 2018) and lentil (De Bezada 1980; Nozzolio and De Bezada 1984; Mirali et al. 2016). The presence of chlorophyll, being a green pigment, can often be estimated through visual inspection. However, some phenolic compounds are colourless (Mirali et al. 2016) and may require biochemical profiling to detect. Therefore, the only alternatives to determining the potential for colour deterioration would require waiting for seed to age before evaluating colour visually (which is time consuming and devalues the seed in the process) or using genetic markers. The results of this study suggest that these methods are not necessary to detect seed coat colour quality following ageing, because phenotyping a population within a (relatively) short timeframe will provide information on the colour quality of seed, relative to other lines, within a population.

3.3.2 The Genetics of Seed Coat Colour

Previous work, conducted by Davey (2007), demonstrated that green lentil seed coat colour quality is highly heritable ($H^2 = 0.82$). Building off this previous work, this study used different methods of data acquisition (BELT) and colour quantification (CIE L*a*b* scores) to corroborate the findings of the work of Davey (2007). The estimates of H^2 in this study, as derived from the "variability" package in R (Popat et al. 2020), were 0.54 for L*, 0.59 for a*, and 0.58 for b*. These values are not as high as the value reported by Davey (2007) but the differences in quantifying colour ("desirability" versus CIE L*a*b* scores) between the two studies render comparisons difficult.

QTL mapping revealed that there are four significant regions (SRs) on the LR-06 genome that are associated with CIE L*a*b scores (Figure 3.5). These regions occurred in more than half of the imaging cycles for individual traits. All four SRs were present in all three ages ("Fresh,"

"Mid," and "Old") of imaging cycles. These regions are not associated with a change in colour but are associated with the L*a*b* scores at single time points. While QTL mapping was performed on ED and ΔL*a*b* scores, there were no SRs for colour deterioration identified.

Figure 3.6 depicts the L*, a* and b* scores of the different allelic combinations at the four SRs in one imaging cycle (i.e., Sutherland-2021-Fresh). This trend is ubiquitous for all site-years and imaging cycles (Appendix I). The genotypes associated with each SR, at each imaging cycle, show significant differences ($p < 0.01$) between L*a*b* scores for the relevant traits, where the F alleles are always associated with more desirable L*, a*, or b* scores, and the opposite is true for U alleles. This is true for imaging cycles that did not show significant associations when performing QTL analysis and helps provide further evidence that the markers within these regions are strongly linked to (or located within) regions on the genome that are responsible for colour quality.

3.3.3 Potential Confounding Factors and Limitations

3.3.3.1 Parental Contamination

Following genotypic analysis, it was determined that the lines that were labelled as the parents of the LR-06 population (1294m-23 and 1048-8R) were mislabelled. This became clear when assembling a linkage map, in which the parents possessed contrasting alleles that were not segregating in the rest of the population. Because the 2019 seed was used for genotyping and seeds grown in 2021 and 2022 were sourced or derived from the 2019 seed, all subsequent years were presumed to be compromised and omitted from analyses.

3.3.3.2 Cotyledons Versus Seed Coat

Because green lentils are typically graded and consumed whole, the additive effects of seed coat and cotyledon (if meaningful) are perceived by the consumer together. Therefore, there is value in evaluating the colour of both together. Since it was not possible to make a distinction between the colours of cotyledons and seed coats, it is not clear whether the QTL are related to the seed coats or cotyledons or both. All LR-06 lines possess yellow cotyledons; however, the cotyledons may vary subtly in intensity of yellow or in visibility through the seed coat. An attempt was made to determine if carotenoid content differed between the parental lines; however, the contamination issues (as previously discussed) rendered these results meaningless.

Jackson et al. (2021) demonstrated that lentil cotyledon colours can change over time when exposed to different light treatments (ultraviolet, blue light, green light, red light, or full-visible). However, these changes were low in yellow cotyledon lentils, and only considered to be visually perceptible for seeds exposed to blue light. Furthermore, the change in cotyledon colours was only found to be significant for Δb* and not for ΔL* or Δa*. These results from Jackson et al. (2021) could explain the minor changes that are detected in b* in this present study. The results of this present study indicate that Δb* plays very little role in colour deterioration of whole seed, whereas ΔL* and Δa* are much more significant.

Green lentils are graded and consumed whole, thus validating the need for evaluating the overall colour deterioration of the seed. However, separating seed coats from the cotyledons of LR-06 population, imaging the cotyledons, and performing QTL mapping would be a method of confirming that SRs related to b* are due to cotyledon changes.

3.3.3.3 Environment

Site-years and replications within site-years played a role in seed coat colour quality (Appendix C). However, these differences don't appear to impact the relative position of lines within a population across ages (Figures 3.3 and 3.4). The implication is that regardless of the environment in which seeds are grown, the highest and lowest performing lines will be consistent and there really is only a need for phenotyping from one location to identify the best from the worst.

Correlations between L*a*b* scores and phenology were typically weak, if present at all (Appendices G and H). DTF did not have any correlations that had an absolute value above –0.3. The presence of slightly negative correlations between DTM and a* scores means that later maturity equated to a more desirable (lower) a* score, however, this may not have been because of a direct relationship between the two. De Bezada (1980) determined that higher temperatures and light tend to deteriorate lentil seed coats more rapidly that when stored in dark, cool and anoxic conditions. Plots at each site-year were hand-harvested all at once in a single day (aside from Sutherland 2021, which was harvested across two days). This means that plots that matured early remained in the field while the other plots continued to mature, meaning they were subjected to the elements longer. Therefore, the desirability of seed (with respect to a*) may

relate to the days that seed is left in the field post maturity and a timely harvest is recommended for better quality colour.

3.3.4 Conclusions and Recommendations to Breeders

The first hypothesis, that L* scores would decrease and that a* scores would increase with aging is strongly supported by this work (Figures 3.1 and 3.2). While b* was omitted from the initial hypothesis, it apparently increases as seed ages. CIE L*a*b* is a three-dimensional colour space and the quantification of colour requires coordinates at each of the axes. Within the LR-06 population, b* does not appear to deteriorate as much (Figure 3.2) or have the same impact on ED as L* and a* (Table 3.2). This is consistent with the idea that seed coat colour quality and deterioration of LR-06 (and perhaps more broadly, green lentil) can be quantified with only two of the CIE colour coordinates. However, caution must be taken when attempting to apply this approach to other populations or market classes. While the differences in b* were minimal in LR-06, they were still present. In cases where a* and b* both show high variation (as one would expect with red lentil, which has a red-orange colour), hue or chroma (which both relate a* and b* together) might be better methods of quantifying colour.

Linear mixed-effect modelling revealed that the colour quality of “Old” seed is predictable based on the colour quality of “Fresh” seed, thus supporting hypothesis two (Figure 3.3). While the environment did play a role in seed coat colour quality between locations, the lines remained in similar positions (relative to the rest of the population). The implications of this are that phenotyping of green lentil seed coat colour quality can occur in one environment and at one location to reveal the desirability of lines within a population relative to known checks. However, while the phenotyping of seed coat quality can be accomplished at any time, it is important to perform it quickly to minimalize the amount of deterioration that could occur while phenotyping is ongoing.

The results of this work support hypothesis three for the absolute L*, a*, and b* scores. These were moderately heritable and appear to be strongly associated with four regions of the genome (Figure 3.5), regardless of the age at which they were screened. Taken together, this provides evidence that pre-disposition to colour quality, following aging, is multi-genic and controlled by distinct regions of the genome.

A distinction between colour quality and deterioration must be made. The changes in L*, a* and b* scores over time (ΔL*, Δa* and Δb*), while differing between lines, did not have enough variation or consistency to be detectable in QTL analysis. Consequently, this work provides evidence that the use of markers may not be necessary to determine the highest quality of seed coat colour within a population. Instead, a breeder may simply phenotype a population, at a single time point, and be able to determine the relative quality of a seed coat following aging. However, the markers that were identified within this study may still have utility when performing complex or multi-parent crosses where the genotypes are not necessarily discernible from phenotypes because of heterozygous conditions.

3.3.5 Future Work

While this work investigated the genetics of seed coat colour quality and deterioration in green lentil, it did not investigate the biochemical mechanisms of seed coat colour deterioration. The abundance of previous works that draw links between colour deterioration and polyphenols in pinto bean (Beninger et al. 2005; Duwadi et al. 2018) and lentil (De Bezada 1980; Nozzolillo and De Bezada 1984; Mirali et al. 2016) suggest that polyphenolic profiling on selected (or all) lines within the LR-06 population is worthy of consideration.

The Canadian Grain Commission (2021) defines seed coat colour quality based on a subjective scale that would benefit from the establishment of colour standards, based on an objective colour scoring system (such as CIE L*a*b*), which would allow for a higher level of precision when evaluating colour than what is currently used. Furthermore, the use of objective colour scoring allows for the rapid phenotyping of seed samples through automation. Thus, saving time and money in evaluating seed colour quality. However, the use of such a system will require additional work in establishing standards of colour quality within a colour scale. This study associated higher L* and lower a* scores with a higher colour quality, but future research should focus on working with the Canadian Grain Commission in establishing CIE L*a*b* scores that are associated with desirable and undesirable green lentil seed.

4 REFERENCES

Alpaslan D, Dudu TE, Şahiner N, Aktas N. 2020. Synthesis and preparation of responsive poly (dimethyl acrylamide/gelatin and pomegranate extract) as a novel food packaging material. Mater Sci Eng C Mater Biol Appl. 108:110339. doi:10.1016/j.msec.2019.110339.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Soft. 67(1). doi:10.18637/jss.v067.i01.

Beninger CW, Gu L, Prior RL, Junk DC, Vandenberg A, Bett KE. 2005. Changes in polyphenols of the seedcoat during the after-darkening process in pinto beans (*Phaseolus vulgaris* L.). J Agric Food Chem. 53(20):7777–7782. doi:10.1021/jf050051l.

Beninger CW, Hosfield GL. 2003. Antioxidant activity of extracts, condensed tannin fractions, and pure flavonoids from *Phaseolus vulgaris* L. seedcoat color genotypes. J Agric Food Chem. 51(27):7879–7883. doi:10.1021/jf0304324.

Bett KE, Vandenberg A [Funding application]. 2018. Enhancing the value of lentil variation for ecosystem survival (EVOLVES). Agricultural Development Fund.

Canadian Grain Commission [Internet]. 2021. Lentils: Grading factors. [Accessed 2021 Sept 28]. Available from: https://www.grainscanada.gc.ca/en/grain-quality/official-grain-grading-guide/18-lentils/grading-factors.html.

Davey BF. 2007. Green seedcoat colour retention in lentil (*Lens culinaris*) [MSc thesis]. Saskatoon (SK): University of Saskatchewan.

De Bezada MV. 1981. Effect of environmental conditions on lentil seeds [MSc thesis]. Ottawa (ON): University of Ottawa.

DellaValle DM, Vandenberg A, Glahn, RP. 2013. Seedcoat removal improves iron bioavailability in cooked lentils: studies using an in vitro digestion/Caco-2 cell culture model. J Agric Food Chem. 61(34):8084–8089. doi:10.1021/jf4022916.

Dueñas M, Hernández T, Estrella I. 2002. Phenolic composition of the cotyledon and the seedcoat of lentils (*Lens culinaris* L.). Eur Food Res Technol. 215(6):478–483. doi:10.1007/s00217-002-0603-1.

Duwadi K, Austin RS, Mainali HR, Bett K, Marsolais F, Dhaubhadel S. 2018. Slow darkening of pinto bean seedcoat is associated with significant metabolite and transcript differences related to proanthocyanidin biosynthesis. BMC Genom. 19(1):260. doi:10.1186/s12864-018-4550-z.

Elessawy FM, Vandenberg A, El-Aneed A, Purves RW. 2021. An untargeted metabolomics approach for correlating pulse crop seedcoat polyphenol profiles with antioxidant capacity and iron chelation ability. Mol. 26(13):3833. doi:10.3390/molecules26133833.

Emami MK, Sharma B. 2000. Inheritance of black testa colour in lentil (*Lens culinaris* Medik.). Euphytica. 115(1):43-47. doi:10.1023/A:1003917115138.

Emery KJ, Volbrecht VJ, Peterzell DH, Webster MA. 2017. Variations in normal color vision. VI. factors underlying individual differences in hue scaling and their implications for models of color appearance. Vis Res. 141:51-65. doi: 10.1016/j.visres.2016.12.006.

Erskine W, Muehlbauer FJ, Ashutosh S, Sharma B. 2009. The lentil: botany, production and uses. Wallingford (Oxon): CAB International. 380 p.

FAOSTAT [Internet]. 2021. Food and Agriculture Organization of the United Nations; [Accessed 2021 Feb 25]. Available from: https://www.fao.org/faostat/en/#data/QCL/visualize.

Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Earth Environ Sci Trans. 52(2): 399-433.

Futuyma DJ. 2013. Evolution. 3rd ed. Sunderland (MA): Sinauer Associates, Inc. Publishers. 349 p.

Gebremedhin A, Li Y, Shunmugam ASK, Sudheesh S, Valipour-Kahrood H, Hayden MJ, Rosewarne GM, Kaur S. 2024. Genomic selection for target traits in the Australian lentil breeding program. Front Plant Sci. 14. doi:10.3389/fpls.2023.1284781.

Gillooly M, Bothwell TH, Torrance JD, MacPhail AP, Derman DP, Bezwoda WR, Mills W, Charlton RW, Mayet F.1982. The effects of organic acids, phytates and polyphenols on the absorption of iron from vegetables. Br J Nutr. 49(3):331–342. doi:10.1079/BJN19830042.

Government of Saskatchewan [Internet]. 2021. 2020 Specialty crop report. [Accessed 2021 Feb 25]. Available from: https://publications.saskatchewan.ca/api/v1/products/111437/formats/125055/download.

Halcro K, McNabb K, Lockinger A, Socquet-Juglard D, Bett KE, Noble SD. 2020. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. Plant Methods. 16(1):1-13. https://doi.org/10.1186/s13007-020-00591-8.

Hansen TF. 2006. The evolution of genetic architecture. Annu Rev Ecol Evol Syst. 37:123–157. doi:https://doi.org/10.1146/annurev.ecolsys.37.091305.110224.

Hunt RG. 1995. The reproduction of colour (vol. 4). England: Fountain Press. 12 p.

Hurrell RF, Reddy M, Cook JD. 1999. Inhibition of non-haem iron absorption in man by polyphenolic-containing beverages. Br J Nutr. 81(4):289–295. doi:10.1017/S0007114599000537.

Jackson N, Noble SD, Vandenberg A. 2021. Evaluating the effect of light exposure and seed coat on lentil cotyledon color by computer vision. Legum. sci. 3(4):e98. doi:10.1002/leg3.98.

Johnson CR, Thavarajah D, Thavarajah P. 2013. The influence of phenolic and phytic acid food matrix factors on iron bioavailability potential in 10 commercial lentil genotypes (*Lens culinaris* L.). J Food Compos Anal. 31(1):82-86. doi:10.1016/j.jfca.2013.04.003.

Junk-Knievel DC, Vandenberg A, Bett KE. 2008. Slow darkening in pinto bean (*Phaseolus vulgaris* L.) seedcoats is controlled by a single major gene. Crop Sci. 48(1):189–193. doi:10.2135/cropsci2007.04.0227.

Joslyn MA, Mackinney G. 1938. The rate of conversion of chlorophyll to pheophytin. J Am Chem Soc. 60(5):1132–1136. doi:10.1021/ja01272a037.

Kilbride PE, Hutman LP, Fishman M, Read JS. 1986. Foveal cone pigment density difference in the aging human eye. Vis. Res. 26(2):321–325. doi:10.1016/0042-6989(86)90029-5.

Kislev ME, Nadel D, Carmi I. 1992. Epipalaeolithic (19,000 BP) cereal and fruit diet at Ohalo II, Sea of Galilee, Israel. Palaeobot. Palynol. 73(1–4):161–166. doi:10.1016/0034-6667(92)90054-K.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest package: tests in linear mixed effects models. J Stat Softw. 82(13):1–26. doi: 10.18637/jss.v082.i13.

Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121(1):185–199. doi:10.1093/genetics/121.1.185.

Larraín RE, Schaefer DM, Reed JD. 2008. Use of digital images to estimate CIE color coordinates of beef. Int Food Res. 41(4):380–385. doi:10.1016/j.foodres.2008.01.002.

Li H, Hearne S, Bänziger M, Li Z, Wang J. 2010. Statistical properties of QTL linkage mapping in biparental genetic populations. Hered. 105(3):257–267. doi:10.1038/hdy.2010.56.

Liber M, Duarte I, Maia AT, Oliveira HR. 2021. The history of lentil (*Lens culinaris* subsp. *culinaris*) domestication and spread as revealed by genotyping-by-sequencing of wild and landrace accessions. Front Plant Sci. 12:628439. doi:10.3389/fpls.2021.628439.

Markovic I, Ilic J, Markovic D, Simonovic V, Kosanic N. 2013. Color measurement of food products using CIE L* a* b* and RGB color space. J Eng Des. 4(1):50-53.

Martins RC, Silva CLM. 2002. Modelling colour and chlorophyll losses of frozen green beans (*Phaseolus vulgaris*, L.). Int J Refrig. 25(7):966–974. doi:10.1016/S0140-7007(01)00050-0.

Mendel G. 1866. Experiments in Plant Hybridization.

Meng L, Li H, Zhang L, Wang J. 2015. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in bi-parental populations. Crop J. 3:269-283. doi:10.1016/j.cj.2015.01.001.

Mirali M, Purves RW, Vandenberg A. 2016. Phenolic profiling of green lentil (*Lens culinaris* Medic.) seeds subjected to long-term storage. Eur Food Res Technol. 242(12):2161-2170. doi:10.1007/s00217-016-2713-1.

Moran JF, Klucas RV, Grayer RJ, Abian, J, Becana, M. 1997. Complexes of iron with phenolic compounds from soybean nodules and other legume tissues: prooxidant and antioxidant properties. Free Radic Biol Med. 22(5):861–870. doi:10.1016/S0891-5849(96)00426-1.

Nozzolillo C, De Bezada MV. 1984. Browning of lentil seeds, concomitant loss of viability, and the possible role of soluble tannins in both phenomena. Can J Plant Sci. 64(4):815-824. doi:10.4141/cjps84-113.

Popat R, Patel, R and Parmar D. 2020. Variability: genetic variability analysis for plant breeding research. R package version 0.1.0. https://CRAN.R-project.org/package=variability.

Ramsay L, Koh CS, Kagale S, Gao D, Kaur S, Haile T, Gela TS, Chen LA, Cao Z, Konkin DJ, Toegelová H, Doležel J, Rosen BD, Stonehouse R, Humann JL, Main D, Coyne CJ, McGee RJ, Cook DR, Penmetsa RV, Vandenberg A, Chan C, Banniza S, Edwards D, Bayer PE, Batley J, Udupa SM, Bett KE. 2021. Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. bioRxiv. [Accessed 2023 Jan 27]. https://knowpulse.usask.ca/genome-assembly/Lcu.2RBY.

R Core Team. 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sandberg A. 2002. Bioavailability of minerals in legumes. Brit J Nutr. 88(S3):281–285. doi:10.1079/BJN/2002718.

Saskatchewan Crop Insurance Corporation [Internet]. 2017. AgriStability Saskatchewan price list 2017. [Accessed 2024 June]. Available from: https://www.scic.ca/uploads/agristability-files/as-price-list-2017.xlsx.

Sherwood L, Hillar K, Yancey PH. 2013. Animal physiology: from genes to organisms, 2nd ed. Boston (MA): Brooks Cole Cengage Learning. 242 p.

Singleton VL, Cilliers JJ. 1995. Phenolic browning: a perspective from grape and wine research. ACS Symposium Series. Washington DC. https://doi.org/10.1021/bk-1995-0600.ch003.

Stonehouse R, Tomita A, Bett KE. 2022. Molecular marker assay from DNA extraction protocol for the USASK Pulse Molecular Biology Laboratory [lab protocol]. Saskatoon (SK): University of Saskatchewan. [Accessed 2024 Feb] https://knowpulse-knowledgebase.github.io/Laboratory-Protocols/02-Micro-CTAB-DNA-Extration-Procedure-For-Fresh-Material/index.html.

Vaillancourt RE, Slinkard, AE. 1992. Inheritance of new genetic markers in lentil (*Lens* Miller). Euphytica. 64(3):227–236. doi:10.1007/BF00046053.

Vandenberg A, Slinkard, AE. (1990). Genetics of seedcoat color and pattern in lentil. J Hered. 81(6):484–488. doi:10.1093/oxfordjournals.jhered.a111030.

Wu D, Sun DW. 2013. Colour measurements by computer vision for food quality control – a review. Trends Food Sci. 29(1):5–20. doi:10.1016/j.tifs.2012.08.004.

Xu S. 2022. Quantitative genetics. Cham: springer international publishing. 19 p.

5 APPENDICES

Appendix A

Table A. Summary of number of images of individual seeds that were saved following the removal of potential contaminants for each imaging cycle and the date range in which imagery occurred.

Imaging Cycle	Images Saved	First Image Acquisition Date	Final Image Acquisition Date
2019 Sutherland Old	88180	2021-03-25	2021-05-12
2019 SPG Old	89494	2021-05-13	2021-08-11
2021 Sutherland Fresh	88841	2021-12-17	2022-01-17
2021 Rosthern Fresh	88219	2022-01-18	2022-02-02
2021 Sutherland Mid	91664	2022-05-31	2022-07-19
2021 Rosthern Mid	84051	2022-07-20	2022-08-18
2022 Rosthern Fresh	93744	2023-01-04	2023-02-12
2022 Sutherland Fresh	95532	2023-03-14	2023-03-28
2021 Sutherland Old	90467	2023-06-05	2023-06-29
2021 Rosthern Old	90578	2023-07-11	2023-07-21
Total	900770		

Appendix B

Table B. The R^2 and p-values of the intercepts (Int.) and slopes of models fit using the lm function in base R to determine if seed coat colours deteriorated during imaging cycles while seeds were in cold storage.

	L*			a*			b*		
Imaging Cycle	R^2	P-value Int.	P-value Slope	R^2	P-value Int.	P-value Slope	R^2	P-value Int.	P-value Slope
2019 SPG Old	0.06	< 0.01	< 0.01	0.09	< 0.01	< 0.01	0.09	< 0.01	0.16
2019 Sutherland Old	0.06	< 0.01	< 0.01	0.02	< 0.01	< 0.01	≈ 0	< 0.01	0.1
2021 Rosthern Fresh	≈ 0	< 0.01	0.37	≈ 0	< 0.01	0.48	0.01	< 0.01	0.06
2021 Rosthern Mid	0.19	< 0.01	< 0.01	0.02	< 0.01	< 0.01	0.01	< 0.01	< 0 .05
2021 Rosthern Old	0.02	< 0.01	< 0.01	0.02	< 0.01	< 0.01	≈ 0	< 0.01	0.78
2021 Sutherland Fresh	0.04	< 0.01	< 0.01	≈ 0	< 0.01	0.23	0.01	< 0.01	< 0. 05
2021 Sutherland Mid	≈ 0	< 0.01	0.96	≈ 0	< 0.01	0.13	0.05	< 0.01	< 0.01
2021 Sutherland Old	0.03	< 0.01	< 0.01	0.02	< 0.01	< 0.01	0.02	< 0.01	< 0.01
2022 Rosthern Fresh	0.04	< 0.01	< 0.01	0.15	< 0.01	< 0.01	0.08	< 0.01	< 0.01
2022 Sutherland Fresh	0.08	< 0.01	< 0.01	≈ 0	< 0.01	0.59	≈ 0	< 0.01	< 0.01

Appendix C

Table C. The ANOVA results of L*a*b* scores between genotype (G); site-years (SY); interactions between G and SY; replications within site-years (REP); and residuals (Res.) for "Fresh," "Mid," and "Old" plots.

			L*			a*			b*		
		Df	Sum Sq	Mean Sq	Pr(>F)	Sum Sq	Mean Sq	Pr(>F)	Sum Sq	Mean Sq	Pr(>F)
Fresh	G	159	1882	11.8	< 0.01	2079	13.1	< 0.01	1053	6.6	< 0.01
	SY	3	2722	907.2	< 0.01	1194	398.1	< 0.01	5793	1931	< 0.01
	SY/ REP	8	58	7.2	< 0.01	222	27.7	< 0.01	91	11.4	< 0.01
	G* SY	477	315	0.7	< 0.01	433	0.9	< 0.01	183	0.4	< 0.01
	Res.	1256	618	0.5		395	0.3		268	0.2	
Mid	G	159	1195	7.5	< 0.01	1311	8.25	< 0.01	376	2.	< 0.01
	SY	1	859	858.6	< 0.01	175	174.8	< 0.01	51.3	51	< 0.01
	SY/ REP	4	0.3	0.1	0.975	6	1.37	< 0.01	3.8	0.95	< 0.01
	G* SY	159	83	0.5	0.867	97	0.61	< 0.01	28	0.18	< 0.01
	Res.	593	359	0.6		171	0.29		59	0.1	
Old	G	159	2439	15.3	< 0.01	1504	9	< 0.01	765	4.8	< 0.01
	SY	3	5396	1799	< 0.01	15972	5324	< 0.01	1371	457.1	< 0.01
	SY/ REP	8	338	42	< 0.01	157	20	< 0.01	35	4	< 0.01
	G* SY	477	590	1	< 0.01	548	1	< 0.01	200	0.4	< 0.01
	Res.	1242	956	0.8		814	1		287	0.2	

Appendix D

Table D. The Tukey HSD results of replications within site-years and imaging cycles.

			L*	a*	b*
Year	Age	Location Reps	p-value	p-value	p-value
2022	Fresh	Rosthern 1-2	0.32	< 0.01	< 0.01
		Rosthern 1-3	< 0.01	< 0.01	< 0.01
		Rosthern 2-3	< 0.05	0.66	≈ 1
		Sutherland 1-2	0.17	0.79	0.14
		Sutherland 1-3	0.42	≈ 1	0.90
		Sutherland 2-3	≈ 1	0.86	0.72
2021	Fresh	Rosthern 1-2	0.89	0.69	0.94
		Rosthern 1-3	0.94	0.99	0.44
		Rosthern 2-3	≈ 1	0.30	0.062
		Sutherland 1-2	0.27	0.95	0.95
		Sutherland 1-3	0.09	≈ 1	0.88
		Sutherland 2-3	≈ 1	0.99	≈ 1
	Mid	Rosthern 1-2	≈ 1	0.89	≈ 1
		Rosthern 1-3	≈ 1	≈ 1	0.60
		Rosthern 2-3	≈ 1	0.96	0.30
		Sutherland 1-2	≈ 1	0.86	≈ 1
		Sutherland 1-3	≈ 1	0.95	0.95
		Sutherland 2-3	≈ 1	≈ 1	≈ 1
	Old	Rosthern 1-2	≈ 1	0.68	≈ 1
		Rosthern 1-3	0.98	0.96	0.81
		Rosthern 2-3	0.98	0.99	0.65
		Sutherland 1-2	0.95	≈ 1	≈ 1
		Sutherland 1-3	0.70	0.97	0.96
		Sutherland 2-3	0.99	0.94	0.99
2019	Old	SPG 1-2	< 0.05	< 0.01	0.48
		SPG 1-3	< 0.01	< 0.01	< 0.05
		SPG 2-3	0.15	< 0.01	0.87
		Sutherland 1-2	< 0.05	0.051	< 0.01
		Sutherland 1-3	< 0.01	0.32	0.42
		Sutherland 2-3	< 0.01	0.97	< 0.01

Appendix E

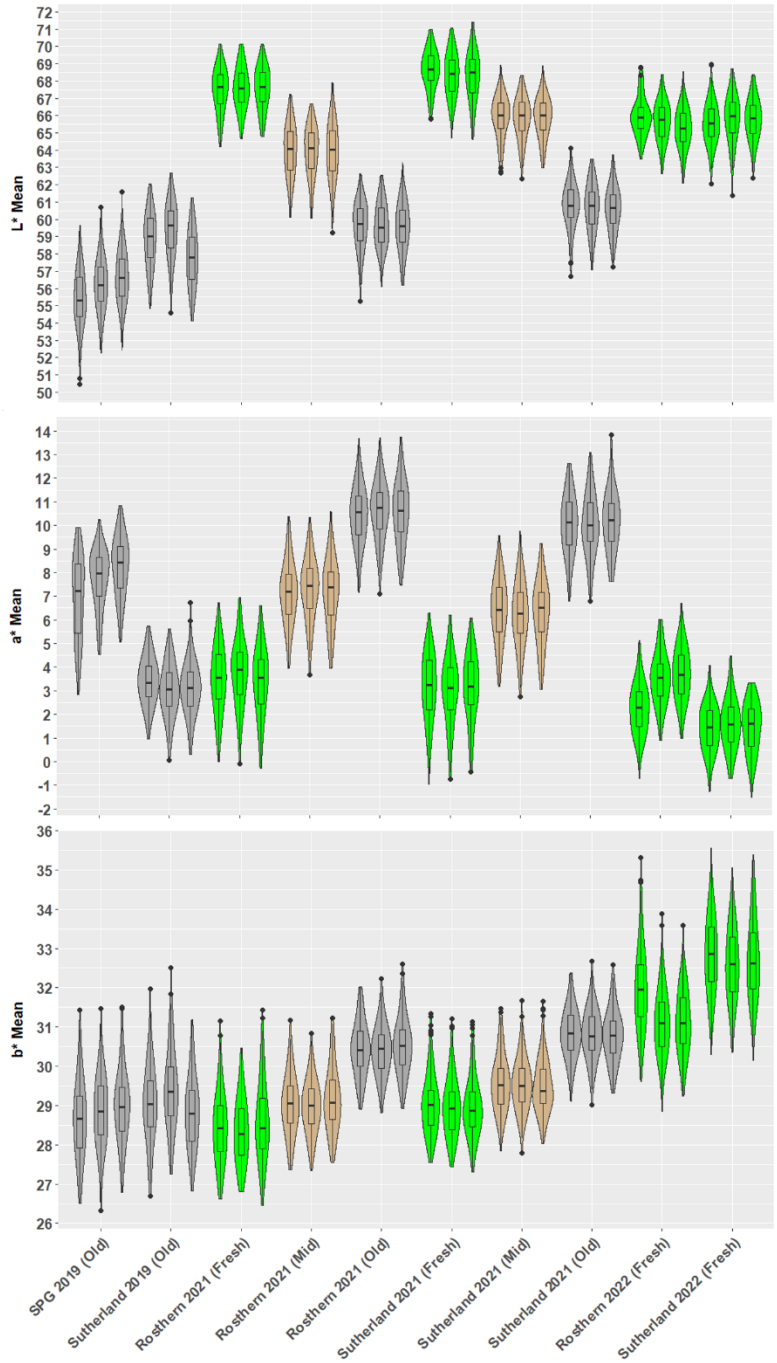

Figure E: The mean CIE L* (top), a* (middle), and b* scores of ≈ 200 images of individual seeds, of 160 Lines of the lentil RIL (LR-06) population grown in three replications in six different site-years. The x-axis depicts the imaging cycle name centered on the 3 replications within that imaging cycle. Imagery occurred when seeds were "Fresh" (≈ 6 months post-harvest; green), "Mid" (≈ 12 months post-harvest; brown), and "Old" (≈ 18 months post-harvest; grey).

Appendix F

Table F. The mean Euclidean Distance (ED) and ΔL*a*b* values of Rosthern and Sutherland 2021 and their percent contribution to the ED and ΔL*a*b* values of POD3 (Period of Deterioration 3).

	Mean				% contribution to POD3			
Location POD	ED	ΔL*	Δa*	Δb*	ED	ΔL*	Δa*	Δb*
Rosthern POD1	5.198	-3.581	3.641	0.5783	48.96	46.22	52.25	31.50
Rosthern POD2	5.662	-4.304	3.307	1.413	52.39	54.08	47.25	70.82
Rosthern POD3	10.81	-7.958	6.999	1.995	100	100	100	100
Sutherland POD1	4.221	-2.544	3.216	0.5433	39.76	32.84	46.15	29.60
Sutherland POD2	6.569	-5.200	3.752	1.294	61.88	67.12	53.84	70.47
Sutherland POD3	10.62	-7.747	6.968	1.836	100	100	100	100

Appendix G

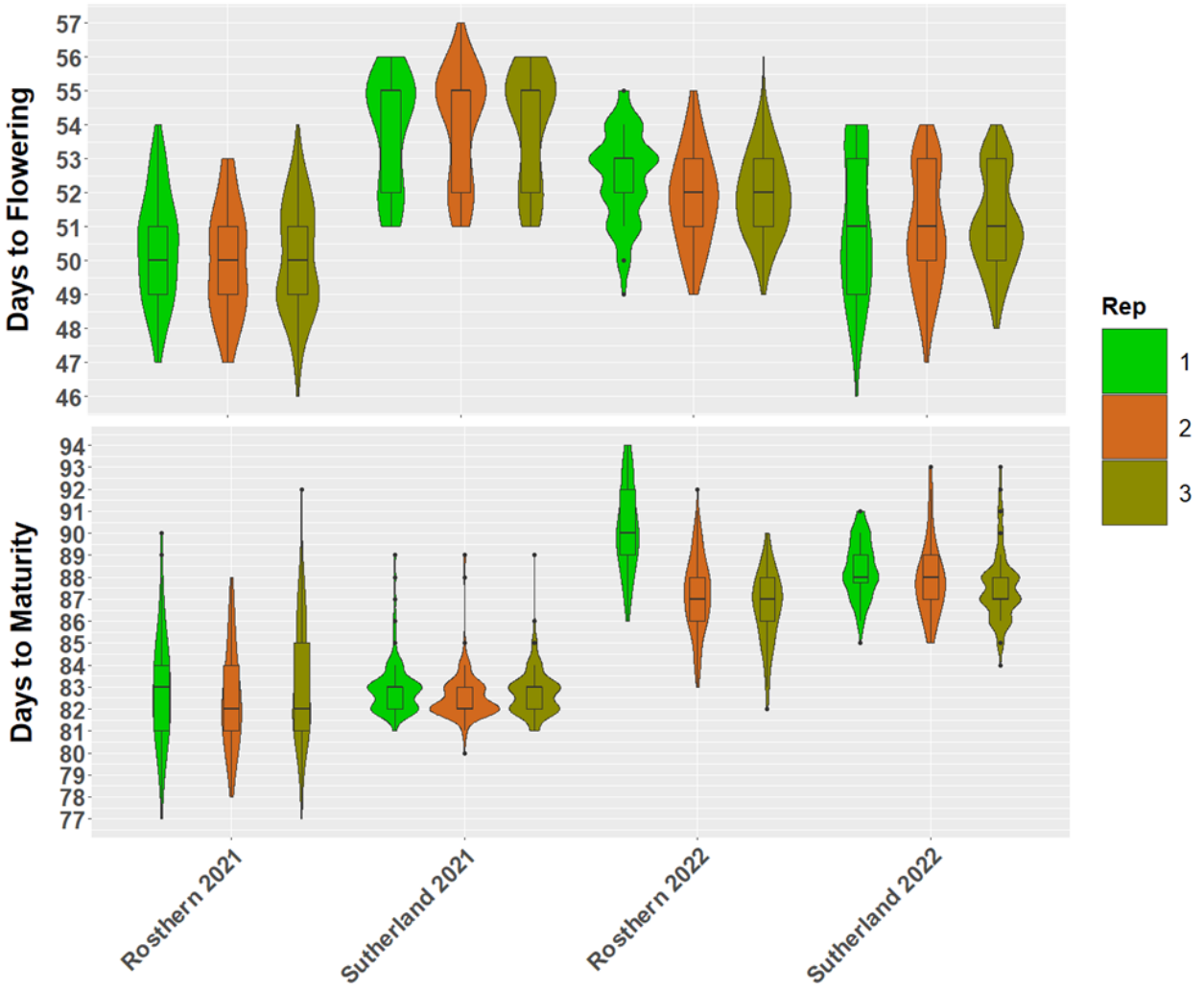

Figure G: The days to flowering (DTF) (top) and the days to maturity (DTM) (bottom) of 160 lines grown in three replications across four site-years (Rosthern 2021, Sutherland 2021, Rosthern 2022, and Sutherland 2022). DTF was defined as the days after planting when 10% of plants in a plot have at least one open flower) and DTM was defined as the days after planting when 10% of plants have 50% of their pods matured.

Appendix H

Table H. The correlation coefficients (cor) of L*a*b* scores with days to flowering (DTF) and days to maturity (DTM).

			L*		a*			b*
	Imaging Cycle	df	cor	p-value	cor	p-value	cor	p-value
DTF	2021 Rost Fresh	474	-0.28	< 0.01	-0.11	< 0.05	-0.06	0.21
	2021 Rost Mid	435	-0.25	< 0.01	-0.16	< 0.01	-0.08	0.09
	2021 Rost Old	476	-0.22	< 0.01	-0.14	< 0.01	-0.16	< 0.01
	2021 Suth Fresh	468	-0.23	< 0.01	-0.17	< 0.05	-0.08	0.077
	2021 Suth Mid	478	-0.24	< 0.01	-0.1	< 0.05	-0.1	< 0.05
	2021 Suth Old	476	-0.24	< 0.01	-0.07	0.15	-0.13	< 0.01
	2022 Rost Fresh	492	-0.27	< 0.01	-0.3	< 0.01	0.031	0.49
	2022 Suth Fresh	492	-0.29	< 0.01	-0.15	< 0.01	0.1	< 0.05
DTM	2021 Rost Fresh	473	-0.28	< 0.01	-0.40	< 0.01	0.31	< 0.01
	2021 Rost Mid	434	-0.27	< 0.01	-0.41	< 0.01	0.23	< 0.01
	2021 Rost Old	475	-0.26	< 0.01	-0.35	< 0.01	0.11	< 0.05
	2021 Suth Fresh	468	-0.11	< 0.05	-0.16	< 0.01	0.13	< 0.01
	2021 Suth Mid	478	-0.07	0.13	-0.21	< 0.01	0.1	< 0.05
	2021 Suth Old	476	-0.12	< 0.01	-0.2	< 0.01	-0.05	0.27
	2022 Rost Fresh	492	≈ 0	0.88	-0.51	< 0.01	0.41	< 0.01
	2022 Suth Fresh	492	-0.33	<0.01	-0.29	< 0.01	0.32	< 0.01

Appendix I

Table I. The means and ANOVA results of lines possessing Favorable (F) versus Unfavorable (U) alleles between the four Significant Regions (SRs) at different seed ages, in which higher L* scores, and lower a* and b* scores are associated with F alleles, whereas lover L* scores, and higher a* and b* scores are associated with U alleles.

SR	Age	Trait	Marker Name	Position	F-Allele	F-Mean	U-Allele	U-Mean	Pr(>F)
1	Fresh	L*	Lcu.2RBY.Chr1.331901891	Left	AA	67.31	GG	66.20	< 0.01
	Fresh	L*	Lcu.2RBY.Chr1.331648200	Right	AA	67.31	CC	66.20	< 0.01
	Mid	L*	Lcu.2RBY.Chr1.331901891	Left	AA	65.52	GG	64.25	< 0.01
	Mid	L*	Lcu.2RBY.Chr1.331648200	Right	AA	65.51	CC	64.25	< 0.01
	Old	L*	Lcu.2RBY.Chr1.331901891	Left	AA	59.36	GG	58.05	< 0.01
	Old	L*	Lcu.2RBY.Chr1.331648200	Right	AA	59.36	CC	58.05	< 0.01
2	Fresh	a*	Lcu.2RBY.Chr1.380727564	Left	CC	2.313	TT	3.114	< 0.01
	Fresh	a*	Lcu.2RBY.Chr1.376986259	Right	TT	2.451	GG	3.106	< 0.01
	Mid	a*	Lcu.2RBY.Chr1.380727564	Left	CC	6.132	TT	7.074	< 0.01
	Mid	a*	Lcu.2RBY.Chr1.376986259	Right	TT	6.325	GG	7.067	< 0.01
	Old	a*	Lcu.2RBY.Chr1.380727564	Left	CC	7.458	TT	8.108	< 0.01
	Old	a*	Lcu.2RBY.Chr1.376986259	Right	TT	7.603	GG	8.103	< 0.01
3	Fresh	b*	Lcu.2RBY.Chr4.479503228	Left	AA	30.17	CC	30.60	< 0.01
	Fresh	b*	Lcu.2RBY.Chr4.481076438	Right	TT	30.18	CC	30.56	< 0.01
	Mid	b*	Lcu.2RBY.Chr4.479503228	Left	AA	29.12	CC	29.46	< 0.01
	Mid	b*	Lcu.2RBY.Chr4.481076438	Right	TT	29.13	CC	29.43	< 0.01
	Old	b*	Lcu.2RBY.Chr4.479503228	Left	AA	29.58	CC	30.03	< 0.01
	Old	b*	Lcu.2RBY.Chr4.481076438	Right	TT	29.57	CC	30.00	< 0.01
4	Fresh	a*	Lcu.2RBY.Chr7.010034595	Left	CC	2.388	AA	3.230	< 0.01
	Fresh	a*	Lcu.2RBY.Chr7.010486621	Right	GG	2.381	AA	3.258	< 0.01
	Mid	a*	Lcu.2RBY.Chr7.010034595	Left	CC	6.303	AA	7.175	< 0.01
	Mid	a*	Lcu.2RBY.Chr7.010486621	Right	GG	6.306	AA	7.196	< 0.01
	Old	a*	Lcu.2RBY.Chr7.010034595	Left	CC	7.478	AA	8.252	< 0.01
	Old	a*	Lcu.2RBY.Chr7.010486621	Right	GG	7.531	AA	8.275	< 0.01
	Fresh	b*	Lcu.2RBY.Chr7.010034595	Left	CC	30.15	AA	30.62	< 0.01
	Fresh	b*	Lcu.2RBY.Chr7.010486621	Right	GG	30.13	AA	30.64	< 0.01
	Mid	b*	Lcu.2RBY.Chr7.010034595	Left	CC	29.02	AA	29.55	< 0.01
	Mid	b*	Lcu.2RBY.Chr7.010486621	Right	GG	29.01	AA	29.56	< 0.01
	Old	b*	Lcu.2RBY.Chr7.010034595	Left	CC	29.59	AA	30.02	< 0.01
	Old	b*	Lcu.2RBY.Chr7.010486621	Right	GG	29.59	AA	30.01	< 0.01