From aerial drone to QTL: Leveraging next-generation phenotyping to reveal the genetics of color and height in field-grown *Lactuca sativa*

Authors

Rens F. Dijkhuizen[1,*], Abraham L. van Eijnatten[1,*], Sarah L. Mehrem[1], Esther S. van den Bergh[2], Jelmer van Lieshout[3], Kiki Spaninks[3], Steven Kaandorp[4], Remko Offringa[3], Marcel Proveniers[2], Guido van den Ackerveken[2], Basten L. Snoek[1,#]

* These two authors contributed equally to this work.

Corresponding author: l.b.snoek@uu.nl

Affiliations

1: Theoretical Biology and Bioinformatics, Institute of Biodynamics and Biocomplexity, Department of biology, Science Faculty, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands

2: Translational Plant Biology, Institute of Environmental Biology, Department of biology, Science Faculty, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands

3: Plant Developmental Genetics, Institute of Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands

4: Bejo Zaden B.V., Trambaan 1, 1749 CZ, Warmenhuizen, The Netherlands

Keywords

Lettuce, Lactuca sativa, drone, UAV, GWAS, natural variation

Abstract

In recent years, the automation of genotyping has significantly enhanced the efficiency of genome-wide association studies. Due to this development, phenotyping is now usually the rate-limiting step, especially in the field. Efforts are now focused on further automating in-field phenotyping. Here we present a GWAS study on 194 field-grown accessions of lettuce (*Lactuca sativa*). These accessions were non-destructively phenotyped at two time points 15 days apart using an unmanned aerial vehicle. Our high throughput phenotyping approach integrates an RGB camera, a multispectral camera to measure the reflectance at 5 wavelengths (blue, green, red, red edge, near-infrared), and precise height estimation. We used the mean and other descriptives such as median, quantiles, minimum and maximum to quantify different aspects of color and height variation in lettuce from the drone images. Using this approach, we confirm several previously described QTLs, now in populations grown under field conditions and identify several new QTLs for plant- height and color.

Introduction

Lettuce (*Lactuca sativa L.*) is one of the most popular salad vegetables worldwide and is an excellent source of dietary minerals, fibers, and phytochemicals with related health benefits (Shi *et al*., 2022). Many different lettuce varieties are grown worldwide, and new varieties are continuously introduced (Wei *et al*., 2021; Zhang *et al*., 2017). Lettuce varieties can be grouped into several genetically and phenotypically distinguishable morphology types, such as Butterhead, Stalk, Crisp or Oilseed (Wei *et al*., 2021). Significant heritable phenotypic variation exists both within and between morphology types. For example, the leaves of different lettuce varieties exhibit a range of colors from green to red or purple (Falcioni *et al*., 2023; Su *et al*., 2020).

Precise phenotyping of color and pigment traits in field-grown lettuce, and the study of underlying genetic pathways, are scientifically and societally relevant, with important implications for consumer health (Falcioni *et al*., 2022). Chlorophyll content is an important determinant of the variation in the greenness of lettuce and has implications for the nutritional state and quality of leafy vegetables (Falcioni *et al*., 2023; Heimler *et al*., 2007). Additionally, chlorophyll itself is considered a good source of several vitamins, important minerals, and essential fatty acids (Shi *et al*., 2022; Yang *et al*., 2022). One gene that has been previously linked to chlorophyll content in lettuce is the transcription factor LsGlk (L. Zhang *et al*., 2022). Conversely, anthocyanins are responsible for the red or purple coloration displayed by many lettuce varieties. *Red Lettuce Leaves* 1 to 4 (*RLL1, RLL2*, *RLL3*, *RLL4*) and anthocyanidin synthase (ANS) are known to have an especially strong effect on anthocyanin content (Su *et al*., 2020; L. Zhang *et al*., 2017). Red lettuce varieties have comparatively higher phenolic and antioxidant levels (Wilson *et al*., 2004), leading to the increasing popularity of red lettuce varieties (Gazula *et al*., 2007).

Color traits are conventionally described using color intensities and vegetation indices (Huang *et al*., 2021; Kureel *et al*., 2022). Widely employed indices to assess chlorophyll content include the Normalized Difference Vegetation Index (NDVI) (Huang *et al*., 2021), the Normalized Difference Red Edge (NDRE) and the Chlorophyll Index Red-Edge (CIRED) (Šiljeg *et al*., 2023). These indices are ratios of reflectance values at specific wavelengths that serve as a proxy for specific plant traits that are not well captured by the individual reflectance values themselves. When measuring color intensities, we commonly rely on the mean or median intensities (Falcioni *et al*., 2023; Kureel *et al*., 2022). However, both descriptions are insensitive to spatial variation. For example, the mean red intensity of a green leaf with some red speckles and a homogeneous slightly red leaf will be the same. This spatial heterogeneity could

potentially be better described by several other descriptives, like quantiles, minimum and maximum.

Another commercially important trait for lettuce is bolting time. Bolting is the transition from the vegetative growth stage to the reproductive growth stage, marked by rapid elongation of the stem (C. Chen *et al*., 2019). Because bolting makes the plant taste bitter, lettuce is typically harvested and consumed in its vegetative growth state. Therefore, delayed bolting and a stable flowering time are desirable traits in lettuce (Leijten *et al*., 2018). Bolting and flowering time are complex traits that have been extensively researched. Flowering time is influenced by ambient temperature, photoperiod, the gibberellic acid pathway, age, and carbohydrate status (Han, Truco, *et al*., 2021). A total of 67 non-overlapping QTLs have been reported for bolting and flowering time (Han, Truco, *et al*., 2021). One especially prominent gene involved in bolting time is Phytochrome C (Rosental *et al*., 2021).

Genome wide-association study (GWAS) is a statistical procedure which can reveal genomic loci associated with phenotypic variation, called quantitative trait loci (QTLs), by associating traits with the allele frequencies of SNPs (Single Nucleotide Polymorphisms) Due to the advancement of next-generation sequencing, the bottleneck for GWAS is no longer genotyping but phenotyping (Spindel *et al*., 2018; Watanabe *et al*., 2017; Xiao *et al*., 2022). Recently, several studies have increased phenotyping efficiency by applying modern drone and sensor technology (Ding *et al*., 2023; Han, Wong, *et al*., 2021; Spindel *et al*., 2018; Ye *et al*., 2023). Drone-assisted measurements provide an opportunity for large-scale in-field phenotyping, using traditional RGB and/or multispectral (MSP) cameras to determine light reflection and harnessing a 2.5D approach to determine plant height. The integration of drone and sensor technologies opens the way for high-throughput and high-resolution phenotyping in field conditions.

In this study, we performed GWAS on plant- color and height traits of 194 *L. sativa* varieties grown in a large field experiment. The traits were obtained using an aerial drone, which imaged the field at two time points 15 days apart using an RGB and a multispectral (MSP) camera to determine light reflection and applying a 2.5D approach to determine plant height. The genotypes of the lettuce varieties were obtained using the assemblies from Wei *et al*., (2021) combined with 67 newly sequenced lines (Van Workum *et al*., 2024). We show that the measured traits form coherent clusters with a shared genetic basis. Our approach is validated through the recovery of several previously identified QTLs, involving members of the anthocyanin pathway, LsGLK and a *Phytochrome C* homolog. Furthermore, we report novel QTLs and candidate genes affecting variation in color and height, demonstrating the power of

drone-assisted phenotyping. We also show the added value of constructing many traits as ratios from the imaging data and describing phenotypes using extended descriptives other than the mean. QTLs can be found dependent on the growth stage of an organism (Bac-Molenaar *et al*., 2016; van Eijnatten *et al*., 2024; Vasseur *et al*., 2014). By phenotyping twice 15 days apart we show that traits from both individual days uncover unique QTLs. Taken together, this study shows that high-throughput, non-destructive, drone-assisted phenotyping can be used to show the interaction between genes, environment, and development under field conditions.

Methods

Fields measurements

For this experiment 194 *Lactuca sativa* accessions (**Supplement Table 1**) were grown under field conditions near Maasbree in the Netherlands. Each accession belongs to a morphology group. These groups are Butterhead, Cos, Crisp, Cutting, Latin, Oilseed and Stalk. We show an example for each group in Figure 1. The accessions shown here are LK061, LK147, LK087, LK183, LK108, LK198 and LK194 respectively. Each accession was grown on two plots, where each plot contained 30 to 40 individual plants. The plants were sowed at a commercial plant nursery on the 25th and 26th of March 2021 and planted in the field on the 27th and 28th of April 2021. The entire field was imaged on two timepoints, on the 11th of June (day-78), and on the 25th of June (day-93). This was done by an unmanned aerial vehicle (UAV), the DJJ Matrice 210 RTK V2, which was equipped with a conventional RGB camera and a multispectral (MSP) camera, the Micasense Altum high resolution. The MSP camera captures blue (475nm, 32nm bandwidth), green (560nm, 27nm bandwidth), red (668nm, 14nm bandwidth), red-edge (717nm, 12nm bandwidth) and near-infrared (NIR 842nm, 57nm bandwidth) light (AgEagle, 2020). Because some plants were removed between day-78 and day-93 for use in destructive measurements (not included in this study), we only examine the average height and color of the plants rather than also looking at the total surface area. The images captured by the drone were stitched together by the DroneWerkers company with geotiff software as well as the estimation of the height by 2.5D software using the overlap and difference in angle between the individual images (DroneWerkers, 2022).

Plot detection

Using the recorded geographic coordinates, individual plots within the field were isolated from the full image. Pixels in the image corresponding to plants were identified by a

thresholding approach using the enhanced vegetation index (EVI). We used different EVI thresholds on the two imaging days, 0.25 and 0.4 respectively, due to differences in the overall brightness of the images between days. These thresholds were visually determined to represent the best tradeoff between not capturing soil pixels while still correctly identifying most plant pixels (**Supplement Figure 1).** The Enhanced Vegetation Index (EVI) worked better for this purpose than NDVI or SR, mainly because the EVI was best at distinguishing between the dark red/purple accessions and soil pixels covered in shadow.

Trait extraction

After isolating the plant pixels per plot, we calculated the traits using the imaging data obtained by the UAV. We calculated these traits over all the plants in a plot (30 - 40), rather than at the level of individual plants. Apart from using the RGB and MSP values as traits, we also used the MSP data to construct several vegetation indices (**Table 1**). These were the normalized difference vegetation index (NDVI), the structure insensitive pigment index (SIPI), the atmospherically resistant vegetation index (ARVI), the chlorophyll index red-edge (CIred), the simple ratio (SR), the enhanced vegetation index (EVI), the normalized difference red edge index (NDRE) and the weighted difference vegetation index (WDVI). We also calculated different ratios of the RGB values and used these as traits (**Supplement Table 2**). We quantified the height by subtracting the mean elevation of the soil pixels of each plot from the height of each plant pixel to account for the unevenness of the terrain. All these traits were calculated for both imaging days and the change in this two-week interval as $\log_2$(day-93/day-78).

Table 1: Vegetation indices used in this study and their formulas. A more complete version can be found in supplement table 2.

Trait	Full name	Formula
ARVI	Atmospherically Resistant Vegetation Index	$\frac{NIR - red - (red - blue)}{NIR + red - (red - blue)}$
CIred	Chlorophyll Index Red Edge	$\frac{NIR}{RedEdge} - 1$
EVI	Enhanced Vegetation Index	$2.5 \cdot \frac{NIR - red}{NIR + 6 \cdot red - 7.5 \cdot blue + 1}$
NDRE	Normalized Difference Red Edge Index	$\frac{NIR - RedEdge}{NIR + RedEdge}$
NDVI	Normalized Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$
SIPI*	Structure Insensitive Pigment Index	$\frac{NIR - blue}{NIR - red}$

SR	Simple Ratio	$\frac{NIR}{red}$
WDVI	Weighted Difference Vegetation Index	$NIR - a \cdot Red;\ \ a = \frac{NIR(soil)}{red(soil)}$

* This index requires light at 800nm, 680nm and 445nm. Since we do not have those exact wavelengths available, we use NIR, red and blue as an approximation.

Extended descriptives

To have a broader description of the present phenotypic variation, we use descriptives to obtain values per accession using the basic R functions (R Core Team, 2023). These are the trimmed means, excluding either the top and bottom 5%, 10% or 40%, the median, the 5%, 10%, 25%, 50%, 75%, 90% and 95% quantile, the minimum and maximum values, and the skewness and kurtosis of the distribution (**Supplement Table 3**).

Phenotype clustering

Many of the traits we quantify are highly similar. To make interpreting this number of traits easier we group them into 10 clusters using K-means clustering from base R (R Core Team, 2023). We visualize these clusters as a network using the R package GGnetwork (Briatte, 2023). In this network all traits with an absolute correlation above 0.8 are connected.

SNP matrix

For genotyping all accessions in our study, we combined raw Illumina short read sequencing data of 130 *L. sativa* accessions from Wei *et al.* (2021) with an additional 67 *L. sativa* accessions from van Workum *et al.* (2024) (European Nucleotide Archive (ENA) at EMBL-EBI, PRJEB63589). SNP calling was done against the Salinas V8 Lettuce reference genome, downloaded from NCBI (GCF_002870075.4, https://www.ncbi.nlm.nih.gov/genome/352)
The Bam files generated in van Workum *et al*. 2024 were used. We added readgroup information to each BAM file to enable using the GATK toolkit (version 4.1.3.0; Poplin *et al*., 2018). The resulting bam files were converted to .cram format for long-term storage using the following command for each LK sample:

```
samtools view --threads 8 -O cram,embed_ref -T reference.fasta -C -o LK.cram LK.bam
```

The variant were called using the readgroup-added and markduped bam files as input for the variantcalling part of the pipeline:

https://github.com/UMCUGenetics/NF-IAP/tree/bqsr_optional_gatk_parameterized

First, we ran the GATK HaplotypeCaller on each sample, which results in a GVCF file which is stored for further use as the input for GATK GenotypeVCF

The following was done on all samples at once. To enable parallel processing, the GVCF files were split into smaller chunks of the genome and all samples are then combined for each chunk (CombineGVCF). Then the combined per chunk GVCFs are Genotyped by GATK GenotypeVCF. After that VariantSelection and filtration is performed (SNP and INDEL) after which these are again combined for all samples for each chunk. This resulted in a combined VCF containing SNP and INDEL for all samples for each chunk. Lastly, all the separate chunks are merged into a single large VCF file.

All default settings were used, unless specified otherwise.

HaplotypeCaller

--params.haplotypecaller.optional = '-ERC GVCF'

VariantFiltration

'SNP' -> "--filter-expression 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0' --filter-name 'snp_filter' --genotype-filter-expression 'DP < 2 || DP > 50' --genotype-filter-name 'dp_fail'"

'INDEL' -> "--filter-expression 'QD < 2.0 || FS > 200.0 || SOR > 10.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0' --filter-name 'indel_filter'"

SelectVariants

'SNP' -> '--select-type SNP --select-type NO_VARIATION'

'INDEL' -> '--select-type INDEL --select-type MIXED'

The SNPS in the obtained VCF were filtered and processed. SNPs were excluded when they had less than 10 of either reference alleles (REF) or alternative alleles (ALT) or more than 10 absent alleles (ABS). Then SNPs were grouped based on linkage. This reduces the size of the data as well as the multiple testing burden. SNPs were inspected for possible groups/clumps in batches of 2000 SNPs. Starting at the first SNP in the list of all SNPs and the first of each following batch of 2000 SNPs, each following SNP that differs from the initial SNP in less than 5 positions, out of 198, were grouped. These SNPs, the SNP under inspection and the SNPs grouped to that SNP) were discarded from the list. Then a new first SNP was considered, and so on, the process repeated until all SNPs were assigned to a group. Positions of each SNP within a group were recorded and can be traced. The allele distribution of the SNP with the highest MAF was taken as the representative. The resulting matrix, with 0 for REF, 1 for HET and 2 for ALT, was used for GWAS.

GWAS

GWAS was performed on all traits using R (version 4.3.1) (R Core Team, 2023) and the lme4QTL package (Ziyatdinov *et al*., 2018). SNPs were obtained from the raw sequences from Wei *et al*. (2021). and from newly sequenced lines, see above (Van Workum *et al*., 2024). The SNP covariance matrix (R-function cov(), **Supplement Table 4**) was used to correct for population structure. For each SNP, we calculated the p-value representing the chance that the SNP is linked to the variation in the measured trait and did this for all traits. To correct for multiple testing, we only consider SNPs with a $-\log_{10} pvalue > 7$ significantly linked to a trait. Broad-sense heritability for each trait was calculated by taking ratio of the between genotype variance and the total variance, using the mean square values as measures of variance.

QTL selection

From the GWAS results we selected and reported a few prominent loci. For the mean traits we selected any locus where at least 3 SNPs in a window of 5 mega-basepairs (Mbp) have a - $\log_{10}$ (p) score > 8. The lower and upper boundary of the regions was determined by searching for the most distant SNPs with a score of at least 8 in 20 Mb window around the SNP with the highest score. For the extended descriptives we applied stricter criteria. In this case, we selected any locus within a 5 Mbp window, that had at least one SNP with - $\log_{10}$ (p) > 12 and at least two other SNPs with - $\log_{10}$ (p) >11. The threshold for the boundaries was set at - $\log_{10}$ (p) > 11.

Code and data availability

Scripts used for this study are available upon request. Data is available upon request

Results

Phenotypic variation

To explore the possibilities of drone-assisted in-field phenotyping, we grew 194 lettuce accessions (**Supplement Table 1**) near nature area "de Peel" in the Netherlands (**Figure 1**). All data was obtained with the help of an unmanned aerial vehicle (UAV). Drone phenotyping was done on the 11th of June 2021 (day-78) and on the 25th of June 2021 (day-93). We quantified 75 traits from this data (**Supplement Table 5**). These traits consist of the RGB and MSP values, several common vegetation indices and height. In addition, we also used several color ratios as traits, the green-blue ratio, red-green ratio, red-blue ratio, green + blue over red ratio, red + blue over green ratio and the red + green over blue ratio (**Supplement Table 2**). First, we

performed GWAS on the mean traits and later we investigated the added benefit of taking several extended descriptives such as the quantiles or the standard deviation.

Substantial variation in these 75 traits was observed across the seven different lettuce morphology types, both within and between days (**Figure 1, Figure 2, Supplement Figure 2, 3)**. For example, cos, cutting, and stalk lettuce types had the largest increase in height between the two measuring days, whereas the oilseed type had the smallest increase (**Figure 2A**). The red edge trait, which describes the reflection on the edge of visible to infra-red light, was found to increase from day-78 to day-93. Oilseed types were found to have the lowest values whereas butterhead types had the highest values, especially at day-78 (**Figure 2B**). The green/blue ratio, which describes the color of the leaves was found to increase from day-78 to day-93 for almost all accessions. The negative values are associated with purple-colored accessions, most of which are found in cutting lettuce (**Figure 2C**).

Most of the phenotypic variation was found to be highly heritable (**Supplement Figure 4, 5, Supplement Table 6**). The average heritability on day-78 and day-93 is 82.7% and 84.1% respectively. Whereas the change between both days is slightly less to hardly heritable with an average heritability of 61.4%. This is most noticeable in traits describing the ratios between the different colors. From the different trait descriptions, the minimum and skewedness had the lowest heritability's, yet the maximum and higher quantiles often show a slightly higher heritability than the mean trait values. Overall, these 194 lettuce varieties encapsulate considerable heritable phenotypic variation which enables the investigation of the genetic architecture of these traits.

Figure 1: Seven different types of lettuce from the perspective of the unmanned aerial vehicle. One accession per lettuce type is shown. The name of the type is shown on top of each panel.

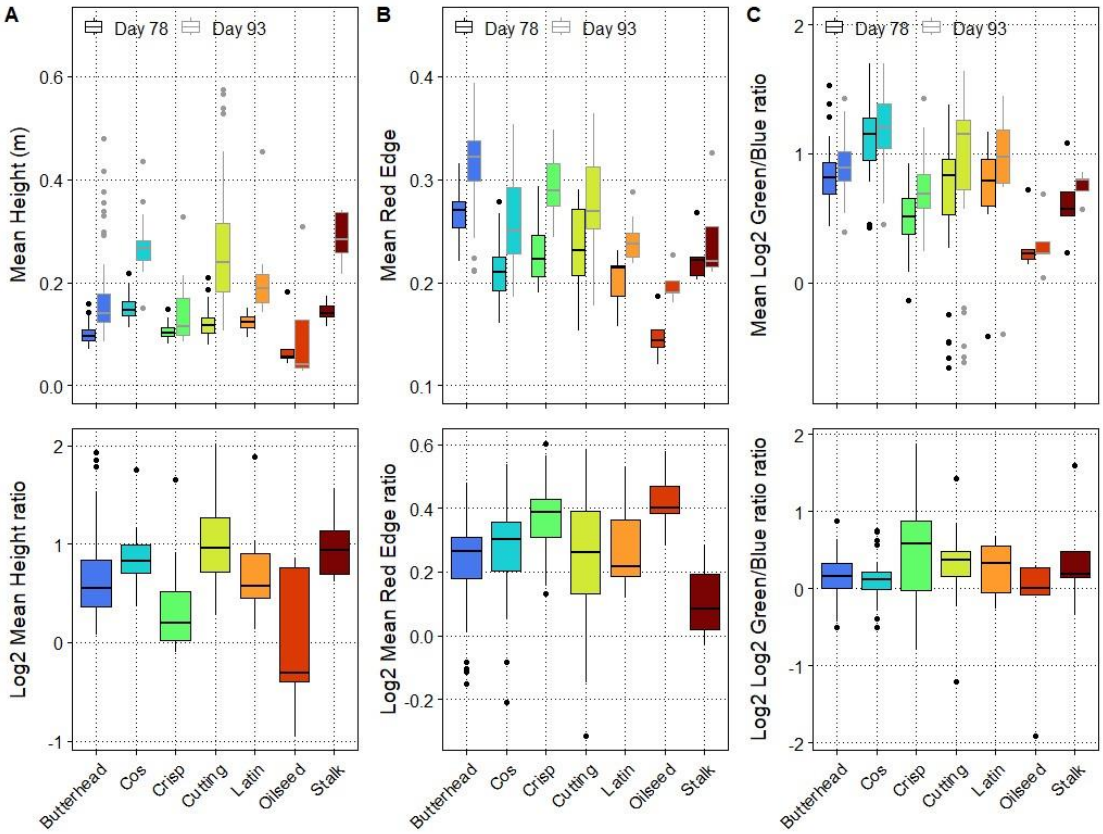


Figure 2: Variation in selected traits between seven lettuce types over two timepoints. **A**) Mean height of seven lettuce types over two time points **B**) Mean Red Edge values from the Multispectral camera **C**) Mean Green/Blue ratio from the RGB camera. Upper panels show the values at day-78 and day-93 and lower panels show the $\log_2$ ratio between those two days. Lettuce types are indicated underneath and by the different colors.

Correlation between traits

By quantifying the height and color on two days and the change between those days we created 75 traits. We expected many of these traits to be highly correlated and lead to the same QTLs. A high correlation, both positive and negative, likely indicates a shared genetic architecture. To quantify the similarity between the traits, we calculated the Pearson correlation between all traits (**Supplement Table 11**). Many traits showed high correlation, such as height at day-78 and height at day-93. This way, we constructed a correlation network with traits as nodes and correlations exceeding 0.8 as edges (**Figure 3**). Most (68 of 75) of the traits share edges and can be found in clusters. The rest of the traits (7 of 75) are isolated from the rest of the network. Some just fall short of the strict correlation threshold. Others, such as the change in Red-blue ratio are not correlated to any of the other traits (**Supplement Table 11**). We used K-means, an unsupervised clustering method, to divide the network into 10 clusters based on the squared correlation matrix of the traits. (**Figure 3)**. We then gave all these clusters a name representing the most abundant traits in the cluster to make discussion and

interpretation easier (**Table 2**). In general, traits from day-78, day-93, and especially the traits describing the change between the two days tend to correlate more within their respective groups than between groups. We also see that the relative blue (**Figure 3, cluster F**), Red-/green (**cluster E**) and Relative red cluster (**cluster I**) are connected in the network. This is as expected as all these traits are a description of how red, purple, or green the leaves are. Other large clusters are cluster B, formed entirely by color ratio traits, cluster C, consisting mostly of several vegetation indices and cluster D, formed by several traits relating to the greenness of the plant. Clusters of traits likely indicate that these traits share part of their genetic architecture and are partial descriptions of a more complex phenotype. Due to the high correlation between traits we expected to find several shared loci using GWAS as opposed to a unique SNP association pattern for each individual trait.

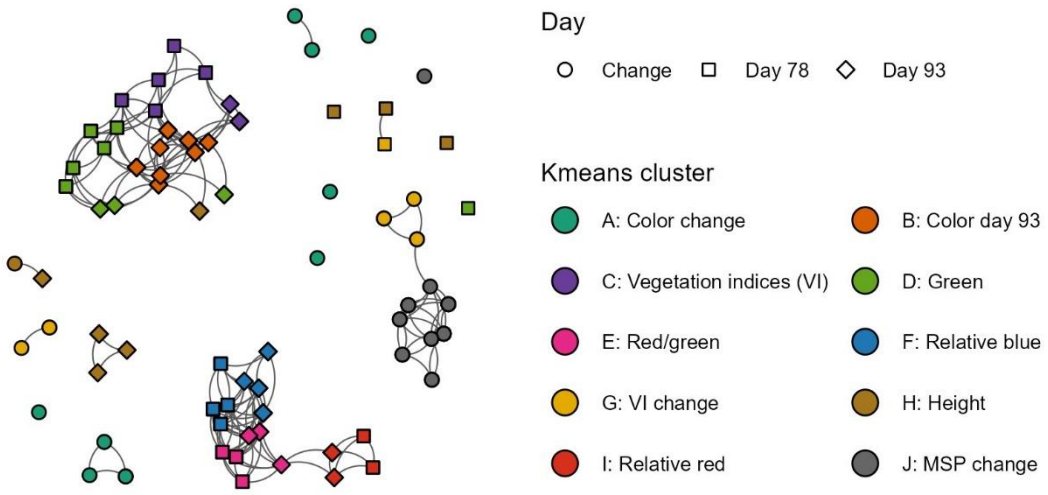


Figure 3: Visualization of correlation between traits using clusters obtained using K-means clustering. K-means clustering on all traits. Each node represents a trait, and each edge represents a correlation > 0.8 between traits. Trait observation type, day-78, day-93 and change as the log2 ratio between the two days are shown by the different shapes. The K-means clusters are indicated by the different colors shown in the legend.

Table 2: Contents of all trait clusters. All traits for clusters A:J sorted by which day the traits were measured. All clusters were given a name that best represents their contents to make discussion easier.

Cluster	Given name	Traits
A	Color change	Change: $\frac{\text{green}}{\text{total color}}$, $\frac{blue}{\text{total color}}$, $\frac{green+blue}{red}$, $\frac{green}{blue}$, $\frac{red}{blue}$, $\frac{red}{\text{total color}}$, $\frac{red+blue}{green}$, $\frac{red+green}{blue}$, $\frac{red}{green}$
B	Color day 93	Day 93: total color, msp1, red, CIred, NDVI, msp3, NDRE, SR
C	Vegetation indices (VI)	Day 78: SR, msp3, NDVI, ARVI, SIPI Day 93: ARVI, SIPI
D	Green	Day 78: green, msp2, total color, red, msp1, msp4

		Day 93: green, msp2, msp4
E	Red / green	Day 78: $\frac{\text{green}}{total\ color}$, $\frac{red+blue}{green}$, $\frac{red}{green}$ Day 93: $\frac{\text{green}}{total\ color}$, $\frac{red+blue}{green}$, $\frac{red}{green}$
F	Relative blue	Day 78: $\frac{blue}{total\ color}$, $\frac{green}{blue}$, $\frac{red}{blue}$, $\frac{red+green}{blue}$ Day 93: $\frac{blue}{total\ color}$, $\frac{green}{blue}$, $\frac{red}{blue}$, $\frac{red+green}{blue}$
G	VI change	Day 78: msp5 Ratio: EVI, ARVI, NDVI, msp5, SIPI
H	Height	Day 78: height, EVI, blue Day 93: height, EVI, blue, msp5, WDVI Change: height
I	Relative red	Day 78: $\frac{green+blue}{red}$, $\frac{red}{total\ color}$ Day 93: $\frac{green+blue}{red}$, $\frac{red}{total\ color}$
J	MSP change	Change: msp1, msp2, msp3, msp4, red, green, blue, total color, SR

Genome wide association

To uncover the polymorphic loci underlying the observed phenotypic variation, we performed single marker/SNP GWAS for all 75 traits. We found significantly associated SNPs ($-\log_{10}(p) > 7$) for 55 traits (**Table 3**). As examples we show the GWAS results for the day-93 Green, Red Edge, and Height traits. (**Figure 4, Supplement Table 7,8,9**). For the variation in height at day-93 we found a locus on chromosome 7 at 164.4 megabasepairs (Mbp) encompassing the previously identified phytochrome C locus associated with bolting and flowering time in lettuce (Rosental *et al*., 2021). This locus is not recovered when only looking at the height at day-78 (**Supplement Figure 6**). This discovery suggests that a substantial number of lettuce varieties started bolting between the first and second day of phenotyping. The green to blue ratio describes whether the color is more green or more purple. This is a good proxy for anthocyanin content. For this trait we recover *RLL2* (chr 5, 86.7 Mbp), *RLL4* (chr 9, 95.1 Mbp) and ANS (chr 9, 152.6 Mbp). All these loci are essential to the typical purple color seen in some lettuce varieties (Su *et al*., 2020). When looking at the red edge intensity we find a locus known as the "pale leaf locus" (chr 4, 106.0 Mbp). This locus harbors the LsGlk gene, which is involved in variation in chlorophyl content (Zhang *et al*., 2022). Recovering these known loci solidifies the reliability of this study and shows the validity of using drones for phenotypic observations.

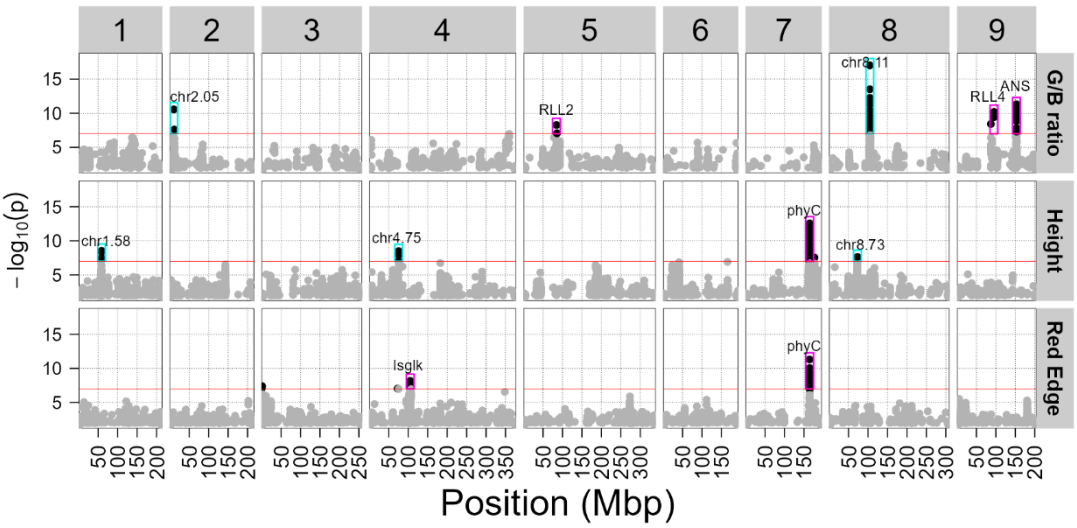


Figure 4: Manhattan plots for G/B ratio, height and Red Edge on day 93 show multiple QTLs. The x-axis represents the genome of *L. sativa* in mega-basepairs (Mbp). Chromosomes are shown on top. The y-axis represents the p-value of the association between SNPs and the specific trait. The red line at -$\log_{10}$(p-value) = 7 represents the multiple testing significance threshold, SNPs above this threshold are shown in black and those below are shown in grey.

To visualize the combined GWAS results of all 55 traits with significant SNPs, we aggregate all SNPs (-log10(p) > 7) into one figure (**Figure 5, Supplement Table 10**), showing each cluster separately. We hypothesized that traits that clustered together based on correlations would have a shared genetic basis. This is largely true, for example the Relative blue (F), Red/Green (E) and Relative red cluster (I) group together phenotypically and show similar GWAS results. The only exception is that the Green cluster (D) correlated mostly with the Color day-93 (B) and VI cluster (C) but shows similar GWAS results to the Red/Green (E) and Relative blue (F) cluster. Prominent QTLs were selected by taking all loci where at least three traits have a -$\log_{10}$ (p) score > 8 within 5 Mbp of each other (**Table 3, Supplement Table 12**). Many prominent QTLs are recovered by several traits from different clusters, such as *RLL2*, *RLL4*, Anthocyanidin synthase and *phyC*. This aligns with expectations as many of the color traits show some overlap and have a high correlation (**Figure 3C**).

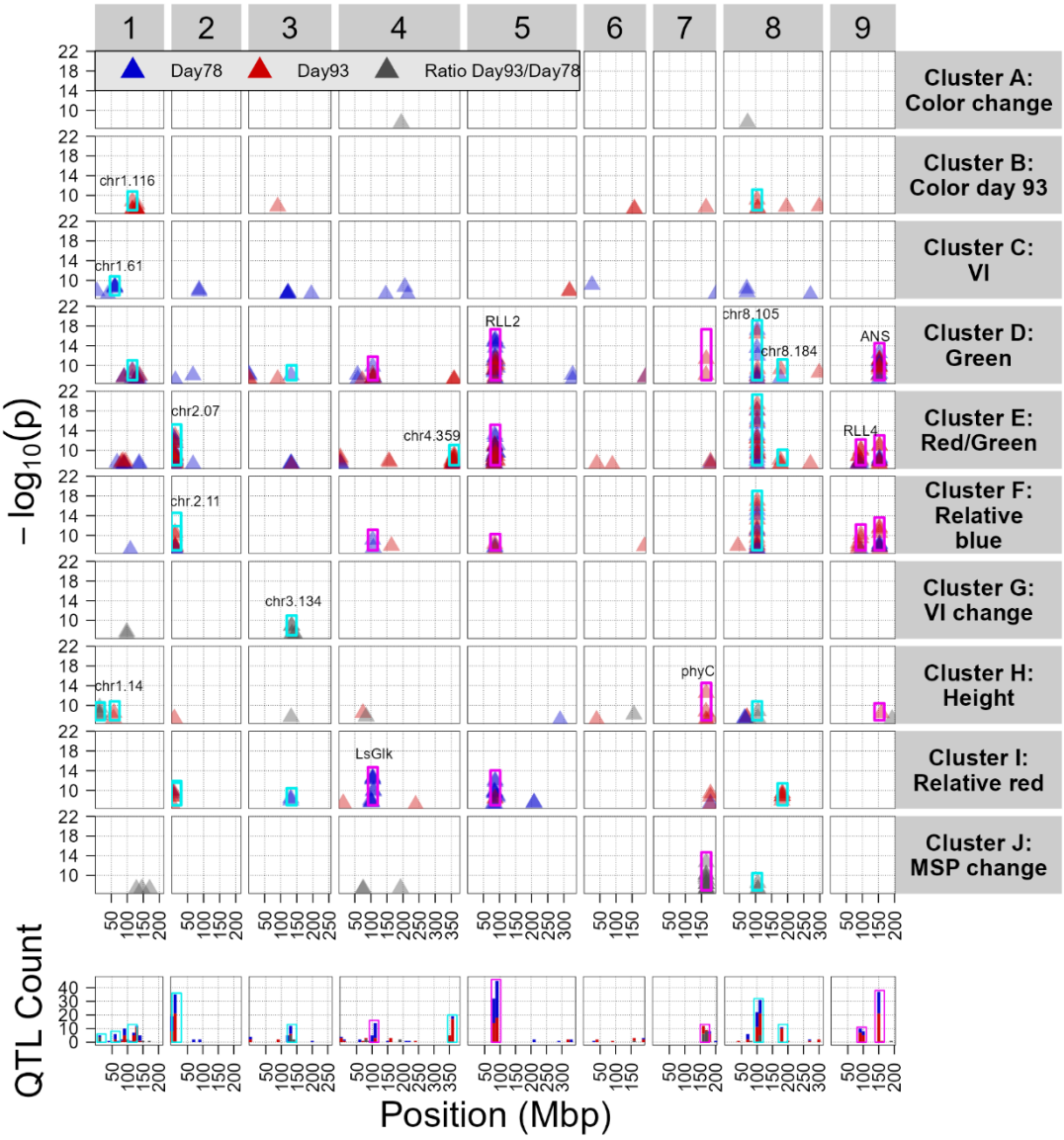


Figure 5. Multi-Manhattan plot of all phenotype clusters. Every row represents one of the clusters, shown at the right. The y-axis represents the -$\log_{10}$ (p-value) of the peak SNPs per trait. Loci that have been previously reported in literature (**Table 3**) are annotated with a purple rectangle and new loci with a cyan rectangle. The color of the points represents whether the traits are from day-78 (blue), day-93 (red), or the difference between the two days (black). The x-axis represents the position on the genome of *L. sativa* in megabasepairs (Mbp). At the bottom a histogram depicts the total amount of significant (-$\log_{10}$(p) > 7) in a 10 Mbp window.

On top of the previously known loci, we also find several that have not been reported before in lettuce. For height at day 93 we report significant loci on chromosome 1 at 58.5 Mbp, chromosome 4 at 75.4 Mbp, and chromosome 8 at 72.8 Mbp (**Figure 4**). The most striking new loci is on chromosome 8 at 105 Mbp for the green to blue ratio and several other traits relating to anthocyanin content. This locus has not been reported in lettuce for any anthocyanin related trait before. When we investigate the most significant SNP further, we find that most accessions that are homozygous for the reference allele show the typical purple color associated with anthocyanin (**Supplement Figure 7**). Close to the most significant SNP (105.48 Mbp) we see the potential candidate gene FQR1 (105.00 Mbp). This gene is described as a flavodoxin like quinone reductase (Laskowski *et al.*, 2002). This protein has already been shown to affect pigmentation in vitro. Another less significant locus can be found on the same chromosome (chr8, 183,81 Mbp) for several traits relating to the green-red ratio **(Supplement table 10)**.

Table 3: All prominent QTLs using mean color and height values. From all the GWAS results we selected prominent QTLs, (all QTLs can be found in **Supplement Table 12**). To qualify a locus needs at least three traits with a score above 8 within 5 Mbp. We reported the statistics of the most trait with the most significant QTL. The first and second columns show the chromosome and position of the most significant SNP of the QTL peak in megabasepairs (Mbp) on the lettuce reference genome v8 (ncbi). The third column shows the SNP's most significant trait. If the locus has been identified before, the source is noted in the source column.

Chromosome	Position Mbp (v8)	Most significant trait	- $\log_{10}$ (p) max	Candidate gene*	Source candidate gene
1	14.06	Relative Red Ratio	9.59		
1	60.95	SIPI day-78	8.82		
1	115.83	Green day-78	9.03		
2	0.70	Red Green ratio day-78	13.25		
2	11.24	Red + Blue over green Ratio day-93	12.92		
3	133.85	NIR Ratio	8.97	GST	(Zhang *et al.*, 2017)
4	106.00	Green + Blue over Red ratio day-78	12.67	LsGlk	(L. Zhang *et al.*, 2022)
4	358.95	Red Green ratio day-93	9.09		
5	86.67	Green day-78	15.34	MYB113/114 (RLL2)	(Su *et al.*, 2020; Zhang *et al.*, 2023)
7	165.02	Red Edge Ratio	12.69	phyC	(Rosental *et al.*, 2021)
8	105.48	Red + Blue over Green ratio day-93	19.33		
8	183.81	Relative Red day-93	9.44		
9	95.11	Red + Blue over Green ratio day-93	10.22	RUP1/2 (RLL4)	(Su *et al.*, 2020)
9	152.57	Green day-78	12.50	ANS	(Zhang *et al.*, 2017)

* For all candidate genes at the locus see supplement table 12. This table includes several suggestions for the new loci identified.

Extended descriptives

In addition to analyzing the mean of the traits, we also investigate several extended descriptives of our measurements to describe different aspects of certain phenotypes, potentially yielding distinct GWAS results. The extended descriptives are the trimmed means, several quantiles, the median, the standard deviation, the minimum and maximum value, and the skewness and kurtosis of the trait distribution (**Supplement Table 3**). To test the usefulness of these descriptives, we compare them to the GWAS results obtained using only the mean trait values. We hypothesized that these descriptives are more sensitive to specific phenotypes than the mean trait value. For example, when a lettuce plant starts bolting this should have a much stronger effect on the maximum height of the plant than the average of plant parts together (**Figure 6A, B**). To test this, we compare the results of all mean height traits with their extended descriptives (**Figure 6C, Supplement Table 14**).

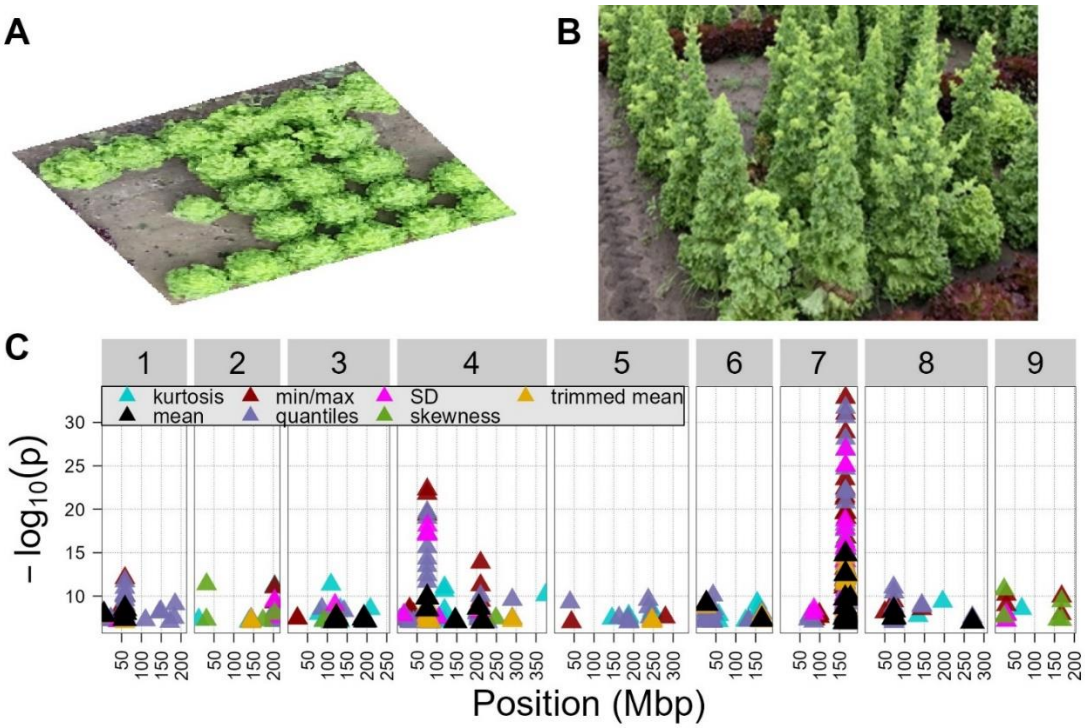


Figure 6. GWAS on traits related to the change of height between day-93 and day-78. A) A lettuce variety showing bolting as seen from a top perspective. B) A lettuce variety showing bolting as seen from a side perspective. C) The GWAS results on all height traits. In black the results of the GWAS on the mean traits and in colors several extended descriptives.

We found that the extended descriptives sometimes represented the traits better than the mean, leading to higher significance in the GWAS. For example, the mean height traits mapped to the *phyC* locus with a -$\log_{10}$ (p-value) of 12.6, yet the maximum height ratio between day-93 and day-78 maps to the same locus with a -$\log_{10}$ (p) score of 32.9 (**Figure 6C)**. Similarly, both analyses identified a locus on chromosome 1 around 60.0 Mbp, with a -$\log_{10}$ (p)

score of 8.5 for the mean height, whereas the QTsL for maximum and quantiles extended descriptives was much more significant with a maximum -$\log_{10}$ (p-value) of 12.1 (**Figure 6C, Supplement Tables 12,13**). Additionally, we uncovered loci not identified when using the mean traits. For example, we recovered a highly significant (-$\log_{10}$ (p) = 22.3) QTL for height at chromosome 4 at 75.5 Mbp that fell short of our strict definition for a prominent QTL when only using mean traits (**Figure 6C**).

Following the same strategy as when focusing on the mean traits we performed a new k-means clustering, combining the mean traits with all extended descriptives (**Supplement Figure 8**). Clustering on all traits is remarkably similar to clustering on only mean traits. Here we also find a height cluster, a red and a green cluster (**Supplement Figure 9**). When used in GWAS we find that the extended descriptives and mean traits share most major QTLs (**Supplement Figure 8**). Notably, using extended descriptives as traits in GWAS resulted often in detecting these shared QTLs with a higher significance. We also recovered the previously reported *RLL3* locus (Su *et al*., 2020), which we did not find using only mean traits, by using the maximum green value on both days (**Supplement Figure 8, Supplement Table 13**). This shows that the utilization of quantiles and maxima is beneficial for identifying novel QTLs. However, the standard deviation, skewness and kurtosis appear to give unreliable results as these descriptives have a low heritability and fail to uncover clear loci (**Supplement Figure 5, 10:19**) while necessitating a stricter significance threshold. In summary, extended descriptives can be useful, especially for GWAS with lower statistical power or when working on complex phenotypes.

Discussion

In recent years, advancements in genotyping technologies have alleviated the bottleneck previously posed in GWAS studies (Spindel *et al*., 2018; Watanabe *et al*., 2017). Now, phenotyping is often considered to be the limiting step, especially in field conditions. In this study we grew 194 lettuce accessions in a field setting and employed drone-assisted imaging and automatic phenotyping to unravel the genetic underpinnings of many diverse color and height traits. Our GWAS uncovered several previously known and several new loci associated with these traits. Apart from using the conventional mean of the traits for GWAS, we also showed that by using several extended descriptives to describe plant traits we recover many loci with a higher significance and even uncover some new loci. In addition, we also show that by using remote sensing to non-destructively phenotype multiple days we can uncover genetic loci that selectively leave a clearer signal in one developmental stage compared to another. By doing this we show a clear interaction between genes, environment, and development.

Confirmed cases

Multiple studies have investigated the genetic background of color and height traits in lettuce before (Falcioni *et al*., 2023; Rosental *et al*., 2021; Su *et al*., 2020; Wei *et al*., 2021). Thus, many of the loci we uncover have been discovered before. Our analysis identified QTLs at *RLL2*, *RLL3*, and *RLL4* (Su *et al*., 2020; Z. Zhang *et al*., 2023) and ANS, the latter being responsible for the pen ultimate step during anthocyanin synthesis (Wilmouth *et al*., 2002; L. Zhang *et al*., 2017). These QTLs are associated with several traits describing the purple color in the plant, which is a proxy for anthocyanin content. We also uncover QTLs at *LsGlk*, a gene that plays a pivotal role in chloroplast development and thus is vital for the typical green color present in many lettuce varieties (Chen *et al*., 2016; L. Zhang *et al*., 2022). For height-related traits, especially the rapid change in height before flowering known as bolting, we identify a QTL mapping to the well-known bolting gene Phytochrome C (Rosental *et al*., 2021). A previous study investigating flowering time and anthocyanin content in lettuce under field conditions already uncovered *phyC*, *RLL2* and *ANS* (Wei *et al*., 2021). We reproduce these QTLs in addition to QTLs for *RLL3, RLL4* and identify multiple novel QTLs. There are several possible explanations for the fact that our study finds additional QTLs. One explanation is simply a difference in population, where additional segregating alleles affecting this trait are present. Another possible reason is that we measure phenotypes at 2 timepoints and many QTLs are only found at specific developmental stages or in certain environmental conditions (Bac-Molenaar *et al*.,

2016; van Eijnatten *et al*., 2024; Vasseur *et al*., 2014). This might also be an explanation for why we do not find a QTL for RLL1. The use of extended descriptives also helps capture different types of phenotypic variation, uncovering novel QTLs . We would not have been able to recover *RLL3* in this experiment using only the mean traits.

New cases

In addition to uncovering several previously reported QTLs, our analysis uncovered previously unreported QTLs. For several novel QTLs we identified candidate genes examining the *Arabidopsis* homologs near the QTLs (**Supplement Table 12**). On chromosome 4 at 75.5 Mbp we find a QTL mostly relating to height and the change in height between both days. Many of these traits also map to phytochrome C (**Figure 4, supplement table 13)**. Close to this QTL are 2 potential candidate genes. Closest is VOZ1, at 76.2 Mbp, a gene known to promote the flowering transition in *Arabidopsis* (Celesnik *et al*., 2013). At 77.1 Mbp is a homolog of SPL3, another gene affecting flowering timing in Arabidopsis (Yamaguchi *et al*., 2009). This QTL on chromosome 4 related to height traits overlaps with a previous QTL identified for flowering time in a RIL population (Han, Truco, *et al*., 2021).

We identified a highly significant QTL on chromosome 8, 105.48 Mbp (**Figure 5**). This QTL consists mostly of several traits relating to the blue to green ratio on day 93. Examining the allele distribution of the most significant SNP in this QTL, we see that most accessions that are homozygous for the alternative allele show the typical purple anthocyanin color. This is a clear indication that this allele is involved in anthocyanin distribution since only a handful of all 194 accessions have this color. (**Supplement Figure 7**). A potential candidate gene for this QTL might be the homolog of Arabidopsis gene FQR1 at 105 Mbp. FQR1 encodes a flavodoxin like quinone reductase which has already been shown to affect pigmentation in vitro (Laskowski *et al*., 2002). Another candidate gene in this QTL is the R2R3 MYB transcription factor MYB119. This transcription factor is known to be involved in the regulation of anthocyanin production (Cho *et al*., 2016)(**Supplement Table 12**).

Timepoints

In most GWAS studies plants are only phenotyped at one timepoint, here we conducted imaging at 78 and 93 days old. We imaged twice because we expected different genes to be relevant depending on the developmental stage of the plant. Here we report that, as expected, some QTLs are unique to traits from either day-78 or day-93. The QTL containing the *LsGlk* gene is almost exclusive to day-78, while the QTL containing the *phyC* gene is

completely exclusive to day-93 and the ratio between the two days. We expected to find the *phyC* gene when looking at the change in height. However, the *phyC* locus is also a QTL for traits like the red-edge reflection and suggests that bolting influences plant-color. Furthermore, we discovered a QTL consisting only of a few ratio-between-days-traits on chromosome 1 at 14.1 Mbp, and a QTL exclusive to day-93 on chromosome 4 at 359.0 Mbp. These results show that similar traits are impacted by different loci based on the growth stage of the plant. This observation is especially striking since the two imaging days are only 15 days apart, which is a relatively short time span for lettuce growth and development. With the rise of automatic phenotyping, this kind of multi-day-, multi-phenotype- data is likely to become more abundant. This enables the discovery of more developmental stage-dependent-QTLs.

Extended Descriptives

In this study we expanded the conventional trait analysis by incorporating various descriptives beyond the mean: the minimum and maximum values, standard deviation, skewness, kurtosis and several quantiles. We have not found any other studies describing traits in this manner. Some of the reported loci would not have been uncovered without these extended descriptives. Especially the maximum and many of the low or high quantiles proved useful to recover known QTLs with a higher significance or even uncover completely new ones. This shows that these descriptives are more sensitive to spatial variation in leaf color or plant height than the mean or median. The skewness, minimum, and kurtosis showed lower heritability than the other descriptives (**Supplement Figure 5**). Although these descriptives did result in some significant SNPs, these usually did not form clear peaks, (**Supplement Figure 6:17**) necessitating stricter significance requirements. Together, this shows that using some extended descriptives can help to discover more relevant genetic variation.

Conclusion

In summary, we used UAV phenotyping to perform GWAS on plant- color and height in lettuce in a large field experiment. We confirmed several previously known QTLs in field conditions as well as identifying several novel QTLs. We demonstrate the advantage of using extended descriptive statistics for phenotypes, revealing several loci that would be missed if we relied only on mean trait values. Additionally, we showed that phenotyping at multiple time points can uncover time-dependent QTLs. This finding highlights that non-destructive phenotyping using a drone can be used to show interaction between genes, environment, and development under field conditions.

Author contributions

Remko Offringa, Marcel Proveniers, Guido van den Ackerveken and Basten L. Snoek conceived the study. Kiki Spaninks, Remko Offringa, Marcel Proveniers, Guido van den Ackerveken, Sarah L. Mehrem, Jelmer van Lieshout, Esther van den Bergh, and Steven Kaandorp performed the experiment. Bram van Eijnatten, Remko Offringa, Sarah L. Mehrem, and Basten L. Snoek extracted the data and performed the investigation. Rens Dijkhuizen, Bram van Eijnatten, and Basten L. Snoek wrote the manuscript with input from all co-authors.

Acknowledgements

We thank the Netherlands Plant Eco-phenotyping Centre (NPEC) for their help with operating the Drone and first steps in data exchange. We also thank Flip Mulder from the Utrecht Bioinformatics Expertise Core (UBEC) for help with generating the VCF.

Funding

This publication is part of the LettuceKnow project (with project number 1.2 of the research Perspective Program P19-17 which is (partly) financed by the Dutch Research Council (NOW; TTW) and the breeding companies BASF, Bejo Zaden B.V., Limagrain, Enza Zaden Research & Development B.V., Rijk Zwaan Breeding B.V., Syngenta Seeds B.V., and Takii and Company Ltd.

References

Bac-Molenaar, J. A., Granier, C., Keurentjes, J. J. B., & Vreugdenhil, D. (2016). Genome-wide association mapping of time-dependent growth responses to moderate drought stress in Arabidopsis. *Plant, Cell & Environment*, *39*(1), 88–102. https://doi.org/10.1111/PCE.12595

Briatte, F. (2023). *Geometries to Plot Networks with "ggplot2" [R package ggnetwork version 0.5.12]*. https://CRAN.R-project.org/package=ggnetwork

Celesnik, H., Ali, G. S., Robison, F. M., & Reddy, A. S. N. (2013). Arabidopsis thaliana VOZ (Vascular plant One-Zinc finger) transcription factors are required for proper regulation of flowering time. *Biology Open*, *2*(4), 424–431. https://doi.org/10.1242/BIO.20133764

Chen, C., Huang, W., Hou, K., & Wu, W. (2019). Bolting, an important process in plant development, two types in plants. *Journal of Plant Biology*, *62*(3), 161–169. https://doi.org/10.1007/S12374-018-0408-9/METRICS

Chen, M., Ji, M., Wen, B., Liu, L., Li, S., Chen, X., Gao, D., & Li, L. (2016). GOLDEN 2-LIKE transcription factors of plants. *Frontiers in Plant Science*, *7*(OCTOBER2016), 1509. https://doi.org/10.3389/FPLS.2016.01509/BIBTEX

Cho, J. S., Nguyen, V. P., Jeon, H. W., Kim, M. H., Eom, S. H., Lim, Y. J., Kim, W. C., Park, E. J., Choi, Y. I., & Ko, J. H. (2016). Overexpression of PtrMYB119, a R2R3-MYB transcription factor from Populus trichocarpa, promotes anthocyanin production in hybrid poplar. *Tree Physiology*, *36*(9), 1162–1176. https://doi.org/10.1093/TREEPHYS/TPW046

Ding, G., Shen, L., Dai, J., Jackson, R., Liu, S., Ali, M., Sun, L., Wen, M., Xiao, J., Deakin, G., Jiang, D., Wang, X., & Zhou, J. (2023). The Dissection of Nitrogen Response Traits Using Drone Phenotyping and Dynamic Phenotypic Analysis to Explore N Responsiveness and Associated Genetic Loci in Wheat. *Plant Phenomics*, *5*. https://doi.org/10.34133/PLANTPHENOMICS.0128/SUPPL_FILE/PLANTPHENOMICS.0128.F2.ZIP

DroneWorkers. (2022). *Home - DroneWerkers*. https://www.dronewerkers.nl/english/

Falcioni, R., Gonçalves, J. V. F., Oliveira, K. M. de, Antunes, W. C., & Nanni, M. R. (2022). VIS-NIR-SWIR Hyperspectroscopy Combined with Data Mining and Machine Learning for Classification of Predicted Chemometrics of Green Lettuce. *Remote Sensing 2022, Vol. 14, Page 6330*, *14*(24), 6330. https://doi.org/10.3390/RS14246330

Falcioni, R., Gonçalves, J. V. F., Oliveira, K. M. de, Oliveira, C. A. de, Demattê, J. A. M., Antunes, W. C., & Nanni, M. R. (2023). Enhancing Pigment Phenotyping and Classification in Lettuce through the Integration of Reflectance Spectroscopy and AI Algorithms. *Plants*, *12*(6), 1333. https://doi.org/10.3390/PLANTS12061333/S1

Gazula, A., Kleinhenz, M. D., Scheerens, J. C., & Ling, P. P. (2007). Anthocyanin Levels in Nine Lettuce (Lactuca sativa) Cultivars: Influence of Planting Date and Relations among Analytic, Instrumented, and Visual Assessments of Color. *HortScience*, *42*(2), 232–238. https://doi.org/10.21273/HORTSCI.42.2.232

Han, R., Truco, M. J., Lavelle, D. O., & Michelmore, R. W. (2021). A Composite Analysis of Flowering Time Regulation in Lettuce. *Frontiers in Plant Science*, *12*, 632708. https://doi.org/10.3389/FPLS.2021.632708/BIBTEX

Han, R., Wong, A. J. Y., Tang, Z., Truco, M. J., Lavelle, D. O., Kozik, A., Jin, Y., & Michelmore, R. W. (2021). Drone phenotyping and machine learning enable discovery of loci regulating daily floral opening in lettuce. *Journal of Experimental Botany*, *72*(8), 2979–2994. https://doi.org/10.1093/JXB/ERAB081

Heimler, D., Isolani, L., Vignolini, P., Tombelli, S., & Romani, A. (2007). Polyphenol content and antioxidative activity in some species of freshly consumed salads. *Journal of Agricultural and Food Chemistry*, *55*(5), 1724–1729. https://doi.org/10.1021/JF0628983/ASSET/IMAGES/LARGE/JF0628983F00002.JPEG

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, *32*(1), 1–6. https://doi.org/10.1007/S11676-020-01155-1/FIGURES/2

Kureel, N., Sarup, J., Matin, S., Goswami, S., & Kureel, K. (2022). Modelling vegetation health and stress using hypersepctral remote sensing data. *Modeling Earth Systems and Environment*, *8*(1), 733–748. https://doi.org/10.1007/S40808-021-01113-8/FIGURES/19

Laskowski, M. J., Dreher, K. A., Gehring, M. A., Abel, S., Gensler, A. L., & Sussex, I. M. (2002). FQR1, a Novel Primary Auxin-Response Gene, Encodes a Flavin Mononucleotide-Binding Quinone Reductase. *Plant Physiology*, *128*(2), 578–590. https://doi.org/10.1104/PP.010581

Leijten, W., Koes, R., Roobeek, I., & Frugis, G. (2018). Translating Flowering Time from Arabidopsis thaliana to Brassicaceae and Asteraceae Crop Species. *Plants 2018, Vol. 7, Page 111*, *7*(4), 111. https://doi.org/10.3390/PLANTS7040111

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Auwera, G. A. Van der, Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178. https://doi.org/10.1101/201178

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.1).

Rosental, L., Still, D. W., You, Y., Hayes, R. J., & Simko, I. (2021). Mapping and identification of genetic loci affecting earliness of bolting and flowering in lettuce. *Theoretical and Applied Genetics 2021 134:10*, *134*(10), 3319–3337. https://doi.org/10.1007/S00122-021-03898-9

Shi, M., Gu, J., Wu, H., Rauf, A., Emran, T. Bin, Khan, Z., Mitra, S., Aljohani, A. S. M., Alhumaydhi, F. A., Al-Awthan, Y. S., Bahattab, O., Thiruvengadam, M., & Suleria, H. A. R. (2022). Phytochemicals, Nutrition, Metabolism, Bioavailability, and Health Benefits in Lettuce—A Comprehensive Review. *Antioxidants 2022, Vol. 11, Page 1158*, *11*(6), 1158. https://doi.org/10.3390/ANTIOX11061158

Šiljeg, A., Marinović, R., Domazetović, F., Jurišić, M., Marić, I., Panđa, L., Radočaj, D., & Milošević, R. (2023). GEOBIA and Vegetation Indices in Extracting Olive Tree Canopies Based on Very High-Resolution UAV Multispectral Imagery. *Applied Sciences 2023, Vol. 13, Page 739*, *13*(2), 739. https://doi.org/10.3390/APP13020739

Spindel, J. E., Dahlberg, J., Colgan, M., Hollingsworth, J., Sievert, J., Staggenborg, S. H., Hutmacher, R., Jansson, C., & Vogel, J. P. (2018). Association mapping by aerial drone reveals 213 genetic associations for Sorghum bicolor biomass traits under drought. *BMC Genomics*, *19*(1), 1–18. https://doi.org/10.1186/S12864-018-5055-5/FIGURES/8

Su, W., Tao, R., Liu, W., Yu, C., Yue, Z., He, S., Lavelle, D., Zhang, W., Zhang, L., An, G., Zhang, Y., Hu, Q., Larkin, R. M., Michelmore, R. W., Kuang, H., & Chen, J. (2020). Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnology Journal*, *18*(2), 479–490. https://doi.org/10.1111/PBI.13213

van Eijnatten, A. L., Sterken, M. G., Kammenga, J. E., Nijveen, H., & Snoek, B. L. (2024). The effect of developmental variation on expression QTLs in a multi parental Caenorhabditis elegans population. *G3 (Bethesda, Md.)*, *14*(2). https://doi.org/10.1093/G3JOURNAL/JKAD273

Van Workum, D.-J. M., Mehrem, S. L., Basten, ;, Snoek, L., Marrit, ;, Alderkamp, C., Dmitry Lapin, ;, Flip, ;, Mulder, F. M., Van Den Ackerveken, G., Dick De Ridder, ;, Eric Schranz, ; M, & Smit, S. (2024). Lactuca super-pangenome reduces bias towards reference genes

in lettuce research. *BioRxiv*, 2024.06.20.599299. https://doi.org/10.1101/2024.06.20.599299

Vasseur, F., Bontpart, T., Dauzat, M., Granier, C., & Vile, D. (2014). Multivariate genetic analysis of plant responses to water deficit and high temperature revealed contrasting adaptive strategies. *Journal of Experimental Botany*, *65*(22), 6457–6469. https://doi.org/10.1093/JXB/ERU364

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., Yano, K., Tokunaga, T., Fujiwara, T., Tsutsumi, N., & Iwat, H. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Frontiers in Plant Science*, *8*, 254051. https://doi.org/10.3389/FPLS.2017.00421/BIBTEX

Wei, T., van Treuren, R., Liu, X., Zhang, Z., Chen, J., Liu, Y., Dong, S., Sun, P., Yang, T., Lan, T., Wang, X., Xiong, Z., Liu, Y., Wei, J., Lu, H., Han, S., Chen, J. C., Ni, X., Wang, J., … Liu, H. (2021). Whole-genome resequencing of 445 Lactuca accessions reveals the domestication history of cultivated lettuce. *Nature Genetics 2021 53:5*, *53*(5), 752–760. https://doi.org/10.1038/s41588-021-00831-0

Wilmouth, R. C., Turnbull, J. J., Welford, R. W. D., Clifton, I. J., Prescott, A. G., & Schofield, C. J. (2002). Structure and Mechanism of Anthocyanidin Synthase from Arabidopsis thaliana and have long-established biomedicinal properties, in-cluding inhibition of cell proliferation and antimutagenic, antimicrobial, antiinflammatory, and antihypertensive. *Structure*, *10*, 93–103.

Wilson, P. E., Morrison, S. C., Hedges, L. J., Kerkhofs, N. S., & Lister, C. E. (2004). Phenolics contribute significantly to higher antioxidant activity of red lettuce compared to green lettuce. *Proceedings of the XXII International Conference on Polphenols*. https://www.researchgate.net/publication/266146321

Xiao, Q., Bai, X., Zhang, C., & He, Y. (2022). Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *Journal of Advanced Research*, *35*, 215–230. https://doi.org/10.1016/J.JARE.2021.05.002

Yamaguchi, A., Wu, M. F., Yang, L., Wu, G., Poethig, R. S., & Wagner, D. (2009). The MicroRNA-Regulated SBP-Box Transcription Factor SPL3 Is a Direct Upstream Activator of LEAFY, FRUITFULL, and APETALA1. *Developmental Cell*, *17*(2), 268–278. https://doi.org/10.1016/j.devcel.2009.06.007

Yang, X., Gil, M. I., Yang, Q., & Tomás-Barberán, F. A. (2022). Bioactive compounds in lettuce: Highlighting the benefits to human health and impacts of preharvest and postharvest practices. *Comprehensive Reviews in Food Science and Food Safety*, *21*(1), 4–45. https://doi.org/10.1111/1541-4337.12877

Ye, Y., Wang, P., Zhang, M., Abbas, M., Zhang, J., Liang, C., Wang, Y., Wei, Y., Meng, Z., & Zhang, R. (2023). UAV-based time-series phenotyping reveals the genetic basis of plant height in upland cotton. *The Plant Journal*, *115*(4), 937–951. https://doi.org/10.1111/TPJ.16272

Zhang, L., Qian, J., Han, Y., Jia, Y., Kuang, H., & Chen, J. (2022). Alternative splicing triggered by the insertion of a CACTA transposon attenuates LsGLK and leads to the development

of pale-green leaves in lettuce. *The Plant Journal*, *109*(1), 182–195. https://doi.org/10.1111/TPJ.15563

Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., Yan, C., Jia, Y., Larkin, R. M., Lavelle, D., Truco, M. J., Chin-Wo, S. R., Michelmore, R. W., & Kuang, H. (2017). RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nature Communications 2017 8:1*, *8*(1), 1–12. https://doi.org/10.1038/s41467-017-02445-9

Zhang, Z., van Treuren, R., Yang, T., Hu, Y., Zhou, W., Liu, H., & Wei, T. (2023). A comprehensive lettuce variation map reveals the impact of structural variations in agronomic traits. *BMC Genomics*, *24*(1). https://doi.org/10.1186/s12864-023-09739-x

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., & Soria, J. M. (2018). lme4qtl: Linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics*, *19*(1). https://doi.org/10.1186/s12859-018-2057-x

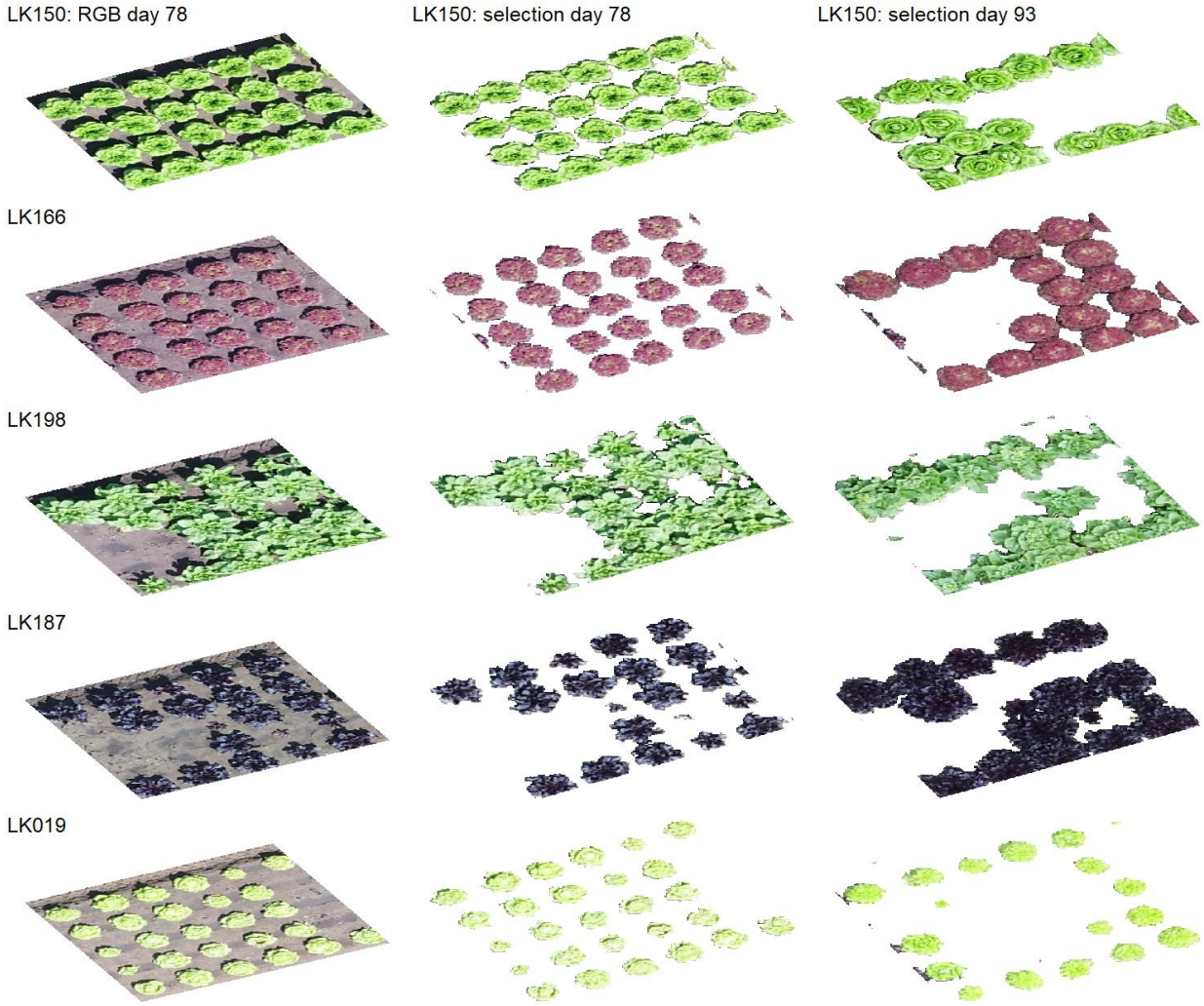
Supplementary material

Supplement tables

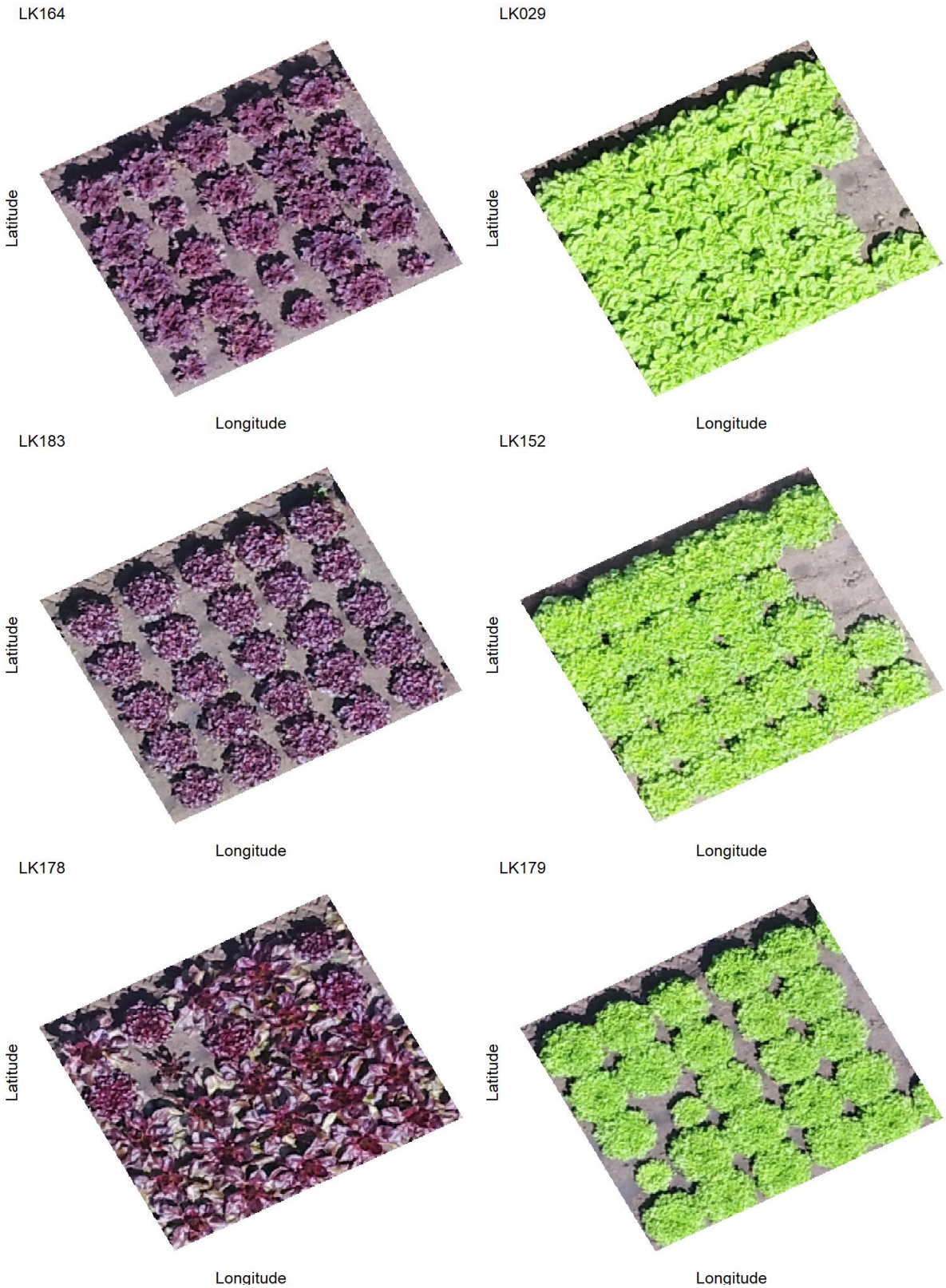
Supplement table 1: LK_lines
Supplement table 2: Pheno_definitions
Supplement table 3: Statistical_definitions
Supplement table 4: kinship
Supplement table 5: Phenotypes
Supplement table 6: Heritability
Supplement table 7: Height.mean.day2
Supplement table 8: gbrat.mean.day2
Supplement table 9: rededge.mean.day2
Supplement table 10: pvalues
Supplement table 11: Correlation matrix
Supplement table 12: QTLs.fig5
Supplement table 13: QTLs.all
Supplement table 14: height.traits

Supplement figures

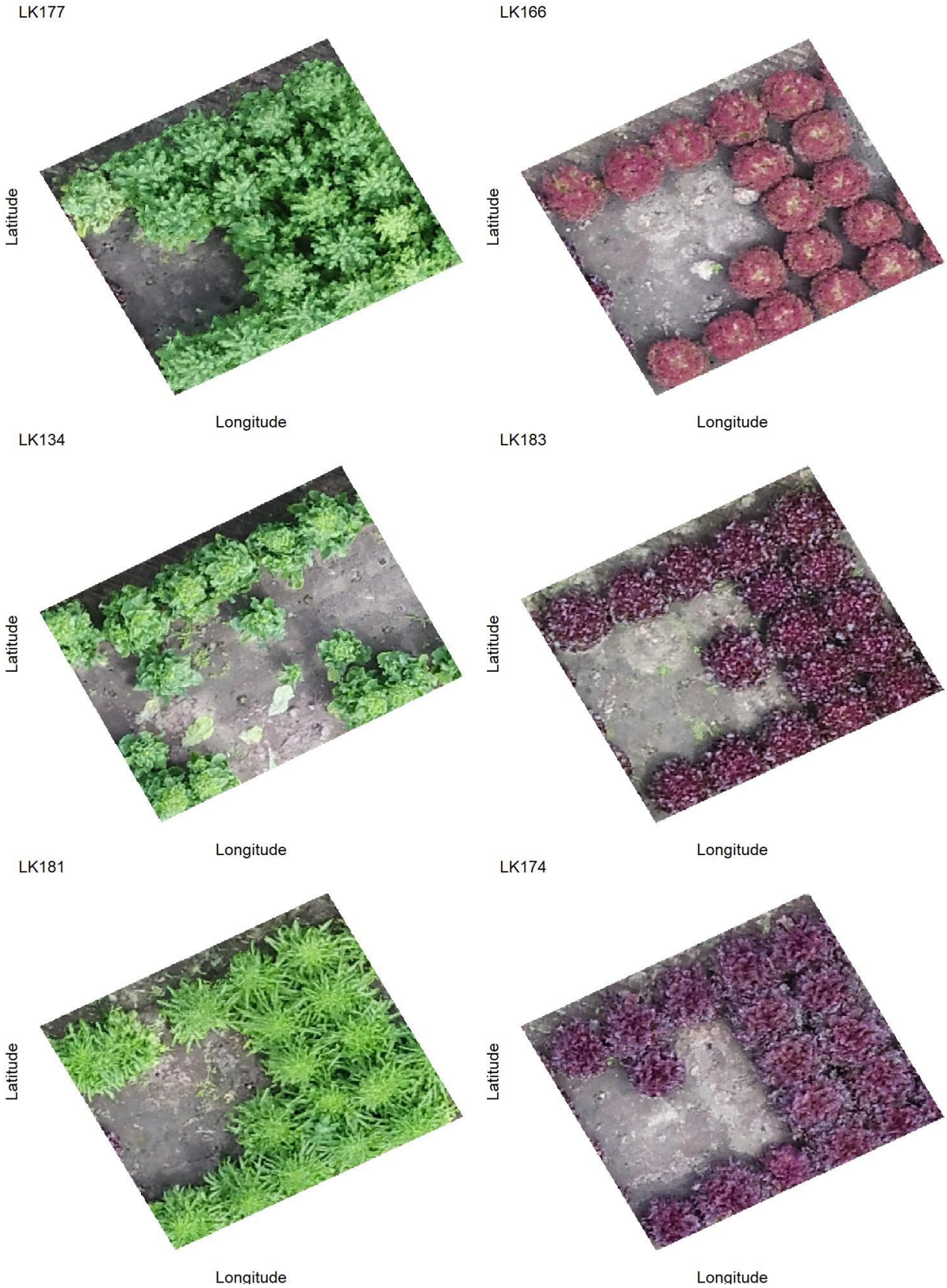
Supplement figures



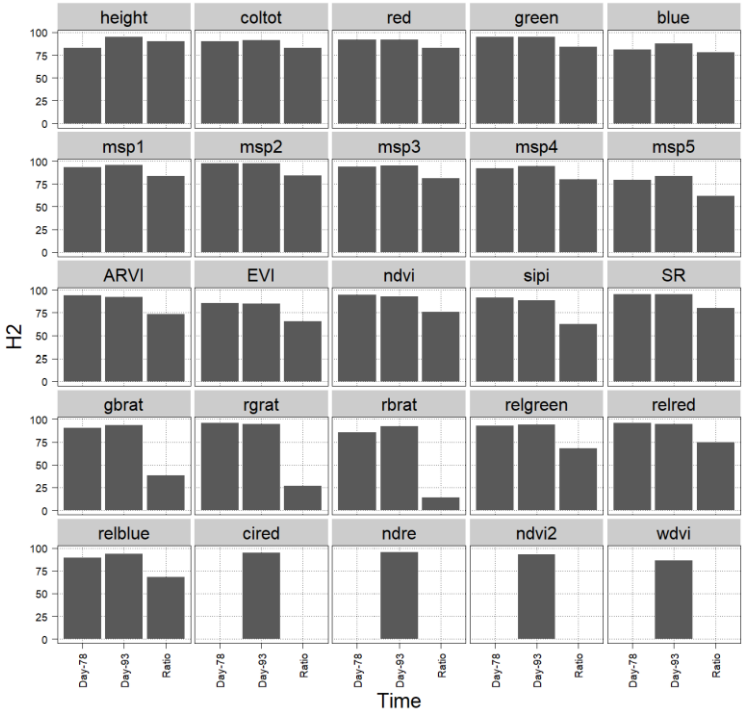
Supplement Figure 1. EVI selection for five accessions. The first column shows the RGB image for the accessions on day 78. The second column shows the plant pixels after thresholding on day 78 with an EVI threshold of 0.25. The third column shows a similar thresholding for day 93 with a threshold of 0.4.



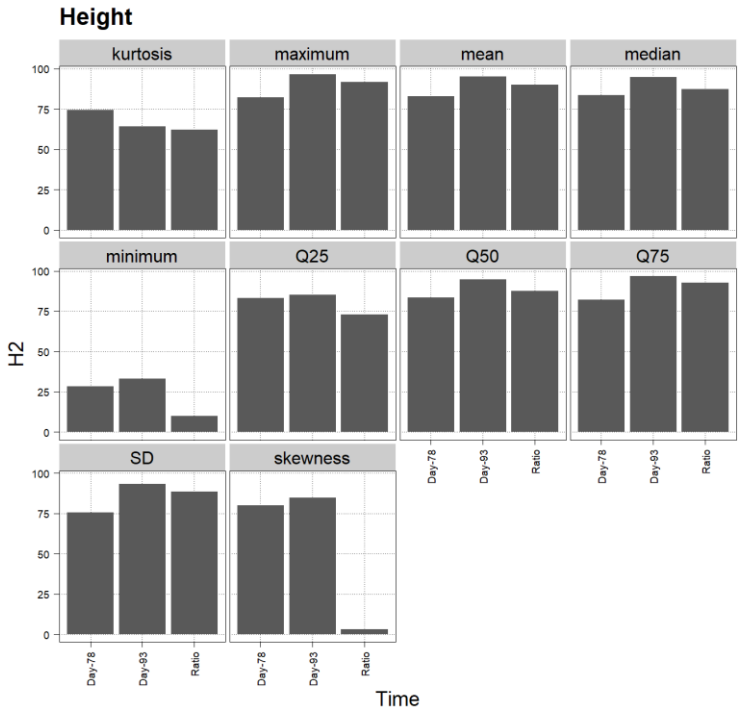
Supplemental Figure 2: Three accessions with extremely low green blue ratio on the left and extremely high on the right.



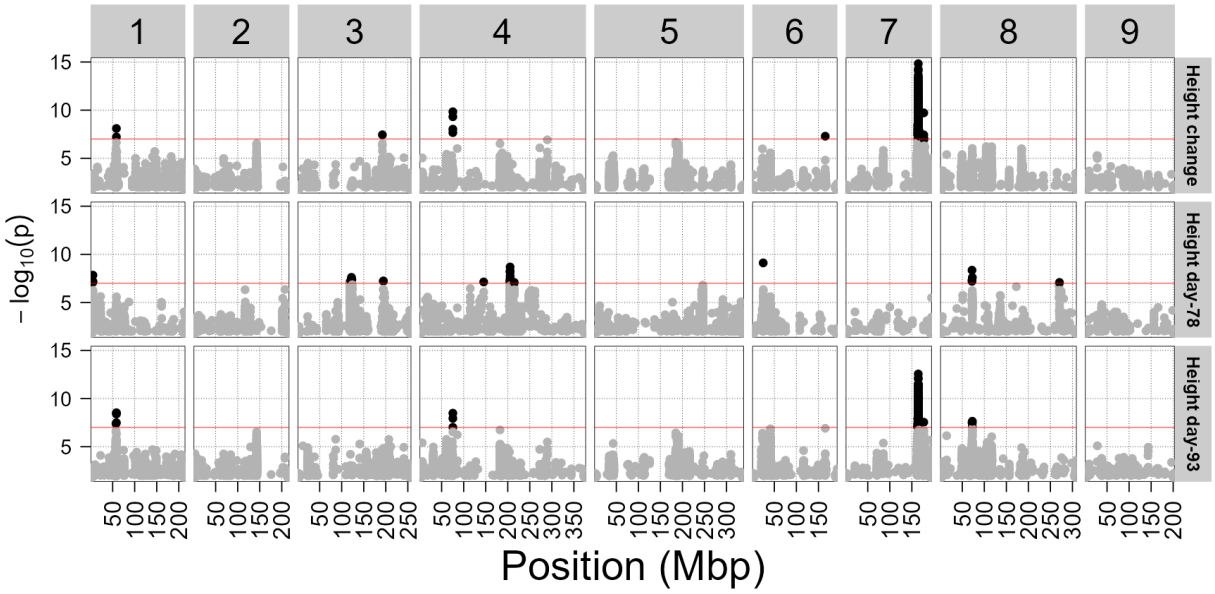
Supplemental Figure 3: Three accessions with extremely low relative red on the left and extremely high on the right.



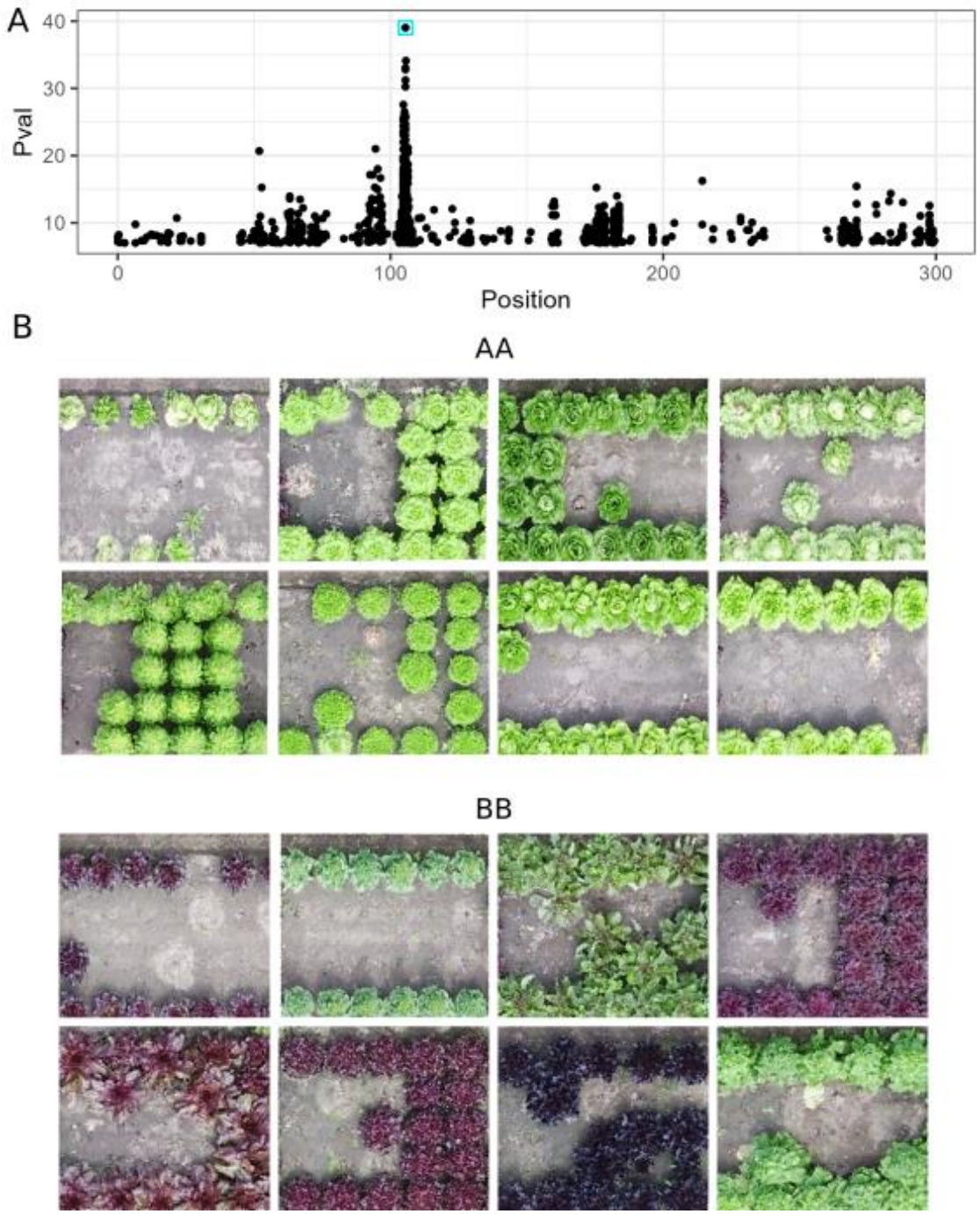
Supplement Figure 4: Broad sense heritability of all mean phenotypes. On day-78, day-93 and the ratio of the trait between those two days.



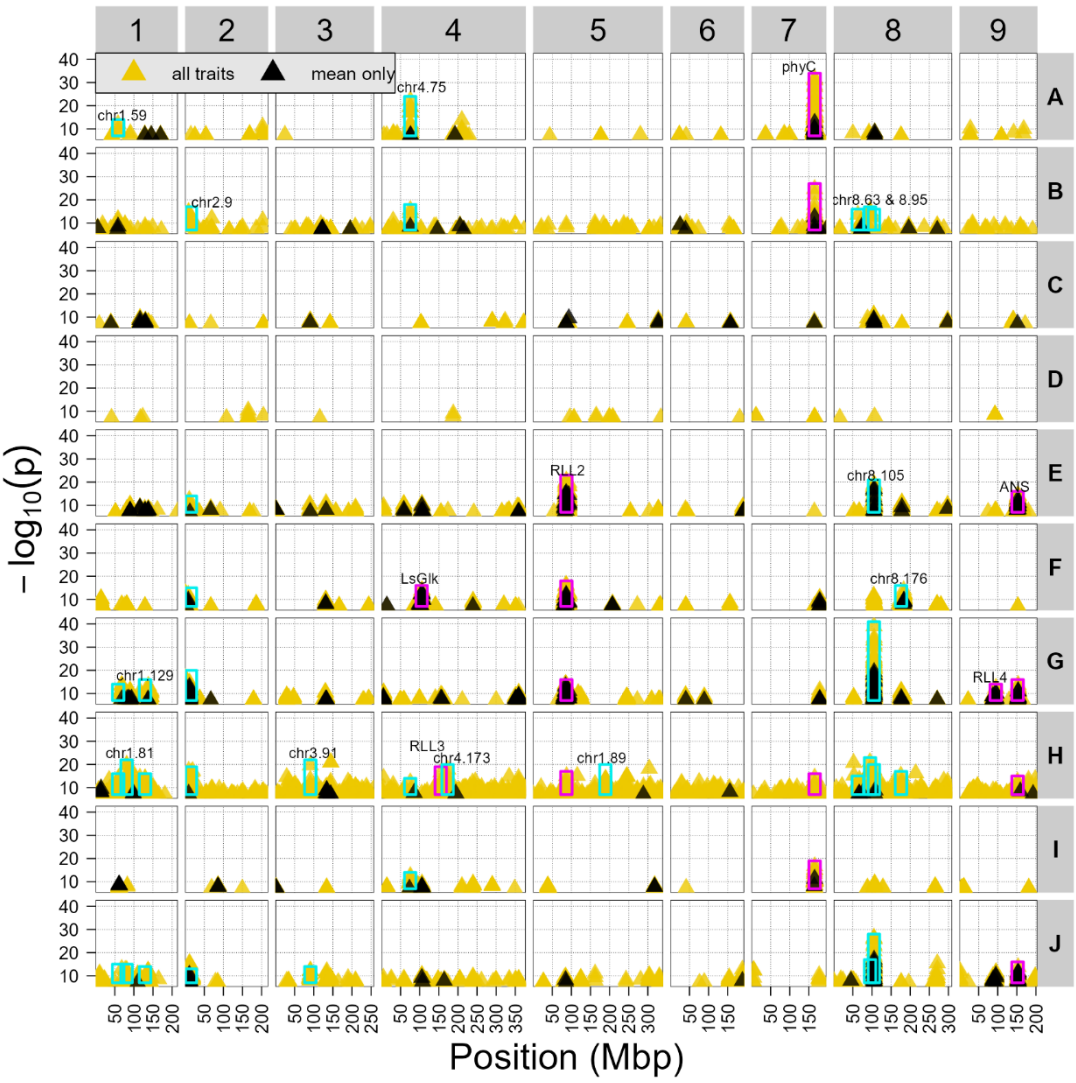
Supplement Figure 5: Broad sense heritability of all descriptives of the height trait. On day-78, day-93 and the ratio of the trait between those two days.



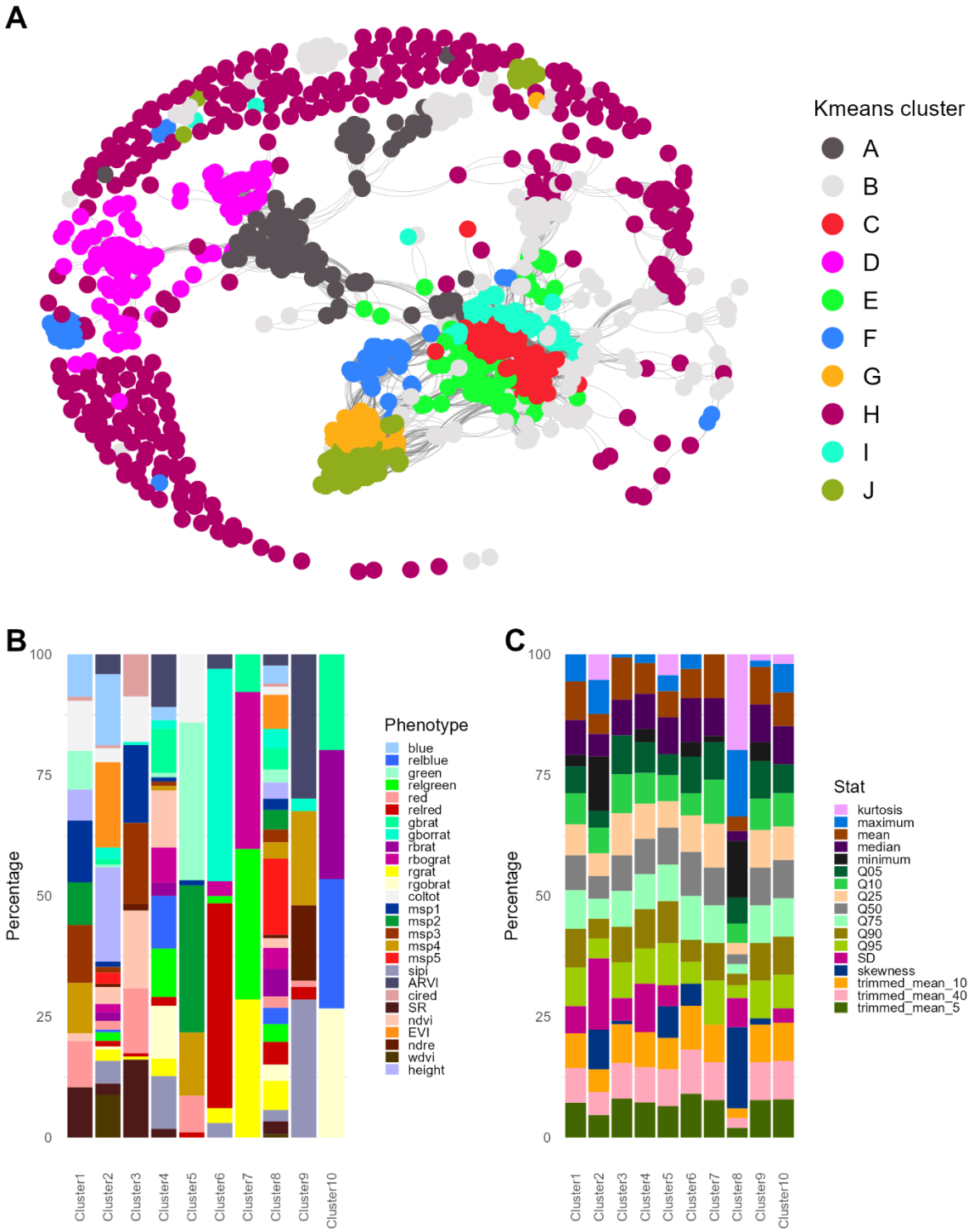
Supplemental Figure 6: Comparison of height on both days and the change between both days. The x-axis represents the genome of *L. sativa* in megabasepairs (Mbp). Chromosomes are shown on top. The y-axis represents the likelihood a SNP is associated to the variation in the specific trait. The red line at -$\log_{10}$(p-value) = 7 represents the multiple testing significance threshold, SNPs above this threshold are shown in black and those below are shown in grey. The traits shown here are the height on day-78, the height on day-93 and the absolute change in height between those days. The peak on chromosome 7 on Phytochrome C is present on day-93 and the change between both days, but absent on day-78.



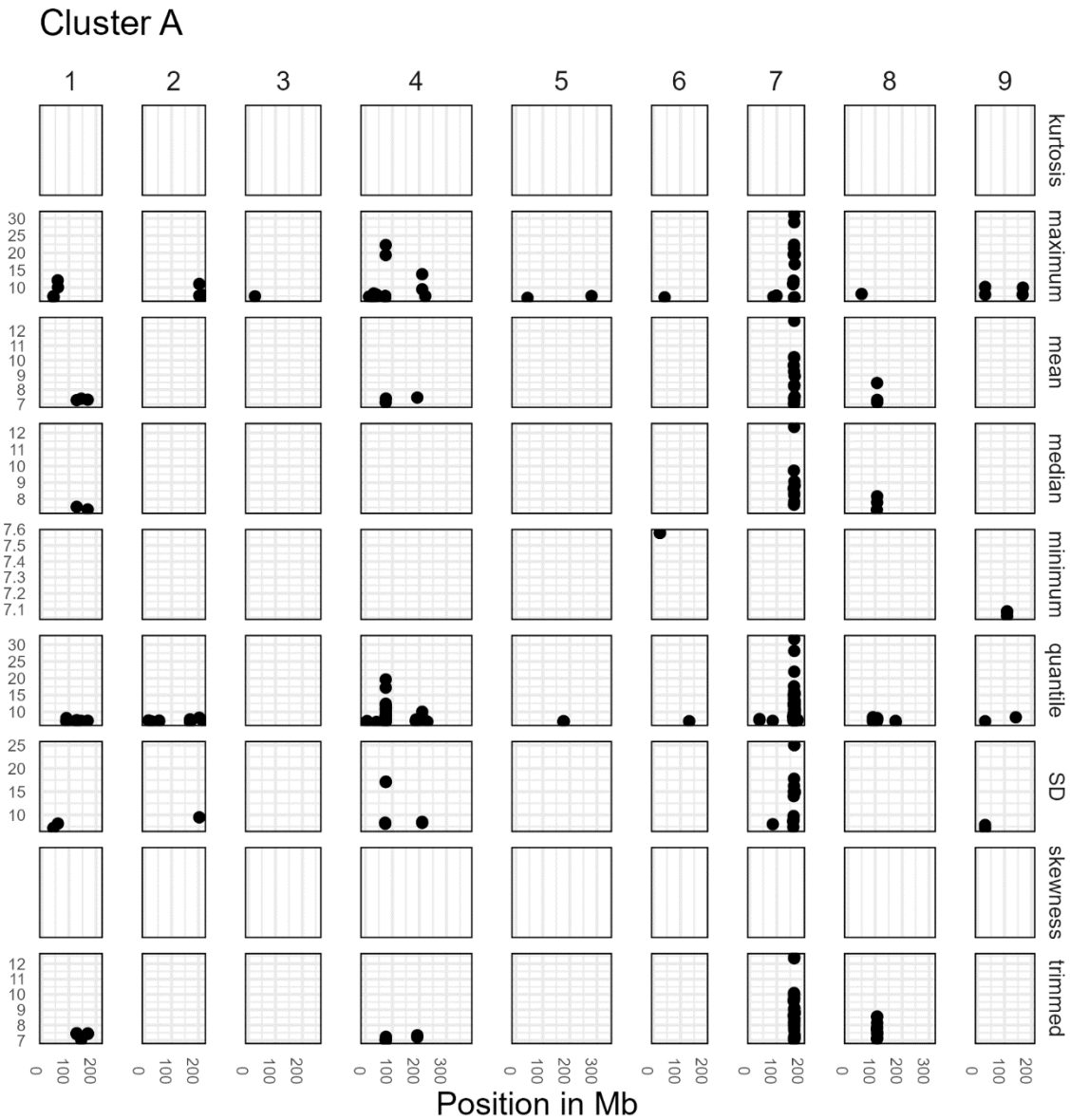
Supplemental Figure 7: The phenotypes causing the QTL on chromosome 8. A) A zoom in on all results for chromosome 8 with the position in megabasepairs on the x-axis and the -log10(pvalue) on the y-axis. A highly significant QTL is present at 105 Mbp circled in teal. B) Phenotypes relating to the significant SNP. When looking at 8 random accessions homozygous for the reference allele we see normal green accessions. When examining all 8 accessions homozygous for the alternative allele 5 out of 8 accessions show the typic purple anthocyanin color.



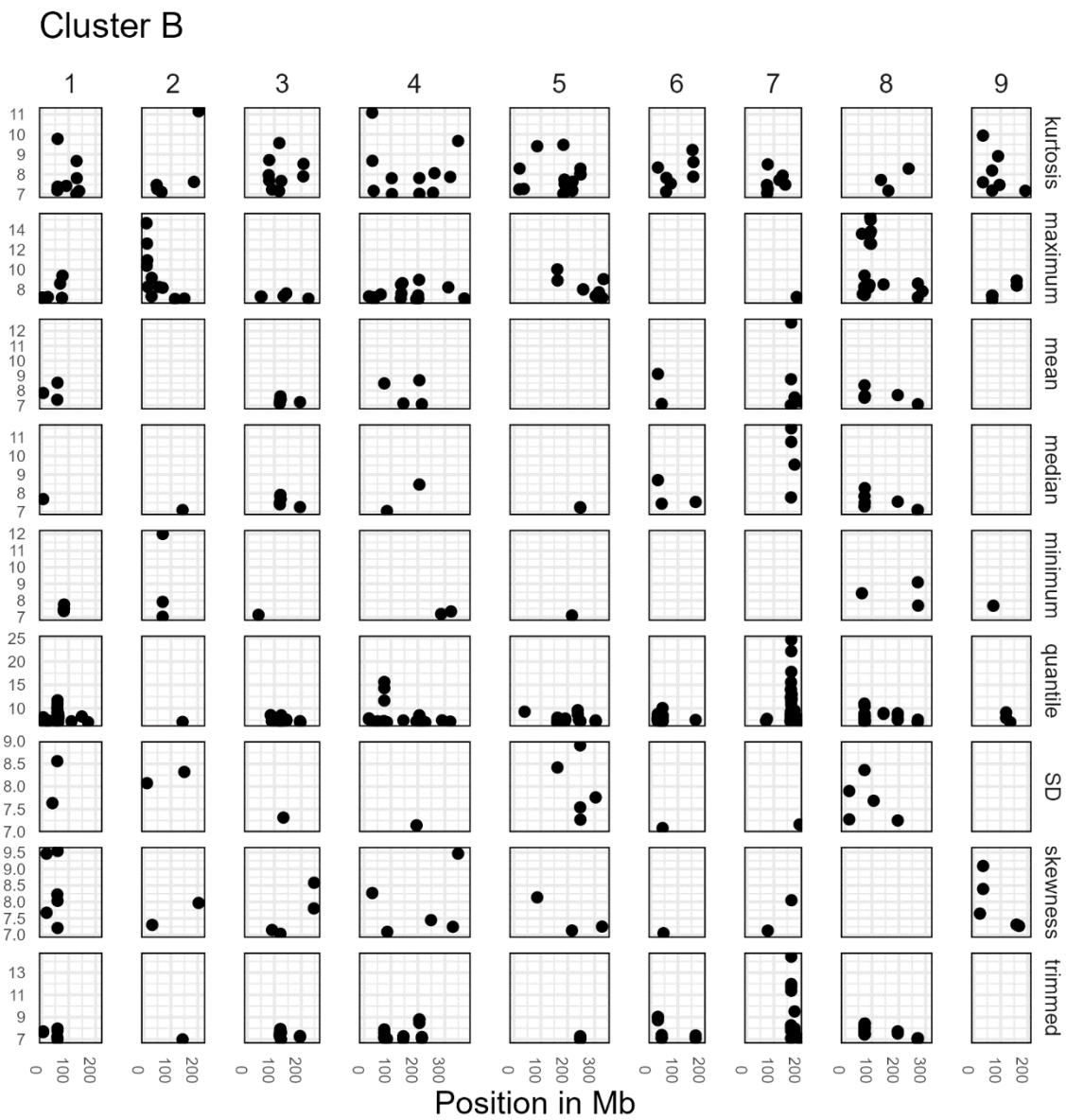
Supplemental Figure 8: Comparison of using only the mean traits or using many extended descriptives. Every row represents a cluster. Clustered on all traits. The black dots represent a mean trait, while the yellow dots represent any of the statistical summaries.



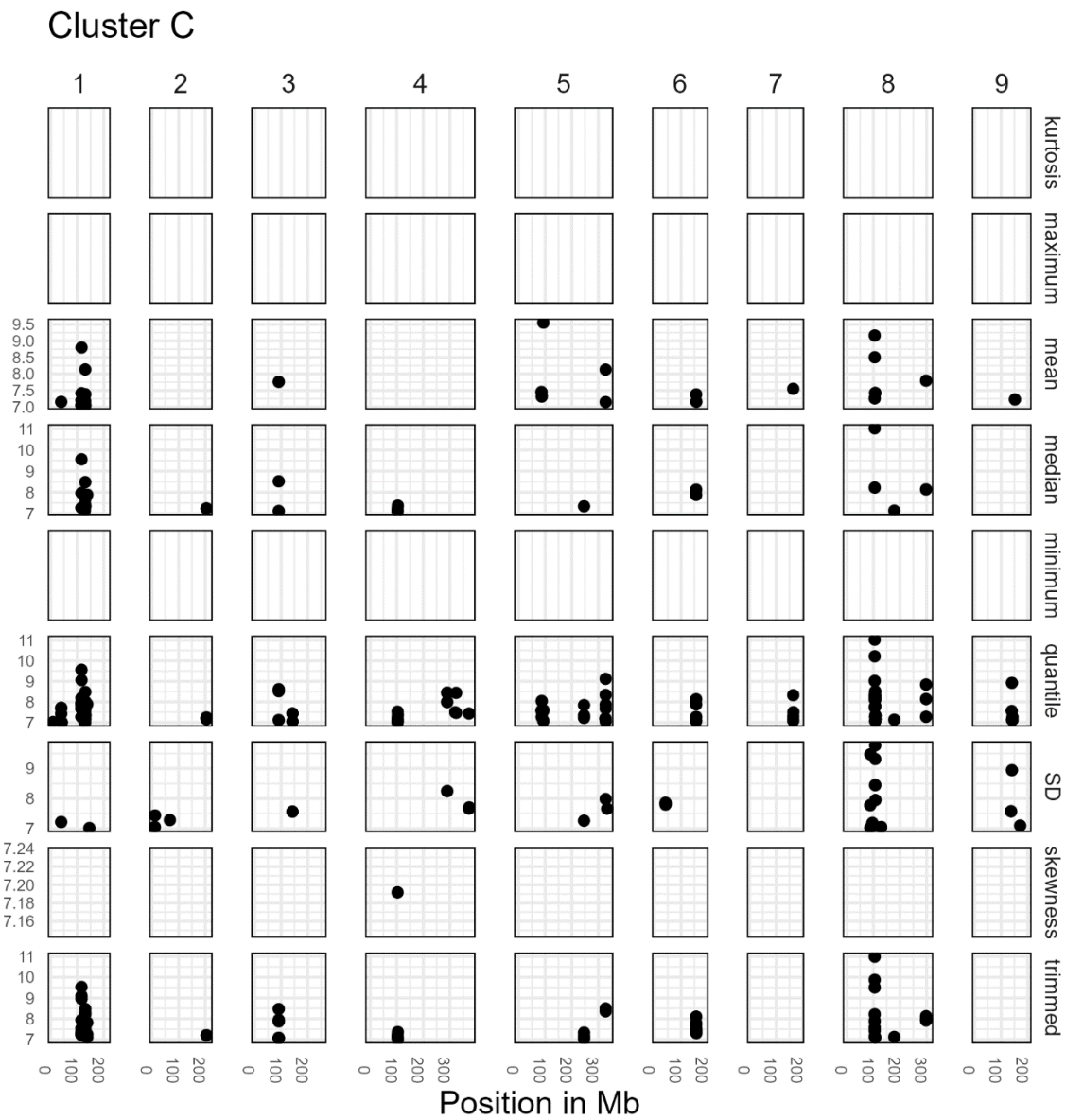
Supplemental Figure 9: Details about the clustering on all traits, including extended descriptives. A) Kmeans clustering on all traits. B) Phenotypic distribution over all clusters. C) Distribution of descriptives over clusters.



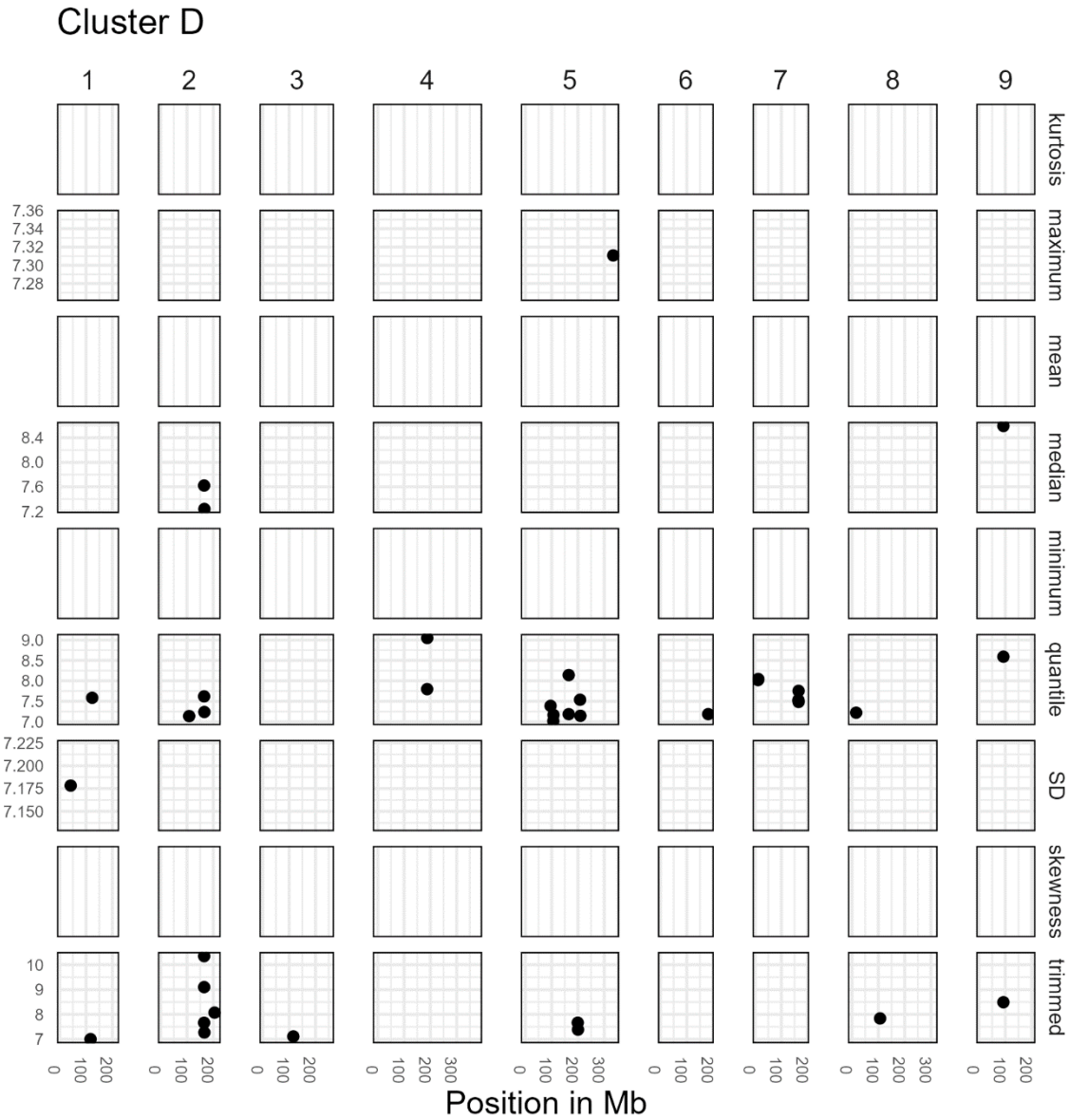
Supplemental Figure 10: The different summary statistics for cluster A plotted separately.



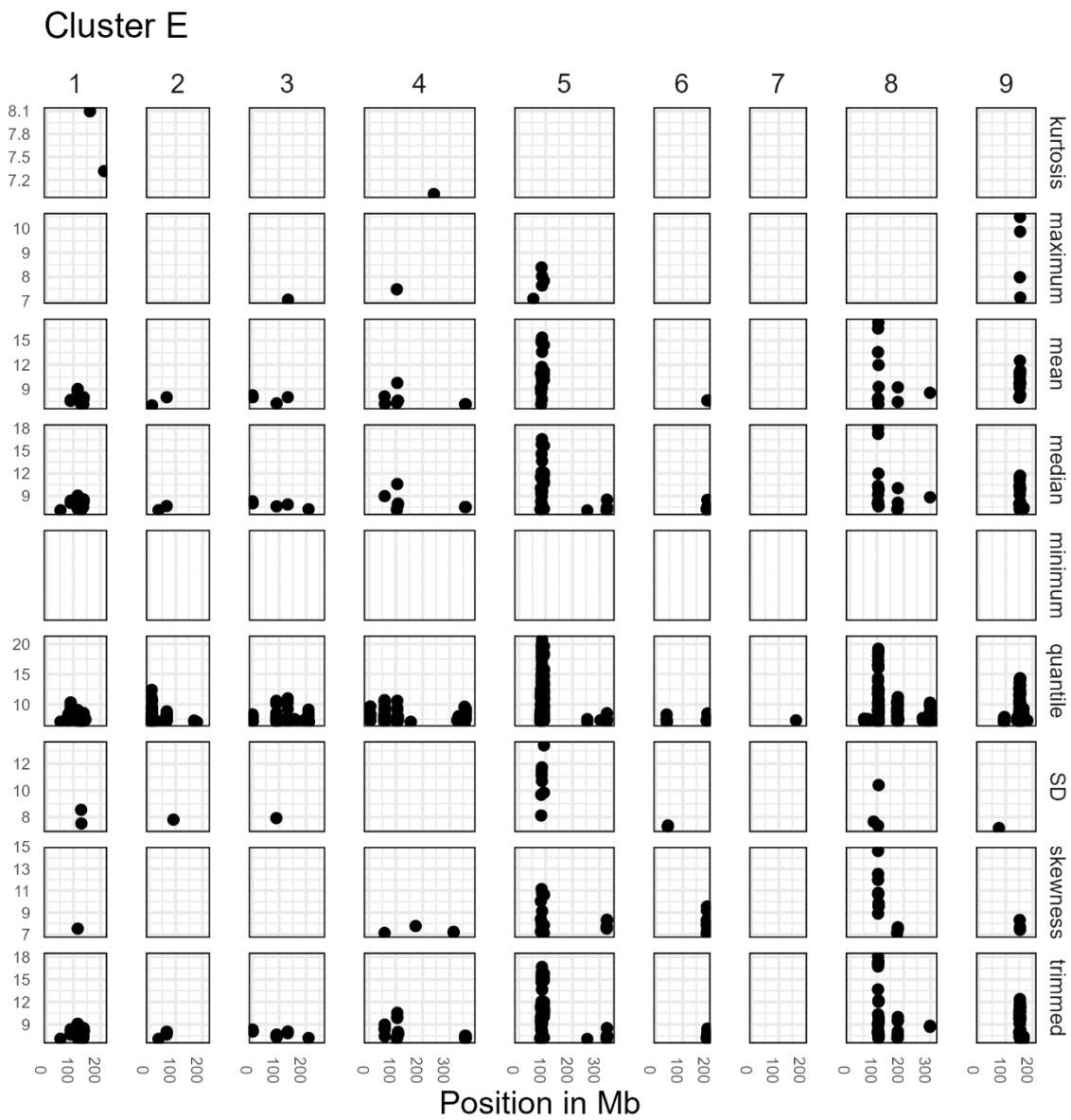
Supplemental Figure 11: The different summary statistics for cluster B plotted separately.



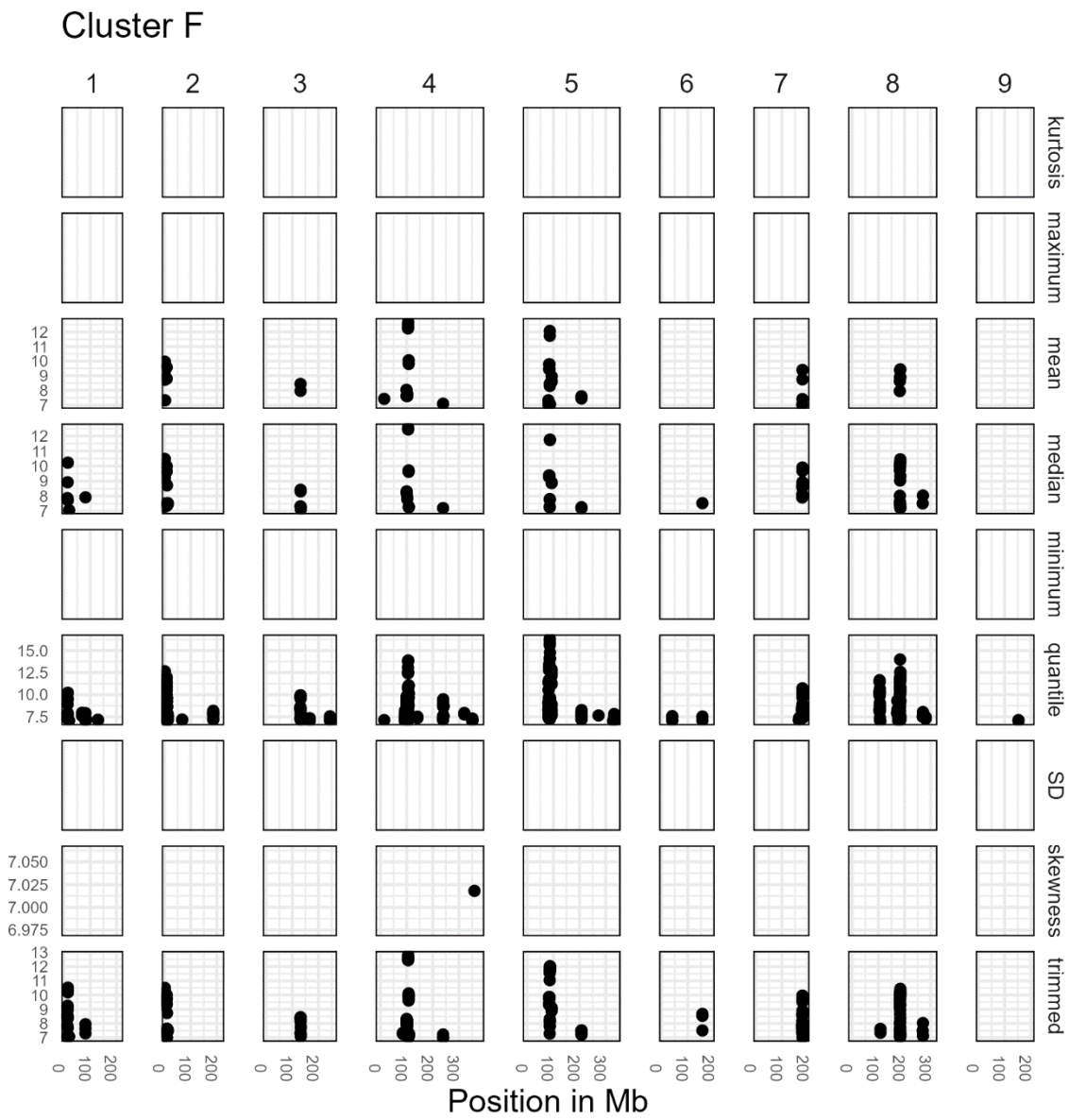
Supplemental Figure 12: The different summary statistics for cluster C plotted separately.



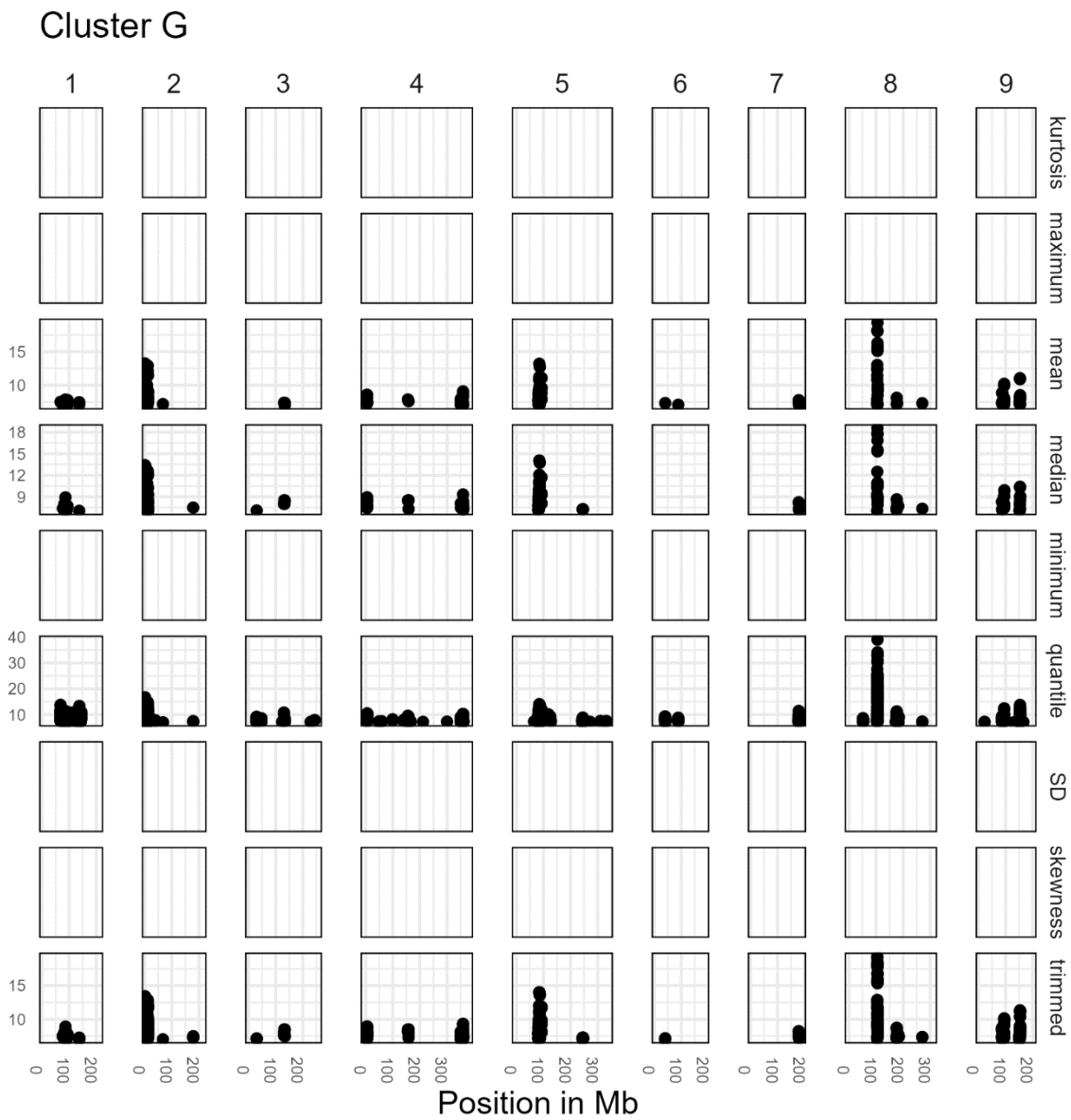
Supplemental Figure 13: The different summary statistics for cluster D plotted separately.



Supplemental Figure 14: The different summary statistics for cluster E plotted separately.

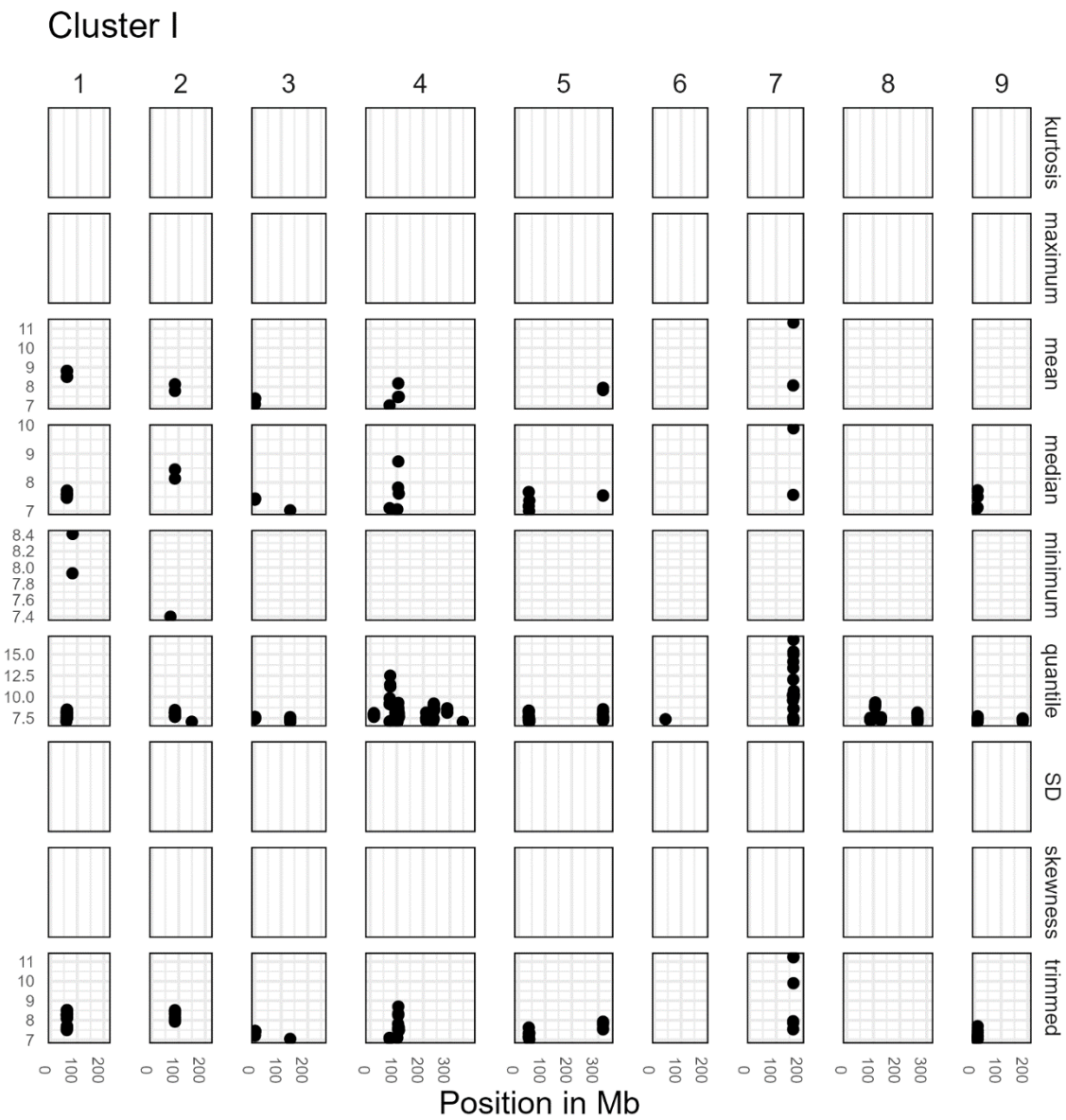


Supplemental Figure 15: The different summary statistics for cluster F plotted separately.

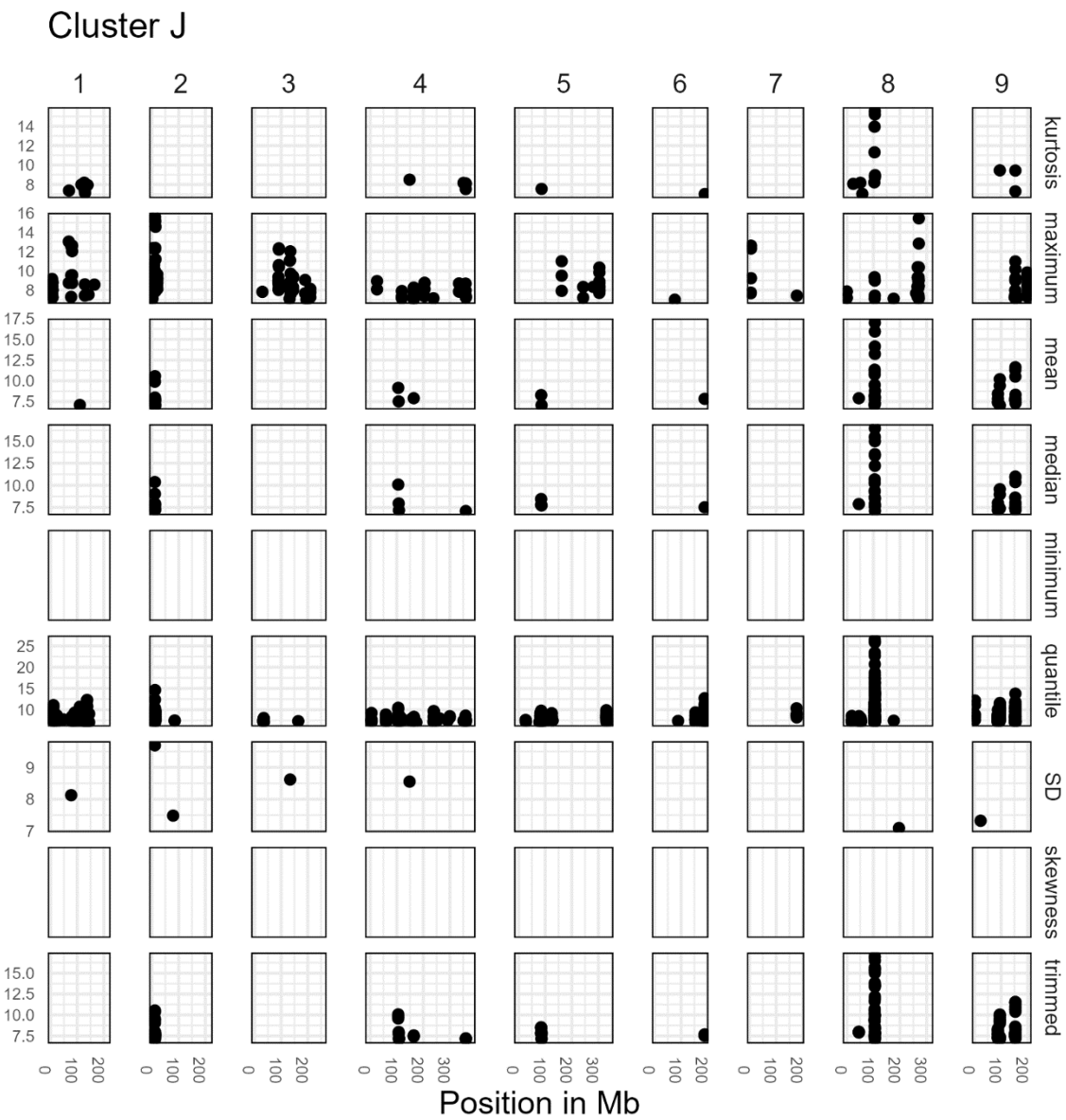


Supplemental Figure 16: The different summary statistics for cluster G plotted separately.

Supplemental Figure 17: The different summary statistics for cluster H plotted separately.



Supplemental Figure 18: The different summary statistics for cluster I plotted separately.



Supplemental Figure 19: The different summary statistics for cluster J plotted separately.