Summer 5-23-2024

# The stability of epigenetic variants that can act as loci causing phenotypic change

Raul Faburrieta

# THE STABILITY OF EPIGENETIC VARIANTS THAT CAN ACT AS LOCI CAUSING PHENOTYPIC CHANGE

by

RAUL A. FABURRIETA

A thesis submitted in partial fulfillment of
the requirement for the degree of
Master of Science
Department of Biology

Joshua Banta, Ph.D., Committee Chair

College of Arts and Science

The University of Texas at Tyler
May 2024

The University of Texas at Tyler
Tyler, Texas

This is to certify that the Master's Thesis of

RAUL FABURRIETA

has been approved for the thesis requirement on
March 26th, 2024
for the Master of Biology degree

Approvals:

I dedicate this manuscript to my Parents, Maria Gonzalez, and Felipe Faburrieta, and to my late brother Esau Gonzalez

ACKNOWLEDGMENTS

I would like to thank UT-Tyler undergraduate students Alejandro Ibarra-Narvaez, Ally McGrath, Star Hightower, Olivia Saucedo, Nallely Vasquez, and Alyssa Blanton for their efforts with planting, rearing, and data collection. I would also like to thank Christopher Conneen, Karley parker and Emmanuel Velazquez for their help with data collection. I would also like to thank my committee members Dr. Greenwold and Dr. Kellner for their guidance and advice on this study. I would like to thank Dr. Christina Richards for her insight trough her published work. I would like to thank my advisor and committee chair Dr. Banta for his guidance, insight, and patience with me throughout both my graduate and undergraduate career and for his encouragement when I needed it.

Table of Contents

List of Figures

List of Tables

Abstract

Epigenetic variations are a possible source of heritable phenotypic variation. In this study I focus on phenotypic alterations seen in epigenetic Recombinant Inbred Lines (epiRILs) of *Arabidopsis thaliana*. These epiRILs allow me to study the effects differentially methylated regions (DMRs) have on phenotypic variance. In a study performed in 2014 by Cortijo et al., they found that DMR's affect flowering time and root length when grown under greenhouse conditions. In this study, I replicated the Cortijo et al. (2014) study, with some changes, to see whether the same significant eQTL regions are found. I found that, some of the eQTLs that were found in the Cortijo et al. (2014) study overlapped with those in this study. While there were some discrepancies, this could be due to insufficient power to detect the eQTL regions that were missed, as well as differences in the experimental conditions between my study and Cortijo et al.(2014), the fact that I found any eQTLs at all suggests that the epigenotypes of the epiRILs are largely the same as when they were constructed. Otherwise, I would not have found any significant eQTL regions at all. Overall, this work adds weight to the observation that methylation changes can be heritable and stable across generations, and that these changes can alter phenotypes -- all of the ingredients needed for evolution -- independent of any DNA sequence changes.

# CHAPTER 1

## Introduction and general information

Deoxyribonucleic acid (or DNA) encodes an organism's hereditary blueprint for cellular function and organismal development (Alberts et al. 2002). Each living organism that is the product of sexual reproduction contains its unique DNA, the product of the diploid gametes that come together to form the zygote. But molecular mechanisms beyond the DNA molecules themselves (known as epigenetics) also contribute to changes in cellular function and organismal development. As explained in (Banta and Richards 2018), epigenetics refers to chemical modifications of chromatin or transcribed DNA that can influence gene activity and expression without changes in DNA sequence (Jablonka and Raz 2009; Kilvitis et al. 2014). Epigenetics focuses on the molecular changes to the DNA, rather than focusing on which nucleotides are comprising the DNA code. For instance, an epigenetic alteration would not affect the genetic code at a particular location on a chromosome, as the nucleotide would not change, but the nucleotide would still be altered chemically in some way. In this example, the DNA still contains the same genetic code, but with a small chemical alteration to a specific nucleotide.

There are three main forms of epigenetic process: (1) cytosine methylation, (2) post-translational modification of histone proteins and remodeling of chromatin, and (3) RNA-based mechanisms. Cytosine methylation is the most studied epigenetic mechanism (Rapp and Wendel 2005). It is a reversable covalent modification of DNA. Cytosine methylation is the addition of a methyl group to a cytosine nucleotide, although it is important to note that DNA methylation can also occur to any of the other nucleotides (Rehm 2018). These methyl groups

change the biophysical characteristics of DNA but do not actually change the DNA sequence (Gibney and Nolan 2010). DNA methylation can affect the phenotype of each organism differently, and it is completely dependent on the way the methyl group interacts with the protein components of the chromosome (Jablonka and Raz 2009). This is particularly important as the cell reads the instructions of the altered cytosine nucleotide differently or sometimes not at all (Rapp and Wendel 2005). Histone modification is an epigenetic process that alters the proteins that form chromatin. DNA is wrapped around histones that behave as a scaffold, allowing the DNA to wrap around them. This allows the DNA to be tightly packed and stored in the nucleosome(Henikoff and Smith 2015). In the nucleosome, histone tails can undergo a variety of modifications, known as post-translational modifications (PTMs), that occur after the initial translation of the mRNA to make the histones and are thus not coded in the histones' DNA sequences or the downstream transcripts (Henikoff and Smith 2015). These post translational modifications can impact cellular function and ultimately the phenotype of an organism without any changes to the genotype (Henikoff and Smith 2015). RNA-based mechanisms involve non-coding infrastructural RNAs such as tRNAs, rRNAs, small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). They have a wide range of sizes. They are involved in translation and splicing and function by sequence-specific recognition of RNA substrates and also as catalysts. Their role in epigenetics is that they can facilitate the other types of epigenetic effects, such as cytosine methylation and histone modifications, as well as modify chromatin in their own right (Wei et al. 2017). Despite having been known for some time, RNA-based mechanisms are the least well understood(Gibney and Nolan 2010).

Epigenetics inheritance has also been labeled as soft inheritance, in which the environment allows for favorable alleles compared to hard inheritance in which genetic changes happen randomly and are stable through generations (Richards 2006; Robertson and Richards 2015). Some biologists emphasize that epigenetic mutations can be environmentally caused, and that they can be re-set before reaching the germ line as an objection to claims that epigenetic changes contribute to heritability and evolution (Robertson and Richards 2015). In recent years, epigenetic differences in traits have been found to be inherited in plants and animals, providing an alternative form of heritability other than genetic inheritance (Cubas et al. 1999; Feng et al. 2010; Verhoeven et al. 2010; Daxinger and Whitelaw 2012; Alonso et al. 2014; Cortijo et al. 2014; Robertson and Richards 2015). This has led some to argue that epigenetic inheritance should be considered in evolutionary theory (Banta and Richards 2018). This argument is based on expressed phenotypic traits over several generations caused by epimutations divorced from genetic inheritance (Jablonka and Lamb 2015). For instance, in a study performed using *Linaria vulgaris,* the symmetry of the petals was compared in a population of genetically identical organisms (Cubas et al. 1999). They found that an epimutation at the *L-CYC* gene, which did not affect the genetic sequence, induced a peloric mutant, causing the flower to exhibit radial symmetry as opposed to the wild-type (WT) bilateral symmetry (Cubas et al. 1999).

A different study using *Helleborus foetidus* found a connection between phenotypic traits and cytosine methylation (Alonso et al. 2014). They took random samples of *Helleborus foetidus* grown in the wild and they recorded fecundity and size related traits at the time of collection (Alonso et al. 2014). After collection DNA was extracted from these samples, the researchers then accounted for genetic background variance and then the percentage of

cytosine methylation was taken for all the sampled plants (Alonso et al. 2014). They found that individual cytosine variation accounted for seven of the nine traits considered (Alonso et al. 2014). They found that the individuals that were hypomethylated were larger in size and fecundity (Alonso et al. 2014).

(Cortijo et al. 2014) Set out to study the inheritance of epimutations in **R**ecombinant **I**nbred **L**ines (RILs) of *Arabidopsis thaliana* over several generations and under different conditions. They used a RIL population developed by (Johannes et al. 2009) that segregates almost exclusively for differentially methylated positions (DMPs) – differences in whether individual cytosine nucleotides in the DNA are methylated or not, and does not segregate for nucleotide differences per se. Thus the RILs are nearly genetically identical.
These **epi**genetic RILs (epiRILs) were derived from two near-isogenic parental lines: the Columbia wild-type laboratory strain, and a mutant for the gene *DECREASED DNA METHYLATION 1* (*DDM1*) that exhibits 70% less methylation genome-wide than wild-type plants (Cortijo et al. 2014). Backcross progeny from the initial cross were selected to be *DDM1*/*DDM1* homozygotes and epiRILs were derived from six rounds of single-seed descent from these backcross progeny. Thus, the epiRILs are nearly isogenic in terms of the DNA sequences, and segregate mostly for DMPs, not DNA sequence-based polymorphisms. The epigenotypes at each epilocus are homozygous, due to the highly-selfing mating system of *A. thaliana*. The epiRILs show variation and often high heritability for many traits relating to growth and morphology, including plant height, flowering time, and primary root length.(Johannes et al. 2009; Cortijo et al. 2014; Kooke et al. 2015) used the epiRILs to study flowering time and primary root length. Once they had grown the plants, they measured

flowering time in the field- and greenhouse-grown plants and they measured root length in the climate-controlled plants. They created a genetic map covering about 81.9 percent of the total genome. They then used **epi**genetic **Q**uantitative **T**rait **L**ocus mapping (epiQTL mapping) to find significant epiQTLs (epiQTLs that were found to affect phenotypic traits). They found multiple epiQTLs accounting for variation in these traits, and accounting for 60 – 90% of the heritability. These epiQTLs were found on 3 different chromosomes and were found throughout the organisms in each of the different conditions. Their findings were significant as they demonstrated that differently methylated regions can be passed on from parents to offspring and may act as epiQTLs controlling phenotypic variation, allowing for an additional form of mutation and variation other than just genetic. These findings demonstrate that not all epigenetic mutations reset during gametogenesis.

Yet there is controversy about how important epigenetic factors are to heritable traits and evolution (Banta and Richards 2018). For instance, in Slatkin's (2009) modeling study, the risk of disease was conceptualized as a product of both genetic and epigenetic loci (Slatkin 2009). He proposed that with each generation, there exists a chance for the reset of individual epigenetic states for every offspring. Consequently, not all epigenetic variants endure in subsequent generations, leading to a diminished likelihood that relatives will share similar epigenetic profiles despite their genetic relatedness. Slatkin's (2009) findings suggested that the impact of epigenetic factors on disease heritability is relatively minor. However, he cautioned that this conclusion hinged heavily upon the persistence durations of heritable epialleles. Day and Bonduriansky (2011) add more nuance to this picture, by presenting a model of non-genetic inheritance that is dependent on alterations in transmission of the non-genetic

information from parent to offspring (reproductive transmission), and modifications to the non-genetic information occurring within the parent generation (survival transmission)(Day and Bonduriansky 2011). Relatedly, Tal et a. (2010) extend the well-known phenotypic variance equation ($V_P = V_G + V_E$) to include a term that accounts for epigenetic transmissibility. This all creates a picture where epigenetic variants can only impact evolution if they are transmitted to offspring with some fidelity (Tal et al. 2010).

My work seeks to understand how well the findings from Cortijo et al.'s (2014) study with the *Arabidopsis* epiRILs can be replicated. If the epigenetic variants in the epiRILs are transmitted with relatively high fidelity, then the same epiQTLs should be recovered in a replicated study. If, however, the methylation states of the epiRILs are re-set or change frequently, then they should not be reliably transmitted from the parent plant to the offspring, and therefore the same epiQTLs from previous research should not be recovered when different seeds, even from the same parent, are used. I am also expanding my study to include additional phenotypes not mapped by Cortijo et al. (2014): rosette diameter, the number of basal branches, the number of lateral branches, and fruit number. Thus, my study will add new information about the epigenetic architecture of *Arabidopsis* traits, separate from genetic variation.

# CHAPTER 2

## Materials and methods

*Arabidopsis thaliana* is a small weedy plant from the mustard family (Shimizu and Purugganan 2005; Richards et al. 2012). It has become a model organism used in genetic, evolutionary, and ecological studies (Koornneef et al. 2004; Shimizu and Purugganan 2005; Banta et al. 2007). It serves well as a study organism due to its short life span, self-mating characteristics, and small genome (Banta et al. 2007). Furthermore, its full genetic code is accessible (Unseld et al. 1997).

The *Arabidopsis* epiRILs from Cortijo et al. (2014) were ordered from the Versailles Arabidopsis Stock Center in September 2018. These epiRILs were the same 123 lines that were used by Colomé-Tatché et al. 2012) and Cortijo et al. (2014) in order to reveal segregating methylation polymorphisms that explained heritable effects for root length and flowering time. These epiRILs were derived from two near-isogenic parental lines: the Columbia wild-type laboratory strain, and a mutant for the gene *DECREASED DNA METHYLATION 1* (*DDM1*) that exhibits 70% less methylation genome-wide than wild-type plants (Johannes et al. 2009). An F1 individual was backcrossed to the Columbia parental line, and then F1 individuals homozygous for the *DDM1* mutation were selfed for six generations. The epiRILs are nearly isogenic in terms of the DNA sequences, and segregate mostly for DMPs (differently methylated positions), not DNA sequence-based polymorphisms (Johannes et al. 2009). They have 126 DMPs that serve as markers, with an average spacing of 3.45 cM (0.804 Mb) that covers 89.1 percent of the genome. The epigenotypes at each epilocus are homozygous, due to the highly-selfing mating system of *A. thaliana*. The epiRILs show variation and often high heritability for traits relating to

growth and morphology, including plant height, flowering time, and primary root length (Johannes et al. 2009).

The seeds I used were from the same generation of seeds as those from Cortijo et al. (2014) (Christine Camilleri, Versailles Arabidopsis Stock Center, personal communication), and thus they were siblings to the seeds used in Cortijo et al. (2014). The epiRILs were grown in a walk-in, controlled chamber at the University of Texas at Tyler. Five replicate plants of each of 123 different epiRILs were grown, for a total of 615 plants, although a small quantity never flowered, or died, or were destroyed before all measurements could be taken. The photoperiod and temperature conditions of Cortijo et al. (2014) were matched by growing the plants in a 16-hour day length with a temperature of 20° C – 22°C and an 8-hour nighttime period with a temperature of 16° C – 18°C. Flowering time was measured as the interval from planting until the first emergence of the primary flowering stalk, or inflorescence. Also measured at the same time was the rosette's diameter, a proxy for maximum adult size (González et al. 2020). Finally, several inflorescence architecture traits were measured once the plants reached senescence: 1) the number of basal inflorescences, which refers the inflorescences growing out of the rosette; 2) number of lateral branches, which refers to the secondary and higher-order branches stemming from the basal inflorescences; and 3) the number of fruits on all inflorescences and branches, which serves as an estimate of lifetime reproductive fitness (see Banta and Pigliucci 2010 for more information)

Epigenetic Quantitative Trait Locus (epiQTL) mapping was performed identically to standard QTL mapping, except that the markers were cytosine methylation variants (methylated or not). Correlations between epigenetic variants and phenotypic variance were

tested using Haley Knott regression as implemented by the "scanone" function of the QTL package (Broman et al. 2003) in R (R Core Team 2023). We used the same genetic map from (Colomé-Tatché et al. 2012) was used in the Cortijo et al. (2014) study. Genome-wide significance was determined empirically for each trait using 1000 permutations of the data, corresponding to a genome-wide false positive rate of 5%. Phenotypes more than 3 SD from the mean were removed. The significant epiQTLs for flowering time in my dataset were compared to the significant epiQTLs for flowering time in the Cortijo et al. (2014) dataset.

For flowering time epiQTLs that were significant in my dataset but not in Cortijo et al.'s dataset, and vice versa, we performed a post-hoc analysis of the power to detect the epiQTLs using the "powercalc" function of the qtlDesign package (Sen et al. 2007) in R. In order to calculate power, it was necessary to calculate the environmental and genetic variances for the flowering time phenotypes (both mine and Cortijo et al.'s 2014). This required treating the epiRILs as a random effect, which inflates the estimates of genetic variance (Sen et al. 2009). Thus the power estimate will only be an upper bound, and the actual power to detect the epiQTL will be lower than this estimate; if it is reported that there is low power to detect an epiQTL, the actual power to detect that epiQTL will be even lower. Following (Cohen 1992), power values of 0.8 and higher are considered to be sufficient to conclude that there was enough power to detect an effect of the size of interest. The effect size of interest corresponded to the effect size of the marker that had the highest LOD score within the epiQTL's 95% confidence interval. According to (Cohen 1992) , "a medium effect of 0.5 is visible to the naked eye of a careful observer. A small effect of 0.2 is noticeably smaller than medium but not so small as to be trivial. A large effect of 0.8 is the same distance above the medium as

small is below it." We used these conventions when determining if the effect sizes of the epiQTLs were small, medium, or large.

The other traits were also mapped, although they were not compared to Cortijo et al. (2014) because Cortijo et al. (2014) did not measure those traits.

# CHAPTER 3

## Results

As previously stated, this study overlaps with Cortijo et al. (2014) in measuring flowering time. To establish a baseline, I re-analyzed the flowering time data collected by Cortijo et al. (2014) but this time using my own pipeline. Their script was not made public and therefore could not be used, and so I had to re-create it myself. They reported that they used the "scanone" function of the r/QTL package in R, and that they used Haley-Knott regression, but they did not provide more details. I found through trial and error that, in order for my pipeline to give similar results, I had to modify it to remove outlier epiRILs whose mean flowering times were more than 3 SD from the mean (Figure 1 and 2). When this was performed, the results yielded eQTLs consistent with the ones Cortijo et al. (2014) reported when using their pipeline (as reported in their paper) on Chromosomes 1 (chr 1), 4 and 5 (Table 1). This establishes the accuracy of my analytical pipeline.

There were two outlying epiRILs in the Cortijo et al. (2014) dataset whose mean flowering times were more than 3 SD from the mean of all epiRILs, and therefore their data was removed to genetically map flowering time in the Cortijo et al. (2014) dataset: epiRIL60 and epiRIL98. There was one outlying epiRIL in my dataset whose mean flowering time was more than 3 SD from the mean of all epiRILs, and therefore its data was removed to map flowering time in my dataset: epiRIL60. Additionally, there were two epiRILs in my dataset that germinated and reached the rosette stage but did not flower: epiRIL98 and epiRIL122. Thus, there were two epiRILs in the Cortijo et al. (2014) dataset, and three epiRILs in my dataset that were excluded. One of the two epiRILs that I excluded from my dataset, epiRIL98, did not

flower in my dataset and was an outlier for late flowering in the Cortijo et al. (2014) dataset. Another of the epiRILs that I excluded from my dataset, epiRIL60, was an outlier for late flowering in both my dataset and in the Cortijo et al. (2014) dataset. The third epiRIL that I excluded from my dataset, epiRIL122, did not flower in my study but flowered within 3 SD of the mean flowering time of all epiRILs in the Cortijo et al. (2014) study. For the other traits, I excluded the following epiRILs from my datasets because there was no data or too much missing data: for basal inflorescence number: epiRIL98, epiRIL114, epiRIL122, and epiRIL315; for fruit number: epiRIL71, epiRIL98, epiRIL122, and epiRIL508; for lateral branch number: epiRIL98, epiRIL122, epiRIL297, and epiRIL305; for rosette diameter: epiRIL98 and epiRIL122.

The second step was to compare the epiQTLs for flowering time in my dataset to the flowering time epiQTLs found by Cortijo et al. (2014) (after re-analyzing their data using my analytical pipeline for consistency.) My flowering time phenotype and their flowering time phenotype had similar epiQTLs at chr 1 and chr 4 (Table 1; Figure 3). Additionally, my flowering time phenotype had a significant epiQTL on chr 2 that was not present in the Cortijo et al. (2014) dataset. Conversely, Cortijo et al.'s (2014) flowering time phenotype had an epiQTL on chr 5 that was not present in my dataset (Table 1; Figure 3).

The power analysis showed that the discrepancies between my results and Cortijo et al. (2014) could be due to a lack of power. In the instance where I found a significant epiQTL in my flowering time dataset that was not present in the Cortijo et al. (2014) flowering time dataset, the peak LOD score on that chromosome in the Cortijo et al. (2014) dataset occurred within the 95% epiQTL confidence interval position in my dataset (the peak LOD score on chr 2 for Cortijo et al.'s flowering time data was at 13.25 cM, and epiQTL confidence interval on chr 2 in my

dataset was between 0.56 and 13.25 cM; Tables 1 and 2 and Figures 3 and 4). Yet the power to detect that epiQTL in Cortijo et al.'s (2014) dataset was quite low (0.055), due to a small effect size (0.47) as compared to the very large effect size of the epiQTL in Cortijo et al. (2014) (1.75) (Table 2). Similarly, in the instance where Cortijo et al. (2014) found a significant epiQTL in their flowering time dataset that was not present in my flowering time dataset, the peak LOD score on that chromosome in my dataset occurred at exactly the same location as the peak LOD score on that chromosome in their dataset (Tables 1 and 2). Yet the power to detect that epiQTL in my dataset was insufficient (0.68), at best, despite a large effect size (1.093). Lastly, rosette diameter and the inflorescence architecture traits from my dataset were mapped. This data was not compared to the Cortijo et al. (2014) dataset, as they did not collect this data. There was a significant epiQTL on chr 1 for the number of basal inflorescences, as well as a significant epiQTL on chr 1 for the number of fruits; the 95% confidence intervals for the epiQTLs of these two traits overlapped with each other, as well as with the 95% confidence interval for flowering time (Table 1; Figures 5 and 6). There was also a significant epiQTL for rosette diameter on chr 5 (Table1; Figures 5 and 6). Interestingly, it overlaps with the epiQTL interval that Cortijo et al. (2014) found for flowering time on chr5 (Table 1). There were no significant epiQTLs for lateral branch number.

# CHAPTER 4

## Discussion

Understanding the role of epigenetics in evolution is important, because epigenetic changes can be stable and inherited, and they can have different consequences for phenotypic variation and evolution than DNA nucleotide-based changes to the genome (Banta and Richards 2018). The evolution of high-throughput genome sequencing technologies has ushered in a new era in evolutionary biology, and by combining quantitative genetics with genomic approaches, researchers can now delve deeper into longstanding evolutionary questions (Roff 2007; Hill 2012; Banta and Richards 2018). Therefore, it is important that epigenetic mechanisms should not be excluded from this new and emerging enterprise (Banta and Richards 2018). Yet it is not well understood how accurately epigenetic information is passed on to the germ line. If parents can reliably transmit their epigenetic states to their offspring, and if those epigenetic states result in phenotypic differences, then epigenetic information is important for understanding evolutionary processes (Banta and Richards 2018). My study aims to see if the results from a study on the influence of epigenetic changes to phenotypic variation can be replicated using seeds coming from the siblings of the plants used in the original study. If the parent plants reliably transmit epigenetic information to the offspring, then the same results should be discovered when re-creating the study using different offspring from the same parents. If, however, epigenetic information is re-set or otherwise not reliably transmitted from parents to offspring, then there is no reason to believe that the same results should be revealed when the study is replicated using sibling plants.

I grew my own plants, and used my own independent analytical pipeline, to map flowering time epiQTLs in the same epiRIL population used by Cortijo et al. (2014). The seeds that I used were siblings to the ones used in the Cortijo et al. (2014) study. Furthermore, I re-analyzed the Cortijo et al. (2014) flowering time data (under greenhouse conditions) using my analytical pipeline so that I could directly compare their findings to mine. I found epiQTLs for flowering time consistent with those reported by Cortijo et al. (2014). The epiQTL intervals I found on chr 1 and chr 4 overlapped with those found by Cortijo et al. (2014). This is consistent with the same epi-allelic variants causing differences in flowering time in both studies (mine and theirs). Of course, it is possible that different, linked epi-lociare responsible for the epiQTL intervals I found on chr1 and chr4, as compared to the epiQTL intervals that Cortijo et al. (2014) found, but this seems less likely; how new epigenetic variants with just the right effects could newly arise in just the right plants in my study, or some different, pre-existing epigenetic variants within those epiQTL intervals could newly have an influence on flowering time in my study, is not clear . It seems more parsimonious to assume that I replicated epiQTL intervals consistent with the ones in Cortijo et al. (2014) because they are underlain by the same causative elements as the ones in Cortijo et al. (2014). Then again, the fact that the environment used in my study (a growth chamber) was different from the one used by Cortijo et al. (2014) (a greenhouse) could introduce an element of phenotypic plasticity to my results that may make it more plausible that different, pre-existing epigenetic variants within those epiQTL intervals could have had an effect on flowering time in my study, whereas they had not had an effect in the Cortijo et al. (2014) study. I note that the photoperiod and temperature conditions that I used were the same as the ones reported by Cortijo et al. (2014), but the way

those conditions are experienced by the plants could be different in a greenhouse versus a growth chamber (Choi et al. 2015).

There were discrepancies between my study and Cortijo et al. (2014) on chr 2 and chr 5. On chr 2, I found a significant epiQTL region associated with flowering time that was not present in the Cortijo et al. (2014) dataset. The power analysis suggests that this could be due to differences in power to detect an effect in the two datasets (Cortijo et al.'s and mine), owing to differences in the effect size of the causative epigenetic change(s) in the two datasets. While the peak marker on chr 2 in the Cortijo et al. (2014) dataset was within the epiQTL interval that I found in my dataset, the effect of the peak marker in that interval was found to be much smaller in the Cortijo et al. (2014) than in my dataset, and consequently, the power for Cortijo et al. (2014) to detect an effect in that epiQTL interval in their dataset was much less than in my dataset. Why the effect of the epiQTL in that interval was so much larger in my dataset as compared to the Cortijo et al. (2014) dataset is unclear. Possible explanations are detailed in the last paragraph of the Discussion.On chr. 5, Cortijo et al. (2014) found a significant epiQTL region associated with flowering time that was not present in my dataset. This appears to be a power issue, since the exact same peak on chr 5 is found in the Cortijo et al. (2014) dataset and in my dataset (albeit with different LOD scores). While there was some power to detect an effect within that epiQTL interval in my study, it was insufficient, even though the effect size of the peak epiQTL marker in that interval was large in my study. In fact, the peak on chr 5 approached the significance threshold in my dataset but did not quite reach it. Therefore, I consider this finding to be largely consistent with the findings from Cortijo et al. (2014) for

flowering time, even if, strictly speaking, I did not find a significant epiQTL for flowering time on chr5 in my dataset.

I documented epiQTLs in this epiRIL population for rosette diameter, the number of basal inflorescences, and fruit number, which are traits that have never been epigenetically mapped before. The eipQTL intervals for basal inflorescence number and fruit number are on chr1 and overlap with one another and with the epiQTL interval for flowering time in my dataset. This opens the intriguing possibility that all three epiQTL intervals are underlain by the same causative epi-locius or epi-loci, representing an example of epigenetic pleiotropy (Chebib and Guillaume 2021). There may be a very important epigenetic variant in this chromosomal region that influences multiple phenotypic properties, from flowering timing to the development of basal inflorescences produced and the consequent number of fruits that are possible. Flowering time genes often have strong pleiotropic effects in *Arabidopsis*, and can influence plant architecture (Tonsor et al. 2005; Auge et al. 2019), so this scenario is plausible. Of course, it is also possible that different but linked epi-loci are causing the overlapping epiQTL intervals observed for these traits on chr1 (Chebib and Guillaume 2021). The epiQTL interval that I found for rosette diameter is on chr 5 overlaps with the epiQTL interval that Cortijo et al. (2014) found for flowering time, so it is also possible that there is epigenetic pleiotropy at play here, too, although the coincident epiQTL intervals could be due to coincidental linkage of different causative epi-loci on chr5 instead.

It is important to also consider the effect that transposable elements (TEs) may have had in our study. For this purpose, I refer to the Cortijo et al. (2014) study. In order to determine if variants in the eipRILs were caused by DMR rather than TEs that may have

occurred in the ddm1-2 parental line, they re-sequenced 52 out of the 123 epiRILs used in their study. They found two segregating TE insertions in chr1, and two shared insertions in chr4, within the significant epiQTL intervals for their traits. After further analysis, they found that these TE insertions were not consistently inherited. Furthermore, they determined that the phenotypic effects presented by the TE insertions were of much weaker effect than those presented by significant epiQTLs overall, leading them to believe that any effects of the TEs were small and unlikely casual drivers of the observed epiQTLs. Thus, any effects of TEs can probably be discounted in my study, although following up on this would be an interesting topic of future research.

The possible reasons for discrepancies between my study and Cortijo et al. (2014) could be due to phenotypic plasticity, as already mentioned, since my environmental conditions (growth chamber) were not the same as Cortijo et al.'s (2014) (greenhouse). Another possible source of the discrepancies is changes to the methylation of some of the markers as compared to what was reported by Cortijo et al. (2014) and what we used as the basis for our epiQTL mapping. This is possible to some degree, but I do not think that all of the methylation states are different from what Cortijo et al. (2014) assumes, or else I would not have recovered any significant epiQTLs. Nevertheless, resequencing of the epiRILs is warranted to verify which markers remain as Cortijo et al. (2014) reported and which ones have actually changed in their methylation status. Yet another possible source of discrepancy is that not all of the epiRILs used for mapping in the Cortijo et al. (2014) dataset were used for mapping in my dataset. Specifically, there was one additional epiRIL used in the Cortijo et al. (2014) dataset that was not used in my dataset, because it did not flower in my dataset. Besides that one epiRIL

difference, however, the other epiRILs were in common between my dataset and the Cortijo et al. (2014) dataset. This last possibility is the least likely, however, because when we excluded epiRIL122 from the Cortijo et al. (2014) dataset, thereby making the list of epiRILs used by my dataset and the Cortijo et al. (2014) datset a perfect match, there was no noticeable difference to the results as compared to the original results.

**Concluding remarks**

A major goal of my study was to evaluate if the epiQTLs for flowering time previously found by Cortijo et al. (2014) in the epiRIL mapping population would be recovered when I replicated their study using sibling seeds. For the most part, this was borne out, although it seems I lacked sufficient power to detect the epiQTL that they found on chr5. This is signifcant because it demonstrates that the epigenetic variants of the epiRILs must be at least somewhat faithfully transmitted from parents to offspring. Otherwise, my study would not have found similar epiQTL intervals to those found by Cortijo et al. (2014) on seeds that were siblings to the ones that they used. In fact, there is no reason why I would find any epiQTL intervals at all, if in fact the epigenetic information were not being at least somewhat faithfully transmitted from parents to offspring; the supposed epi-alleles present at a locus would not match what is found in my plants, if they did not inherit the epi-alleles as expected. Therefore, my study suggests thatthe epi-loci used in this mapping population are at least partially faithfully transmitted to the offspring, and therefore can serve as a source of evolutionary important phenotypic variance that is separate from DNA sequence-based genetic variance.

I also documented a new epiQTL for flowering time on chr2. For the first time, I also documented epiQTLs for basal inflorescence number, fruit number, and rosette diameter.

Future work could develop strategies to fine-map and positionally clone these significant epiQTL regions (Jaganathan et al. 2020) ,to determine the precise epigenetic variants accounting for the significant effects on the phenotypes.

**Figures**

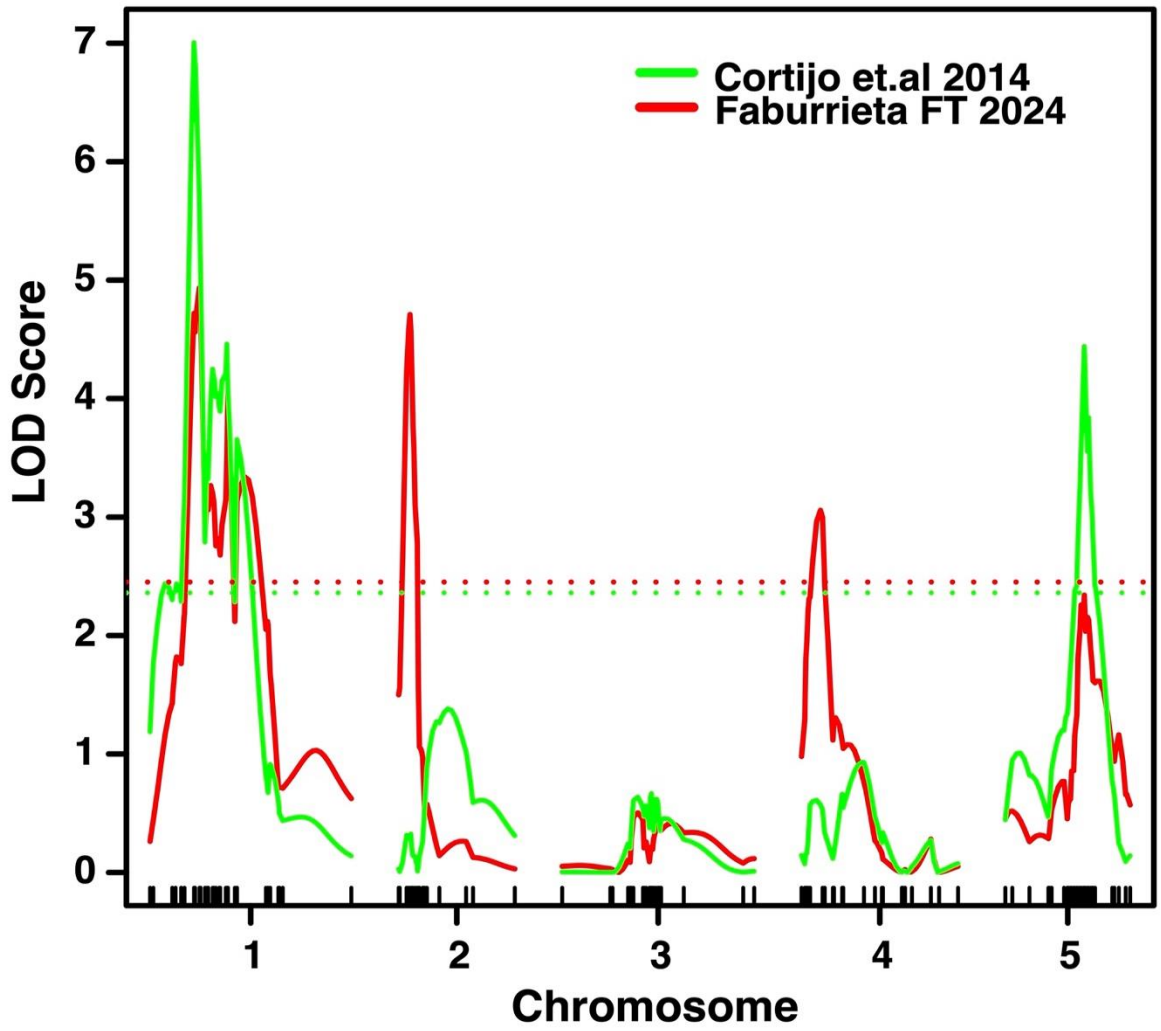


**Figure 1:** eQTL mapping profiles for two flowering time measurements: Under greenhouse conditions in the **Cortijo et al. (2014)** study and under controlled growth chamber conditions in the current study **(Faburrieta FT 2024).** The outlier phenotypes were not removed prior to mapping. The x-axis shows position on chromosome 1, 2, 3, 4, or 5, in centimorgans. The y-axis shows the LOD score, which for every marker is the $\log_{10}$ of the ratio of the probability that an epiQTL is present to the probability that an epiQTL is absent. The dotted lines are the permutation thresholds to determine significance, with peaks above the threshold indicating that the association between the epi-marker and the phenotype at that locus is significant.

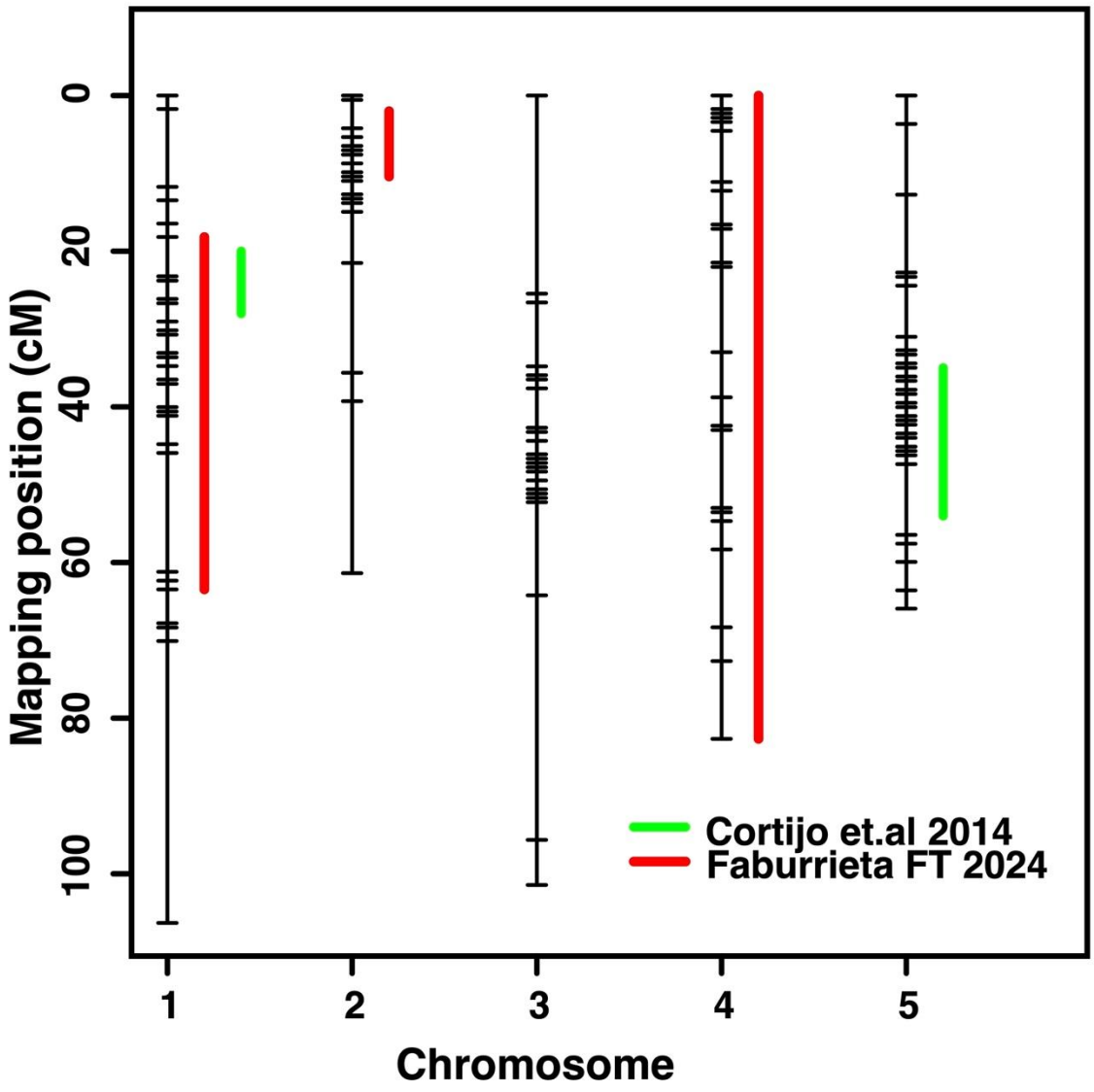


**Figure 2.** A genetic linkage map showing the 95% confidence intervals for the eQTL regions for flowering time from the **Cortijo et al. (2014)** greenhouse study and the current study conducted under controlled environmental conditions **(Faburrieta FT 2024).** The outlier phenotypes were not removed prior to mapping. The mapping positions of the epiQTL intervals are positioned according to the location on the chromosome in centimorgans.

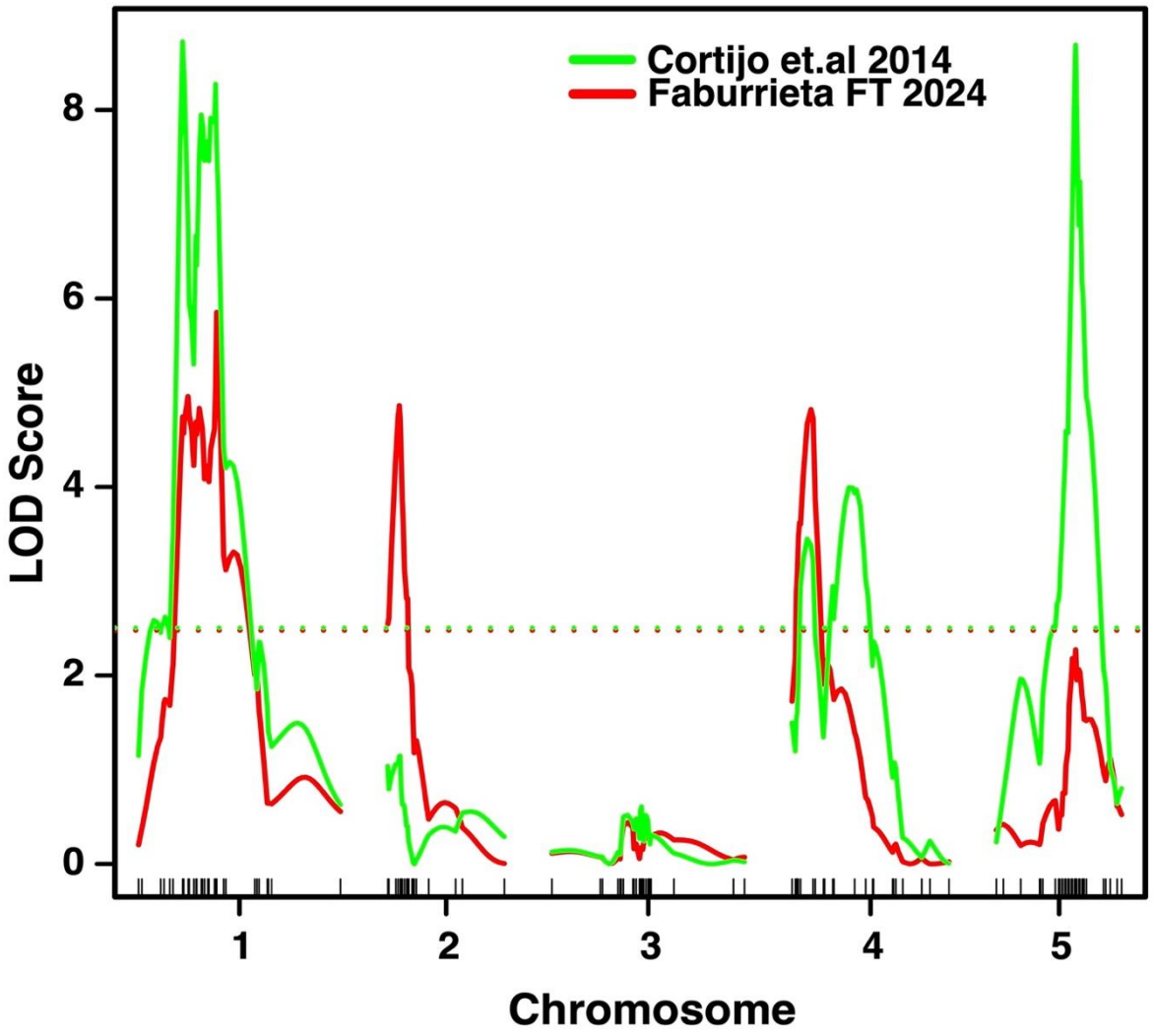


**Figure 3:** eQTL mapping profiles, for two flowering time measurements: Under greenhouse conditions in the **Cortijo et al. (2014)** study and under controlled growth chamber conditions in the current study **(Faburrieta FT 2024).** The outlier phenotypes were removed prior to mapping**.** The x-axis shows position on chromosome 1, 2, 3, 4, or 5, in centimorgans. The y-axis shows the LOD score, which for every marker is the $\log_{10}$ of the ratio of the probability that an epiQTL is present to the probability that an epiQTL is absent. The dotted lines are the permutation thresholds to determine significance, with peaks above the threshold indicating that the association between the epi-marker and the phenotype at that locus is significant.

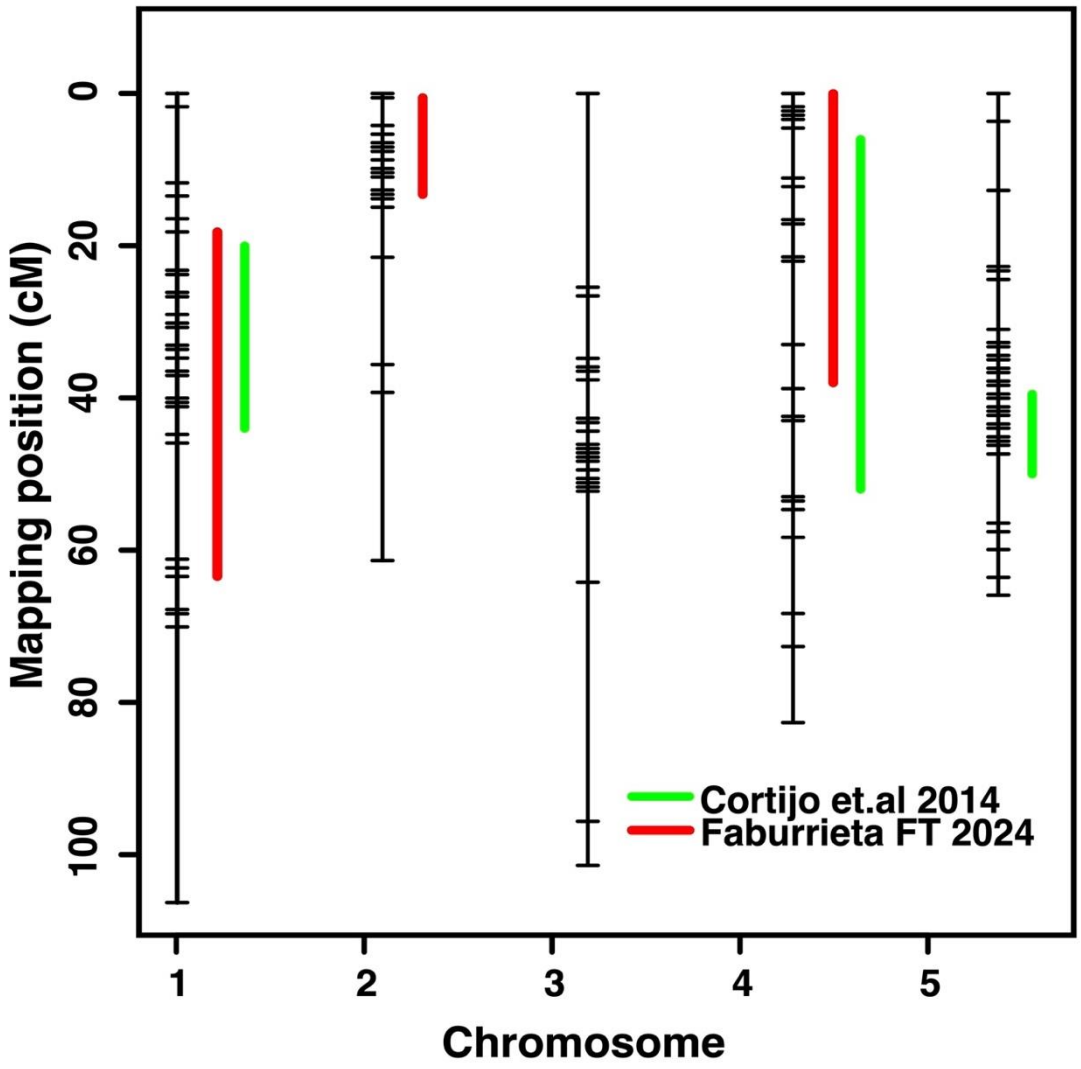


**Figure 4.** A genetic linkage map showing the 95% confidence intervals for the eQTL regions for flowering time from the **Cortijo et al. (2014)** greenhouse study and the current study conducted under controlled environmental conditions **(Faburrieta FT 2024).** The outlier phenotypes were removed prior to mapping**.** The mapping positions of the epiQTL intervals are positioned according to the location on the chromosome in centimorgans.

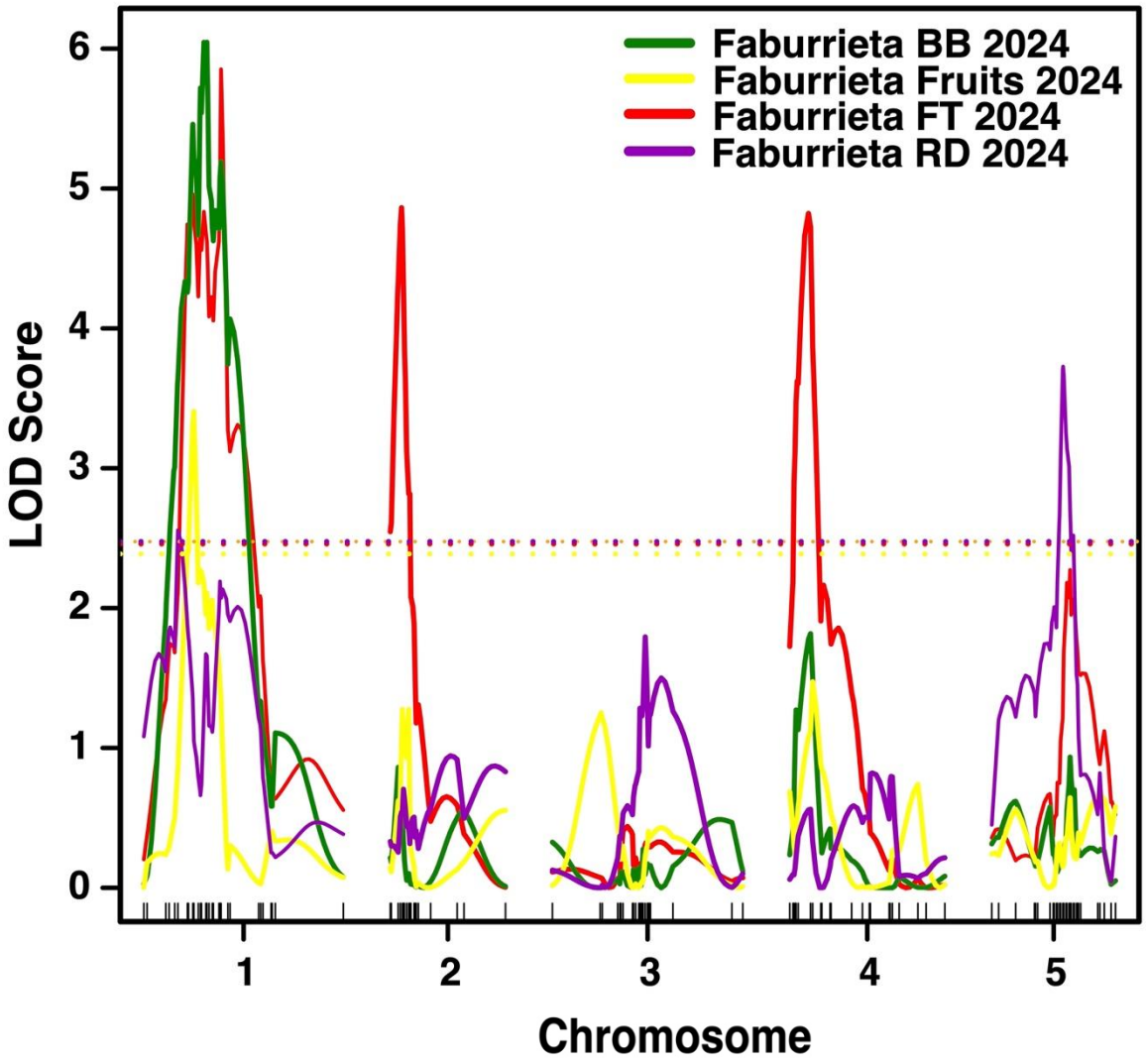


**Figure 5:** eQTL mapping profiles of flowering time various traits in the current study under growth chamber conditions: number of basal inflorescences, or branches (**Faburrieta BB 2024**), the number of fruits (**Faburrieta Fruits 2024**), flowering time (**Faburrieta FT 2024**), rosette leaf number (**Faburrieta RLN 2024**) and rosette diameter (**Faburrieta RD 2024**). The x-axis shows position on chromosome 1, 2, 3, 4, or 5, in centimorgans. The y-axis shows the LOD score, which for every marker is the $\log_{10}$ of the ratio of the probability that an epiQTL is present to the probability that an epiQTL is absent. The dotted lines are the permutation thresholds to determine significance, with peaks above the threshold for indicating that the association between the epi-marker and the phenotype at that locus is significant.

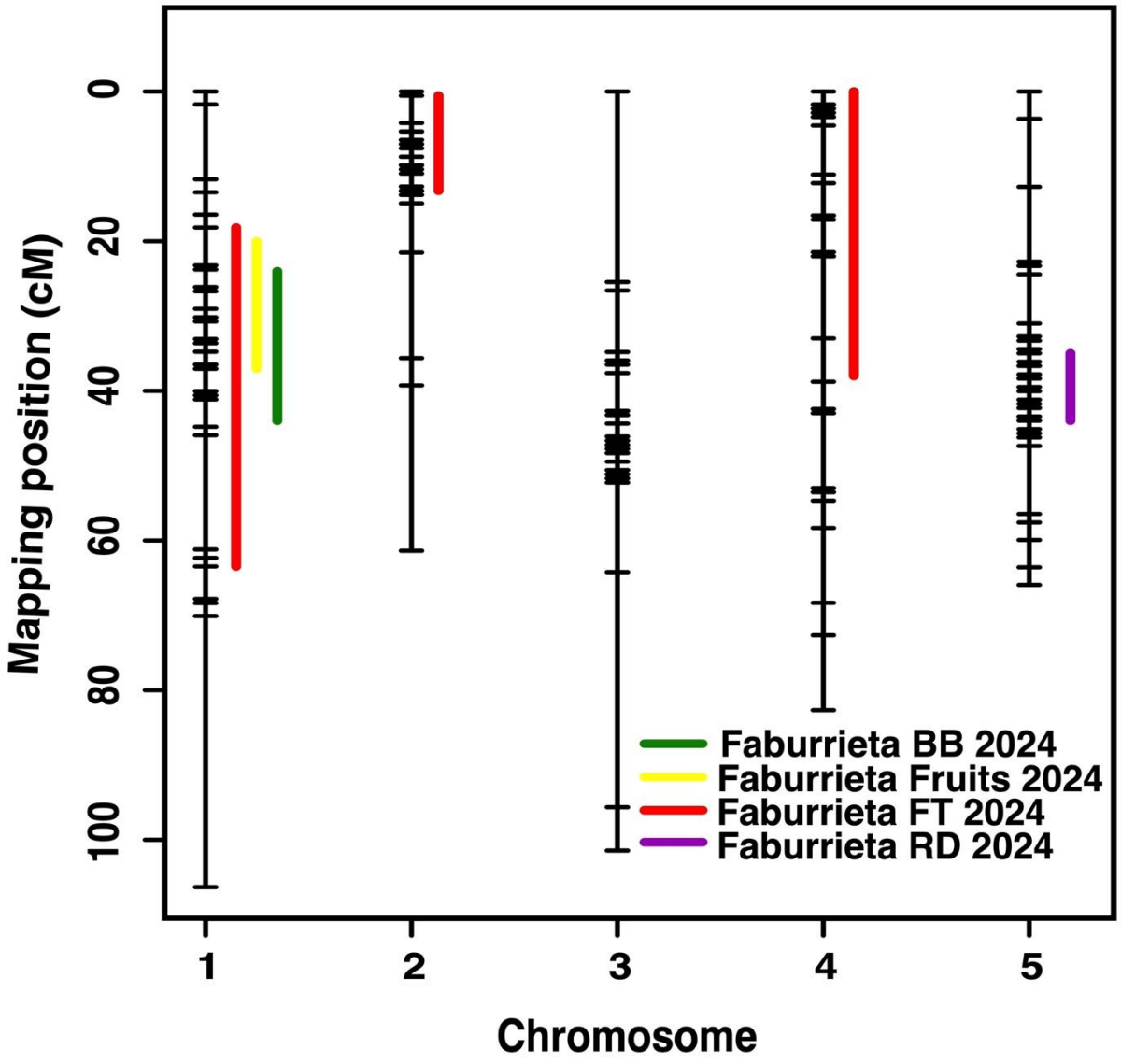


**Figure 6.** A genetic linkage map showing the 95% confidence intervals for the eQTL regions from the current study showing architectural traits data and flowering time data: number of basal branches (**Faburrieta BB 2024**), the number of fruits (**Faburrieta Fruits 2024**), flowering time **Faburrieta FT 2024**), and rosette diameter (**Faburrieta RD 2024**). The mapping positions of the epiQTL intervals are positioned according to the location on the chromosome in centimorgans.

## Tables

| Mapping Study | Phenotype | Environment | Pipeline | Significant epiQTL Regions | | | |
|---|---|---|---|---|---|---|---|
| | | | | | Position (cM) | | |
| | | | | Chromosome | 95% Confidence Interval | | |
| Faburrieta (2024) | Julian Flowering Date | Growth Chamber | Faburrieta (2024) | 1 | 18.18 | 41.15 | 63.46 |
| Faburrieta (2024) | Julian Flowering Date | Growth Chamber | Faburrieta (2024) | 2 | 0.56 | 5.34 | 13.25 |
| Faburrieta (2024) | Julian Flowering Date | Growth Chamber | Faburrieta (2024) | 4 | 0.00 | 10.00 | 38.00 |
| Faburrieta (2024) | Basal Inflorescence Number | Growth Chamber | Faburrieta (2024) | 1 | 24.00 | 33.63 | 44.00 |
| Faburrieta (2024) | Fruit Number | Growth Chamber | Faburrieta (2024) | 1 | 20.00 | 26.69 | 37.05 |
| Faburrieta (2024) | Rosette Diameter | Growth Chamber | Faburrieta (2024) | 5 | 34.97 | 38.00 | 43.98 |
| Cortijo et al. (2014) | Julian Flowering Date | Greenhouse | Faburrieta (2024) | 1 | 20.00 | 40.59 | 44.00 |
| Cortijo et al. (2014) | Julian Flowering Date | Greenhouse | Faburrieta (2024) | 4 | 6.00 | 34.00 | 52.00 |
| Cortijo et al. (2014) | Julian Flowering Date | Greenhouse | Faburrieta (2024) | 5 | 39.49 | 41.73 | 50.00 |
| Cortijo et al. (2014) | Julian Flowering Date | Greenhouse + Field (Average) | Cortijo et al. (2014) | 1 | 37.04 | 40.59 | 42.00 |
| Cortijo et al. (2014) | Julian Flowering Date | Greenhouse + Field (Average) | Cortijo et al. (2014) | 4 | 22.02 | 30.00 | 42.00 |
| Cortijo et al. (2014) | Julian Flowering Date | Greenhouse + Field (Average) | Cortijo et al. (2014) | 5 | 38.35 | 41.73 | 47.36 |

**Table 1.** The significant epiQTL markers found in my dataset and the Cortijo et al. (2014) dataset, showing the point estimate of the most significant marker as well as the positions of the 95% confidence interval for the significant association. The mapping study from which the results were derived is indicated (either my dataset or Cortijo et al.'s 2014 dataset), as is the phenotype that was mapped, the environment it was mapped in, and the pipeline that was used to analyze the results. I used my own pipeline to analyze my data as well as to re-analyse the Cortijo et al. (2014) data, but I also compared the results from my pipeline to the results that Cortijo et al. (2014) published. The Cortijo et al. (2014) detailed results were only published for the combined greenhouse + field phenotypes, which is why those results are presented here from her published study.

| Julian Flowering Date | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Study** | **Chromosome** | **Position** | **n** | **Replicates** | **Effect size** | **% variance** | **Power** | **Threshhold** | **LOD** |
| Faburrieta (2024) | 2 | 5.34 | 120 | 5 | 1.75 | 20.61 | 0.99 | 2.34 | 4.73 |
| Cortijo et al. (2014) | 2 | 6.47 | 121 | 6 | 0.66 | 6.1 | 0.28 | 2.46 | 1.14 |
| Faburrieta (2024) | 5 | 41.73 | 120 | 5 | 1.09 | 9.18 | 0.56 | 2.49 | 2.27 |
| Cortijo et al. (2014) | 5 | 41.73 | 121 | 6 | 1.51 | 25.11 | 0.99 | 2.41 | 8.69 |

**Table 2.** Post-hoc power analysis of eQTL regions that were significant in one study (either Cortijo et al. 2014 or the current study) but not the other one. Shown is the power to detect an effect in the study where the effect was found versus the power to detect an effect in the study where the effect was not found. The maximal marker effect size within the significant eQTL interval was used to calculate the power to detect an effect in that epiQTL intervalpower. Presented are the positions of the markers with the highest LOD scores within the epiQTL regions, the number of distinct epiRILs that were used, and the number of replicates of each epiRIL that were used. Also presented are the effect sizes, the percent of the phenotypic variance explained by the significant epi-marker, the power to detect an effect of that size, and the LOD score of the maximal marker in that epiQTL interval and the threshold that was used to determine significance (where above that threshold indicates significance).

Literature Cited

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. *The Structure and Function of DNA.*

Alonso C, Pérez R, Bazaga P, Medrano M, Herrera CM. 2014. Individual variation in size and fecundity is correlated with differences in global DNA cytosine methylation in the perennial herb Helleborus foetidus (Ranunculaceae). *Am J Bot.* 101(8):1309–1313. doi:10.3732/ajb.1400126.

Auge GA, Penfield S, Donohue K. 2019. Pleiotropy in developmental regulation by flowering-pathway genes: is it an evolutionary constraint? *New Phytol.* 224(1):55–70. doi:10.1111/nph.15901.

Banta JA, Dole J, Cruzan MB, Pigliucci M. 2007. Evidence of local adaptation to coarse-grained environmental variation in Arabidopsis thaliana. *Evolution.* 61(10):2419–2432. doi:10.1111/j.1558-5646.2007.00189.x.

Banta JA, Richards CL. 2018. Quantitative epigenetics and evolution. *Heredity*. 121(3):210–224. doi:10.1038/s41437-018-0114-x.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 19(7):889–890. doi:10.1093/bioinformatics/btg112.

Chebib J, Guillaume F. 2021. Pleiotropy or linkage? Their relative contributions to the genetic correlation of quantitative traits and detection by multitrait GWA studies. *Genetics.* 219(4). doi:10.1093/genetics/iyab159.

Choi HG, Moon BY, Kang NJ. 2015. Effects of LED light on the production of strawberry during cultivation in a plastic greenhouse and in a growth chamber. *Sci Hortic.* 189:22–31. doi:10.1016/j.scienta.2015.03.022.

Cohen J. 1992. Statistical Power Analysis. Current Directions in Psychological Science. 1(3):98–101. doi:10.1111/1467-8721.ep10768783.

Colomé-Tatché M, Cortijo S, Wardenaar R, Morgado L, Lahouze B, Sarazin A, Etcheverry M, Martin A, Feng S, Duvernois-Berthet E, et al. 2012. Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. Proc Natl Acad Sci USA. 109(40):16240–16245. doi:10.1073/pnas.1212955109.

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, Caillieux E, Hospital F, Aury J-M, Wincker P, et al. 2014. Mapping the epigenetic basis of complex traits. *Science.* 343(6175):1145–1148. doi:10.1126/science.1248127.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature.* 401(6749):157–161. doi:10.1038/43657.

Daxinger L, Whitelaw E. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet.* 13(3):153–162. doi:10.1038/nrg3188.

Day T, Bonduriansky R. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. Am Nat. 178(2):E18-36. doi:10.1086/660911.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. *Science.* 330(6004):622–627. doi:10.1126/science.1190614.

Gibney ER, Nolan CM. 2010. Epigenetics and gene expression. *Heredity.* 105(1):4–13. doi:10.1038/hdy.2010.54.

González R, Butković A, Rivarez MPS, Elena SF. 2020. Natural variation in Arabidopsis thaliana rosette area unveils new genes involved in plant development. *Sci Rep.* 10(1):17600. doi:10.1038/s41598-020-74723-4.

Henikoff S, Smith MM. 2015. Histone variants and epigenetics. Cold Spring Harb Perspect Biol. 7(1):a019364. doi:10.1101/cshperspect.a019364.

Hill WG. 2012. Quantitative genetics in the genomics era. *Curr Genomics.* 13(3):196–206. doi:10.2174/138920212800543110.

Jablonka E, Lamb MJ. 2015. The inheritance of acquired epigenetic variations. *Int J Epidemiol.* 44(4):1094–1103. doi:10.1093/ije/dyv020.

Jablonka E, Raz G. 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol.* 84(2):131–176. doi:10.1086/598822.

Jaganathan D, Bohra A, Thudi M, Varshney RK. 2020. Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor Appl Genet.* 133(5):1791–1810. doi:10.1007/s00122-020-03560-w.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P, et al. 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* 5(6):e1000530. doi:10.1371/journal.pgen.1000530.

Kilvitis HJ, Alvarez M, Foust CM, Schrey AW, Robertson M, Richards CL. 2014. Ecological epigenetics. *Adv Exp Med Biol.* 781:191–210. doi:10.1007/978-94-007-7347-9_10.

Kooke R, Johannes F, Wardenaar R, Becker F, Etcheverry M, Colot V, Vreugdenhil D, Keurentjes JJB. 2015. Epigenetic basis of morphological variation and phenotypic plasticity in Arabidopsis thaliana. *Plant Cell.* 27(2):337–348. doi:10.1105/tpc.114.133025.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in Arabidopsis thaliana. *Annu Rev Plant Biol.* 55:141–172. doi:10.1146/annurev.arplant.55.031903.141605.

Rapp RA, Wendel JF. 2005. Epigenetics and plant evolution. *New Phytol.* 168(1):81–91. doi:10.1111/j.1469-8137.2005.01491.x.

Rehm J. 2018. Commentary on Liang et al. (2018): The potential impact of medical cannabis on public health with respect to reducing prescription opioid use and associated harm. *Addiction.* 113(11):2071–2072. doi:10.1111/add.14408.

Richards CL, Rosas U, Banta J, Bhambhra N, Purugganan MD. 2012. Genome-wide patterns of Arabidopsis gene expression in nature. *PLoS Genet.* 8(4):e1002662. doi:10.1371/journal.pgen.1002662.

Richards EJ. 2006. Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet.* 7(5):395–401. doi:10.1038/nrg1834.

Robertson M, Richards C. 2015. Non-genetic inheritance in evolutionary theory - the importance of plant studies. Non-Genetic Inheritance. 2(1). doi:10.1515/ngi-2015-0002.

Roff DA. 2007. A centennial celebration for quantitative genetics. *Evolution*. 61(5):1017–1032. doi:10.1111/j.1558-5646.2007.00100.x.

Sen S, Johannes F, Broman KW. 2009. Selective genotyping and phenotyping strategies in a complex trait context. *Genetics.* 181(4):1613–1626. doi:10.1534/genetics.108.094607.

Sen S, Satagopan JM, Broman KW, Churchill GA. 2007. R/qtlDesign: inbred line cross experimental design. *Mamm Genome.* 18(2):87–93. doi:10.1007/s00335-006-0090-y.

Shimizu KK, Purugganan MD. 2005. Evolutionary and ecological genomics of Arabidopsis. *Plant Physiol.* 138(2):578–584. doi:10.1104/pp.105.061655.

Slatkin M. 2009. Epigenetic inheritance and the missing heritability problem. *Genetics.* 182(3):845–850. doi:10.1534/genetics.109.102798.

Tal O, Kisdi E, Jablonka E. 2010. Epigenetic contribution to covariance between relatives. *Genetics.* 184(4):1037–1050. doi:10.1534/genetics.109.112466.

Tonsor SJ, Alonso-Blanco C, Koornneef M. 2005. Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell Environ.* 28(1):2–20. doi:10.1111/j.1365-3040.2004.01264.x.

Unseld M, Marienfeld JR, Brandt P, Brennicke A. 1997. The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nat Genet.* 15(1):57–61. doi:10.1038/ng0197-57.

Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* 185(4):1108–1118. doi:10.1111/j.1469-8137.2009.03121.x.

Wei J-W, Huang K, Yang C, Kang C-S. 2017. Non-coding RNAs as regulators in epigenetics (Review). *Oncol Rep.* 37(1):3–9. doi:10.3892/or.2016.5236.