UPPSALA
UNIVERSITET

Correlation between flowering time, circadian rhythm and gene expression in *Capsella bursa-pastoris*

Johanna Nyström

Degree project in biology, Bachelor of science, 2013
Examensarbete i biologi 15 hp till kandidatexamen, 2013
Biology Education Centre and Department of Ecology and Genetics, Uppsala University
Supervisor: Karl Holm

UPPSALA
UNIVERSITET

Correlation between flowering time, circadian rhythm and gene expression in *Capsella bursa-pastoris*

Johanna Nyström

Degree project in biology, Bachelor of science, 2013
Examensarbete i biologi 15 hp till kandidatexamen, 2013
Biology Education Centre and Department of Ecology and Genetics, Uppsala University
Supervisor: Karl Holm

Abstract

The ability to adapt to the yearly environmental changes in their habitat is an important fitness trait for plants; even more so as most land plants are immobile. When information about day length and temperature is integrated with one another, the plants can estimate the time of the year and are therefore able to synchronize the flowering event to the most suitable environmental conditions of the year. For better understanding of how plants adapt to the environmental changes in their local habitat, knowledge about how differential gene expression affects phenotypic characters important for the adaption, such as flowering time and circadian period length, is desirable. The fact that *C. bursa-pastoris* is widely distributed among different kinds of habitat and exhibits natural variation in flowering time and period length makes it a suitable plant in which to study the genes that, through the circadian clock, regulate flowering time and period length. In this study differential gene expression amongst individuals in a worldwide sample, consisting of two regions, of *C. bursa-pastoris* has been investigated. In general, a correlation between differential gene expression and variable flowering times and period lengths of the individuals was found, but the correlation was stronger between differential gene expression and flowering time than between differential gene expression and period length. The two regions were found to be differentiated in terms of gene expression suggesting demographic differentiation, and the genes essential to flowering time seemed to differ between the regions.

Contents

Introduction

In all places of the world, the environmental conditions change over the year. Since land plants are immobile, the ability to adapt to the environmental changes in their habitat is an important fitness trait. Two often dramatically altering environmental conditions are temperature and day length.

The ability to detect and adapt to seasonal changes in day length is referred to as photoperiodism [1]. Photoreceptors sensitive to the ratio of red and far-red light measures the length of darkness (night) during a 24-hour cycle and thereby also measures the length of sunlight (day) [1]. When this information is integrated with information about temperature, the plants can estimate the time of the year and are therefore able to synchronize the flowering event to the most suitable environmental conditions of the year. In other words, the duration of certain day lengths combined with certain temperatures signal to the plant to shift from vegetative to reproductive development [1].

The response to inputs about day length and temperature and hence the timing of flowering is controlled by the circadian clock, an endogenous clock that integrates environmental cues with the organism's inner circadian rhythm. The inner rhythm is in fact generated by the circadian clock, which will maintain a free-running rhythm if the environmental cues does not change [1]. The rhythm can be described by three parameters: period length, phase and amplitude (figure 1) [2]. Amplitude is defined as half the peak-to-bottom distance of a circadian clock output, while phase is the time of the day when a given event of the clock output occurs, the peak for example, and is usually measured in zeitgeber time (ZT). Dawn is defined as ZT 0 [2]. Period length is defined as the time it takes to complete one cycle, which is approximately 24 hours in circadian rhythms as it is determined by the light/dark cycle [2]. However, since the day length varies greatly over a year in some parts of the world it has been found that the free-running period length can vary within a species, and often it has been found that the more the day length varies over a year, the longer the period length is [3]. The greatest variation in day length is found at the highest latitudes, resulting in a latitudinal cline in period length. Flowering time also varies within a species [4]. Since the circadian clock to a large extent determines flowering time, the variation has been found to correlate with circadian rhythm and day length in some populations where long period length and a highly variable day length correlates with increased flowering time [3].

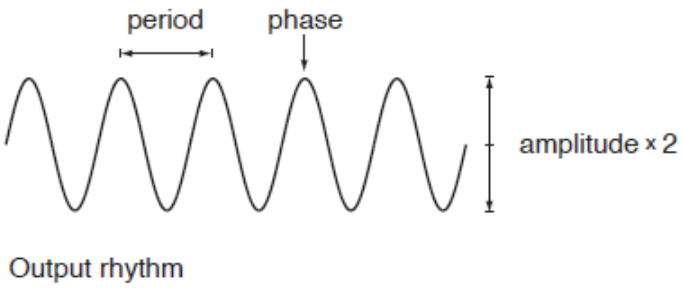

Figure 1. The three parameters of the endogenous circadian rhythm: amplitude, phase, and period length. Modified from [2].

Genetic regulation of the plant circadian clock

In the well-studied model plant *Arabidopsis thaliana*, the circadian clock is regulated by a complex feedback system. The system includes a feedback loop that fluctuates daily where the expression of *TIMING OF CAB EXPRESSION 1 (TOC1)* together with an unknown component stimulates the expression of *CLOCK ASSOCIATED 1 (CCA1)* and *LATE ELONGATED HYPOCOTYL (LHY)*, which in turn inhibits the expression of *TOC1* [5]. This feedback loop is involved in regulation of the circadian outputs as a response to changes in temperature, and the activity of *CCA1* and *LHY* have been found to have prolonging effects on period length [1, 6]. An important circadian regulated output in *A. thaliana* is the expression of *FLAVIN-BINDING FACTOR 1 (FKF1)* which interacts with a repressor of *CONSTANS (CO)* [7]. When the repressor is inhibited, expression of *CO* initiates which in turn activates the expression of *SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* through the expression of *FLOWERING LOCUS T (FT)* [8]. Together with the core clock genes, these three components are all a part of the circadian clock regulated pathway that is involved in determining flowering in higher plants [8].

Local adaptation

Local adaptation is an evolutionary process that acts on individuals within their habitat, i.e. locally [9]. Which traits that are fitness-related vary between habitats, therefore genotypes well fitted to the local habitat and its environmental changes will be selected for in that specific habitat [9]. As a result, individuals of local populations generally will have higher fitness in their own habitat than will individuals originating in other habitats [9]. Therefore, the variation seen in flowering time and period length between populations of different origins could possibly be due to local adaptation, where the populations are adapted to their own local habitat.

Capsella bursa-pastoris (Brassicaceae)

Capsella bursa-pastoris is a selfing tetraploid that is closely related to the model plant *A. thaliana* and belongs to the genus *Capsella* along with the outcrossing diploid *Capsella grandiflora* and the selfing diploids *Capsella rubella* and *Capsella orientalis* [10, 11]. *C. bursa-pastoris* is an annual weed, and in contrast to the other species of *Capsella*, it is one of the most widespread flowering plants in the world [12]. The plant can be found in Europe, Asia, Australasia, Africa and America [12]. It grows on latitudes between the equator in Kenya, although on relatively high elevations, to 71°N in Norway and on altitudes from the sea-level to 5900 m in the northwest Himalaya [12].

The origin of *C. bursa-pastoris* remains uncertain. There have been many attempts to determine whether or not the tetraploidy of *C. bursa-pastoris* is the result of a single genome duplication event including a single ancestor or if it is the result of multiple events and has multiple ancestors. Alleles of *C. rubella* have been found in the genome of some *C. bursa-pastoris* individuals located in the southern Europe. However, genetic introgression has been suggested as the most likely explanation to this since *C. bursa-pastoris* individuals from China where *C. rubella* does not occur lack *C. rubella* alleles, and also the alleles shared between *C. bursa-pastoris* and *C. rubella* is identical to near-identical, implying that the alleles were introduced in the *C. bursa-pastoris* gene pool recently [11, 13].
The fact that *C. bursa-pastoris* is widely distributed among different kinds of habitat and exhibits much natural variation in flowering time and period length make it a suitable plant in

which to study the genes that, through the circadian clock, regulate flowering time and period length. In an earlier study, where an early flowering accession from Puli, Taiwan (PL) was compared with a late flowering accession from Härnösand, Sweden (SE.14) several genes involved in the regulation of circadian clock outputs, such as *TOC1* and *CCA1*, were found to be differentially expressed between the accessions [4]. However, not only do these accessions represent different flowering times, period lengths and latitudinal origins, they are also originating from very different parts of the world. Therefore the comparison might be confounding since it is difficult to sort out whether the differential gene expression and phenotypes are primarily a result of local adaptation to different latitudes or if it has a demographic source as well. In order to clarify these issues, several individuals of *C. bursa-pastoris* with early flowering time and short period length needs to be tested against several individuals with late flowering time and long period length, respectively, in different parts of the plant's distribution.

Research aims

The aim of this project was to investigate the correlation between differential gene expression and variation in flowering time and period length in a worldwide sample of *C. bursa-pastoris*. Furthermore, genes related to variation in flowering time and period length was investigated.

More specifically:

I. Investigate how gene expression differs between individuals of *C. bursa-pastoris* with early and late flowering time.
II. Investigate how gene expression differs between individuals of *C. bursa-pastoris* with short and long period length.
III. Determine the genes whose differential gene expression could explain the variation seen in flowering time and period length.

Materials and methods

Plant sample

Plant material was gathered from 24 individuals located in two demographic and genetically differentiated regions: region I being Europe, Russia, the Middle East, North Africa and North America while region II samples originate from China and Taiwan (figure 2) [14]. All individuals in region I were growing within a 31°N – 63°N latitudinal range, had a flowering time (defined as number of days between seed germination and the first flower to blossom) ranging from 36 – 108 days and a period length between 24.5 – 27.8 hours. The individuals in region II were growing within a 23°N – 46°N latitudinal range, had a flowering time ranging from 27 – 91 days and a period length between 23.8 – 26.4 hours.

The plant material was grown and sampled in growth chambers under equal conditions, a 16h light and 8h dark photoperiod in 20°C.

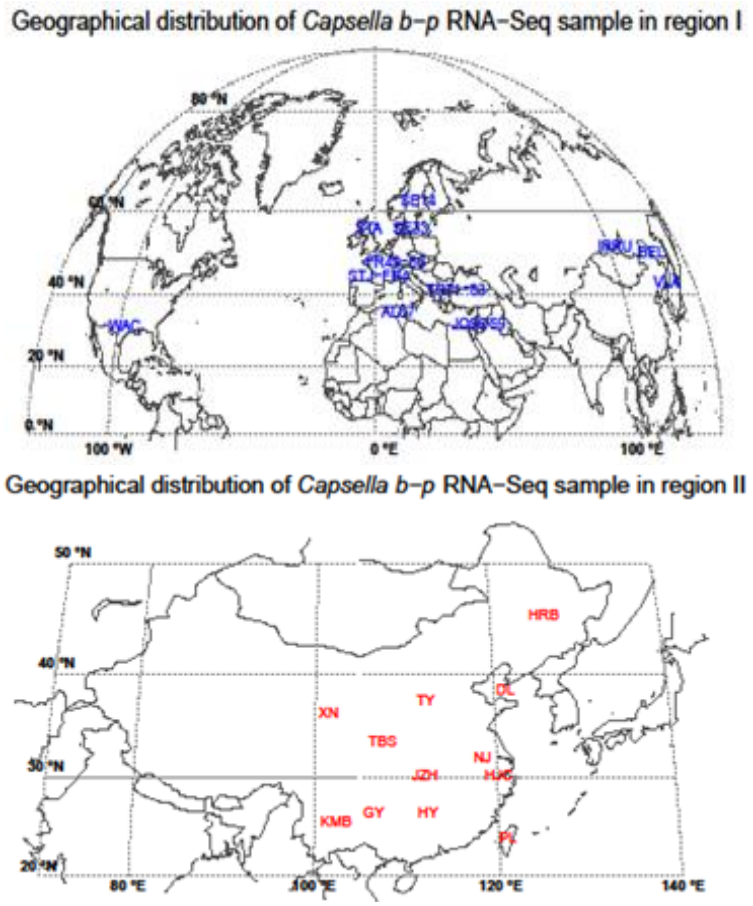

Figure 2. The locations of the 12 individuals belonging to region I (upper) and the locations of the 12 individuals belonging to region II (lower) [3].

Transcriptome sequencing and data count

When the seedlings were 10 days old, at ZT 8, the total RNA from green leaves, stems and roots was extracted. After cDNA synthesis the full transcriptome (RNA-seq) was sequenced on the *Illumina* platform at SciLife, Stockholm [15].

RNA-seq, the sequencing of the transcriptome using Next Generation Sequencing (NGS), reveals the set of expressed sequences in any biological tissue at the time of sampling. This since it produces paired-end reads where each read is assigned to a locus, represented by the existing transcripts in the tissue. In other words, the number of reads assigned to a locus correlates with the amount of that specific transcript present in the tissue at time of sampling [16]. NGS is a high-throughput sequencing method that produces billions of short reads, where two paired-end reads covering the same gene make up one count. The reads were 100 base pairs (bp) and the sequencing produced 35-100 000 000 reads/individual. The quality of the reads was controlled and the reads of low quality were filtered away. The remaining reads were mapped to the genome of the closely related reference species *C. rubella* [17]. With default settings of the mapping algorithm, about 97 % of the *C. bursa-pastoris* raw reads mapped to the genome of *C. rubella*, suggesting a very high similarity between the species. Out of 26521 genes annotated in *C. rubella*, the mean number of genes covered in the transcriptomes of the *C. bursa-pastoris* samples was 23493. The input data used for subsequent statistical analyses were the normalized number of counts per loci that mapped to *C. rubella*. The total amount of counts were 740 374 017.

Estimation of differential gene expression

DESeq is a software package from Bioconductor for use in the statistical platform R [16, 18, and 19]. The software allows for testing differentially expressed genes between individuals and groups of individuals. It applies a negative binomial distribution, instead of Poisson distribution [16]. This is motivated by the standard deviation when testing for differential gene expression being greater than what is assumed when using a Poisson distribution [16]. In the Poisson distribution the standard deviation is the same as the mean, while in a model based on a negative binomial distribution, the standard deviation is estimated from the data [16].

When testing samples in a count data set against each other, the effective library size needs to be estimated. This is due to the possibility that the libraries of the samples have been sequenced to different depths [20]. The estimation of the effective library size is also known as normalisation of the count data [20].

The variance seen between two counts depends on two things: sample-to-sample variation, also known as dispersion, and the uncertainty in measuring concentrations in number of reads, known as shot noise [20]. The dispersion can be estimated in three steps in DESeq before testing for differential gene expression. Initially, the dispersion value for each gene is estimated. Secondly, a curve is fitted to the estimates and finally each gene is assigned with a dispersion value somewhere between the per-gene dispersion value and the fitted value, depending on how conservative the estimation should be [20]. If the numbers of replicates are many, recommended at least n = 7, the least conservative mode of estimation can be used. DESeq also offers a very conservative form of estimation where all dispersions not aligned on the curve are considered as noise [20]. However, in this experiment the number of replicates was considered enough to use the default estimation, where dispersions below the fitted curve is defined as noise whereas dispersions above is considered as significant [20]. This estimation can be considered to be somewhere between the conservative and the least conservative mode of estimation [20].

Euclidean distance is the distance between two points, in this case between two samples when all dimensions, i.e. all genes expressed, are taken into account [20]. Through a Euclidean

heatmap the euclidean distances between the samples is visualized [20]. In similar ways Principal Component Analysis (PCA) plots integrates all information about the samples and enables visualization of clustering effects within the data set on a two or three dimensional scale [20]. The result of the visualizations in heatmaps and PCA plots is important in order to detect possible errors in sample handling and to decide on how to set up contrasts for testing.

Binomial testing in DESeq tests gene expression between two individuals or user-defined groups based on the above visualization and quality checks. In this investigation tests were performed between individuals and groups with differences in flowering time and circadian period length. After the test, lists of significantly differentially expressed genes and genes that are up- or downregulated, measured in mean of normalised counts, in one of the test groups compared to the other can be obtained. All lists include p-values for each gene that are adjusted for multiple testing with the Benjamini-Hochberg procedure, which controls the false discovery rate (FDR) [20, 21]. The lists also include fold change and ($\log_2$) fold change for each gene, where fold change is the difference in mean normalised counts of a gene between the two groups compared [20]. ($\log_2$) fold change is the logarithm (to basis 2) of the fold change, and enables an easier comparison in expression differences between the groups, where a ($\log_2$) fold change value of 0 equals no difference in mean normalised counts between the groups (fold change 1), a ($\log_2$) fold change value of 1 or -1 equals a 2 times difference, and so forth [20].

The results were merged with annotation data from the genome of *C. rubella*, since the raw reads of *C. bursa-pastoris* mapped to about 97 % of the *C. rubella* genome. The annotation data of the genome of *C. rubella* in turn is based on its homology with *A. thaliana*. This allows for further study of the genes biological functions in databases such as the Database for Annotation, Visualization and Integrated Discovery (DAVID). DAVID facilitates functional annotation and analysis of gene lists [22]. It integrates annotation data from its own and several other knowledge bases with the supplied gene list and a chosen suited reference gene list, in this case *A. thaliana* since the biological knowledge about its genome is widely studied. DAVID then generates different visualization tools that enable effective exploring of the biological knowledge associated to the gene list of interest.

Functional Annotation Table is an easy but non-statistical way to get information about all the biological functions connected to a specific gene on the uploaded list [23]. DAVID lists all functional annotations associated with a gene grouped together, this for the entire gene list. Functional Annotation Table is a suitable visualization tool when only interested in specific gene's biological functions. In Functional Annotation Clustering all similar biological functions are clustered together, and information about which genes that are associated to the cluster can be obtained [23]. This visualization tool is optimal when only interested in genes associated to a certain biological function and not in a specific gene's functional annotation. For each clustering a Group Enrichment Score is declared. The score is a mean of the clustering member's p-value and is used to rank the biological significance of the group, where low p-values will result in a high Group Enrichment Score. This equals an enrichment of the group, i.e. that more genes on the uploaded list are associated to that specific clustering than would be expected by chance [23]. The p-value of a clustering member, i.e. an annotation term, is also known as an EASE-score, which is a more conservative version of a Fisher's Exact Test p-value [23].

Gene Lists

Genes that were noted in DAVID as being related to flowering, circadian rhythm, photoperiodism, responsive to stimulus of gibberellic acid or cold were collected and ordered in gene lists. The annotation data accompanied with the ids and names of the genes in the lists was based on information derived from *The Annotation Information Resource (TAIR)* (www.arabidopsis.org), which often offers further more information about the gene function than DAVID [24].

For some genes directly related to flowering time or circadian rhythm, the level of expression was noted for all 12 individuals in region I and II, respectively. The difference in level of gene expression was then visualized in plots and the correlation between flowering time and expression levels was tested with a Spearman's rank-order test.

Results

Data visualization

The entire normalised count data set was visualized in a Euclidean heatmap and PCA plot with the individuals divided into region I and region II according to their demographic location (figure 3). The heatmap showed distinctive clustering between region I and II, with the individual HRB.132 as the only exception. HRB.132 was defined as belonging to region II but clustered with region I. Possibly HRB.132 could have been spread from region I to region II unintentionally by someone, or there could have been a mix-up of the sample labels. However, since the reason for this clustering pattern remained unknown, HRB.132 was not included in the tests. The clustering in the PCA plot followed the same pattern as in the heatmap but also showed two clusters within region I on the PC2 axis, but not on the PC1 axis. This pattern was considered more in detail when gene expression related to flowering time and period length was investigated.

Distinctive clustering can be a motivation to handle the data as several populations instead of one big population, where each population is considered as a replicate within a group of populations. Since there were bigger differences between than within regions, the count data could not be considered as homogenous but rather as two differentiated groups. Therefore, when testing differentially expressed genes related to flowering time and period length, region I and II were tested separately.

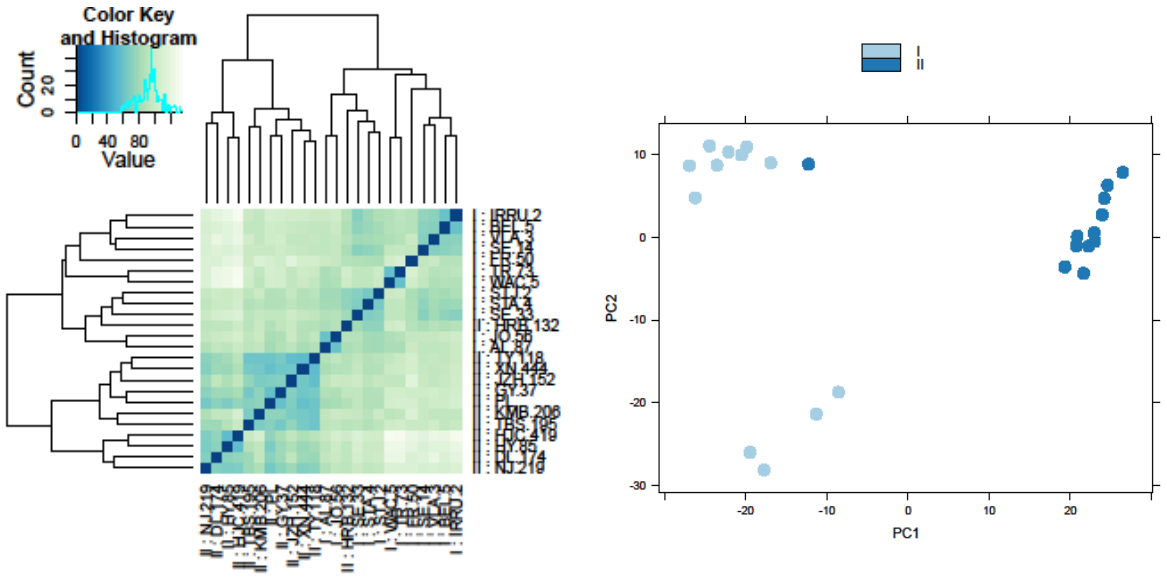

Figure 3. Euclidean Heatmap of individuals in region I and II (left) and PCA plot of region I versus region II (right).

Gene expression patterns in relation to flowering time

Region I

The individuals in region I were divided into three groups, with four individuals in each, depending on if they were considered having an early, intermediate or late flowering time. These groups were studied in a Euclidean heatmap and PCA plot (figure 4). The heatmap

revealed clustering of individuals with late and intermediate flowering time with one intermediate, individual TR.73, as an exception. All early flowering time individuals except STJ.2 clustered together. A clustering pattern of late flowering time individuals was also clear in the PCA plot. Here all late flowering individuals clustered and all but one early flowering time individual clustered on the PC1 axis. No definitive clustering was observed along the PC2 axis.

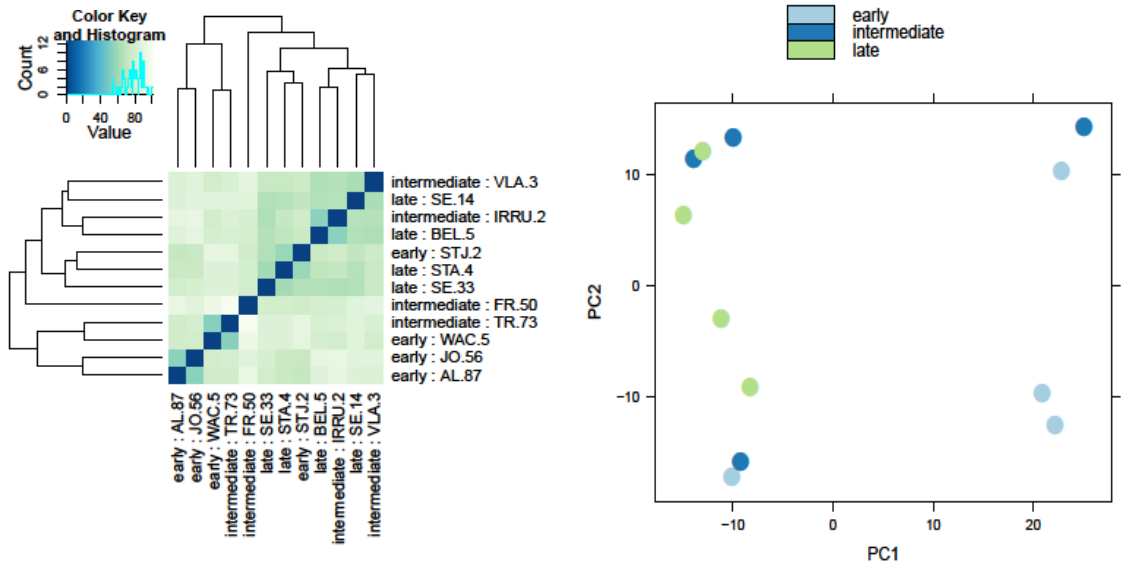

Figure 4. Euclidean heatmap (left) and PCA plot (right) of overall gene expression patterns among individuals in region I grouped according to flowering time (early, intermediate, late).

Intermediate samples were excluded from the test because including these could result in an overlap of gene expression levels, and therefore gradually differentially expressed genes could go unnoticed. This created two test groups with distinctively differentiated flowering time.

129 significantly differentially expressed genes were found. Among these, 51 genes had lower mean normalised counts in late flowering individuals compared to early flowering individuals and 78 genes had the opposite outcome. Three genes had zero counts in one of the test groups, two in late flowering individuals and one in early flowering individuals. The reason for this remains unclear. Possibly it could be due to gene silencing resulting in no expression of the genes, or the expression level of these genes could have been so low at time of sampling that the transcripts were missed during sequencing. From this study it is impossible to tell, but the mean normalised counts of two of the genes was considered as few in the individuals they were expressed in, supporting the latter thesis. The mean normalised counts of the third gene was about 100, which corresponds to a notable difference in mean expression level between early flowering individuals and individuals with late flowering. None of these genes were related to flowering time or circadian rhythm.

Out of the 129 significantly differentially expressed genes, five genes related to flowering time and circadian rhythm where found through the Functional Annotation Table in DAVID. Three of these genes were noted as directly related to flowering time and circadian rhythm; *FKF1*, *LUX* and *ATHAP2A* (table 1). *FKF1* is clock-controlled and regulates transition to

flowering, while *LUX* is required for a normal rhythm of multiple clock outputs during constant light or darkness [24]. *LUX* is co-regulated with *TOC1* and repressed by *LHY* and *CCS1* [24]. Both of them were upregulated in individuals with early flowering time than those with late flowering time. *ATHAP2A*, upregulated in late flowering individuals, is involved in the regulation of timing of transition from vegetative to reproductive phase in that the expression of *ATHAP2A* delays flowering [24, 25].

Although they were not noted in DAVID as being related to flowering time or circadian rhythm, *SIGMA FACTOR 5 (ATSIG5)* and *GERMIN-LIKE PROTEIN 3 (GER3)* could be of interest. *ATSIG5* was upregulated in late flowering individuals and responds to red and far red light while *GER3*, on the opposite, had more normalised counts in early than late flowering individuals, and not only responds to light and cold but is expressed more during evenings [24]. Both response to red and far red light and higher expression levels of a gene during a certain part of the day or night could imply involvement of the circadian clock.

Table 1. *List of genes related to flowering time and circadian rhythm that are significantly differentially expressed between individuals with early and late flowering time in region I.*
Negative (log_2) fold change (FC) equals higher expression level in early flowering individuals. All p-values are adjusted for multiple testing with the Benjamini-Hochberg procedure [21].

Gene number	Gene name	Biological function	FC	p-value	Reference
AT1G68050	FKF1	Circadian rhythm, flower development	-2.025	0.013	[24]
AT3G46640	LUX	Circadian rhythm	-1.878	0.006	[24]
AT5G12840	ATHAP2A	Transition from vegetative to reproductive phase	1.189	0.007	[24]
AT5G20630	GER3	Light and cold response, higher levels of RNA in the evening	-1.124	0.016	[24]
AT5G24120	ATSIG5	Blue, red and far red light response	0.959	0.010	[24]

Region II

The individuals in region II were similarly divided into three groups, with four individuals in each, depending on if they were considered having an early, intermediate or late flowering time. The groups were studied in a Euclidean heatmap and PCA plot (figure 5). In the heatmap all individuals with early flowering time clustered together and all with late flowering time except individual TBS.195 clustered together. However, the PCA plot revealed this pattern to be diffuse. Intermediate flowering time individuals showed no clustering effect at all, and when comparing the visualization of region II to that of region I it was clear that there were no evident clustering at all in region II (figure 4). The heatmap and PCA plot both showed that individual HRB is distinctively different from rest of the individuals in region II, verifying the observations made when the entire data set was visualized (figure 3).

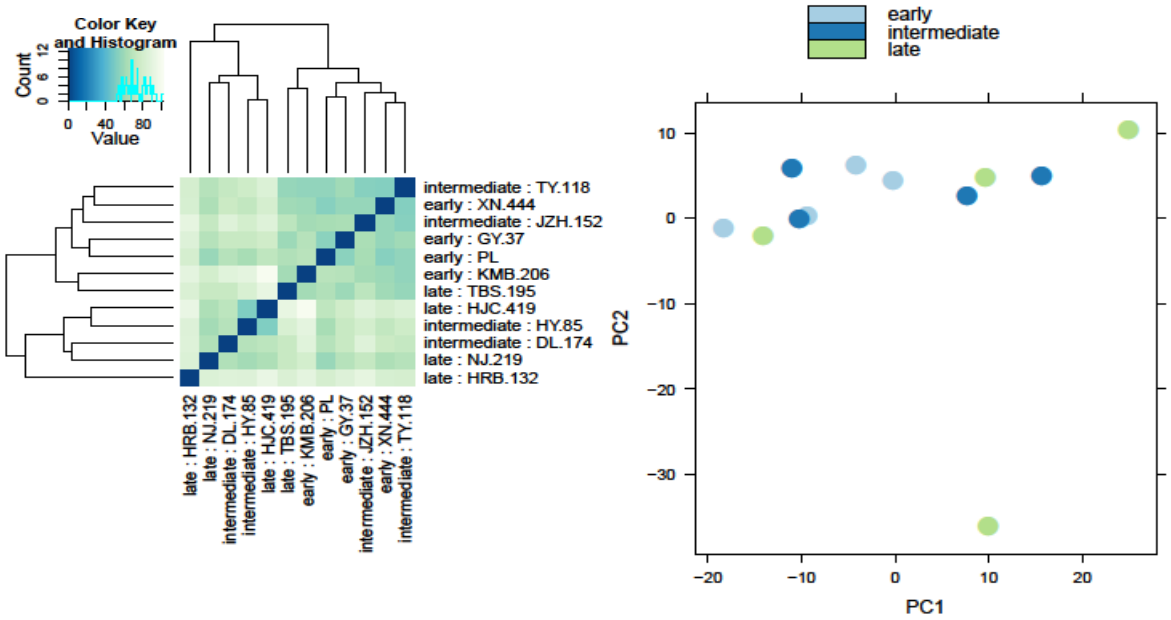

Figure 5. Euclidean heatmap (left) and PCA plot (right) of overall gene expression patterns among individuals in region II grouped according to flowering time (early, intermediate, late).

As for the region I samples, only the four individuals with the earliest flowering time were tested against the four individuals with the latest flowering time.

29 significantly differentially expressed genes were found. Among these, seven were downregulated in late flowering individuals compared with early flowering individuals and for 22 of the genes the outcome was the opposite. All genes were expressed in both test groups.

One gene, *SOC1*, was upregulated in early flowering individuals and noted as being directly related to flowering time and circadian rhythm in the Functional Annotation Table visualization tool of DAVID (table S1) [23]. *SOC1* acts downstream of *FT* and promotes flowering [8, 24]. Over-expression of *SOC1* counteracts late flowering and stimulates the phase transition from vegetative to reproductive stages [24].

Comparison between regions I and II

When testing for differential gene expression between groups of early and late flowering individuals, there were more genes upregulated in late flowering individuals compared to early flowering individuals in both regions. A few genes related to flowering time or the circadian rhythm was found, five and one in region I and II, respectively. The expression level of three of these; *FKF1*, *LUX* and *SOC1*, measured in normalised counts, was noted and visualized in plots for all 12 individuals in region I and II, respectively.

The difference in level of *FKF1* expression could clearly be visualized in region I, were the expression level of *FKF1* was significantly negatively correlated with flowering time (Spearman: $\rho = -0.818$, p-value = 0.002) (figure 6). In region II no correlation could be found (Spearman: $\rho = -0.159$, p-value = 0.640).

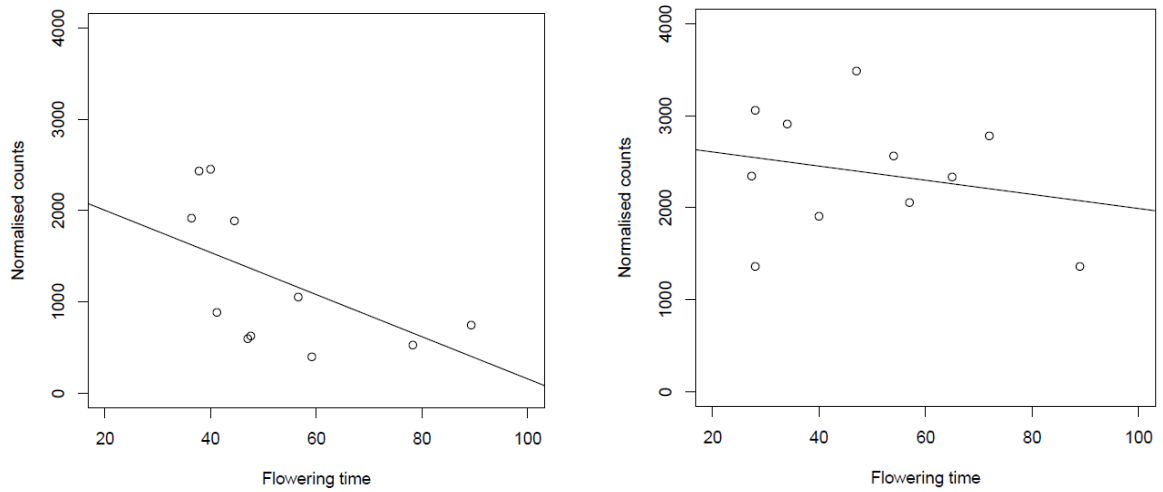

Figure 6. Normalised expression level of FKF1 in all 12 individuals in region I (left, p-value = 0.002) and all 12 individuals in region II (right, p-value = 0.640).

Differential levels of *LUX* expression could be seen in region I, were the expression level significantly correlated with early flowering time (Spearman: ρ = -0.832, p-value = 0.001) (figure 7). In region II no correlation could be found (Spearman: ρ = 0.091, p-value = 0.790).

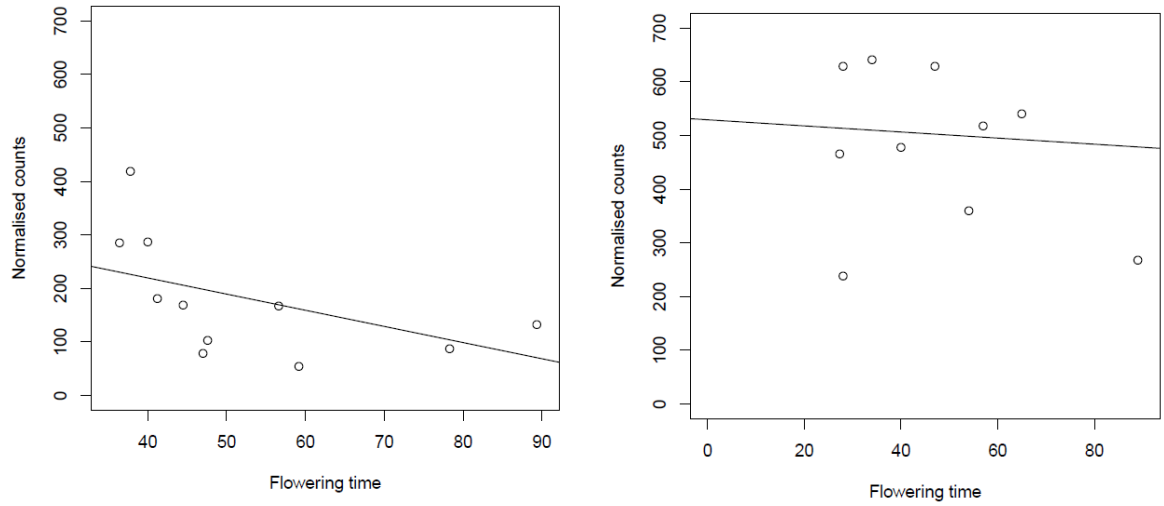

Figure 7. Normalised expression levels of *LUX* in all 12 individuals in region I (left, p-value = 0.001) and all 12 individuals in region II (right, p-value = 0.790).

The difference in normalised expression levels of *SOC1* was significantly correlated with differential flowering times in region II (Spearman: ρ = -0.843, p-value = 0.001) (figure 8). In region I no significant correlation could be detected (Spearman: ρ = 0.0140, p-value = 0.974).

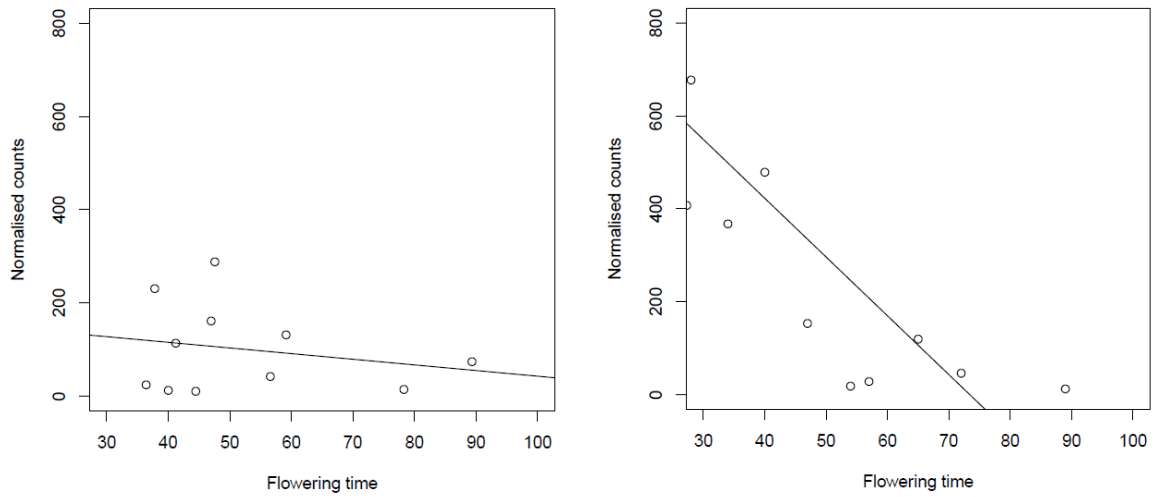

Figure 8. Normalised expression levels of *SOC*1in all 12 individuals in region I (left, p-value = 0.974) and all 12 individuals in region II (right, p-value = 0.001).

Gene expression patterns in relation to the circadian period length

Region I

The individuals in region I were divided into three groups depending on if they were considered having a short, intermediate or long period length and then studied in a Euclidean heatmap and PCA plot (figure 9). The heatmap showed a clustering pattern similar to that of the clustering according to their flowering time, except there were different individuals deviating from the general pattern between the plots (figure 4, figure 9). All individuals with intermediate period length clustered with the individuals with long period length, with WAC.5 as the only exception instead of TR.73 as in the flowering time clustering pattern. All individuals with short period length except IRRU.2 clustered, while in the flowering time heatmap the individual deviating from the early flowering time cluster was STJ.2. The same pattern between individuals with short and long period length was also detected along the PC1 axis in the PCA plot. No clustering could be seen along the PC2 axis.

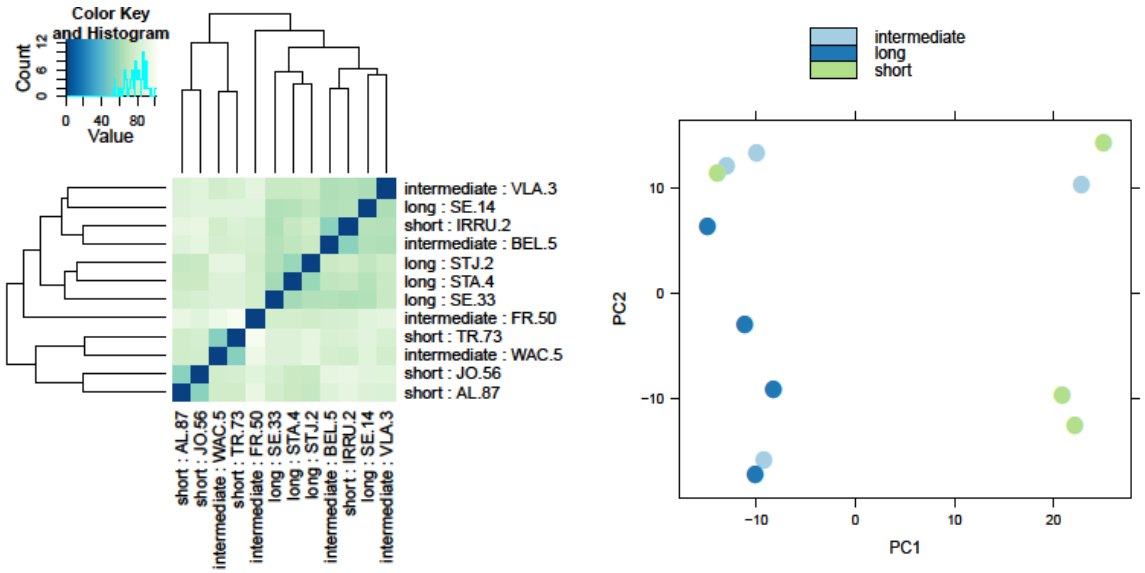

Figure 9. Euclidean heatmap (left) and PCA plot (right) of overall gene expression patterns among individuals in region I grouped according to period length (short, intermediate, long).

To get a clearer view on differential gene expression in individuals with different period lengths only the individuals with extreme period lengths, short and long, was included in the test.

74 significantly differentially expressed genes were found. Out of these, 40 genes were downregulated in individuals with a long period length compared to individuals with a short period length and for 34 genes the relationship was the opposite. One gene had zero counts in individuals with long period length and one in individuals with short period length. The mean normalized counts were however considered as few in the individuals they were expressed in, and were not related to flowering time or circadian rhythm.

No genes directly related to flowering time and period length was documented in the Functional Annotation Table visualization tool of DAVID [23]. However, *MATERNAL EFFECT EMBRYO ARREST 14 (MEE14)* was noted to be involved in embryo development ending in seed dormancy and it was upregulated in individuals with short period length compared to individuals with long period length (table S2) [24].

Region II

For the same reasons as above, the individuals in region II were divided into three groups depending on if they were considered having a short, intermediate or long period length and then studied in a Euclidean heatmap and PCA plot (figure 10). The heatmap showed a clustering pattern even weaker than that according to the flowering time of the individuals, where all early flowering individuals clustered (figure 5). Here, all individuals with short period length except HJC.419 clustered. However, no clustering pattern at all was detected in the PCA plot. Once again, individuals HRB.132 was noted as distinctively different from rest of the individuals in region II.

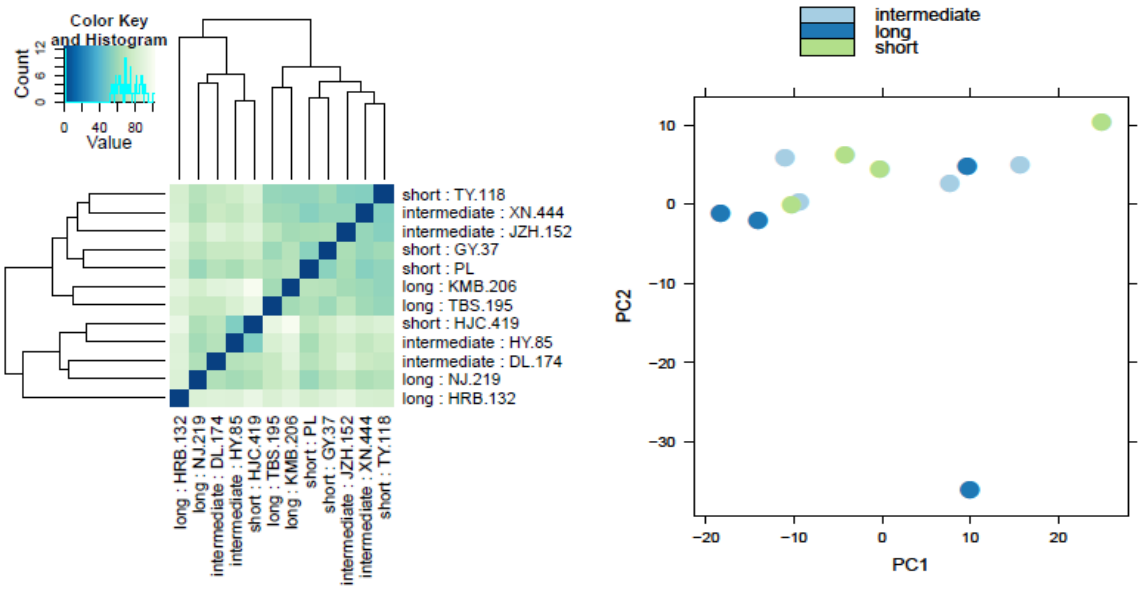

Figure 10. Euclidean heatmap (left) and PCA plot (right) of overall gene expression patterns among individuals in region II grouped according to period length (short, intermediate, long).

Individuals with intermediate period length were excluded from the binomial test. 1 significantly differentially expressed gene was found. This gene was not directly related to flowering time or the circadian rhythm, but involved in response to salt stress [23, 24]. The gene was upregulated in individuals with short period length compared to individuals with long period length (p = 1.038 e-30).

Comparison between regions I and II

When testing for differential gene expression between groups of individuals with short and long period length, the number of differentially expressed genes was greater in region I. In fact, only one gene was found to be differentially expressed in region II, while in region I the amount was 74. Here, there were more genes upregulated in individuals of short period length compared to those of long period length than genes having the opposite relationship.

Differential gene expression between two individuals – PL and SE.14

Finally, differential gene expression between the individual samples PL and SE.14 was tested. PL belongs to region II, has an early flowering time and a short period length while SE.14 belongs to region I and has a late flowering time and a long period length. The testing was motivated by the possibility to compare these results to an earlier microarray study on differential gene expression between PL and SE.14 [4]. Here differential gene expression between several replicates of PL and SE.14 was investigated, and the result were verified by testing differential gene expression between two samples from North America with more similar flowering times [4]. However testing between only SE.14 and PL accessions could possibly have yielded confounding results, since the test was not only on accessions with different phenotypic characters but with entirely different origins (table S3). By treating region I and II separately as done in the tests of this study this has been corrected for. The testing between PL and SE.14 was also motivated by the possibility to compare results when

running tests with several technical replicates of the same samples (microarray study [4]), independent biological replicates but not technical replication (this study) and with no replicates at all (this test).

309 significantly differentially expressed genes were found. For 22 genes, count data was found only in SE.14 and 16 genes only in PL. However, none of these genes were annotated as related to flowering or circadian rhythm in previous studies or DAVID [23]. In the earlier study 1642 genes were found to be significantly differentially expressed between the accessions [4].

Among the 309 significantly differentially expressed genes, 14 genes related to flowering time and period length was documented in the Functional Annotation Clustering visualization tool of DAVID (table S3) [22, 23]. Four genes that were found to be differentially expressed between groups of individuals in region I and II respectively were also found to be significantly differentially expressed between PL and SE.14. In region I *FKF1* and *LUX* were upregulated in individuals with early flowering time, *MEE14* in individuals with a short period length and *ATSIG5* in individuals with late flowering time (table 1). *SOC1* was upregulated in early flowering individuals of region II (table S1). This corresponds with *FKF1*, *LUX*, *SOC1* and *MEE14* being upregulated in PL and *ATSIG5* in SE.14, since PL was considered having an early flowering time and short period length while SE.14 had a late flowering time and a long period length.

Out of these genes only *SOC1* was found in the earlier study, here also upregulated in PL (table 2) [4]. Also in line with the earlier study, *LHY* and *CCA1* were found to have higher expression levels in SE.14 compared to PL [4]. *LHY* and *CCA1* encode proteins that have overlapping functions in a regulatory feedback loop that is closely associated with the circadian rhythm and prolongs its period length [4, 6].

Table 2. *List of genes related to flowering time or circadian rhythm significantly differentially expressed between accessions in both studies.*
Negative (log_2) fold change (FC) equals higher expression level in PL. All p-values are adjusted for multiple testing with the Benjamini-Hochberg procedure [21].

Gene number	Gene name	Biological function	Reference	This study		Earlier study [4]	
				FC	p-value	FC	p-value
AT1G01060	LHY	Circadian rhythm	[24]	5.7	3.99e-16	4.3	2.73e-03
AT2G46830	CCA1	Circadian rhythm	[24]	3.8	1.36e-08	3.0	5.19e-03
AT2G45660	SOC1	Promotes flowering, responds to gibberellic acid	[24]	-3.7	0.008	-2.5	4.87e-04

Three other transcription factors involved in regulation of circadian rhythm outputs were noted; *CIRCADIAN 1 (CIR1)*, *LHY/CCA1-LIKE 1 (LCL1)* and *LHY/CCA1-LIKE 5 (LCL5)*

and they all had more counts in SE.14 than PL [24]. The expression of *CIR1* increases tolerance to freezing stress before and after cold acclimation while *LCL1* and *LCL5* carry biological functions similar to that of *LHY* and *CCA1* [24, 26]. *CO*, upregulated in PL, acts upstream of *FT* and is involved in the promotion of flowering during long days [24]. *AGAMOUS-LIKE 2 (AGL2)* encodes a MADS-box transcription factor that is involved in flower and ovule development, in that it prevents indeterminate growth of flower meristems and ensures proper development of petals, stamens and carpels [24]. *AGL2* was upregulated in SE.14 compared with PL. *PSEUDO-RESPONSE REGULATOR 5 (APRR5)* is a transcriptional repressor of *LHY* and *CCA1* and it was upregulated in PL compared with SE.14. Mutation in this gene affects several circadian-associated biological procedures, such as flowering time in long day conditions and the sensitivity of the red light response in seedlings [24]. Although the function of *CATALASE 2 (CAT2)* is not directly related to flowering or period length, mRNA levels have been noted as higher during early morning implying some kind of circadian regulation of the gene [24]. *CAT2* is upregulated in SE.14.

Besides the genes that were found in both studies, expression of the core clock gene *TOC1,* which is inhibited by the expression of *LHY* and *CCAI,* was found to be upregulated in PL in the microarray study [4]. Also the expression of several genes in the Gibberellic Acid (GA) pathway, a pathway closely related to flower development, was found to be differentially expressed between the accessions where genes involved in the synthesis of gibberellic acid were downregulated in SE.14 compared to PL while genes encoding repressors of the pathway were upregulated [4].

Discussion

In order to gain better understanding of how plants adapt to the environmental changes in their local habitat, knowledge about how differential gene expression affects phenotypic characters important for the adaption, such as flowering time and circadian period length, is desirable. The fact that *C. bursa-pastoris* is widely distributed among different kinds of habitat and exhibits natural variation in flowering time and period length make it a suitable plant in which to study the genes that, through the circadian clock, regulate flowering time [4, 12]. Therefore, in this study differential gene expression amongst individuals in a worldwide sample of *C. bursa-pastoris* where these fitness related traits vary has been investigated.

Demographic histories of region I and II

The study shows that in terms of overall gene expression patterns, region I and II are differentiated groups. This is evident when the groups are visualized in a heatmap, where individuals of region I cluster together and individuals of region II cluster together (figure 3). The result is supported by an earlier study where the two regions were noted as having different amount of genetic variation [3]. Region I has greater genetic variation than region II, possibly due to an introgression event of *C. rubella* which is located in region I but not in II [10]. Also, the individuals of region I originate from an area closer to the putative place of origin of *C. bursa-pastoris* and could therefore supposedly be older while the individuals of region II could be an offspring population with an initially smaller sample size, resulting in lower levels of overall genetic variation (figure 2) [27]. This would also result in a difference in genetic variation between the regions and could explain more differential gene expression between individuals in region I than in region II, since a greater genetic variation within a population allows for greater variation in gene expression.

Gene expression patterns according to flowering time

In general, clustering of early flowering individuals and clustering of late flowering individuals is evident when visualizing overall gene expression patterns among individuals with different flowering times in region I (figure 4). Furthermore, 129 genes are significantly differentially expressed between early and late flowering individuals in region I, in which three genes are of certain interest since their biological functions could at least partially explain differential flowering time. Two of these genes, *FKF1* and *LUX*, are closely associated with the circadian clock and the expression of the genes promotes flowering [24]. They are upregulated in early flowering individuals and are negatively correlated with time to flowering (table 1, figure 6 and 7). *ATHAP2A*, upregulated in late flowering individuals, encodes a protein that regulates the timing of transition from vegetative to reproductive phase as it inhibits flowering [24].

In region II there is no clustering of early flowering individuals and of late flowering individuals when overall gene expression patterns among individuals with different flowering time in region II is visualized (figure 5). However, 29 genes are significantly differentially expressed between early and late flowering individuals. Among these, *SOC1*, a gene encoding a protein crucial to the promotion of flowering is upregulated in early flowering individuals and the expression of the gene is negatively correlated with time to flowering (table S1, figure 8).

According to these results, differential gene expression seems to be correlated with different flowering times among individuals of *C. bursa-pastoris.* The fact that *ATHAP2A* is upregulated in late flowering individuals compared to early flowering individuals while *FKF1, LUX* and *SOC1* are upregulated in early flowering individuals and negatively correlated with time to flowering supports this hypothesis.

When comparing differential expression of genes involved in the regulation of flowering time between the regions, different genes seems to be essential to flowering time in region I compared to region II. *FKF1* and *LUX* is significantly negatively correlated with flowering time in region I, but not in region II (figure 6 and 7). Similarly, *SOC1* is significantly negatively correlated with flowering time, but only in region II (figure 8). This finding supports the hypothesis that the two regions have different demographic histories, where different genes may have evolved to be crucial in the regulation of flowering time. However further data and analyses are warranted in order to draw any certain conclusions.

Gene expression patterns according to the circadian period length

A general clustering pattern where individuals with short period length cluster and individuals with long period length cluster is evident when visualizing overall gene expression patterns among individuals with different period lengths in region I (figure 4). Here, 74 differentially expressed genes between individuals of short and long period length is found, while in region II only one gene is differentially expressed. Also, there are no clustering patterns between individuals of short and long period length when overall gene expression patterns among individuals with different period length in region II is visualized (figure 5).

The period length of individuals in region II is more homogenous (data not shown). However, one needs to consider the fact that the range of latitude in the groups of individuals belonging to region II is narrower than that in region I (figure 2). Together with the difference in genetic variation between the regions mentioned earlier, this could at least partially explain the greater number of differentially expressed genes between individuals of short and long period length in region I compared to region II.

The number of significantly differentially expressed genes between individuals with different period length is less than between those of different flowering time. Even though there are differentially expressed genes between individuals of short and long period length, the difference is evident only in region I and only in genes that have no relation to either flowering time or the circadian rhythm.

Micro local adaptation

In the two heatmaps of overall gene expression patterns in region I there are distinctive clustering patterns except for two individuals, TR.73 and IRRU.2, which deviate from the pattern (figure 4 and 9). This could be due to small scale local adaptation. For example, the two individuals may have been growing in and sampled from an area in their local habitat exposed to more sunlight or with a higher mean day temperature, resulting in a somewhat different environment than places nearby. The same way that plants adapt to their local habitats these individuals could have adapted to the specific environment at the place in which they grow, and may therefore not have the same gene expression pattern as the other individuals in that particular habitat. Of course, this applies to all clustering results but since

there is a general pattern where individuals of similar phenotypic characters cluster together when overall gene expression is visualized in region I and since one can always expect a small amount of individual variation, the avoidance from the general clustering pattern of these individuals was not investigated further.

Differential expression between individuals PL and SE14

In this study, 309 genes were differentially expressed between PL and SE.14 while in the earlier microarray study there were 1642 differentially expressed genes [4]. However, no replicates where used for the testing on differential gene expression between the accessions in this study resulting in low statistical power to detect small differences in gene expressions between the accessions, while in the microarray studies several replicates were used, most likely enough to achieve statistical significance even in the small differences. This could explain the difference in amount of significantly differentially expressed genes between the studies.

Three genes closely related to the circadian clock and regulation of flowering time, *LHY, CCAI* and *SOC1,* are found in both studies. The results of the two studies correlate well; the genes are upregulated in the same accessions in both studies and the fold changes are similar (table 2).

Besides these, several different genes related to flowering time and the circadian clock are found only in one of the studies, such as *LHY/CCA1*-like genes in this study and GA-pathway associated genes in the microarray study [4]. Either way, the studies agree with one another in that genes with function in the promotion of flowering and shortening of the period length are upregulated in PL, while genes having the opposite effect are upregulated in SE.14 [4]. Furthermore this puts even more support to the hypothesis of differential gene expression being correlated with flowering time and period length. However, one can only carefully draw conclusions from comparison between only two accessions. The amount of differentially expressed genes between SE.14 and PL are several times greater than those between region I and II. As mentioned earlier, SE.14 and PL do not only have different phenotypic characters but entirely different origins, which confound the result of the test.

Conclusions

Region I and II are two genetically and possibly demographically differentiated groups where region I has a greater genetic variation [3]. In this study this is evident in that gene expression between groups of individuals with different phenotypic characters in region I differs more than between groups of individuals in region II. In general there is differential gene expression between individuals of different flowering times and period lengths in this worldwide sample of *C. bursa-pastoris*, but this difference is greater between individuals of different flowering times than of different period lengths. In fact, no differential gene expression except one gene could be found between individuals of short and long period length in region II. However, even though there is less genetic variation in this region, the shorter range of latitude in region II compared to region I need to be considered before drawing any conclusions.

The genes essential to flowering time seem to differ between the two regions, since the genes that promote flowering is negatively correlated to flowering time in one of the regions but not the other (figure 6, 7, and 8).

When comparing differential gene expression between two accessions of different phenotypic characters and origins, a greater amount of differential gene expression is found. This result is however confounded due to differences not only in phenotypic flowering time and period length but also in origin. Therefore it is difficult to sort out which genes that are differentially express due to phenotypic characters and which are differentially expressed due to different demographic histories. By testing for differential gene expression with several biological replicates and by treating region I and II separately the confounding effect has been corrected for in this study.

For further knowledge about correlation between differential gene expression and flowering time and period length in *C. bursa-pastoris* individuals, more studies where groups of individuals with different flowering times and period lengths are tested against each other is desirable. Most wanted is further in-depth knowledge about genes essential to flowering time and period length in populations with different genetic backgrounds. The study of the demographic history of *C. bursa-pastoris* will sort out the differences in genetic variation among populations of *C. bursa-pastoris* and clarify which populations that should be tested separately. Thereafter, testing of differential gene expression between groups of individuals with different phenotypic characters within a population will clarify which genes are essential to flowering time and period length in that specific population.

Acknowledgements

I wish to express my sincerest thanks to Karl Holm for his thorough, educational and ever so enthusiastic supervising. Thank you. I wish also to thank the staff, Martin Lascoux and Ulf Lagercrantz in particular, at the Department of Ecology and Genetics; *Plant ecology and evolution* for sharing their data. Lastly, to my dear friend Sara Kurland, thank you for always being there.

References

[1] RAVEN PETER H., EVERT RAY F., and EICHHORN SUSAN E. 2005. Biology of Plants. 7th ed. W. H. Freeman and Company Publishers, New York.

[2] ROBERTSON, MCCLUNG C. 2006. Plant circadian rhythm. The Plant Cell 18:792-803.

[3] HOLM K., GOULD P. D., HALL A., LASCOUX M., and LAGERCRANTZ U. 2010. Natural variation in circadian rhythm in a worldwide sample of *Capsella bursa-pastoris* (Brassicaceae). Manuscript, Uppsala University.

[4] SLOTTE T., HOLM K., MCINTYRE L. M., LAGERCRANTZ U., and LASCOUX M. 2007. Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (Brassicaceae). Plant Physiology 145:160-173

[5] ALABADÍ D., OYAMA T., YANOVSKY M. J., HARMON F. G., MÁS P., and KAY S. A. 2001. Reciprocal regulation between *TOC1* and *LHY/CCS1* within the *Arabidopsis* circadian clock. Science 293:271-277.

[6] LU S. X., KNOWLES S. M., ANDRONIS C., ONG M. S., and TOBIN E. M. 2009. *CIRCADIAN CLOCK ASSOCIATED 1* and *LATE ELONGATED HYPOCOTYL* function synergistically in the circadian clock of *Arabidopsis*. Plant Physiology 150:834-843.

[7] SAWA M., NUSINOW D. A., KAY S. A., and IMAIZUMI T. 2007. *FKF1* and *GI-GANTEA* complex formation is required for day-length measurement in *Arabidopsis*. Science 318:261-5.

[8] TURCK F., FORNARA F., and COUPLAND G. 2008. Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. Annual Review of Plant Biology 59:573-94.

[9] KAWECKI T.J. and EBERT D. 2004. Conceptual issues on local adaptation. Ecology Letters 55:1225-1241.

[10] HINTZ M., BARTHOLMES C., NUTT P., ZIERMANN S., HAMEISTER S., NEUFFER B., and THEISSEN G. 2006. Catching a 'hopeful monster': shepherd's purse (*Capsella bursa-pastoris)* as a model system to study the evolution of flower development. Journal of Experimental Biology 56:3531-42.

[11] HURKA H., FRIESEN N., GERMAN D. A., FRANLE A., and NEUFFER B. 2012. 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). Molecular Biology 21:1223-1238.

[12] AKSOY A., DIXON J. M, and HALE W. H. G. 1998. *Capsella bursa-pastoris* (L.) Medikus (*Thlapsi bursa-pastoris* (L.), *Bursa bursa-pastoris* (L.) Shull, *Bursa pastoris* (L.) Weber). Journal of Ecology 86:171-186

[13] SLOTTE T., CEPLITIS A., NEUFFER B., HURKA H., and LASCOUX M. 2006. Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. American Journal of Botany 93:1714-17

[14] SLOTTE T., HUANG H., LASCOUX M., and CEPLITIS A. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). Molecular Biology and Evolution 25:1472-1481.

[15] SCILIFELAB. Stockholm, Sweden. (http://www.scilifelab.se/)

[16] ANDERS S., HUBER W. 2010. Differential expression analysis for sequence count data. Genome Biology 11:R106.

[17] WRIGHT S. 2013. University of Toronto. In press

[18] GENTLEMAN R., CAREY V. J., BATES D. M., BOLSTAD B., DETTLING M., DUDOIT S., ELLIS B., GAUTIER L., GE Y., ET AL. 2004. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biology 5: R80.

[19] R CORE TEAM. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org/.

[20] ANDERS S. and HUBER W. 2012. Differential expression of RNA-seq data the gene level – the DESeq package. European Molecular Biology Laboratory (EMBL). Heidelberg, Germany.

[21] BENJAMINI Y. and HOCHBERG Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 57:289-300.

[22] DENNIS JR G., SHERMAN B. T., HOSACK D. A., YANG J., GAO W., LANE H C., and LEMPICKI R. A. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 4:R60

[23] HUANG D.W., SHERMAN B.T., and LEMPICKI R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 4:44-57. [PubMed]

[24] LAMESH P., BERARDINI T.Z., LI D., SWARBECK D., WILKS C., SASIDHARAN R., MULLER R., DREHER ., ALEXANDER D.L., GARCIA-HERNANDEZ M., KARTHIKEYAN A.S., LEE C.H., NELSON W.D, PLOETZ L., SINGH S., WENSEL A. and HUALA E. 2011. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucl. Acids Res. Doi: 10.1093/nar/gkr1

[25] WENKEL S., TURCK F., SINGER K., GISSOT L., LE GOURRIEREC J., SAMACH A., and COUPLAND G. 2006. *CONSTANS* and the *CCAAT Box Binding Complex* share a functionally important domain and interact to regulate flowering of *Arabidopsis.* The Plant Cell 18:2971-2984.

[26] GUAN Q., WU J., ZHANG Y., JIANG C., LIU C., CHAI C., and ZHU J. 2013. A DEAD box RNA helicase is critical for pre-mRNA splicing, cold-responsive gene regulation, and cold tolerance in Arabidopsis. The Plant Cell 25:342-356.

[27] CEPLITS A., SU Y., LASCOUX M. 2005. Bayesian inference of evolutionary history from interaction chloroplast microsatellites in the cosmopolitan weed *Capsella bursa-pastoris* (Brassicaceae). Molecular Ecology, 14:4221-4233.

Supplementary material

Table S1. *List of genes related to flowering time and circadian rhythm that are significantly differentially expressed between individuals with early and late flowering time in region II.*
Negative (log_2) fold change (FC) equals higher expression level in early flowering individuals. All p-values are adjusted for multiple testing with the Benjamini-Hochberg procedure [22].

Gene number	Gene name	Biological function	FC	p-value	Reference
AT2G45660	SOC1	Promotes flowering, responds to gibberellic acid	-3.594	0.030	[24]

Table S2. *List of genes related to flowering time and circadian rhythm that are significantly differentially expressed between individuals with short and long period length in region I.*
Negative (log_2) fold change (FC) equals higher expression level in individuals with short period length. All p-values are adjusted for multiple testing with the Benjamini-Hochberg procedure [22].

Gene number	Gene name	Biological function	FC	p-value	Reference
AT2G15890	MEE14	Embryo development ending in seed dormancy	-1.675	< 0.001	[24]

Table S3 and S4 on the following page.

Table S3. *List of genes related to flowering time and circadian rhythm that are significantly differentially expressed between PL and SE.14.*
Negative (log_2) fold change (FC) equals higher expression level in PL. All p-values in this study are adjusted for multiple testing with the Benjamini-Hochberg procedure [22].

Gene number	Gene name	Biological function	FC	p-value (adjusted)	Reference
AT1G01060	LHY	Circadian rhythm	5.726	3.993e-16	[24]
AT1G68050	FKF1	Circadian rhythm, flower development	-3.419	2.009e-05	[24]
AT2G45660	SOC1	Promotes flowering, responds to Gibberellic acid	-3.665	0.008	[24]
AT2G46830	CCS1	Circadian rhythm	3.798	1.362e-08	[24]
AT3G09600	LCL5	Circadian rhythm	4.142	2.040e-07	[24]
AT3G46640	LUX	Circadian rhythm	-3.347	0.012	[24]
AT4G35090	CAT2	mRNA levels high in morning	3.051	6.001e-06	[24]
AT5G02840	LCL1	Circadian rhythm	2.939	< 0.001	[24]
AT5G15800	AGL2	Flower development	6.720	1.980e-05	[24]
AT5G15840	CO	Circadian rhythm, regulation of flowering	-4.063	0.038	[24]
AT5G24120	ATSIG5	Response to blue, red and far red light	2.267	0.005	[24]
AT5G24470	APRR5	Circadian rhythm, flowering	-2.443	0.003	[24]
AT5G37260	CIR1	Circadian rhythm	2.812	0.003	[24]

Table S4. *Phenotypic data characterizing SE.14 and PL.*

Id	Origin	Region	Latitude	Longitude	Flowering time	Period length
PL	Taiwan	II	23.97	120.95	27.3	24.23
SE.14	Sweden	I	62.64	17.94	107.5	27.84