LUND UNIVERSITY

Inbreeding affects gene expression differently in two self-incompatible Arabidopsis lyrata populations with similar levels of inbreeding depression.

Menzel, Mandy; Sletvold, Nina; Ågren, Jon; Hansson, Bengt

[Link to publication](#)

Inbreeding affects gene expression differently in two self-incompatible *Arabidopsis lyrata* populations with similar levels of inbreeding depression

Mandy Menzel[1,2], Nina Sletvold[3], Jon Ågren[3], Bengt Hansson[1*]

Affiliation:

[1] Molecular Ecology and Evolution Lab, Department of Biology, Lund University, Sweden.

[2] Respiratory Immunopharmacology, Department of Experimental Medical Science, Lund University, Sweden.

[3] Plant Ecology and Evolution, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Sweden.

* Corresponding author: Bengt Hansson, Department of Biology, Lund University, Ecology Building, Sölvegatan 37, Lund 223 62, Sweden. E-mail: Bengt.Hansson@biol.lu.se.

Abstract

Knowledge of which genes and pathways are affected by inbreeding may help understanding the genetic basis of inbreeding depression, the potential for purging (selection against deleterious recessive alleles) and the transition from outcrossing to selfing. *Arabidopsis lyrata* is a predominantly self-incompatible perennial plant, closely related to the selfing model species *A. thaliana*. To examine how inbreeding affects gene expression, we compared the transcriptome of experimentally selfed and outcrossed *A. lyrata* originating from two Scandinavian populations that express similar inbreeding depression for fitness ($\partial \approx 0.80$). The number of genes significantly differentially expressed between selfed and outcrossed individuals were 2.5 times higher in the Norwegian population (≈ 500 genes) than in the Swedish population (≈ 200 genes). In both populations a majority of genes were up-regulated upon selfing ($\approx 80\%$). Functional annotation analysis of the differentially expressed genes showed that selfed offspring were characterized by (i) up-regulation of stress-related genes in both populations, and (ii) up-regulation of photosynthesis-related genes in Sweden but down-regulation in Norway. Moreover, we found that reproduction- and pollination-related genes were affected by inbreeding only in Norway. We conclude that inbreeding causes both general and population-specific effects. The observed common effects suggest that inbreeding generally up-regulates rather than down-regulates gene expression and affects genes associated with stress response and general metabolic activity. Population differences in number of affected genes and in effects on the expression of photosynthesis-related genes show that the genetic basis of inbreeding depression can differ between populations with very similar levels of inbreeding depression.

Keywords: Inbreeding, gene expression, stress-related genes, *Arabidopsis lyrata*, RNA-Seq

Introduction

The fitness consequences of inbreeding were investigated already by Darwin in order to understand why outcrossing is the prevalent mechanism of reproduction in nature. He found self-fertilization to be highly disadvantageous since the inbred progeny were often less viable and less fertile; a phenomenon known as inbreeding depression (Darwin 1876). Inbreeding depression results from exposure of recessive harmful alleles due to increased homozygosity at inbreeding, and may be the product of several mechanisms. According to the most generally accepted model, the dominance or partial dominance model, deleterious recessive mutations, which occur at low frequencies throughout the genome, will be detrimentally expressed as inbreeding increases homozygosity. In contrast, the overdominance model suggests that balancing selection favours heterozygous genotypes over homozygote ones, and that fitness is therefore reduced during inbreeding (Charlesworth and Charlesworth 1987). Other models assume (i) that closely linked genes with different deleterious recessive alleles in repulsion will result in lower fitness in homozygotes compared to heterozygotes in an apparently overdominant fashion (pseudo-overdominance), or (ii) that unlinked deleterious alleles can compensate each other in a positive epistatic manner in hybrids, causing heterosis (Charlesworth and Willis 2009). Thus, there are several possible mechanisms resulting in the expression of inbreeding depression upon selfing.

However, the negative consequences of inbreeding can in some situations be more than balanced by the transmission advantage and the reproductive assurance that a selfing mutant may enjoy in an otherwise outcrossing population. Selfing allows the transmission of two parental copies of genes to the progeny through female function compared with only one copy per parent among outcrossing individuals. This is advantageous because a higher proportion of own genetic material will be transmitted (Fisher 1941), but also because locally adapted genotypes will not be broken up by recombination in each generation. Moreover, reproductive assurance may favour the evolution of selfing in environments with few suitable mates and where pollinators are scarce (Jarne and Charlesworth 1993). The direction of mating system evolution is thus expected to depend on the magnitude of inbreeding depression, but also on environmental context. Indeed the magnitude of inbreeding depression is much stronger in regularly outcrossing species than in obligate selfers (Barrett and Charlesworth 1991; Husband and Schemske 1996; Charlesworth and Willis 2009; Winn, et al. 2011).

Although the fitness consequences of inbreeding are relatively well studied in some systems (Husband and Schemske 1996; Keller and Waller 2002), little is known about the effects of inbreeding on gene expression (Kristensen et al. 2010; Hansson et al. 2014). Recent developments in

'omic' technologies and systems biology allow new approaches for studying inbreeding-induced gene expression variation. For example, transcriptomics and proteomics allow visualisation of gene expression and protein changes in response to inbreeding and may thus help unravelling pathways affected. Furthermore, SNP arrays and QTL mapping can assist in finding the deleterious alleles causing inbreeding depression. Transcriptomics studies of inbreeding effects in *Drosophila melanogaster* have found that inbreeding mainly affects response to stress and fundamental metabolic processes (Kristensen et al. 2010). One of the most common effects is increased expression of heat shock proteins, which are known to bind to unstable proteins and facilitate correct protein folding (Gething and Sambrook 1992). The induction of heat shock proteins is likely to be a general stress response that buffers the negative effects of expression of deleterious mutations causing protein misfolding (Ayroles et al. 2009). It is also well documented that inbreeding increases sensitivity towards environmental stress in inbred individuals (Bijlsma et al. 2000), suggesting that the up-regulation of stress response genes is a general response to both genetic and environmental stress (Sorensen et al. 2003). Stress responses are associated with energetic costs, and inbred individuals in general show a lower metabolic efficiency and an up-regulation of genes involved in metabolic processes (Kristensen et al. 2006; Ayroles et al. 2009). However, so far it is not known to what extent changes in gene expression reflect (a) direct effects of deleterious mutations in the corresponding genes, and (b) indirect downstream responses to buffer the negative effects of stress caused by exposure of other deleterious alleles, respectively.

Several transcriptomics and genome-wide expression studies of inbreeding have been conducted using *Drosophila*, but there are few corresponding studies of plants. Therefore, how inbreeding affects gene expression is still largely unknown. This is unfortunate because knowledge of which genes and pathways are affected by inbreeding is of importance for understanding central questions in evolutionary and conservation biology, including the genetic basis of inbreeding depression, the potential for purging (selection against deleterious recessive alleles) and the transition from outcrossing to selfing (Bijlsma et al. 2000; Charlesworth and Willis 2009; Kristensen et al. 2010). Purging and transition to selfing are for instance less likely when inbreeding depression is caused by many genes with weakly deleterious effects and when gene interaction is pronounced (Charlesworth et al. 1990, 1993; Ayroles et al. 2009).

In the present study, we investigate the effects of inbreeding on gene expression in experimentally selfed and outcrossed *Arabidopsis lyrata*, a perennial herb in the family Brassicaceae and a close relative of the model organism *A. thaliana* (Kaul et al. 2000). Experimental plants originated from two self-incompatible, Scandinavian *A. lyrata* populations, which both express severe inbreeding depression for overall fitness measured across two growing seasons in a

glasshouse (ð≈0.80), with similar timing and magnitude of inbreeding depression across life stages (Sletvold et al. 2013). The experimentally induced inbreeding level, i.e. selfing (f = 0.5), will on average reduce heterozygosity by 50% relative to the outcrossed controls (f = 0). Therefore, any effect of the experiment is likely to be caused by increased homozygosity and expression of partly recessive alleles in the selfed individuals. We measure gene expression levels in the leaves of young plants (12 weeks old) because we want to test inbreeding effects at an early life stage. We apply transcriptome and pathway analyses and identify a large number of genes and biological functions that are sensitive to inbreeding. We also conduct gene interaction network analyses of the genes that were significantly affected by inbreeding. The results are discussed in light of the similar magnitude of inbreeding depression in the two study populations (Sletvold et al. 2013) and the fact that a mating-system transition from self-incompatibility to self-fertility has occurred in several *Arabidopsis* lineages, including in *A. thaliana* and some North American *A. lyrata* populations (Clauss and Koch 2006; Mable and Adam 2007).

Material and Methods

Study species, crossings and RNA extraction

Arabidopsis lyrata ssp. *petraea* has a disjunct, patchy distribution in Europe (Jalas 1994), and occurs across a wide altitudinal range in habitats characterized by moderate soil disturbance. In 2007, seeds were collected from 50 plants in each of two large Scandinavian *A. lyrata* ssp. *petraea* populations (>2000 individuals): an alpine population in Norway (Spiterstulen: 1100 m a.s.l., 61°41'N, 8°25'E), and a coastal population in Sweden (Stubbsand; <5 m a.s.l., 63°13'N, 18°58'E). The populations are differentiated at microsatellite loci (Gaudeul et al. 2007), as well as in phenotypic traits such as flowering phenology (Sandring et al. 2007), leaf trichome production (Løe et al. 2007) and tolerance to drought (Sletvold and Ågren 2012). Progeny raised from field-collected seeds were grown for one generation and randomly outcrossed in the greenhouse. Detailed descriptions of the laboratory conditions are given elsewhere (Sletvold et al. 2013) and will only briefly be mentioned here.

In 2008, seeds were produced through controlled self- and cross-pollination, respectively. To circumvent the self-incompatibility system, flowers were self-pollinated at the bud stage (Cabin et al. 1996; Kärkkäinen et al. 1999; Kachroo et al. 2002). Each cross-pollinated flower received pollen from two random donors within the source population. In 2009, nine replicates per cross type of 24 maternal parents per population were planted in 5.5x5.5 cm pots. In each pot, two seeds were planted. The pots were stratified for four days at 4°C in a dark cold-room and subsequently transferred to a growth room and kept at 20°C 16h day (150 µEm^{-2}s^{-1})/16°C 8h night. After 3 weeks,

seedlings were thinned to one per pot. After ca. 12 weeks, one rosette leaf per experimental plant was collected in RNALater (Qiagen), and the sample was frozen. At this stage, all plants were still vegetative; outbred progeny were on average twice the size of inbred progeny in both populations (∂ = 0.50; effect of crosstype $P < 0.0001$ in two-way ANOVA), but there was no significant difference between populations in plant size (effect of population $P = 0.30$, population × crosstype $P = 0.95$ (Sletvold et al. 2013)).

To verify the success of experimental selfing at the bud stage, maternal plants and their progeny were genotyped at 7 microsatellite loci (Clauss et al. 2002; Woodhead et al. 2007; Sletvold et al. 2013). DNA was extracted and the microsatellites PCR-amplified in two touchdown multiplexes using Qiagen Multiplex PCR Kit (PCR conditions given in (Sletvold et al. 2013)). The PCR products from the two multiplexes were pooled and then separated based on length and primer-labelling using ABI Prism 3730 capillary Sequencer (Applied Biosystems). Genotypes were scored with GeneMapper 4.0 (Applied Biosystems). Parentage analyses were performed to verify the experimental selfing by matching the genotype at the seven loci of each progeny to the genotype of the mother. To verify selfing we used a strict criterion, i.e. no mismatches were allowed between the progeny and the mother at any loci. Among the outcrossed progeny not a single individual had a perfect match to their mothers, which indicates high power to detect contamination during the selfing experiment with this set of microsatellites. The overall proportion of contaminants in the whole greenhouse population (consisting of more that a thousand plants of which 848 were tested) was 0.15, and ranged from 0 to 0.78 within maternal lines.

RNA extraction and sequencing

In total, 24 individual seedlings (progeny) that were the products of outcrossing (12 individuals; 5 Norwegian and 7 Swedish) or of verified selfing (12 individuals; 5 Norwegian and 7 Swedish), respectively, were selected for RNA sequencing. Total RNA extraction, DNAse treatment and rRNA reduction as well as RNA quality control were performed by LGC Genomics (Berlin, Germany). LGC Genomics also conducted mRNA isolation and cDNA synthesis as well as Illumina library preparation, indexing and library quality control. LGC genomics then performed cluster generation and sequencing using Illumina HiSeq2000 (100 bp paired-end).

Finding differentially expressed genes

Raw reads of the transcriptome were assembled for each sample using TopHat (version 2.04 (Trapnell et al. 2009)), a spliced read mapper, by using default parameters and specifying the *Arabidopsis lyrata* reference genome and the corresponding gene model annotation file (both obtained from Joint Genome Institute (JGI (Grigoriev et al. 2012)). Afterwards, read counts were

extracted from the TopHat output using the python script HTSeq (http://www-huber.embl.de/users/anders/HTSeq/) and analysis of differential expression was performed using the Bioconductor package DESeq (Anders and Huber 2010) as outlined in the manual performing the following comparisons: Norway selfed vs. Norway outcrossed and Sweden selfed vs. Sweden outcrossed.

HTSeq counts the number of reads aligning to the mRNA of a gene by using the coordinates of the annotation file as guidance. In the approach used by DESeq to determine differential expression, the analysis is based on the assumption that the more reads are counted for one mRNA, the higher the corresponding gene is expressed. First, DESeq estimates the effective library size of each sample and divides the read counts for each gene of a sample by the value of this sample's effective library size to obtain a common scale and enable comparison. In the next step, the dispersion for replicates of one condition is estimated, which is used as an estimate of the square of the coefficient for biological variation. The final step is the call for differential expression between two conditions (here: selfed vs. outcrossed (Anders and Huber 2010)), which is followed by filtering for significant genes (here a false discovery rate (FDR) of 0.05 was chosen as threshold for significance).

A detailed description of all analysis steps and computer code can be found in the Supplementary Material 1: A-C.

Annotation and GO term enrichment

In our functional annotation analyses, we first considered genes that were significantly differentially expressed in both Norway and Sweden. Then, we examined the functional annotation of genes being differentially expressed in one population only.

The functional annotation analysis was performed using the standalone version of InterProScan (version 5RC1 (Zdobnov and Apweiler 2001)) as shown in Supplementary Material 1:D and by extracting information from TAIR (The Arabidopsis Information Resource, http://www.arabidopsis.org/). For gene ontology (GO) annotation of the selected genes, Blast2GO (Conesa et al. 2005) was used and GO term enrichment analysis was performed using the Cytoscape (version 2.8.3; (Shannon et al. 2003)) plugins BiNGO (Maere et al. 2005) and EnrichmentMap (Merico et al. 2010). The annotation file used for BiNGO was obtained from JGI (Araly1_goinfo_FilteredModels6.tab) (Grigoriev et al. 2012) and modified according to the plugin's requirements. Hypergeometric test and FDR of 0.05 were chosen for enrichment analysis. An enrichment map of the GO terms was generated using the plugin EnrichmentMap with the following statistical parameters: p-value cut-off of 0.001, Q-value cut-off of 0.05 and Jaccard coefficient with an overlap coefficient cut-off of 0.25.

Pathway visualisation

For pathway visualisation we used MapMan (version 3.5.1 (Thimm et al. 2004)), a program that assigns genes to functional categories (biological functions), so called BINs, which can then be displayed in pathway images. Since a mapping file needed for the functional categorization does not exist for *A. lyrata* but for the close relative *A. thaliana*, *A. lyrata* genes were assigned to their *A. thaliana* orthologs. A fasta file containing focal genes' protein sequences was uploaded to orthoMCL (web version; http://orthomcl.org/cgi-bin/OrthoMcl Web.cgi?rm=proteomeUploadForm), which uses a Markov clustering algorithm to find orthologs. Afterwards, ortholog identifiers were translated into Arabidopsis Genome Initiative (AGI) identifiers and together with their appertaining log2 fold changes for selfing orthologs loaded in MapMan for pathway mapping. Supplementary Material 1:E shows an example of the input mapping file.

Gene interaction networks

In order to identify genes that interact with the genes found to be differentially expressed in each of the populations, gene interaction networks were created. Differentially expressed *A. lyrata* genes were assigned to their orthologs in *A. thaliana*, since gene interactions are better documented for *A. thaliana* than for *A. lyrata*. The *A. thaliana* ortholog ID was used together with the expression fold changes (i.e. the expression ration between selfed and outcrossed progeny) from the present experiment as input for the PSICQUIC (Proteomics Standard Initiative Common Query InterfaCe (Aranda et al. 2011)) client inbuilt in Cytoscape. PSICQUIC allows the search of multiple resources for molecular interactions. For our data IntAct (Kerrien et al. 2012), BAR (The Bio-Analytic Resource for Plant Biology), BIND (Bader et al. 2003), MINT (Licata et al. 2012) and UniProt (UniProt 2014) were queried and from the database outputs a union network was constructed and interactions were restricted to first neighbour interactions. Nodes, representing the differentially expressed genes, were coloured by fold change.

Distribution of differentially expressed genes on the *A. lyrata* genome

In order to display the distribution of the differentially expressed genes of each population on the *A. lyrata* genome, their chromosomal locations were retrieved by BioMart (Smedley et al. 2009) and plotted in Circos (version 0.67-5 (Krzywinski et al. 2009)).

Statistical analysis

If not stated otherwise statistical analyses were performed in R (http://www.r-project.org/).

Results

Mapping statistics and clustering of samples

The number of reads and mapping statistics for the transcriptome assembly of each of the 24 samples are given in Supplementary Table 1.

Clustering by Euclidean distance among individuals based on their count data revealed that the transcriptomes of population Sweden and Norway differed from each other and largely clustered separately whereas selfed and outcrossed progeny within populations were more weakly separated, in particular in the Sweden population (Figure 1).

Differential gene expression between selfed and outcrossed individuals

Differential gene expression between selfed and outcrossed individuals was analysed for each population separately. In Norway, 507 genes were significantly differentially expressed between selfed and outcrossed progeny, whereas in Sweden only 195 genes were differentially expressed (Figure 2A, B). The Venn diagram in Figure 2C shows the overlap between these comparisons. Sixty-two differentially expressed genes were common to both comparisons. Among those genes, a majority showed up-regulation in selfed progeny, while quite a few show up-regulation in one population (Sweden) and down-regulation in the other (Norway; Figure 2D). A visual inspection showed that the significantly differentially expressed genes were distributed all over the genome in both populations (Figure 3).

The following analyses of functional annotation, GO term enrichment, and pathways affected were first conducted on the 62 genes that were differentially expressed in both populations (shown in Figure 2C), and second on the genes that were significantly differentially expressed in one population only (445 genes in the Norwegian population, and 133 genes in the Swedish population; Figure 2C).

Annotation and GO term enrichment of genes that were significantly differentially expressed in both populations

Functional annotation analyses allowed the assignment of domains and protein families to nearly all genes that were significantly affected by inbreeding in both the Norwegian and the Swedish population. Genes without an annotation included genes (or predicted genes) of unknown function, and some relatively unspecific genes, like transcription factors (Supplementary Table 2). It is worth noting that some protein families and predicted domains had more than one gene assigned; like cytochrome P450 (5 genes), glycosyl hydrolases (4 genes) and chlorophyll A-B binding proteins (6

genes) to mention a few (Supplementary Table 2). Furthermore the effect of selfing on the expression profiles of many of the genes with multiple entries differed between the Norway and Sweden populations. Inbreeding was associated with over-expression of photosynthesis-related genes in Sweden, whereas the opposite was true in the Norway population (Supplementary Table 2).

Annotation of GO terms and their enrichment showed that two categories of genes were particularly common among the differentially expressed genes (Figure 4). A large portion of the genes was associated with stress responses and there were also many photosynthesis-related genes. "Response to stimulus" was with 41.2% of assigned GO terms the largest group of biological processes, and "plastid" was with 24% of assigned GO terms the largest group of cell compartments within the category cellular component (Figure 4A). Enrichment of significant GO terms based on this annotation showed a similar outcome (Figure 4B).

Pathway visualisation of genes that were differentially expressed in both populations

The genes that were differentially expressed in both populations mapped to different pathways, but could be assigned to two major categories. These categories are "biotic stress" (Figure 5) and "chloroplast" (Figure 6). A closer examination of these categories showed that genes falling into the category of biotic stress were predominantly up-regulated in selfed compared to outcrossed progeny in both populations (Figure 5A,B). Photosynthesis-related genes on the other hand were down-regulated among selfed offspring in the Norwegian population (Figure 6A), but up-regulated in the Swedish population (Figure 6B).

Annotation, GO term enrichment and pathway visualization of genes that were differentially expressed in one population only

We list the annotation to domains and protein families of the genes that were significantly differentially expressed between selfed and outcrossed progeny only in Norway in Supplementary Table 3 and only in Sweden in Supplementary Table 4.

GO annotation, enrichment and pathway analyses showed that the genes that were differentially expressed in only one population to a large extent shared the patterns detected for genes being differentially expressed in both populations: a large proportion of genes was associated with biotic stress (up-regulation after selfing in both populations) and photosynthesis (down-regulation after selfing in Norway; up-regulation in Sweden) (Supplementary Figure 1-3). However, there were also some novel patterns detected, most notably that genes related to reproduction and pollination were affected by inbreeding, and only so in the Norwegian population (Supplementary Figure 1).

10

All significant GO terms, their significance levels and associated differentially expressed genes are listed in Supplementary Table 5 for the Norwegian population and in Supplementary Table 6 for the Swedish population.

Gene interaction networks

Mapping of the differentially expressed *A. lyrata* genes of both populations to their *A. thaliana* orthologs resulted in 476 matches for the Norwegian population and 177 matches for the Swedish population, of which 224 and 104, respectively, were involved in interactions with other genes (Supplementary Figures 4-5). Among the genes that were significantly differentiated in both populations, 33 were involved in interactions (results not shown).

Discussion

This study has demonstrated that inbreeding affects the expression of a large number of genes in two self-incompatible populations of *Arabidopsis lyrata*. Many of the identified genes are involved in gene interactions or are transcription factors and proteases likely to have very general effects, consistent with the earlier finding of strong reductions in several components of fitness following inbreeding (Sletvold et al. 2013). Moreover, the two populations differed substantially in expression patterns affected. The number of differentially expressed genes was approximately 2.5 times higher in the Norwegian compared to the Swedish population, and genes associated with photosynthetic activity were down-regulated in selfed progeny in the Norwegian population but up-regulated in the Swedish population. That the two populations differ at least partly in how inbreeding affects gene expression contrasts the situation on the level of reproduction: the two populations show almost identical inbreeding depression on several fitness components (Sletvold et al. 2013).

Population differences in expression patterns

The comparison of selfed and outcrossed offspring identified 507 differentially expressed genes in the Norwegian population, compared with only 195 differentially expressed genes in the Swedish population. Furthermore, clustering by Euclidean distance revealed a separate grouping of individuals from Norway and Sweden, respectively, regardless of selfing. These population differences corroborate the pronounced differentiation observed between Scandinavian populations in a previous marker-based study (Gaudeul et al. 2007), and in traits such as trichome production (Løe et al. 2007) and tolerance to drought (Sletvold and Ågren 2012). Thus, differentiation throughout the genome, which is likely due to different demographic histories and lack of gene flow between these populations as well as differences in landscape structure and selection regimes

(mountains in Norway vs. coastal habitat in Sweden), can potentially explain why these populations at least partly differ in the genetic basis of inbreeding depression.

Several factors may influence the number of inbreeding sensitive genes detected in a population. First, short-term inbreeding effects are expected to be stronger in older and more genetically variable populations because such populations should have had time to accumulate more deleterious recessives alleles. However, SNP calling detected more SNPs in the Swedish than in the Norwegian population (methods and results given in Supplementary Material 2, Supplementary Figure 6, Supplementary Table 7). Furthermore, a previous study observed a higher diversity at microsatellite loci in the Swedish population (Gaudeul et al. 2007). This suggests that the Swedish population should have had more time to accumulate recessive mutations and thus be more sensitive to inbreeding; a prediction opposite to what we observe. Second, the magnitude of within-population variance in gene expression should affect the number of significant differences detected. However, the variance in expression was higher in the Norwegian than in the Swedish population (Supplementary Table 8), and this could thus not explain why fewer significant differences were detected in the latter. Finally, it may be that the population difference does not directly reflect a difference in the number of deleterious alleles exposed, but rather a difference in the number of "repair-genes" induced as a secondary response (Kristensen et al. 2010). More stress-response genes may be active at the 12-week stage in the Norwegian population, because of differences in seasonal phenology of growth and in the proportion of plants flowering in the first year. All plants were vegetative at the time of sampling and gene expression was measured in leaves, but eight weeks later a higher proportion of plants had produced flowers in the Norwegian populations than in the Swedish population (Sletvold et al. 2013). At the time of sampling, gene expression (in particular for genes associated with amino acid and sugar synthesis) differed between outcrossed progeny of the two populations (data not shown), which is consistent with a difference in the physiological states of plants.

Differences between selfed and outcrossed offspring

In total, 62 of the significantly differentially expressed genes were common to both populations, whereas 445 genes were differentially expressed only in Norway and 133 only in Sweden. The functional annotation analyses identified two major underlying biological functions: (a) biotic stress response and photosynthesis. Moreover, among the genes that were significantly affected by inbreeding only in Norway, genes related to reproduction and pollination were represented.

The inbreeding effect on gene expression that we have detected in this study is likely to have been caused by a reduction in heterozygosity across the genome produced by the experimental

treatment (selfing; inbreeding coefficient, f = 0.5). Indeed, the selfed progeny had much lower level of heterozygosity than had the outcrossed plants both (i) at the seven loci we used for parentage analyses (H_{selfed} = 29%; $H_{outbred}$ = 52%; p < 0.001; data from 358 selfed and 490 outcrossed plants included in a larger experiment (Sletvold et al. 2013)), and (ii) at genome-wide SNPs estimated from the present RNA sequences (SNP calling, see Supplementary Material 2; Norway: H_{selfed} = 15%, $H_{outcrossed}$ = 28%, t-test: p < 0.0001; Sweden: H_{selfed} = 18%, $H_{outcrossed}$ = 26%, t-test: p < 0.003). Another type of variation in heterozygosity occurs due to random segregation of a finite number of chromosome parts. This will make also individuals with exactly the same inbreeding coefficients to differ somewhat in heterozygosity. However, such variation is much less pronounced than the variation in heterozygosity between outcrossed and selfed progeny (Chakraborty 1981; Hansson and Westerberg 2008; Hill and Weir 2011; Schraiber et al. 2012), and therefore we did not evaluate expression differences between individuals within crossing type in this study. Furthermore, random segregation induces variation in heterozygosity over the genome (even within an individual). However, also this variation is expected to be marginal in relation to the overall heterozygosity effect of the selfing treatment. In line with this reasoning, the significantly differentially expressed genes were distributed all over the genome (Figure 3).

In both populations, stress response-related genes were over-expressed in the selfed compared to the outcrossed progeny. That selfing causes a stress-related expression response and affects metabolic functions is in agreement with studies in *Drosophila melanogaster* (Kristensen et al. 2005; Ayroles et al. 2009). It is likely that many of these genes are up-regulated to alleviate the negative effects of exposure of other deleterious alleles, paralleling the role of heat shock proteins documented both in *Drosophila* (Kristensen et al. 2010) and *Arabidopsis* (Seki et al. 2002). The up-regulation of stress genes is thus not necessarily a direct cause of inbreeding depression, but could rather represent a secondary response to the expression of the genetic load, that helps restore protein function and stability. In *Drosophila*, there is considerable overlap in genes found to be up-regulated upon inbreeding versus genes up-regulated by other 'internal stresses', such as oxidative stress and aging (Landis et al. 2004; Kristensen et al. 2010), and also by external factors, such as heat and chemicals (Sorensen et al. 2003). Also in *Arabidopsis*, there is overlap among stress-inducible genes responding to oxidative stress and different environmental factors such as drought, salinity, cold and wounding (Seki et al. 2002; Swindell 2006). This indicates that a large part of the expression pattern induced by inbreeding represents a general stress response. This may be due to similar biological damage caused by different stresses, or a similar demand on metabolism and energetic resource allocation (Bubliy and Loeschcke 2005). In *A. lyrata*, there was a 50% growth reduction among inbred progeny still alive at the 12-week stage (Sletvold et al. 2013), indicating severe

deleterious effects that would induce the expression of many metabolic pathways that control protein functionality.

The expression of resistance (R) genes can be associated with a cost (Tian et al. 2003), and the down-regulation of some R genes in selfed progeny (see Figure 5; Supplementary Table 2, protein ID 475633) may reflect their lower resource status. However, we also found some R genes to be up-regulated among selfed progeny (e.g. cytochrome P450; Supplementary Table 2), which is consistent with a study performed in ragged robin (*Lychnis flos-cuculi*) showing that inbreeding increases the activity of pathogenesis-related proteins such as kitinases and ß-1,3-glucanases (Leimu et al. 2012). Furthermore, there are parallels between pathogen recognition and self-incompatibility in the recognition and rejection of self and non-self. Both mechanisms largely rely on receptor-like kinases, like S-receptor kinases (SRKs) in the self-incompatible pathway, and pattern recognition receptors (PRRs), which recognize pathogen-associated molecular patterns (PAMPs) in innate immunity of plants (Sanabria et al. 2008). Previous studies (Pastuglia et al. 1997, 2002) found SRKs in *Brassica* and *Arabidopsis* to be up-regulated upon response to biotic stress like wounding and pathogen infection, suggesting that they might play a role in processes other than self-incompatibility. In this study, such a SRK gene was up-regulated in the selfed progeny of both populations. Since our data originated from vegetative tissue, it is possible that SRK has other functions in *A. lyrata* than its known role in the self-incompatibility system.

Photosynthesis was another major biological function affected by inbreeding in both populations, whereas effects on genes related to reproduction and pollination were observed only in the Norwegian population. The effect of inbreeding on the expression of photosynthesis related genes differed between the two populations. While selfed offspring had lower expression of photosynthesis-related genes in the Norwegian population; the opposite was true in the Swedish population. Effects of inbreeding on the expression of photosynthesis-related genes may reflect a reduced ability to match gene expression to prevailing environmental conditions, whereas the difference in the direction of the effect may be related to inherent differences in the seasonal pattern of growth between the two populations. The Norwegian population experiences a shorter growing season at its site of origin, mainly due to its higher altitude, and may at the 12-week stage slow down energy accumulation in preparation for winter. In addition, in the present experiment, a higher proportion of the 12-week plants may have been in transition to initiate the formation of reproductive structures in the Norwegian population compared to the Swedish population (eight weeks later, 26% of the Norwegian and 13% of the Swedish plants had produced flowers in the outcross treatment (Sletvold et al. 2013)). This could potentially contribute to the difference observed in the expression of photosynthesis-related genes, but also explain why an effect of selfing

on the expression of genes related to reproduction and pollination was observed only in the Norwegian population.

Conclusion

Whether deleterious recessive alleles can be purged from a population, thereby reducing inbreeding depression and facilitating mating-system transitions from outcrossing to selfing, is a central question in conservation and evolutionary biology. Both purging and transition to selfing are less likely if inbreeding depression is caused by many genes with weakly deleterious effects (Charlesworth et al. 1990, 1993; Ayroles et al. 2009). We found that many genes were significantly differentially expressed after selfing in *A. lyrata* (507 genes in Norway and 195 in Sweden), and that at least half of them interacted closely with other genes. Moreover, some of the inbreeding-sensitive genes were identified as transcription factors and proteases and are therefore likely to be located upstream in functional pathways and have regulatory functions. This indicates that the observed differences in gene expression may reflect the accumulation of a large number of interacting and functionally important recessive mutations in these populations, which would in turn render purging and transition to selfing unlikely. Nevertheless, transition from self-incompatibility to self-fertility has occurred in several *Arabidopsis* lineages, including during the formation of *A. thaliana* as well as in some North American populations of *A. lyrata* (Clauss and Koch 2006; Mable and Adam 2007). This indicates that some of the genetic load has been purged in those lineages. Thus, it is possible that the wide inbreeding effects on gene expression that we have detected are at least partly driven by an extensive network of secondary responses, where pathways may be influenced by the expression of relatively few key genes with deleterious mutations. In *Drosophila* it was found that at least 38% of the differentially expressed genes during inbreeding were affected by such indirect network effects (Ayroles et al. 2009). This highlights the difficulties in inferring the actual number of genes that are directly affected by inbreeding from transcriptome-wide gene expression analyses.

The similar magnitude of inbreeding depression (Sletvold et al. 2013) suggests that the opportunity for deleterious mutations to accumulate has been similar in the two self-incompatible study populations. In line with this, we found considerable population overlap in the expression of genes associated with stress response and general metabolic activities. The genes that were significantly affected by inbreeding in the same direction in both populations are particularly strong candidates for being inbreeding-sensitive genes in *A. lyrata*. We also found striking differences between the populations in the patterns of gene expression affected by selfing, where more genes were significantly affected by inbreeding in Norway, and photosynthesis-related genes were up-regulated in Norway but down-regulated in Sweden. Thus, the present study indicates that the

genetic basis of inbreeding depression partly differs between the two populations, despite very similar effects of inbreeding on fitness (Sletvold et al. 2013).

Acknowledgement

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E, Dana JM, De Las Rivas J, Dumousseau M, Galeota E, et al. 2011. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat Methods* 8:528-529.

Ayroles JF, Hughes KA, Rowe KC, Reedy MM, Rodriguez-Zas SL, Drnevich JM, Caceres CE, Paige KN. 2009. A genomewide assessment of inbreeding depression: gene number, function, and mode of action. *Conserv Biol.* 23:920-930.

Bader GD, Betel D, Hogue CW. 2003. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* 31:248-250.

Barrett SCH, Charlesworth D. 1991. Effects of a change in the level of inbreeding on the genetic load. *Nature* 325:522-524.

Bijlsma R, Bundgaard J, Boerema AC. 2000. Does inbreeding affect the extinction risk of small populations?: predictions from *Drosophila*. *J Evol Biol.* 13:502-514.

Bubliy OA, Loeschcke V. 2005. Correlated responses to selection for stress resistance and longevity in a laboratory population of *Drosophila melanogaster*. *J Evol Biol.* 18:789-803.

Cabin RJ, Evans AS, Jennings DL, Marshall DL, Mitchell RJ, Sher AA. 1996. Using bud pollinations to avoid self-incompatibility: Implications from studies of three mustards. *Can J Bot.* 74:285-289.

Chakraborty R. 1981. The distribution of the number of heterozygous loci in an individual in natural populations. *Genetics* 98:461-466.

Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Syst.* 18:237-268.

Charlesworth D, Morgan MT, Charlesworth B. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* 44:1469-1489.

Charlesworth D, Morgan MT, Charlesworth B. 1993. Mutation accumulation in finite outbreeding and inbreeding populations. *Genetical Res.* 61:39-56.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nat Rev Genet.* 10:783-796.

Clauss MJ, Cobban H, Mitchell-Olds T. 2002. Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). *Mol Ecol.* 11:591-601.

Clauss MJ, Koch MA. 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci.* 11:449-459.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.

Darwin CR. 1876. The effects of cross and self fertilization in the vegetable kingdom. London: John Murray.

Fisher RA. 1941. Average excess and average effect of a gene substitution. *Ann Hum Genet.* 11:53-63.

Gaudeul M, Stenøien HK, Ågren J. 2007. Landscape structure, clonal propagation, and genetic diversity in Scandinavian populations of *Arabidopsis lyrata* (Brassicaceae). *Am J Bot.* 94:1146-1155.

Gething MJ, Sambrook J. 1992. Protein folding in the cell. *Nature* 355:33-45.

Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, et al. 2012. The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res.* 40:D26-D32.

Hansson B, Naurin S, Hasselquist D. 2014. Does inbreeding affect gene expression in birds? *Biol Lett.* 10:20140648.

Hansson B, Westerberg L. 2008. Heterozygosity-fitness correlations within inbreeding classes: local or genome-wide effects? *Conserv Genet.* 9:73-83.

Hill WG, Weir BS. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Res.* 93:47-64.

Husband BC, Schemske DW. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* 50:54-70.

Jalas JS, J. 1994. Atlas Florae Europaeae: Distribution of Vascular Plants in Europe, Vol. 1–10. Helsinki, Finland.: The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.

Jarne P, Charlesworth D. 1993. The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Annu Rev Ecol Syst.* 24:441-466.

Kachroo A, Nasrallah ME, Nasrallah JB. 2002. Self-incompatibility in the Brassicaceae: Receptor-ligand signaling and cell-to-cell communication. *Plant Cell* 14:S227-S238.

Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY, et al. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. *Trends Ecol Evol.* 17:230-241.

Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, et al. 2012. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 40:D841-846.

Kristensen TN, Pedersen KS, Vermeulen CJ, Loeschcke V. 2010. Research on inbreeding in the 'omic' era. *Trends Ecol Evol.* 25:44-52.

Kristensen TN, Sorensen P, Kruhoffer M, Pedersen KS, Loeschcke V. 2005. Genome-wide analysis on inbreeding effects on gene expression in *Drosophila melanogaster*. *Genetics* 171:157-167.

Kristensen TN, Sorensen P, Pedersen KS, Kruhoffer M, Loeschcke V. 2006. Inbreeding by environmental interactions affect gene expression in *Drosophila melanogaster*. *Genetics* 173:1329-1336.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19:1639-1645.

Kärkkäinen K, Kuittinen H, van Treuren R, Vogl C, Oikarinen S, Savolainen O. 1999. Genetic basis of inbreeding depression in *Arabis petraea*. *Evolution* 53:1354-1365.

Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, Tavare S, Tower J. 2004. Similar gene expression patterns characterize aging and oxidative stress in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 101:7663-7668.

Leimu R, Kloss L, Fischer M. 2012. Inbreeding alters activities of the stress-related enzymes chitinases and beta-1,3-glucanases. *PLoS ONE* 7:e42326.

Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E, et al. 2012. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40:D857-861.

Løe G, Toräng P, Gaudeul M, Ågren J. 2007. Trichome production and spatiotemporal variation in herbivory in the perennial herb *Arabidopsis lyrata*. *Oikos* 116:134-142.

Mable BK, Adam A. 2007. Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. *Mol Ecol.* 16:3565-3580.

Maere S, Heymans K, Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448-3449.

Merico D, Isserlin R, Stueker O, Emili A, Bader GD. 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* 5:e13984.

Pastuglia M, Roby D, Dumas C, Cock JM. 1997. Rapid induction by wounding and bacterial infection of an S gene family receptor-like kinase gene in *Brassica oleracea*. *Plant Cell* 9:49-60.

Pastuglia M, Swarup R, Rocher A, Saindrenan P, Roby D, Dumas C, Cock JM. 2002. Comparison of the expression patterns of two small gene families of S gene family receptor kinase genes during the defence response in *Brassica oleracea* and *Arabidopsis thaliana*. *Gene* 282:215-225.

Sanabria N, Goring D, Nurnberger T, Dubery I. 2008. Self/nonself perception and recognition mechanisms in plants: a comparison of self-incompatibility and innate immunity. *New Phytologist* 178:503-513.

Sandring S, Riihimäki MA, Savolainen O, Ågren J. 2007. Selection on flowering time and floral display in an alpine and a lowland population of *Arabidopsis lyrata*. *J Evol Biol.* 20:558-567.

Schraiber JG, Shih S, Slatkin M. 2012. Genomic tests of variation in inbreeding among individuals and among chromosomes. *Genetics* 192:1477-1482.

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, et al. 2002. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31:279-292.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13:2498-2504.

Sletvold N, Mousset M, Hagenblad J, Hansson B, Ågren J. 2013. Strong inbreeding depression in two Scandinavian populations of the self-incompatible perennial herb *Arabidopsis lyrata*. *Evolution* 67:2876-2888.

Sletvold N, Ågren J. 2012. Variation in tolerance to drought among Scandinavian populations of *Arabidopsis lyrata*. *Evol Ecol.* 26:559-577.

Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A. 2009. BioMart--biological queries made easy. *BMC Genomics* 10:22.

Sorensen JG, Kristensen TN, Loeschcke V. 2003. The evolutionary and ecological role of heat shock proteins. *Ecol Lett.* 6:1025-1037.

Swindell WR. 2006. The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics* 174:1811-1824.

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37:914-939.

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J. 2003. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423:74-77.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105-1111.

UniProt C. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42:D191-198.

Winn AA, Elle E, Kalisz S, Cheptou PO, Eckert CG, Goodwillie C, Johnston MO, Moeller DA, Ree RH, Sargent RD, et al. 2011. Analysis of inbreeding depression in mixed-mating plants provides evidence for selective interference and stable mixed mating. *Evolution* 65:3339-3359.

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Cardle L, Gibby M, Powell W. 2007. Development of EST-derived microsatellite markers for *Arabidopsis lyrata* subspecies *petraea* (L.). *Mol Ecol Notes* 7:631-634.

Zdobnov EM, Apweiler R. 2001. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847-848.

Figure Legends

Figure 1: Clustering of individuals based on count data. The two *A. lyrata* populations Sweden and Norway with the treatments selfed and outcrossed were clustered by Euclidean distance in DESeq. Individuals from Norway are shown in grey and individuals from Sweden in black, with plant ID indicating the identity of the maternal parent and whether the plant was a selfed (S) or outcrossed (W) progeny.

Figure 2: Differential expression of selfed and outcrossed preogeny and its overlap between the populations. Differential expression of genes was calculated using DESeq with a FDR and an adjusted p-value of 0.05. **(A)** Scatter plot of log2 fold change vs. mean of normalized counts for the comparison selfed vs. outcrossed in Norway showing in red 507 significantly differentially expressed genes between the conditions, **(B)** Scatter plot for the comparison selfed vs. outcrossed in Sweden, resulting in 195 significantly differentially expressed genes. The Venn diagrams represent the overlap of significantly differentially expressed genes between the two populations with focus on **(C)** the overlap of genes differentially expressed within-Norway and within-Sweden and **(D)** the overlap according to genes up- and down-regulated in the selfed progeny of a given population.

Figure 3: Distribution of the differentially expressed genes in each population on the *A. lyrata* genome. Genes that were significantly differentially expressed genes in Norway are shown in blue and those differentially expressed in Sweden in orange. Their chromosomal locations were retrieved by BioMart and plotted in Circos. Numbers refer to the *A. lyrata* chromosome number.

Figure 4: Gene ontology (GO) annotation and enrichment analysis of the genes that were significantly differentially expressed in both populations. (A) Distribution of GO annotation for the candidates performed using Blast2GO. Percentages indicate the percentage of candidate sequences assigned to each subcategory of either biological process or cellular component. **(B)** Enrichment map, built upon GO annotation (Cytoscape plugins BiNGO and EnrichmentMap), providing insights into GO terms that are highly significant and their interactions.

Figure 5: Results from biotic stress pathway analyses in the (A) Norwegian and (B) Swedish population, respectively. The analyses were based on genes that were significantly differentially expressed in both populations. Assignment of the genes into pathways was achieved by ortholog Markov clustering against *A. thaliana* and mapping of orthologs into pathways. Blue signifies a positive log2 fold change in the selfed compared to the outcrossed (up-regulation), whereas red signifies a negative log2 fold change in the selfed compared to the outcrossed (down-regulation).
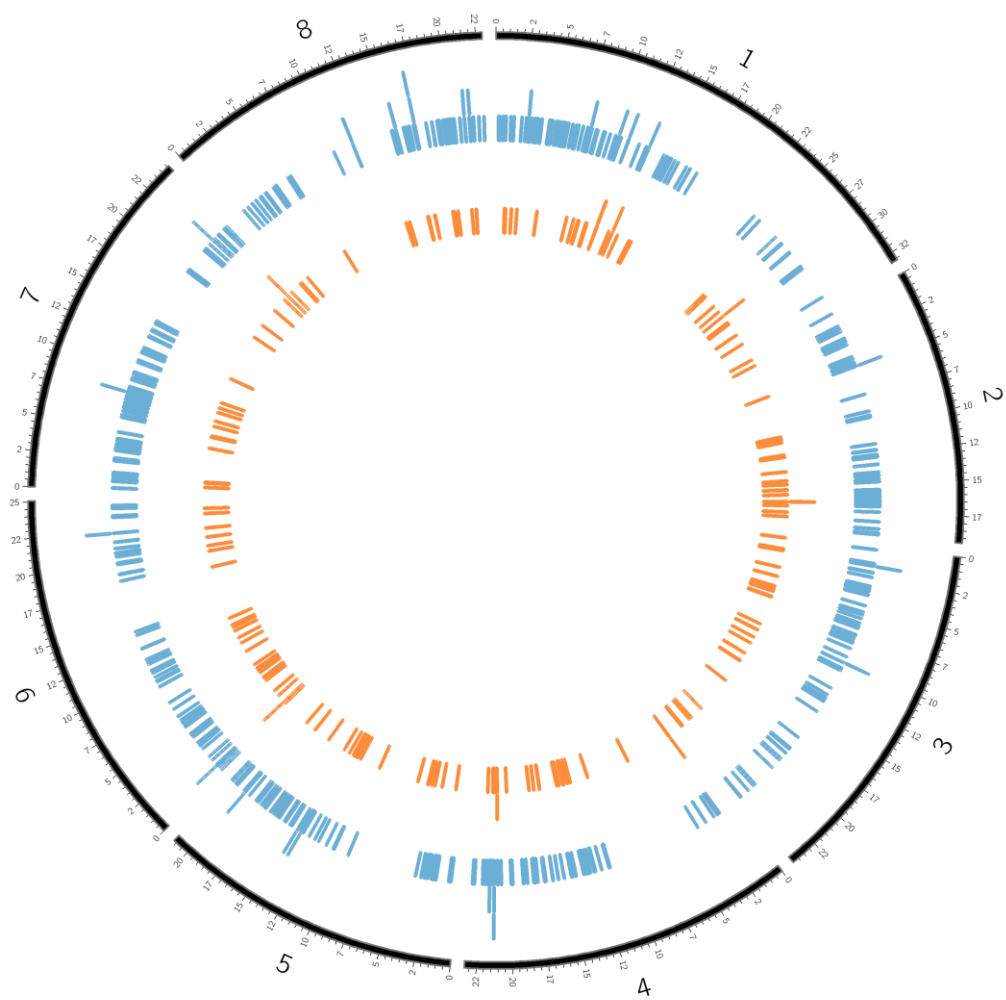
Figure 6: Results from chloroplast pathway analyses in the (A) Norwegian and (B) Swedish population, respectively. The analyses were based on genes that were significantly differentially expressed in both populations. Assignment of genes into pathways was achieved by ortholog Markov clustering against *A. thaliana* and mapping of orthologs into pathways. Blue signifies a positive log2 fold change in the selfed compared to the outcrossed (up-regulation), whereas red signifies a negative log2 fold change in the selfed compared to the outcrossed (down-regulation).
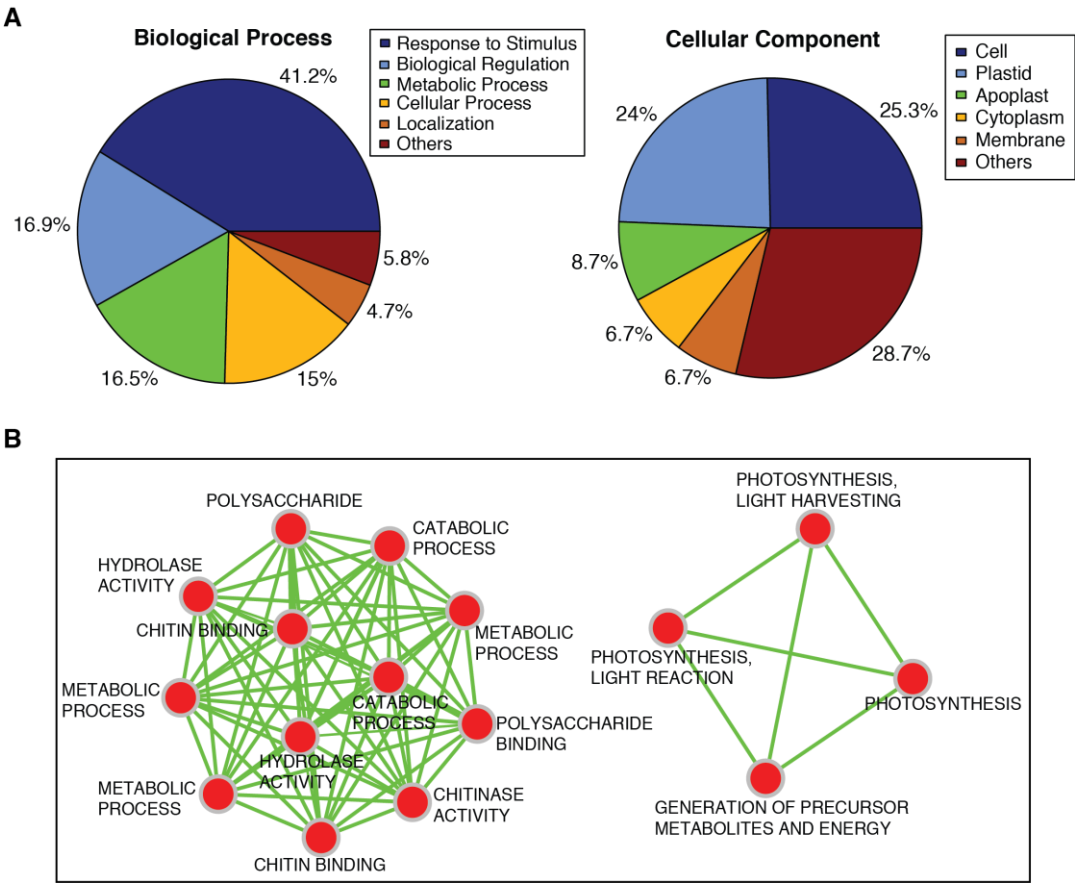
FIGURE 1
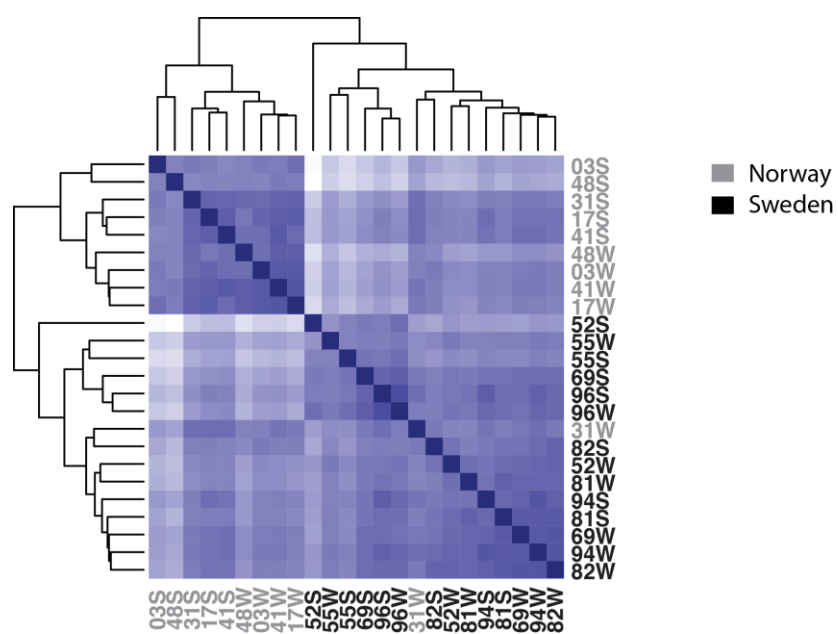
FIGURE 2
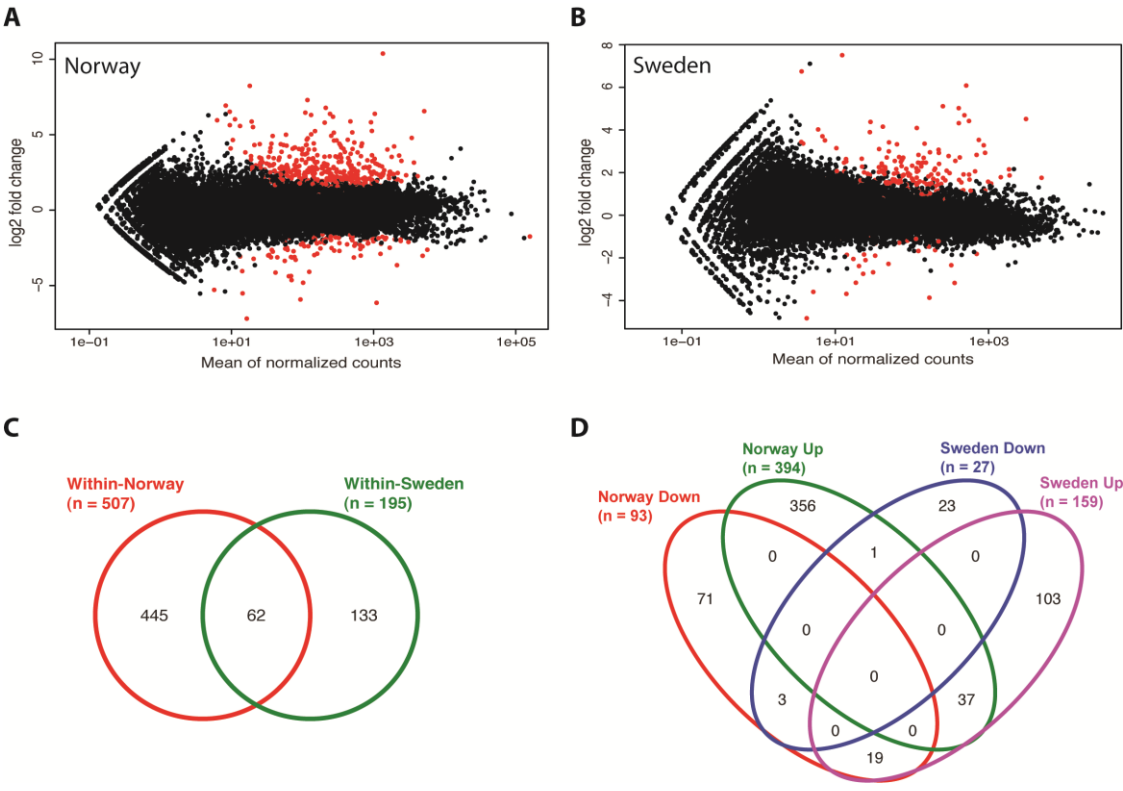
FIGURE 3

FIGURE 4

FIGURE 5

Norway

Sweden

FIGURE 6