Supplementary Material - Content **Page 1**

Supplementary Material 1

A) Transcriptome assembly

The A. lyrata genome was indexed using Bowtie2 (version 2.0.0) (Langmead and Salzberg 2012):

```
bowtie2-build -f Araly1_assembly_scaffolds.fasta Araly
```

Now TopHat was run with default parameters and specifying the index of the A. lyrata genome and a corresponding gene model annotation file (x signifies the sample name):

```
tophat -p 8 -G Araly1_GeneModels_FilteredModels6.gff -o TopHat2 Araly
x_1.fastq x_2.fastq
```

B) Obtaining read counts

Read counts for all samples were obtained from the BAM files in the TopHat output which contain the read alignments. Since HTSeq requires SAM files as input, files needed to be converted and sorted first using SAMtools (version 0.1.18) (Li, et al. 2009):

```
     samtools sort -n accepted_hits.bam hits.sorted
samtools view -h -o hits.sorted.sam hits.sorted.bam
```

Then HTSeq was run with the following command to match read counts to the Arabidopsis lyrata genes:

```
python -m HTSeq.scripts.count -i name hits.sorted.sam Araly1_GeneMod
els_FilteredModels6.gff > counts.txt -s no
```

C) Obtaining differentially expressed genes

Analysis of differential expression was performed using the Bioconductor package DESeq. The read counts per gene were arranged in a table to fit the DESeq script and analysis was carried out according to the vignette with a FDR of 0.05:

```
CountTable ← read.table(“ReadCounts.txt”, header=TRUE, sep=”\t”, row.names=1)
Design ← data.frame(row.names=colnames(CountTable), condition= c(“Within”,
“Selfed”, [...], “Within”, “Selfed”))
```

```
conds ← factor(c(Within”, “Selfed”, [...], “Within”, “Selfed”))

library(“DESeq”)

cds ← newCountDataSet(CountTable, conds)

cds ← estimateSizeFactors(cds)

sizeFactors(cds)

cds ← estimateVaraianceFunction(cds)

res ← nbinomTest(cds, “Within”, “Selfed”)

resSig ← res[res$padj<0.05, ]

write.table(resSig, file=”results.txt”)
```

D) Annotation of candidate genes

The candidate genes were annotated by running InterProScan, and thus, obtaining gene ontology terms, protein families and domains. Input was a protein fasta file of the candidates.

```
interproscan.sh -iprlookup -goterms -pa -f gff3,xml -t p -i
Candidates.fasta -T Temp -b Candidates_InterProScan
```

E) Example of input file for pathway visualisation with MapMan

homolog_id	**log2 fold Norway**	**flag N**	**log 2 fold Sweden**	**flag S**
AT1G11300	2.110	1	1.529	1
AT2G30750	2.780	1	2.709	1
AT4G19810	4.222	1	2.323	1

References:

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Supplementary Table 1: Total number of reads and mapping statistics using TopHat. The upper table summarizes the results for the population Norway, and the lower table the results for the population Sweden. The proportion of properly paired reads and singletons to the *Arabidopsis lyrata* reference genome are based on the amount of mapped reads. ”W” represents outcrossing and ”S” selfing with each number being assigned to a different individual. The TopHat output was parsed using the SAMtools flagstat command.

	03W	**03S**	**17W**	**17S**	**31W**	**31S**	**41W**	**41S**	**48W**	**48S**
Total Reads [M]	22.89	33.20	41.90	33.23	15.51	20.72	32.10	35.33	29.42	23.01
Reads mapped [%]	81.6	84.3	84.9	84.8	92.1	88.5	86.4	82.2	86.6	84.0
Properly paired reads [%]	83.5	87.6	84.8	81.7	82.5	83.3	85.0	81.3	86.2	86.1
Singletons [%]	11.1	9.8	10.1	11.2	10.9	10.5	9.5	12.1	9.0	9.9

	52W	**52S**	**55W**	**55S**	**69W**	**69S**	**81W**	**81S**	**82W**	**82S**	**94W**	**94S**	**96W**	**96S**
Total reads [M]	8.58	5.27	4.08	3.86	31.47	9.30	12.01	28.07	29.34	30.34	25.21	29.70	23.88	27.83
Reads mapped [%]	79.3	79.7	72.5	84.1	85.7	79.0	74.3	85.8	85.7	91.5	86.8	87.4	84.4	86.9
Properly paired reads [%]	80.5	64.1	69.6	57.7	82.4	70.4	79.8	83.1	81.4	81.3	80.9	77.5	68.2	71.2
Singletons [%]	12.2	17.9	15.8	19.7	11.0	14.5	12.6	10.7	11.3	10.6	11.5	12.5	15.3	15.0

Supplementary Table 2: Overview and annotation of genes that were significantly differentially expressed in both populations. The genes were obtained by differential expression analysis using DESeq with the comparisons: Norway selfed vs. Norway outcrossed (within-Norway comparison) and Sweden selfed vs. Sweden outcrossed (within-Sweden comparison). The upper half of the table displays the genes differentially expressed in the same direction in the two populations, whereas the second half adds genes that were differentially in both populations but in different directions. The table shows the annotation as retrieved by performing InterProScan on the samples and public annotation as performed the Arabidopsis Information Resource (TAIR). The last column gives the log2 fold changes in expression in the selfed compared to the non-selfed for both populations.

Protein ID	InterProScan	TAIR Annotation	Reference	Log2 Fold Change in Selfed
879372	PAN-like domain	Apple-like; Bulb-type lectin domain; S-receptor-like serine/ threonine-protein kinase; Serine/threonine-protein kinase, active site; PAN-2 domain; Protein kinase, catalytic domain; S-locus glycoprotein; Serine-threonine/tyrosine-protein kinase; Protein kinase-like domain	(Naithani, et al. 2007)	Norway 2.11 Sweden 1.53
320824	Cytochrome P450	Cytochrome P450, E-class, group I; Cytochrome P450, conserved site; Cytochrome P450	(Schuler 2006)	Norway 2.78 Sweden 2.71
329582	Glycosyl hydrolases family 18	Chitinase II; Glycoside hydrolase, superfamily; Glycoside hydrolase, subgroup, catalytic core; Glycoside hydrolase, family 18, catalytic domain	(Kasprzewska 2003)	Norway 4.22 Sweden 2.32
330253	Late embryogenesis abundant protein	Late embryogenesis abundant protein, LEA-14	(Hundertmark and Hincha 2008)	Norway 5.24 Sweden 3.33
338501				Norway 1.25 Sweden 7.51
340627	Barwin family	Glycoside hydrolase, family 19, catalytic; Barwin; Barwin-related endoglucanase; Barwin, conserved site; Barwin-like endoglucanase	(Svensson, et al. 1992)	Norway 3.10 Sweden 4.09

346169	Chitin recognition protein	Chitin-binding, type 1, conserved site; Glycoside hydrolase, family 19; Lysozyme-like domain; Chitin-binding, type 1; Glycoside hydrolase, family 19, catalytic	(Kasprzewska 2003)	Norway 1.52 Sweden 1.92
348799	Glycosyl hydrolase family 17	Glycoside hydrolase, family 17; Glycoside hydrolase, subgroup, catalytic core; Glycoside hydrolase, superfamily	(Minic 2008)	Norway 2.25 Sweden 2.71
354029	O-methyltransferase	O-methyltransferase, family 2; O-methyltransferase, COMT, imerization; Plant methyltransferase imerization; Winged helix-turn-helix transcription repressor DNA-binding	(Lam, et al. 2007)	Norway 3.04 Sweden 4.32
472778	Berberine and berberine like	FAD-binding, type 2, subdomain 1; Berberine/berberine-like; FAD linked oxidase, N-terminal; FAD-linked oxidase, FAD-binding, subdomain 2; FAD-binding, type 2	(Facchini, et al. 1996)	Norway 1.73 Sweden 2.47
472780	FAD binding domain	FAD-linked oxidase, FAD-binding, subdomain 2; FAD linked oxidase, N-terminal; Berberine/berberine-like; FAD-binding, type 2	(Facchini, et al. 1996)	Norway 5.88 Sweden 2.25
472781	FAD binding domain	FAD-linked oxidase, FAD-binding, subdomain 2; FAD-binding, type 2, subdomain 1; FAD linked oxidase, N-terminal; FAD-binding, type 2; Berberine/berberine-like	(Facchini, et al. 1996)	Norway 3.46 Sweden 1.84
475633	Phloem protein 2	Toll-Interleukin receptor	(Kehr 2006)	Norway -2.97 Sweden -3.87
475766	SAM dependent carboxyl methyltransferase	SAM dependent carboxyl methyltransferase	(Ross, et al. 1999)	Norway 3.09 Sweden 2.83
476458	Trypsin and protease inhibitor	Kunitz inhibitor ST1-like; Proteinase inhibitor I3, Kunitz legume	(Jofuku and Goldberg 1989)	Norway 1.64 Sweden 1.13
477006	VQ motif	VQ	(Cheng, et al. 2012)	Norway 2.36 Sweden 1.36

481863	Glutathione S-transferase, C-terminal domain	Glutathione S-transferase, N-terminal; Glutathione S-transferase, C-terminal-like; Glutathione S-transferase, C-terminal; Thioredo-xin-like fold; Glutathione S-transferase/chloride channel, C-terminal	(Marrs 1996)	Norway 2.89 Sweden 1.64
481994	Cytochrome P450	Cytochrome P450; Cytochrome P450, E-class, group I; Cyto-chrome P450, conserved site	(Schuler 2006)	Norway 6.56 Sweden 4.42
483400	No apical meristem (NAM) protein	No apical meristem (NAM) protein	(Duval, et al. 2002)	Norway 2.99 Sweden 2.02
483526	Initiation factor 2 subunit family	Initiation factor 2B-related	(Pavitt 2005)	Norway 2.80 Sweden 1.90
485691	UDP-glucoronosyl and UDP-glycosyl transferase	UDP-glucuronosyl/UDP-glucosyltransferase	(Ross, et al. 2001)	Norway 3.53 Sweden 1.53
487053	Glutathione S-transferase, C-terminal domain	Glutathione S-transferase, N-terminal; Glutathione S-transferase/ chloride channel, C-terminal; Thioredoxin-like fold; Glutathione S-transferase, C-terminal; Glutathione S-transferase, C-terminal-like	(Marrs 1996)	Norway 4.31 Sweden 1.49
488124	WRKY DNA-binding domain	DNA-binding WRKY	(Cheng, et al. 2012)	Norway 2.70 Sweden 1.67
492486	Glycosyl hydrolases family 28	Pectin lyase fold; Glycoside hydrolase, family 28; Parallel beta-helix repeat; Pectin lyase fold/virulence factor	(Minic 2008)	Norway 2.12 Sweden 1.97
493878	GDSL-like lipase/acylhydrolase	Lipase, GDSL, active site; Lipase, GDSL	(Ling 2008)	Norway 3.25 Sweden 2.09
494390	Papain family cysteine protease	Peptidase C1A, papain C-terminal; Peptidase, cysteine peptidase active site; Proteinase inhibitor I29, cathepsin propeptide; Peptidase C1A, papain	(Shindo and Van Der Hoorn 2008)	Norway 10.38 Sweden 2.60

497196	Short chain dehydrogenase	Short-chain dehydrogenase/reductase, conserved site; Short-chain dehydrogenase/reductase SDR; Glucose/ribitol dehydrogenase; NAD(P)-binding domain	(Kavanagh, et al. 2008)	Norway 2.97 Sweden 3.07
862882	Jacalin-like lectin domain	Mannose-binding lectin	(Barre, et al. 2001)	Norway 3.21 Sweden 3.42
865114	NmrA-like family	NmrA-like; NAD(P)-binding domain	(Stammers, et al. 2001)	Norway 2.80 Sweden 2.55
870854	Chitinase class I	Chitin-binding, type 1; Glycoside hydrolase, family 19, catalytic; Glycoside hydrolase, family 19; Chitin-binding, type 1, conserved site; Lysozyme-like domain	(Kasprzewska 2003)	Norway 2.03 Sweden 2.55
871042	Pectinesterase	Pectin lyase fold/virulence factor; Carbohydrate-binding/sugar hydrolysis domain; Pectin lyase fold; Pectinesterase, active site; Pectinesterase inhibitor; Pectinesterase, catalytic	(Micheli 2001)	Norway 4.02 Sweden 0.72
873781	Probable lipid transfer	Plant lipid transfer protein/Par allergen; Bifunctional inhibitor/plant lipid transfer protein/seed storage; Plant lipid transfer protein/ hydrophobic protein, helical domain; Plant lipid transfer protein/ seed storage/trypsin-alpha amylase inhibitor	(Kader 1996)	Norway 5.15 Sweden 2.47
880659	Brix domain	Brix domain; Anticodon-binding	(Eisenhaber, et al. 2001)	Norway 1.95 Sweden -
886183	Non-haem dioxygenase in morphine synthesis, N-terminal	Oxoglutarate/iron-dependent oxygenase	(Dunwell, et al. 2004)	Norway 2.51 Sweden 1.98

889581	Wall-associated receptor kinase galacturonan-binding	Epidermal growth factor-like, type 3; Serine-threonine/tyrosine-protein kinase; Protein kinase, catalytic domain; EGF calcium-binding; EGF-type aspartate/asparagine hydroxylation site; Concanavalin A-like lectin/glucanase, subgroup; EGF-like calcium-binding; EGF-like calcium-binding, conserved site; Epidermal growth factor-like; Protein kinase-like domain	(He, et al. 1999)	Norway 3.31 Sweden 2.51
890879	Subtilase family	Peptidase S8/S53, subtilisin/kexin/sedolisin; Peptidase S8, subtilisin, His-active site; Protease-associated domain, PA; Peptidase S8, subtilisin, Ser-active site; Peptidase S8, subtilisin-related; Proteinase inhibitor I9, subtilisin propeptide	(Rautengarten, et al. 2005)	Norway 4.02 Sweden 2.32
895371	Thaumatin family	Thaumatin, conserved site; Thaumatin, pathogenesis-related	(Liu, et al. 2010)	Norway 2.37 Sweden 2.83
900481				Norway -4.81 Sweden -3.69
902535	Cytochrome P450	Cytochrome P450, E-class, group I; Cytochrome P450; Cyto-chrome P450, conserved site	(Schuler 2006)	Norway 5.13 Sweden 2.01
906073	Cytochrome P450	Cytochrome P450, E-class, group I; Cytochrome P450; Cyto-chrome P450, conserved site	(Schuler 2006)	Norway 3.57 Sweden 1.64
908138	Protein phosphatase 2C	Protein phosphatase 2C-like; Protein phosphatase 2C; Protein phosphatase 2C, manganese/magnesium aspartate binding site	(Rodriguez 1998)	Norway -2.21 Sweden -1.29
911976	Legume lectin domain	Protein kinase-like domain; Concanavalin A-like lectin/glucanase, subgroup; Serine/threonine-protein kinase, active site; Protein kinase, ATP binding site; Concanavalin A-like lectin/glucanase; Serine/threonine-protein kinase-like domain; Protein kinase, catalytic domain; Legume lectin, beta chain	(Bouwmeester and Govers 2009)	Norway 3.54 Sweden 2.21

472262	Copper binding proteins, plastocyanin/azurin family	Blue (type 1) copper domain; Plastocyanin; Blue (type 1) copper protein; Cupredoxin	(De Rienzo, et al. 2000)	Norway -1.77 Sweden 2.06
892250	Glycosyl hydrolase family 1	Glycoside hydrolase, family 1; Glycoside hydrolase, superfamily; Glycoside hydrolase, subgroup, catalytic core	(Minic 2008)	Norway -3.14 Sweden 2.28
489965	Cathepsin propeptide inhibitor domain (I29)	Peptidase C1A, papain C-terminal; Peptidase, cysteine peptidase active site; Peptidase C1A, papain; Proteinase inhibitor I29, cathepsin propeptide	(Guay, et al. 2000)	Norway -2.27 Sweden 3.36
899321	Chlorophyll A-B binding protein	Chlorophyll a/b binding protein domain; Chlorophyll A-B binding protein; Chlorophyll A-B binding protein, plant	(Leutwiler, et al. 1986)	Norway -2.90 Sweden 1.91
474626	Photosystem I psaG / psaK	Photosystem I PsaG/PsaK domain; Photosystem I PsaG/PsaK protein; Photosystem I PsaG, plant; Photosystem I PsaG/PsaK, plant	(Jensen, et al. 2007)	Norway -2.07 Sweden 1.66
482403	Chlorophyll A-B binding protein	Chlorophyll A-B binding protein; Chlorophyll A-B binding protein, plant; Chlorophyll a/b binding protein domain	(Leutwiler, et al. 1986)	Norway -6.14 Sweden 6.09
491721	Xyloglucan endo-transglycosylase (XET) C-terminus	Glycoside hydrolase, family 16, active site; Xyloglucan endotransglucosylase/hydrolase; Beta-glucanase; Concanavalin A-like lectin/glucanase; Concanavalin A-like lectin/glucanase, subgroup; Xyloglucan endo-transglycosylase, C-terminal; Glycoside hydro-lase, family 16	(Campbell and Braam 1999)	Norway -4.37 Sweden 1.49
899117	Cytochrome P450	Cytochrome P450, conserved site; Cytochrome P450, E-class, group I; Cytochrome P450	(Schuler 2006)	Norway 2.05 Sweden -1.84
330654	SAM dependent carboxyl methyltransferase	SAM dependent carboxyl methyltransferase	(Ross, et al. 1999)	Norway -2.88 Sweden 1.44

480225	Xyloglucan endo-transglycosylase (XET) C-terminus	Glycoside hydrolase, family 16; Glycoside hydrolase, family 16, active site; Concanavalin A-like lectin/glucanase, subgroup; Concanavalin A-like lectin/glucanase; Beta-glucanase; Xyloglucan endo-transglycosylase, C-terminal; Xyloglucan endotransglucosylase/hydrolase	(Campbell and Braam 1999)	Norway -2.70 Sweden 1.86
473203	Chlorophyll A-B binding protein	Chlorophyll A-B binding protein; Chlorophyll A-B binding protein, plant; Chlorophyll a/b binding protein domain	(Leutwiler, et al. 1986)	Norway -2.62 Sweden 3.29
478197	Chlorophyll A-B binding protein	Chlorophyll a/b binding protein domain; Chlorophyll A-B binding protein, plant; Chlorophyll A-B binding protein	(Leutwiler, et al. 1986)	Norway -2.36 Sweden 1.94
472386				Norway -2.02 Sweden 2.43
488319	Carbonic anhydrase	Carbonic anhydrase, prokaryotic-like, conserved site; Carbonic anhydrase	(Moroney, et al. 2001)	Norway -1.95 Sweden 1.75
495512	Chlorophyll A-B binding protein	Chlorophyll A-B binding protein, plant; Chlorophyll A-B binding protein; Chlorophyll a/b binding protein domain	(Leutwiler, et al. 1986)	Norway -2.11 Sweden 1.79
480276		Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor; Bifunctional inhibitor/plant lipid transfer protein/seed Plant lipid transfer protein/hydrophobic protein, helical domain; storage	(Kader 1996)	Norway -3.36 Sweden 0.99
908079		Heat shock protein DnaJ, cysteine-rich domain	(Miernyk 1999)	Norway -2.20 Sweden 2.70
473204	Chlorophyll A-B binding protein	Chlorophyll A-B binding protein, plant; Chlorophyll a/b binding protein domain; Chlorophyll A-B binding protein	(Leutwiler, et al. 1986)	Norway -3.03 Sweden 4.52
496598	Photosystem I reaction centre subunit N (PSAN or PSI-N)	Photosystem I PsaN, reaction centre subunit N	(Jensen, et al. 2007)	Norway -1.94 Sweden 1.16

920397	Norway -1.65 Sweden 1.43

References:

Barre A, Bourne Y, Van Damme EJM, Peumans WJ, Rouge P. 2001. Mannose-binding plant lectins: Different structural scaffolds for a common sugar-recognition process. Biochimie 83:645-651.

Bouwmeester K, Govers F. 2009. Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles. Journal of Experimental Botany 60:4383-4396.

Campbell P, Braam J. 1999. Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. Trends in Plant Science 4:361-366.

Cheng Y, Zhou Y, Yang Y, Chi YJ, Zhou J, Chen JY, Wang F, Fan BF, Shi K, Zhou YH, et al. 2012. Structural and Functional Analysis of VQ Motif-Containing Proteins in Arabidopsis as Interacting Proteins of WRKY Transcription Factors. Plant Physiology 159:810-+.

De Rienzo F, Gabdoulline RR, Menziani MC, Wade RC. 2000. Blue copper proteins: A comparative analysis of their molecular interaction properties. Protein Science 9:1439-1454.

Dunwell JM, Purvis A, Khuri S. 2004. Cupins: the most functionally diverse protein superfamily? Phytochemistry 65:7-17.

Duval M, Hsieh TF, Kim SY, Thomas TL. 2002. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. Plant Mol Biol 50:237-248.

Eisenhaber F, Wechselberger C, Kreil G. 2001. The Brix domain protein family - a key to the ribosomal biogenesis pathway? Trends in Biochemical Sciences 26:345-347.

Facchini PJ, Penzes C, Johnson AG, Bull D. 1996. Molecular characterization of berberine bridge enzyme genes from opium poppy. Plant Physiology 112:1669-1677.

Guay J, Falgueyret JP, Ducret A, Percival MD, Mancini JA. 2000. Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. European Journal of Biochemistry 267:6311-6318.

He ZH, Cheeseman I, He DZ, Kohorn BD. 1999. A cluster of five cell wall-associated receptor kinase genes, Wak1-5, are expressed in specific organs of Arabidopsis. Plant Mol Biol 39:1189-1196.

Hundertmark M, Hincha DK. 2008. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics 9.

Jensen PE, Bassi R, Boekema EJ, Dekker JP, Jansson S, Leister D, Robinson C, Scheller HV. 2007. Structure, function and regulation of plant photosystem I. Biochimica Et Biophysica Acta-Bioenergetics 1767:335-352.

Jofuku KD, Goldberg RB. 1989. Kunitz Trypsin-Inhibitor Genes Are Differentially Expressed during the Soybean Life-Cycle and in Transformed Tobacco Plants. Plant Cell 1:1079-1093.

Kader JC. 1996. Lipid-transfer proteins in plants. Annual Review of Plant Physiology and Plant Molecular Biology 47:627-654.

Kasprzewska A. 2003. Plant chitinases - Regulation and function. Cellular & Molecular Biology Letters 8:809-824.

Kavanagh K, Jornvall H, Persson B, Oppermann U. 2008. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cellular and Molecular Life Sciences 65:3895-3906.

Kehr J. 2006. Phloem sap proteins: their identities and potential roles in the interaction between plants and phloem-feeding insects. Journal of Experimental Botany 57:767-774.

Lam KC, Ibrahim RK, Behdad B, Dayanandan S. 2007. Structure, function, and evolution of plant O-methyltransferases. Genome 50:1001-1013.

Leutwiler LS, Meyerowitz EM, Tobin EM. 1986. Structure and expression of three light-harvesting chlorophyll a/b-binding protein genes in Arabidopsis thaliana. Nucleic Acids Res 14:4051-4064.

Ling H. 2008. Sequence analysis of GDSL lipase gene family in Arabidopsis thaliana. Pak J Biol Sci 11:763-767.

Liu JJ, Sturrock R, Ekramoddoullah AKM. 2010. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep 29:419-436.

Marrs KA. 1996. The functions and regulation of glutathione S-transferases in plants. Annual Review of Plant Physiology and Plant Molecular Biology 47:127-158.

Micheli F. 2001. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. Trends in Plant Science 6:414-419.

Miernyk JA. 1999. Protein folding in the plant cell. Plant Physiology 121:695-703.

Minic Z. 2008. Physiological roles of plant glycoside hydrolases. Planta 227:723-740.

Moroney JV, Bartlett SG, Samuelsson G. 2001. Carbonic anhydrases in plants and algae. Plant Cell and Environment 24:141-153.

Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB. 2007. Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain. Proceedings of the National Academy of Sciences of the United States of America 104:12211-12216.

Pavitt GD. 2005. eIF2B, a mediator of general and gene-specific translational control. Biochemical Society Transactions 33:1487-1492.

Rautengarten C, Steinhauser D, Bussis D, Stintzi A, Schaller A, Kopka J, Altmann T. 2005. Inferring hypotheses on functional relationships of genes: Analysis of the Arabidopsis thaliana subtilase gene family. Plos Computational Biology 1:297-312.

Rodriguez PL. 1998. Protein phosphatase 2C (PP2C) function in higher plants. Plant Mol Biol 38:919-927.

Ross J, Li Y, Lim EK, Bowles DJ. 2001. Higher plant glycosyltransferases. Genome Biology 2.

Ross JR, Nam KH, D'Auria JC, Pichersky E. 1999. S-adenosyl-L-methionine : salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. Archives of Biochemistry and Biophysics 367:9-16.

Schuler MAD, H.; Bilgin, M.; Ali, S. . 2006. Arabidopsis cytochrome P450s through the looking glass: a window on plant biochemistry. Phytochem Rev 5:205-237.

Shindo T, Van Der Hoorn RAL. 2008. Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. Molecular Plant Pathology 9:119-125.

Stammers DK, Ren J, Leslie K, Nichols CE, Lamb HK, Cocklin S, Dodds A, Hawkins AR. 2001. The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. Embo Journal 20:6619-6626.

Svensson B, Svendsen I, Hojrup P, Roepstorff P, Ludvigsen S, Poulsen FM. 1992. Primary Structure of Barwin - a Barley Seed Protein Closely Related to the C-Terminal Domain of Proteins Encoded by Wound-Induced Plant Genes. Biochemistry 31:8767-8770.

Supplementary Table 3: Overview and annotation of genes differential expressed only in population. Genes were obtained by differential expression analysis using DESeq by comparin vs. Norway outcrossed and by excluding the genes that were differentially expressed in both Supplementary Table 2). The table shows the annotation as retrieved by performing InterP samples. The last column gives the log2 fold changes in expression in the selfed compared t population.

Protein ID	InterProScan	Log2 Fold Chang
311814	50S ribosome-binding GTPase	1.88
312360	Protein of unknown function (DUF1262)	2.61
312409	Xyloglucan fucosyltransferase	1.61
312603	Calcium-binding EGF domain	1.76
312863	Flavin-binding monooxygenase-like	3.74
313816	POT family	1.70
313855	Dimerisation domain	3.48
315131	Peptidase C13 family	1.91
315221	Mlo family	1.62
317500	Glycosyl hydrolase family 14	4.16
318019	Chitin recognition protein	3.83
318682	TLC domain	2.10
319239	MatE	2.44
319280	Calponin homology (CH) domain	2.16
319814	Rare lipoprotein A (RlpA)-like double-psi beta-barrel	2.43
321443		2.70
321760	WRKY DNA -binding domain	2.01
322090	Domain of unknown function (DUF4409)	2.36
322563	F-box domain	2.40
322779	Protein of unknown function (DUF707)	2.23
323821	Calcineurin-like phosphoesterase	2.88
324269	Peptidase family S41	2.09
326599	HMGL-like	-3.65
327109	Peroxidase	3.15
327678	Ion transport protein	1.60
327935	Multicopper oxidase	2.13
328029	Cytochrome P450	2.53
329503	Helix-loop-helix DNA-binding domain	-2.18
329563	VQ motif	2.19
329840	Terpene synthase family, metal binding domain	3.03
329886	Glycosyl hydrolases family 17	3.52
330043	Chlorophyll A-B binding protein	3.11
330113	56kDa selenium binding protein (SBP56)	-1.83
331046	Potato inhibitor I family	6.02
331197	Protein kinase domain	1.52
331608	Cytochrome P450	4.66
332024	Cytochrome P450	3.22
332328	Major Facilitator Superfamily	2.41
333079		2.51
334097	Eukaryotic-type carbonic anhydrase	4.32

335144	WRKY DNA -binding domain	2.51
335552	Pathogenesis-related protein Bet v I family	2.88
336359	Zinc finger, C3HC4 type (RING finger)	2.53
337274	Eukaryotic translation initiation factor eIF2A	
337410		2.39
338258	Terpene synthase, N-terminal domain	3.15
339216	WRKY DNA -binding domain	1.76
339498	CCAAT-binding transcription factor (CBF-B/NF-YA) subunit B	2.78
339683	GNS1/SUR4 family	2.66
342045		3.09
343972	Cytochrome P450	1.82
344701	Cation efflux family	2.70
346891	GDSL-like Lipase/Acylhydrolase	2.52
347509	Leucine rich repeat N-terminal domain	-3.11
350231	Annexin	-3.12
352125	2OG-Fe(II) oxygenase superfamily	1.93
352273	Staygreen protein	2.58
353011	Protein kinase domain	2.03
353102	Cytochrome P450	1.89
353319	Arabidopsis phospholipase-like protein (PEARLI 4)	2.73
353505	Alanine dehydrogenase/PNT, N-terminal domain	2.83
353547	Retinal pigment epithelial membrane protein	2.93
354767	ABC-2 type transporter	-2.63
355259		1.68
359589	Domain of unknown function (DUF239)	2.14
470192	Cellulase (glycosyl hydrolase family 5)	3.14
470207	Polyketide cyclase / dehydrase and lipid transport	4.29
470282	ANTH domain	2.28
470382	HAD superfamily, subfamily IIIB (Acid phosphatase)	-3.23
470440	Aldehyde oxidase and xanthine dehydrogenase, a/b hammerhead domain	3.69
470785	Pentatricopeptide repeat domain	2.32
470930	Beta-eliminating lyase	2.58
470959	Copine	3.64
471013	Nicotianamine synthase protein	4.42
471117	Putative serine esterase (DUF676)	3.50
471294	Pyridoxal-phosphate dependent enzyme	2.30
471343	Stigma-specific protein, Stig1	2.57
471443	Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain	-1.83
471490	Regulator of Vps4 activity in the MVB pathway	1.64
471526	Methyladenine glycosylase	2.03
471558	Protein of unknown function (DUF1336)	2.69
471868	NADH(P)-binding	-1.93
471969	TIR domain	1.07
472077	Myb-like DNA-binding domain	1.98
472226	Tubulin C-terminal domain-2012	-1.86

472361	Protein tyrosine kinase	2.55
472375		1.96
472635	Alcohol dehydrogenase GroES-like domain	-1.91
473139		1.94
473529	beta-Glucocerebrosidase 2 N terminal	1.94
473561	Allinase	2.23
473665		2.49
474819	Glycosyl transferase family 8	2.64
474884	WD domain, G-beta repeat	1.75
475045	AP2 domain	-1.45
475178	S-locus glycoprotein family	2.18
475262	Aldo/keto reductase family	2.77
475434		2.42
475462	Bromodomain	8.24
475479		-2.11
475808	Cytochrome P450	2.43
475878	Photosystem II protein Y (PsbY)	-1.70
475879	Pectate lyase	-2.49
475889	Fe-S metabolism associated domain	2.56
476004		4.91
476141	Matrixin	2.64
476150		-2.75
476393	Cupin	-3.01
476550	Trypsin and protease inhibitor	3.34
476582	Pyridine nucleotide-disulphide oxidoreductase	-2.49
476608	chorismate binding enzyme	2.19
476695	FAD binding domain	2.16
476696	ATP-dependent protease La (LON) domain	-1.98
476725	Tubulin C-terminal domain	-1.87
477377	Homeobox associated leucine zipper	2.55
477682	TIR domain	1.71
477692	Trypsin and protease inhibitor	6.31
477741	Chitin recognition protein	2.85
477857	Leucine rich repeat N-terminal domain	1.53
477862	CCAAT-binding transcription factor (CBF-B/NF-YA) subunit B	1.79
478210	Glycosyltransferase sugar-binding region containing DXD motif	2.86
478260	Pectinacetylesterase	4.76
478341	Calcineurin-like phosphoesterase	2.15
478362	Protein of unknown function (DUF563)	4.61
478998		1.57
479069	Asparaginase	3.19
479116	Legume lectin domain	4.48
479421		-2.54
479778	Probable lipid transfer	4.14
479826	Domain of unknown function (DUF3475)	2.00

479837	AP2 domain	6.52
479864	MatE	4.06
480090		1.55
480324	Aminotransferase class I and II	2.55
480389	FAE1/Type III polyketide synthase-like protein	-1.79
480406	Xyloglucan fucosyltransferase	2.29
481137	Helix-loop-helix DNA-binding domain	2.58
481138	Helix-loop-helix DNA-binding domain	1.95
481168	Serine carboxypeptidase	-1.89
481829	Ligand-gated ion channel	0.69
481887	UDP-glucoronosyl and UDP-glucosyl transferase	1.69
482040	PMR5 N terminal Domain	2.81
482050	AP2 domain	1.90
482294	No apical meristem (NAM) protein	2.07
482542	Heavy-metal-associated domain	3.88
482568	Late embryogenesis abundant protein	2.42
482676	Heavy-metal-associated domain	2.08
482736	C2H2-type zinc finger	2.85
482737	Endonuclease/Exonuclease/phosphatase family	3.12
482891		1.64
483093		2.17
483126	Response regulator receiver domain	3.69
483330	Basic region leucine zipper	-2.55
483469	Chitin recognition protein	3.76
483470	Chitin recognition protein	4.79
483607		4.93
483654	Protease inhibitor/seed storage/LTP family	-4.64
483714	C2 domain	2.78
483895	Myb-like DNA-binding domain	2.43
483936	Pectinesterase	3.38
484123	Pyridoxal-dependent decarboxylase conserved domain	3.27
484262	Ureide permease	1.78
484297		5.87
484318	CDP-alcohol phosphatidyltransferase	2.76
484387	Cytochrome P450	6.19
484388	Cytochrome P450	2.00
484460	Alpha/beta hydrolase family	2.10
484885	SAM dependent carboxyl methyltransferase	2.40
484965	Protein tyrosine kinase	2.34
485103	Plant invertase/pectin methylesterase inhibitor	-2.72
485205	Cleavage site for pathogenic type III effector avirulence factor Avr	2.38
485216	Glycosyl hydrolases family 16	3.19
485452	Alpha/beta hydrolase family	2.42
485496	Strictosidine synthase	1.72
486390	Pirin C-terminal cupin domain	-1.75

486477	Glycosyl hydrolase family 1	3.64
486479	Glycosyl hydrolase family 1	1.99
486481	Glycosyl hydrolase family 1	3.07
486484	Ring finger domain	1.78
486698		5.51
486718	ABC transporter	2.64
486733	HCO3- transporter family	1.57
486816	NADH(P)-binding	-2.21
486913	Protein of unknown function, DUF538	3.14
487073	Alpha/beta hydrolase family	1.73
487282	Nicotianamine synthase protein	-4.04
487575	Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	2.46
487602	No apical meristem (NAM) protein	2.89
487785	lactate/malate dehydrogenase, alpha/beta C-terminal domain	-1.86
487844	Myo-inositol-1-phosphate synthase	-2.39
487885	Protein kinase domain	2.79
488048		-2.18
488135	Domain of unknown function (DUF2828)	1.74
488150	GH3 auxin-responsive promoter	2.16
488205		2.03
488488	Lecithin retinol acyltransferase	1.68
489007	FAD binding domain in molybdopterin dehydrogenase	2.42
489094	No apical meristem (NAM) protein	1.54
489237	Remorin, C-terminal region	2.10
489370	Cytochrome P450	1.83
489522	Calmodulin binding protein-like	1.86
489642	Protein kinase domain	2.02
489742	EamA-like transporter family	2.81
489772	Peroxidase	5.83
489775	Peroxidase	3.62
489834	Helix-loop-helix DNA-binding domain	2.01
489946	WRKY DNA -binding domain	1.64
490098	Salt stress response/antifungal	1.37
490355	Cotton fibre expressed protein	-1.83
490500	C2 domain	2.47
490604	Multicopper oxidase	1.75
490753	FKBP-type peptidyl-prolyl cis-trans isomerase	-2.68
490844	VQ motif	4.07
490882	GH3 auxin-responsive promoter	2.79
491113	Papain family cysteine protease	-1.79
491132	Transmembrane amino acid transporter protein	2.58
491170	Alkaline and neutral invertase	2.22
491342		4.67
491541	Cytochrome P450	6.78
491574	B3 DNA binding domain	2.67

492068	PAN-like domain	1.76
492072	GH3 auxin-responsive promoter	2.96
492157	Fructose-bisphosphate aldolase class-I	-2.57
492227	Aldose 1-epimerase	2.42
492254	Fatty acid hydroxylase superfamily	2.54
492327	Ferric reductase NAD binding domain	-5.92
492338	Alpha-amylase C-terminal beta-sheet domain	-2.98
492412	Phosphorylase superfamily	3.58
492419	Protein of unknown function (DUF679	3.31
492497	Sodium/hydrogen exchanger family	2.53
492542	Protein tyrosine kinase	1.91
492547	Salt stress response/antifungal	2.13
492573	Staygreen protein	2.55
492645	Domain of unknown function (DUF3354)	-2.39
493065	ABC transporter transmembrane region	2.00
493212	Putative peptidoglycan binding domain	6.46
493376	VQ motif	2.42
493537	Cytochrome P450	-2.31
493675	HAD-hyrolase-like	-2.23
493689	Myb-like DNA-binding domain	1.84
493869	Glycerophosphoryl diester phosphodiesterase family	2.17
493913		2.09
493970	Leucine rich repeat	3.97
494088	DSBA-like thioredoxin domain	2.30
494099	Proline dehydrogenase	2.14
494160	Peptidase C26	3.04
494895	Receptor family ligand binding region	1.82
495299	CAP160 repeat	-2.22
495625	Probable lipid transfer	4.31
495626	Probable lipid transfer	3.18
495628	Probable lipid transfer	1.92
495629	Probable lipid transfer	2.97
495935	NADH-quinone oxidoreductase cyanobacterial subunit N	-1.62
496034	Protein phosphatase 2C	1.88
496082	Haloacid dehalogenase-like hydrolase	2.80
496327	AP2 domain	1.55
496399	Leucine Rich Repeat	-2.64
496502	Pectate lyase	-1.98
496593	Inositol monophosphatase family	1.49
496594	Inositol monophosphatase family	2.19
496697	WRKY DNA -binding domain	1.92
497144		2.77
497434	PsbP	-1.71
855924	Glycosyl hydrolase family 14	4.65
856363	Protein phosphatase 2C	1.93
856370	Bacterial extracellular solute-binding proteins, family	0.89

	3	
857278		
858339	Salt stress response/antifungal	4.06
858352	MatE	2.16
859362	Cupin	3.83
860334	Cupin	3.49
861088	Protein kinase domain	2.44
861238		3.56
861637	Tubulin C-terminal domain	-1.92
862790	Squalene epoxidase	-3.13
863279	Targeting protein for Xklp2 (TPX2)	4.40
863875	PAN-like domain	1.67
864007	Protein kinase domain	2.06
864008	Salt stress response/antifungal	0.84
865684	No apical meristem (NAM) protein	1.65
865949	F-box associated	2.33
866209		6.93
866365	NPH3 family	2.36
868358		
868935		4.37
869016	Protein kinase domain	3.00
869152		-5.51
870083	ABC-2 type transporter	4.62
870781	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	1.76
870958	Glycosyl hydrolase family 1	3.02
872183	Glycosyl hydrolases family 28	4.33
872280	Glutathione S-transferase, N-terminal domain	2.42
872763	Seven in absentia protein family-2012	3.44
872815	CCAAT-binding transcription factor (CBF-B/NF-YA) subunit B	2.66
873233	Calcineurin-like phosphoesterase	3.40
873301	No apical meristem (NAM) protein	3.32
874670	Cysteine-rich secretory protein family	2.87
874671	Cysteine-rich secretory protein family	4.41
875179		2.27
875778	S-locus glycoprotein family	1.72
876049	NB-ARC domain	1.62
876737	Phosphomannose isomerase type I	1.95
876790	D-mannose binding lectin	1.48
877206	MatE	1.49
877352	Protein kinase domain	2.86
877459	Methyltransferase domain	-2.51
878291	Universal stress protein family	1.79
878877	Mitochondrial carrier protein	-3.00
879378	D-mannose binding lectin	1.92
880720	Carbonic anhydrase	1.35

881866	Protein kinase domain	2.06
883904	Glycolipid transfer protein (GLTP)	2.55
884047	NB-ARC domain	
884048	NB-ARC domain	
884137	Membrane transport protein	3.05
884148	Protein tyrosine kinase	2.94
884156	Leucine Rich Repeat	2.28
884231	Glutathione S-transferase, C-terminal domain	3.18
884253	UDP-glucose 4-epimerase C-term subunit	3.71
884350		-2.28
884488	NB-ARC domain	2.09
885526		-3.07
885699	Ubiquitin-2 like Rad60 SUMO-like	5.63
886302		1.87
886352	Transferase family	1.89
886773		2.70
886846	Legume lectin domain	4.02
887070		2.22
887302	Aminotransferase class I and II	2.91
887802	Fatty acid desaturase	-4.38
888182	NADH:flavin oxidoreductase / NADH oxidase family	3.25
888266	Aminotransferase class IV	1.78
888583	UAA transporter family	2.44
888835	Family of unknown function (DUF706)	3.46
888929	Glyoxalase-like domain	5.80
889326	Cupin	4.38
889331		2.03
889332	Putative lysophospholipase	3.36
889719	CRAL/TRIO domain	2.10
889959	Putative peptidoglycan binding domain	1.71
889962		1.96
890334		
890490	Senescence regulator	2.70
890765	Phosphoribulokinase / Uridine kinase family	-1.93
890825	Isocitrate/isopropylmalate dehydrogenase	
891113		-2.11
891246	Trehalose-phosphatase	2.52
891536	Endonuclease/Exonuclease/phosphatase family	2.48
891603	Plant invertase/pectin methylesterase inhibitor	2.19
891606	Ferritin-like domain	4.07
891956	Isocitrate/isopropylmalate dehydrogenase	5.96
892077	MATH domain	7.30
892640	Protein tyrosine kinase	3.86
893206	Plant invertase/pectin methylesterase inhibitor	4.42
893249	Zinc finger, C3HC4 type (RING finger)	2.73
893410	Terpene synthase family, metal binding domain	5.25

893660	MatE-2012	2.86
893738		
893739	Aminotransferase class I and II	
893984	Photosynthetic reaction centre protein	-1.74
894063	Ribosomal protein S7p/S5e	-4.09
894894	NADH dehydrogenase transmembrane subunit	-1.99
895245	Core-2/I-Branching enzyme	2.10
895333	Gibberellin regulated protein	-3.11
895457	Gamma-thionin family	4.24
895565		-3.29
895881		4.12
895978	Cotton fibre expressed protein	2.73
896009	Mitochondrial carrier protein	2.90
897110	Exo70 exocyst complex subunit	1.62
897210	Aromatic amino acid lyase	1.73
897732	Cytochrome P450	3.41
897928	2Fe-2S iron-sulfur cluster binding domain	-2.19
898119		1.57
898417	Cytochrome P450	3.86
898600	Protein of unknown function (DUF679)	4.05
898836	AP2 domain	2.24
899310		1.00
899622		2.97
899888	Malic enzyme, NAD binding domain	2.80
900345	Peroxidase	3.44
900492		
900807		3.03
901626	Protein kinase domain	2.08
901791	Brix domain	-1.77
901802		3.55
902308	Cellulose synthase	1.85
903017	GDSL-like Lipase/Acylhydrolase	1.94
903112	Haloacid dehalogenase-like hydrolase	1.71
903754	Chitinase class I	1.72
903867	Glycosyl hydrolase family 1	3.53
903868	Glycosyl hydrolase family 1	2.62
904774		3.58
905380	Protein of unknown function (DUF3639)	1.77
906119		
906139	Mitochondrial carrier protein	2.97
906184		
906185		
906354	Protein kinase domain	2.04
906615		
906768	Sodium/hydrogen exchanger family	1.73
906847	Eukaryotic aspartyl protease	-1.74

906967	C2 domain	1.82
906971	Rare lipoprotein A (RlpA)-like double-psi beta-barrel	1.76
907023	Yippee putative zinc-binding protein	4.28
907092		-5.29
907564	Histidine phosphatase superfamily (branch 1)	1.88
907941	PsbP	-1.80
908296	Histidine phosphatase superfamily (branch 1)	-4.14
909775	Glycosyl hydrolase family 1	6.38
910252	Stress responsive A/B Barrel Domain	-3.14
910605	Protein of unknown function, DUF642	-2.41
910767	Exonuclease	2.60
911275		4.09
911410	NAD dependent epimerase/dehydratase family	2.11
912116	Domain of unknown function (DUF4308)	-2.05
912435	F-box domain	2.50
912617		3.44
912862	Possible lysine decarboxylase	3.21
913202		2.70
914188	Trehalase	2.03
914296	Thaumatin family	6.39
914365		2.82
914538	Protein tyrosine kinase	2.39
914854	FMN-dependent dehydrogenase	4.15
915032	Lipase (class 3)	2.23
916148		2.36
916315		
916316		
916350	Protein of unknown function (DUF1997)	5.03
916533	Wall-associated receptor kinase C-terminal	1.67
916973	NB-ARC domain	-7.19
917307	AP2 domain	5.58
917443		3.35
917741	Bacterial extracellular solute-binding proteins, family 3	1.71
918363	E1-E2 ATPase	
918364	haloacid dehalogenase-like hydrolase	
918486		-1.84
918509	UDP-glucoronosyl and UDP-glucosyl transferase	2.85
918581	GH3 auxin-responsive promoter	2.85
919271	Legume lectin domain	2.28
919466		2.65
919817		2.00
920005	Protein kinase domain	1.80
920573		5.00

Supplementary Table 4: Overview and annotation of genes differential expressed only in the Swedish population. Genes were obtained by differential expression analysis using DESeq by comparing Sweden selfed vs. Sweden outcrossed and by excluding the genes that were differentially expressed in both populations (see Supplementary Table 2). The table shows the annotation as retrieved by performing InterProScan on the samples. The last column gives the log2 fold changes in expression in the selfed compared to the non-selfed population.

Protein ID	InterProScan	Log 2 Fold Change in Selfed
315582	Transferase family	1.25
316436	Thaumatin family	2.10
316515	GDSL-like Lipase/Acylhydrolase	1.53
322778	GDSL-like Lipase/Acylhydrolase	-1.24
323258	Cofilin/tropomyosin-type actin-binding protein	-2.75
331145		2.20
335126	Myb-like DNA-binding domain	0.90
337099	HORMA domain	-1.81
344438	F-box domain	-3.12
347816	Leucine rich repeat N-terminal domain	-2.28
351411	Transferase family	2.16
358204	Core histone H2A/H2B/H3/H4	2.03
470302		1.91
470332	Photosystem II Pbs27	1.72
470488	Thioredoxin	1.57
470654	Fatty acid desaturase	2.26
472537		2.07
472779	FAD binding domain	2.73
473113	Ribosomal protein L34	1.50
473523		1.65
474480	Chalcone-flavanone isomerase	3.00
474482	Hsp20/alpha crystallin family	2.46
475984	Plastid and cyanobacterial ribosomal protein	1.15
476914	Peroxidase	1.44
477371	Carbonic anhydrase	1.87
477375	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD	1.63
477690	GDSL-like Lipase/Acylhydrolase	2.00
477705		-1.85
478781	Glycosyl hydrolases family 32 C terminal	2.41
478836	GDSL-like Lipase/Acylhydrolase	1.68
478873	Polyprenyl synthetase	2.65
478916	Uroporphyrinogen decarboxylase (URO-D)	1.38
479077	Major intrinsic protein	2.09
479089	GDSL-like Lipase/Acylhydrolase	1.44
480223	Photosystem II reaction centre X protein (PsbX)	1.18
480563	Major intrinsic protein	1.71
480771		2.12
481081	Gelsolin repeat	3.15
481775	Plant transposon protein	1.81
481893	UDP-glucoronosyl and UDP-glucosyl transferase	-1.64

482547	Domain of unknown function (DUF588)	-1.19
484604	MATH domain	2.93
484691	Uncharacterized ACR, COG1678	1.17
485048	NADH(P)-binding	1.54
485112	Chlorophyll A-B binding protein	2.15
485118	Chitinase class I	2.70
486168	Glycosyl hydrolases family 17	2.93
488045	Hsp20/alpha crystallin family	1.76
488726	D-mannose binding lectin	1.71
489312	Sigma 54 modulation protein / S30EA ribosomal protein	2.04
489446	Glycosyl hydrolase family 1	1.45
489709	Peroxidase	1.37
489817	ATP synthase delta (OSCP) subunit	1.12
489869	Hsp20/alpha crystallin family	2.95
490836	Xyloglucan endo-transglycosylase (XET) C-terminus	1.41
492051	short chain dehydrogenase	3.53
492328	Magnesium-protoporphyrin IX methyltransferase C-terminus	1.59
492755	Oxygen evolving enhancer protein 3 (PsbQ)	1.57
493327	UDP-glucoronosyl and UDP-glucosyl transferase	1.55
493595	Chloroplast protein precursor Ycf15 putative	6.75
494078	Cupin	4.38
494136	Ribulose-1,5-bisphosphate carboxylase small subunit	1.49
494137	Ribulose bisphosphate carboxylase, small chain	1.76
494249	Palmitoyl protein thioesterase	-2.34
494386	Magnesium chelatase, subunit ChlI	1.95
495095		0.99
495160	Fumarase C C-terminus	0.93
495501	short chain dehydrogenase	5.12
495659	Fasciclin domain	1.94
496104	Hsp20/alpha crystallin family	1.63
496228	Heavy-metal-associated domain	1.81
496724	Annexin	-3.18
497157	Glutathione S-transferase, N-terminal domain	1.23
497425		-0.95
856310	NAD(P)-binding Rossmann-like domain	2.64
864354		-2.10
865964	Zinc knuckle	3.37
868649	Glutamine amidotransferase class-I	-1.33
872501	UDP-glucoronosyl and UDP-glucosyl transferase	2.78
876642	Formyl transferase	-2.68
881642	Hsp20/alpha crystallin family	2.17
882492	Cysteine-rich secretory protein family	4.03
883330	Protein kinase domain	1.69
883800	F-box associated	1.93

890625	Berberine and berberine like	2.47
892108		2.29
892609		
892638	Protein tyrosine kinase	-1.47
892824	Core histone H2A/H2B/H3/H4	2.06
892975	Glycosyl hydrolase family 9	1.47
893116	ABC transporter	-4.83
894122		3.90
894159	Late embryogenesis abundant protein	1.48
894808		2.47
894902	Pathogenesis-related protein Bet v I family	1.03
895063	POT family	1.54
898865	Flavin containing amine oxidoreductase	1.82
899019	Alpha/beta hydrolase family	1.59
900301		4.69
900474		1.90
900498		-0.58
900653		
900658		
901851	Rare lipoprotein A (RlpA)-like double-psi beta-barrel	1.85
902030	Myb-like DNA-binding domain	1.55
902799	UDP-glucoronosyl and UDP-glucosyl transferase	1.91
903400	Myb-like DNA-binding domain	-1.68
903677	GDSL-like Lipase/Acylhydrolase	2.52
904536		
905021	Chlorophyll A-B binding protein	3.11
905079		3.44
905420		3.51
905859	DVL family	3.71
905912		
906213	Papain family cysteine protease	-2.38
907497	GNAT acetyltransferase 2	-1.10
908373	Sec1 family	-1.89
908552	AhpC/TSA family	1.98
908912	Ferritin-like domain	-3.59
908994	Core histone H2A/H2B/H3/H4	1.52
909035	Eukaryotic aspartyl protease	-1.22
909242	Leucine rich repeat N-terminal domain	0.79
909291	GH3 auxin-responsive promoter	1.55
909315	Domain of unknown function (DUF3479)	1.15
910104	Cupin	5.02
910242	Male sterility protein	2.18
911648		1.95
912392	Zinc-binding dehydrogenase	2.31
912528		-2.05
914076	Phosphopantetheine attachment site	0.75

914433	Multicopper oxidase	1.62
916700	Palmitoyl protein thioesterase	
920182	Protein of unknown function (DUF626)	4.16

Supplementary Table 5: Statistical overrepresentation of gene ontologies (GOs) in the differentially expressed genes of the Norwegian population. GO overrepresentation of differentially expressed genes was performed with the Cytoscape plugin BiNGO which tests for overrepresentation by hypergeometric testing using a Benjamini-Hochberg correction.

<table>
<tr><th>Description</th><th>p-value</th><th>Genes in test set</th></tr>
<tr><td>carbohydrate binding</td><td>3.8104E-8</td><td>861088|869016|914538|477741|919271|876790|318019|903754|483469|492068|879372|863875|483470|875778|879378|911976|346169|479116|870854|886846|492227</td></tr>
<tr><td>catalytic activity</td><td>6.1502E-5</td><td>494088|891536|480406|903754|861637|885699|476696|881866|490098|880720|903112|334097|494099|486477|486479|327935|476695|475808|487053|328029|476141|899888|901626|870083|911976|486481|903868|862790|903867|331197|326599|482737|493212|475262|487844|471868|484387|484388|481168|897210|870958|471294|484965|909775|914854|346891|863875|315131|488319|492227|876737|483936|332024|329840|470382|475178|484123|471117|879372|890825|884137|472361|879378|353102|489370|481887|318019|313855|496593|496594|893206|910767|900345|491113|892640|915032|471526|318682|898417|481994|920005|489772|493878|494390|476582|889581|489965|319814|492338|479864|873233|485691|490604|472226|481863|496082|884253|914188|908138|472635|893410|492327|472778|884148|861088|856363|914538|476608|864007|483526|864008|312863|472781|472780|317500|492157|890765|352125|329886|891603|475879|493675|493970|918364|918363|888182|491721|489775|902535|470930|485496|355259|492486|484318|353505|493869|489642|476725|893739|348799|918509|906073|492412|477857|327109|877352|899117|496034|483469|480389|329582|902308|483470|475766|323821|343972|878291|353011|470440|891246|485103|331608|496502|865114|858339|320824|907023|486718|869016|493537|471013|908296|903017|887802|324269|480225|338258|906354|496399|485452|484885|483895|875778|346169|478210|480324|491541|474819|887302|492542|872183|919271|492547|889959|888835|354029|478341|871042|473561|350231|489007|497196|870854|312603|911410|886846|487575|493065|888266|487785|876790|487885|492068|487282|312409|479069|471443|485216</td></tr>
</table>

pollination	5.4049E-4	492068|879372|914538|863875|875778|879378|475178
pollen-pistil interaction	5.4049E-4	492068|879372|914538|863875|875778|879378|475178
extracellular region	5.4049E-4	906971|860334|494895|493212|874670|874671|889959|491721|476141|476393|480225|889326|859362|485216
reproduction	1.4655E-3	492068|879372|914538|863875|875778|879378|475178
cell communication	2.0331E-3	492068|879372|914538|863875|875778|879378|475178
proteinaceous extracellular matrix	3.4345E-3	493212|889959|476141
receptor activity	4.1177E-2	917741|884488|475633|471969|892640|494895|856370|493970|477857|477682|496399|481829|884137|884148

Supplementary Table 6: Statistical overrepresentation of gene ontologies (GOs) in the differentially expressed genes of the Swedish population. GO overrepresentation of differentially expressed genes was performed with the Cytoscape plugin BiNGO which tests for overrepresentation by hypergeometric testing using a Benjamini-Hochberg correction.

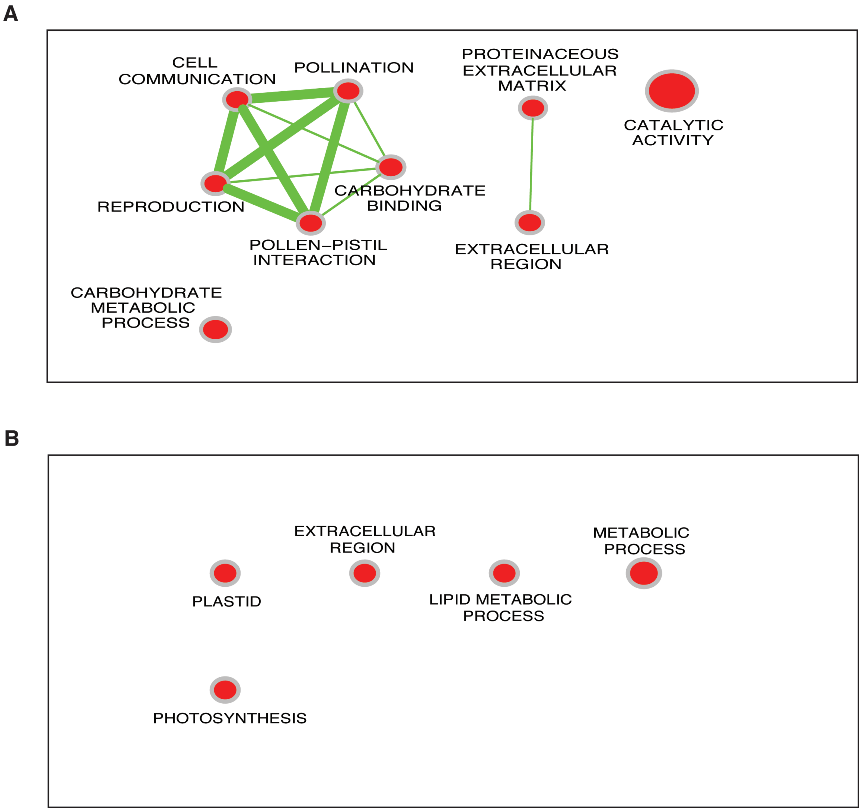
<table>
<tr><th>Description</th><th>p-value</th><th>Genes in test set</th></tr>
<tr><td>photosynthesis</td><td>1.0918E-15</td><td>495512|492051|473204|473203|495501|478197|494386|474626|482403|492755|496598|485112|899321|492328|480223|905021</td></tr>
<tr><td>generation of precursor metabolites and energy</td><td>1.1043E-3</td><td>495512|489817|485112|473204|473203|899321|478197|482403|905021</td></tr>
<tr><td>plastid</td><td>1.1043E-3</td><td>475984|492051|495501|494137|494136</td></tr>
<tr><td>metabolic process</td><td>1.3940E-3</td><td>914076|495501|483526|496598|492755|478916|347816|344438|879372|899321|883330|316515|470654|892975|492051|477375|890879|474626|491721|876642|892638|475984|903677|481893|495160|480225|473113|486168|909035|906213|346169|911976|479089|480223|905021|492486|898865|868649|478197|494386|909315|902799|494249|493878|494390|474480|493327|889581|489965|494137|478873|497196|494136|348799|870854|485691|490836|495512|489817|473204|473203|478836|322778|482403|489312|485118|477690|485112|329582|908138|478781|337099|489446|492328|872501</td></tr>
<tr><td>extracellular region</td><td>2.1284E-3</td><td>490836|480225|905079|910104|901851|491721|494078|882492</td></tr>
<tr><td>lipid metabolic process</td><td>4.1340E-3</td><td>903677|493878|914076|477690|898865|478873|478836|322778|479089|316515|470654</td></tr>
<tr><td>thylakoid</td><td>9.9842E-3</td><td>474626|496598|492755|480223</td></tr>
<tr><td>carbohydrate metabolic process</td><td>2.2840E-2</td><td>490836|892975|492486|492051|495501|491721|485118|480225|486168|329582|478781|348799|346169|489446|870854</td></tr>
<tr><td>external encapsulating structure</td><td>2.8134E-2</td><td>490836|871042|893116|480225|491721</td></tr>
<tr><td>catalytic activity</td><td>3.3568E-2</td><td>909242|495501|483526|489709|472781|472780|470488|478916|347816|879372|883330|890625|914433|476914|865114|316515|470654|497157|892975|477371|492051|320824|477375|487053|491721|876642|892638|903677|481893|495160|480225|486168|909035|906213|902535|346169|911976|479089|492486|898865|868649|494386|481994|902799|912392|354029|494249|493878|494390|871042|474480|893116|493327|889581|489965|494137|497196|494136|348799|496724|870854|485691|490836|908552|489817|906073|478836|481863|322778|899117|485118|477690|329582|908138|478781|472779|488319|475766|337099|472778|489446|492328|872501</td></tr>
</table>

Supplementary Table 7: Overview of SNP calling with GATK based SNP calling. The amount of detected SNPs for each round of SNP calling for each population and treatment is given below. The database for the first SNP call is based on the overlap for each population from the primary SNP calling. Overlap for each population of each round was used as database input for the next round of SNP calling.

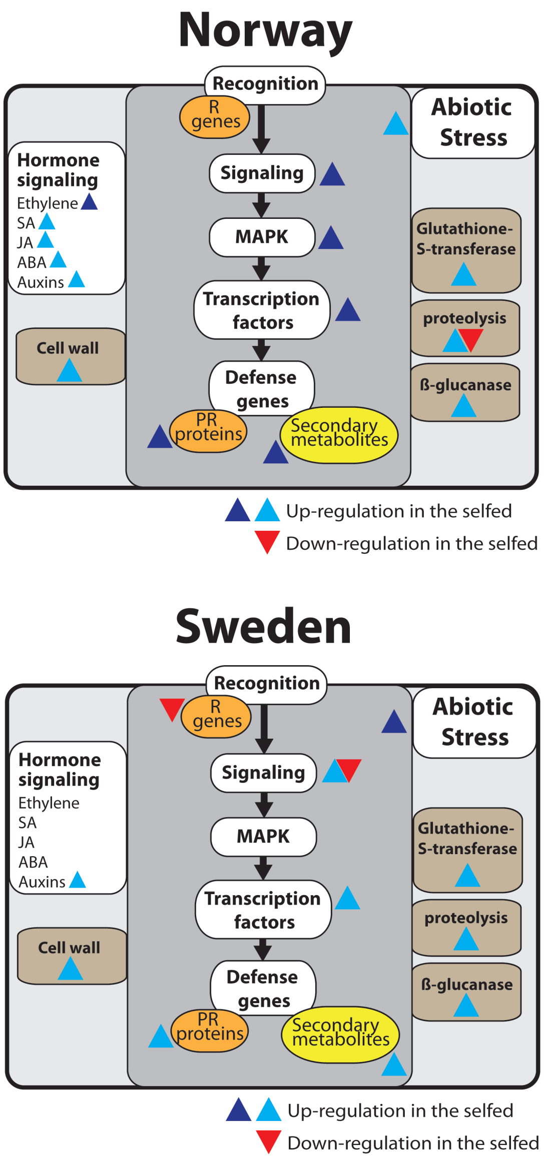
	Norway outcrossed	**Norway selfed**	**Sweden outcrossed**	**Sweden selfed**
Overlap of preliminary SNP calling	214,313	219,626	325,285	336,162
SNP calling in GATK using overlap	752,747	788,590	1,233,356	1,288,294
1st Iteration of SNP calling in GATK	802,937	839,978	1,346,785	1,411,425
2nd Iteration of SNP calling in GATK	803,913	841,003	1,349,520	1,414,490
3rd Iteration of SNP calling in GATK	804,029	841,016	1,349,523	1,414,570
4th Iteration of SNP calling in GATK	804,036	841,107	1,349,556	1,414,578
5th Iteration of SNP calling in GATK	803,985	841,073	1,349,614	1,414,668

Supplementary Table 8: Variance in gene expression in each group of samples. For each gene, population and condition, variances in expression among replicates were calculated and the mean and variance of these variances were determined.

	Norway outcrossed	**Norway selfed**	**Sweden outcrossed**	**Sweden selfed**
Mean of variances	0.831	0.882	0.541	0.669
Variance of variances	3.260	5.694	0.763	2.145

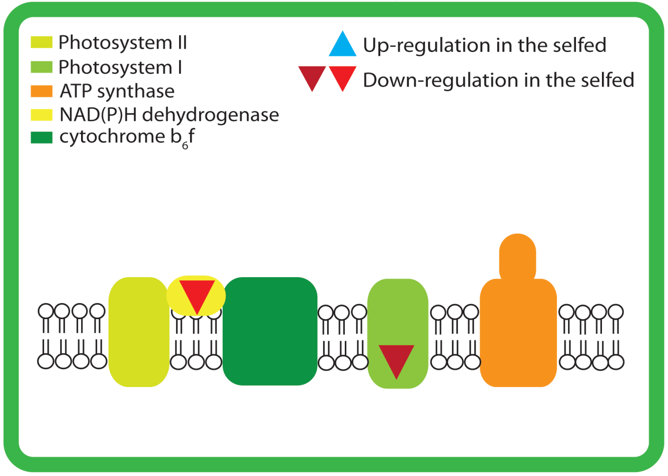


Supplementary Figure 1: Enrichment analysis of the genes that were significantly differentially expressed in selfed and outcrossed progeny only in (A) Norway and (B) Sweden. Enrichment map, built upon GO annotation (Cytoscape plugins BiNGO and EnrichmentMap), providing insights into GO terms that are highly significant and their interactions in each of the populations.

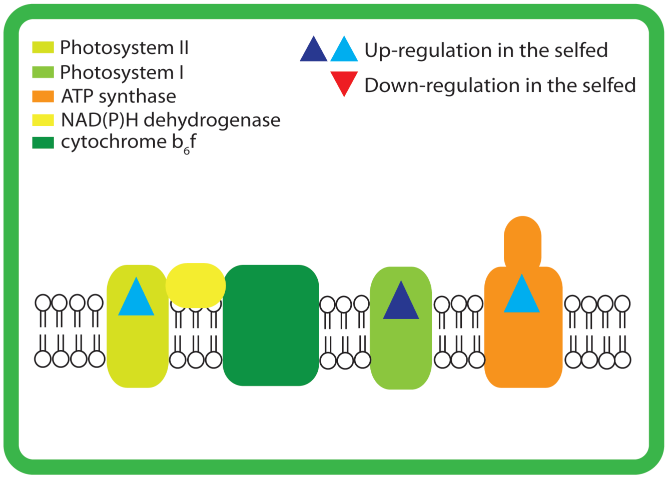


Supplementary Figure 2: Results from biotic stress pathway analyses in the Norwegian and the Swedish population, respectively. The analyses were based on the genes that were significantly differentially expressed only in the respective population. Assignment of the genes into pathways was achieved by 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; dark blue: many genes, light blue: single genes), whereas red signifies a negative log2 fold change in the selfed compared to the outcrossed (down-regulation).

Norway

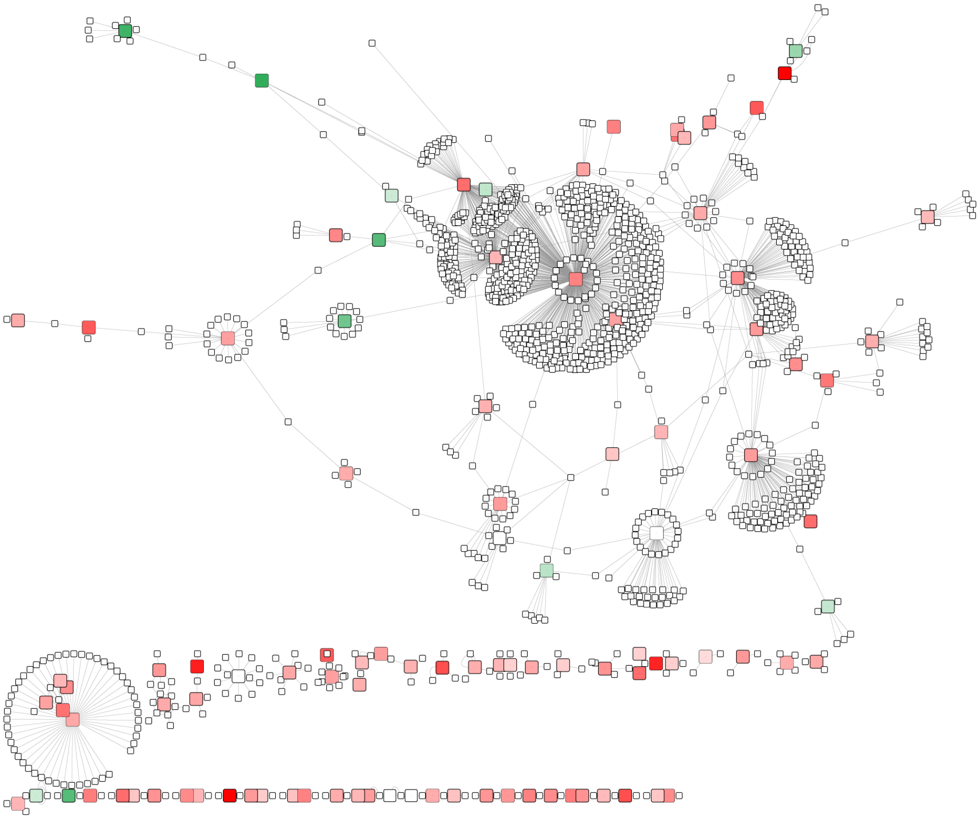


Sweden

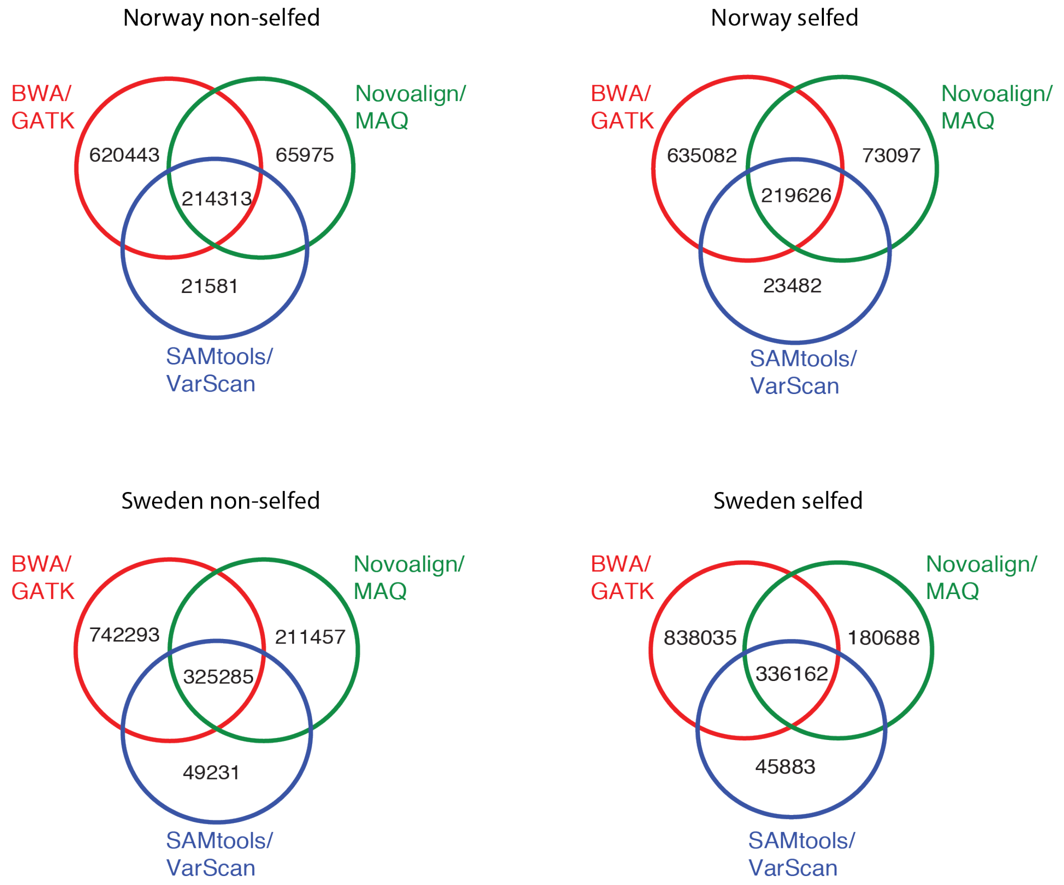


Supplementary Figure 3: Results from chloroplast pathway analyses in the Norwegian and the Swedish population, respectively. The analyses were based on the genes that were significantly differentially expressed only in the respective population. Assignment of genes into pathways was achieved by ortholog 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; dark blue: many genes, light blue: single genes), whereas red signifies a negative log2 fold change in the selfed compared to the outcrossed (down-regulation).

Supplementary Figure 4: Gene interaction network for 224 genes that were significantly differentially expressed in the Norwegian population. Interactions were restricted to first neighbour interactions. Nodes, representing the differentially expressed genes, are coloured by fold change (red/pink: highly/moderately up-regulated in selfed progeny; dark/pale green: strongly/moderately down-regulated in selfed progeny).



Supplementary Figure 5: Gene interaction network for 104 genes that were significantly differentially expressed in the Swedish population. Interactions were restricted to first neighbour interactions. Nodes, representing the differentially expressed genes, are coloured by fold change (red/pink: highly/moderately up-regulated in selfed progeny; dark/pale green: strongly/moderately down-regulated in selfed progeny).



Supplementary Figure 6: Results from SNP calling. SNPs were called using three different methods: alignment with BWA followed by SNP calling with GATK using a SNP database created with Freebayes (red), alignment with Novoalign followed by SNP calling with MAQ (green) and alignment with TopHat followed by SNP calling with SAMtools and VarScan (blue). The overlap of the resulting amount of SNPs for each approach and population and treatment is shown as Venn diagrams.

Supplementary Material 2

SNP calling – Methods and Results

Methods

SNP calling was performed using three different programs and methods:

For the first approach reads were assembled with BWA (version 0.6.2 (Li and Durbin 2009)) and the resulting alignment was used for SNP calling with Freebayes (version 0.9.6; http://arxiv.org/abs/ 1207.3907) which output was used as a database to feed into GATK (version 2.1.13 (McKenna, et al. 2010; DePristo, et al. 2011)). The actual SNP calling on the alignment was then performed with GATK using the Freebayes generated SNP database for each individual. For the final SNP calling step individuals for one treatment and population were merged. GATK takes coordinate sorted alignments as input. These alignments are then locally realigned to avoid areas of misalignment. Afterwards duplicated reads are marked and removed, the best pair is kept. Based on a provided SNP database base quality scores are recalibrated, so that bases which agree with the bases in the SNP database are assigned a high quality score. This gives more confidence to the base calls. Then replicates are input into the UnifiedGenotyper, a multi-sample SNP calling tool that uses a Bayesian genotype likelihood to call SNPs.

For the second approach, reads were aligned against the reference genome with Novoalign (version 2.08.03; http://www.novocraft.com/) and then converted to a format used by the aligner MAQ (version 0.7.1 (Li, et al. 2008)). Assemblies of all individuals were merged for each treatment and population and SNPs were called and filtered using default parameters. First MAQ assembles the consensus of all replicates based on a statistical model. It then calculates a quality score for each base of the consensus. Later SNPs are filtered. The default parameters are as followed: 1) discard SNPs with 3bp flanking region around potential indel, 2) discard SNPs covered by three or fewer reads, 3) discard SNPs covered by no read with mapping quality >60, 4) if three or more SNPs fall into a 10 bp window, discard all, 5) discard SNP with consensus quality <10.

For the third approach the TopHat (version 2.04 (Trapnell, et al. 2009)) alignment was used and fed into SAMtools (version 0.1.18 (Li, et al. 2009)) using the mpileup command, combining the alignment of one treatment and population. Afterwards SNPs were called using VarScan (version 2.3.3 (Koboldt, et al. 2012)) and filtered using a p-value of 0.05.

SAMtool's mpileup function generates a pileup of bases by aligning reads to the reference genome. For each base it computes a base alignment quality (BAQ), the probability that a base is misaligned, to rule out false-positives. VarScan then employs a heuristic/statistic approach to call variants from the pileup that meets certain threshold criteria.

A detailed pipeline for all programs is found below.

The results of the above-mentioned SNP calling procedures were overlapped for one treatment and population and the overlap for selfed and non-selfed was then merged for each population and used as a database for SNP calling with GATK. The resulting SNPs for each treatment and population were again overlapped (overlap of selfed and non-selfed of one population) and low quality SNP predictions were filtered out before a new round of SNP calling with GATK was performed. This procedure was performed five times in total (until the number of SNPs stabilised).

i) SNP calling with BWA, Freebayes and GATK

(Nb! Notation "x" signifies sample name.)

First reads were aligned with BWA.

```
bwa index Araly1_assembly_scaffolds.fasta

bwa aln -t8 Araly1_assembly_scaffolds.fasta x_1.fastq > x_1.sai

bwa aln -t8 Araly1_assembly_scaffolds.fasta x_2.fastq > x_2.sai

bwa sampe Araly1_assembly_scaffolds.fasta x_1.sai x_2.sai x_1.fastq x_2.fastq
> x.sam
```

In the next step the alignment was sorted and indexed with SAMtools.

```
samtools view -bt Araly1_assembly_scaffolds.fasta.fai x.sam > x.bam

samtools sort x.bam > x.sorted

samtools index x.sorted.bam
```

Then read groups unique for each sample were added with Picard tools (version 1.81; http://picard.sourceforge.net).

```
java -jar AddOrReplaceReadGroups.jar INPUT=x.sorted.bam OUTPUT=xpicard.bam
SORT_ORDER=coordinate CREATE_INDEX=true RGPL=illumina   RGSM=x RGLB=x RGPU=x
VALIDATION_STRINGENCY=LENIENT
```

This was followed by local realignment with GATK.

```
java -jar GenomeAnalysisTK.jar -T RealignerTargetCreator -R
Araly1_assembly_scaffolds.fasta -I xpicard.bam -o x.intervals

java -jar GenomeAnalysisTK.jar -T IndelRealigner -R
Araly1_assembly_scaffolds.fasta -I xpicard.bam -targetIntervals x.intervals -
o xrealigned.bam --filter_mismatching_base_and_quals
```

Afterwards duplicates were marked and removed using Picard tools.

```
java -jar MarkDuplicates.jar INPUT=xrealigned.bam REMOVE_DUPLICATES=
true VALIDATION_STRINGENCY=LENIENT AS=true METRICS_FILE=x.dubs
OUTPUT=x.realnDup.bam

samtools index x.realnDup.bam
```

In the next step a SNP database for each individual was created using Freebayes, which in a later step was fed into GATK.

```
freebayes --fasta-reference Araly1_assembly_scaffolds.fasta xpicard.bam >
     x_knownSites.vcf
```

Now base quality recalibration was performed with GATK.

```
java -jar GenomeAnalysisTK.jar -T BaseRecalibrator -I xrealnDup.bam
-R Araly1_assembly_scaffolds.fasta -knownSites x_knownSites.vcf -o
x.recal_data.grp --disable_indel_quals

java -jar GenomeAnalysisTK.jar -T PrintReads -R Araly1_assembly_
scaffolds.fasta -I x.realnDup.bam -BQSR x.recal_data.grp -o x.recal.bam
```

Finally SNPs were called with GATK on merged samples for each treatment and population.

```
java -jar GenomeAnalysisTK.jar -T CombineVariants -R Araly1_assembly_
scaffolds.fasta --variant x1_knownSites.vcf [--variant xn_knownSites
.vcf] -o merged_knownSites.vcf -genotypeMergeOptions UNIQUIFY

java -jar GenomeAnalysisTK.jar -T UnifiedGenotyper -R Araly1_assembly_
scaffolds.fasta -I x1.recal.bam [-I xn.recal.bam] –dbsnp merged_known
Sites.vcf -o merged_GATKvariants.raw.vcf -stand_call_conf 50.0 -stand_
emit_conf 10.0 -dcov 100 -glm BOTH

java -jar GenomeAnalysisTK.jar -T SelectVariants -R Araly1_assembly_
scaffolds.fasta --variant merged_GATKvariants.raw.vcf -selectType SNP -log
merged.SelectSNPs.log -o merged.GATKsnps.raw.vcf
```

ii) SNP calling with Novoalign and MAQ

First reads were aligned with Novoalign.

```
novoindex Araly_reference.nix Araly1_assembly_scaffolds.fasta

novoalign -d Araly_reference.nix -f x_1.fastq x_2.fastq > x.novo
```

Now the alignment needed to be converted to a format used by the MAQ assembler.

```
novoutil n2mhdrs Araly1_assembly_scaffolds.fasta > in.list

novo2maq x.map in.list x.novo
```

In the following step samples were merged according to each treatment and population and were later assembled with MAQ.

```
maq mapmerge merged.map x1.map [xn.map]

maq fasta2bfa Araly1_assembly_scaffolds.fasta Araly_reference.bfa

maq assemble merged.cns Araly_reference.bfa merged.map
```

In the last step SNPs were called and filtered with MAQ using default parameters.

```
maq cns2snp merged.cns > merged.snp

maq.pl SNPfilter merged.snp > merged.filtered.snp
```

iii) SNP calling with SAMtools and VarScan

For this approach the TopHat output (see *Transcriptome assembly* above) was used. This alignment was fed into SAMtools to generate a pileup for each treatment and population.

```
samtools mpileup -f Araly1_assembly_scaffolds.fasta x1.bam [xn.bam] >
merged.mpileup
```

This was followed by SNP calling and filtering with VarScan.

```
java -jar VarScan.jar mpileup2snp merged.mpileup --p-value 0.05 --output-vcf
1 > merged.snp.vcf
```

Results

For preliminary SNP calling three approaches were used as described in Supplementary Method 1E. The number of SNPs called varied largely between the different methods.

Common to all was that they detected a larger amount of SNPs in the selfed of both populations compared to the non-selfed and more SNPs in the population Sweden compared with Norway (Supplementary Figure 1). The overlap included 14313 SNPs common for Norway non-selfed, 219626 SNPs common for Norway selfed, 325285 SNPs common for Sweden non-selfed and 336162 SNPs common for Sweden selfed. Based on these results, the overlap of both treatments for one population was used for further SNP calling with GATK. This GATK calling was performed in five iterations. It could be seen that the differences in the amount of detected SNPs between each round of SNP calling became smaller and smaller and almost converged. After the fifth iteration Sweden showed more SNPs than Norway and the selfed of both populations (~841100 for Norway and ~1414700 for Sweden) showed more SNPs than the non-selfed (~804000 for Norway and 1349600 for Sweden). Supplementary Table 4 gives a detailed overview of the results of each SNP calling round.

References:

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

Koboldt DC, Zhang QY, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research 22:568-576.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research 18:1851-1858.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297-1303.

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